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

Kirill Borusyak, UCL
Peter Hull, University of Chicago

Contents

A Theoretical Appendix

A.1 Randomization Inference

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

Formally, we have the following result on hypothesis testing:

Proposition A1. *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 $\beta = b$,*

$$Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right) \geq 1 - \alpha, \tag{A1}$$

where the acceptance region is constructed for a given b as

$$T_{\alpha/2} = \sup\left\{t \in \mathbb{R} \cup \{-\infty\} : \; Pr(T^* < t \mid y, x, w) \leq \frac{\alpha}{2}\right\} \tag{A2}$$

$$T_{1-\alpha/2} = \inf\left\{t \in \mathbb{R} \cup \{+\infty\} : \; Pr(T^* \geq t \mid y, x, w) \leq \frac{\alpha}{2}\right\}. \tag{A3}$$

Equation (A1) further holds with equality when $T^* \mid (y, x, w)$ is continuously distributed under the null.

Proof. See Appendix C.4. □

This result shows that when shocks are as-good-as-randomly assigned, a test of $\beta = b$ which rejects when $T \notin \left[T_{\alpha/2}, T_{1-\alpha/2}\right]$ has size of exactly α in finite samples provided the test statistic is conditionally continuously distributed under the null. When this distribution is not continuous, the test is still guaranteed to be conservative with a rejection rate of no greater than α.[52] The lower- and upper-bounds of the test region, $T_{\alpha/2}$ and $T_{1-\alpha/2}$, are given by the shock assignment process (Assumption 3) and represent the lower- and upper $\frac{\alpha}{2}$th percentile tails of the known conditional distribution of T^*. With exchangeable shocks, for example, $T_{\alpha/2}$ and $T_{1-\alpha/2}$ are given by the tails

[52] $T \mid w$ 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).

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 (Lehmann and Romano 2006, p. 636).[53] We note that while the previous intuition for such a testing procedure conditioned on ε and w, Proposition A1 establishes correct unconditional coverage of the test. This follows by the law of iterated expectations: the unconditional coverage $Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right)$ is the expectation, across realizations of ε and w, of the controlled conditional coverage $Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right] \mid \varepsilon, w\right)$.[54]

It follows from Proposition A1 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 A1. Then $Pr(\beta \in CI) \geq 1 - \alpha$, with equality if $T^* \mid (y, x, w)$ is continuously distributed under the null.*

Proof. Follows from Proposition A1 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 last possibility providing evidence against correct specification (Imbens and Rosenbaum 2005).

A.2 Consistency of Recentered IVs

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

We study consistency of a recentered IV estimator,

$$\tilde{\beta} = \beta + \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell}, \tag{A4}$$

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

[53] When a random sample of permutations is used, the realized g (i.e. identity permutation) should be added to this sample. The test then remains exact, or slightly conservative because of discreteness (Lehmann and Romano 2006, p. 636; Hemerik and Goeman 2018). In contrast to identification (see footnote 15) randomness of the chosen permutations is important here: non-random permutation sets do not generally guarantee valid inference (e.g. Southworth et al. 2009).

[54] It is instructive to highlight how exactly the knowledge of the shock assignment process matters in Proposition A1. 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.

Assumption A1. *(Relevance):* $\frac{1}{L}\sum_{\ell}\tilde{z}_\ell x_\ell \xrightarrow{p} M \neq 0$.

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

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

Assumption A2. *(Regularity):* $\mathbb{E}\left[\varepsilon_\ell^2 \mid w\right] \leq B$ *for finite* B.

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

Assumption A3. *(Weak IV dependence):* $\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0$.

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

Proposition A2.

(i) *Suppose Assumptions 2, 3, and A1-A3 hold. Then* $\tilde{\beta} \xrightarrow{p} \beta$.

(ii) *Suppose Assumptions 1, 3, and A1-A3 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 A1 with* $T = \frac{1}{L}\sum_{\ell} f_\ell(g,w)(y_\ell - bx_\ell)$ *is consistent, i.e. for any* $b \neq \beta$ *we have* $Pr\left(T \notin \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right) \to 1$.

Proof. See Appendix C.5. □

The key condition of weak IV dependence states that the average absolute value of mutual covariances of the recentered instrument $\tilde{z}_\ell$ converges to zero as L grows. Typically, this would require the number of shocks N to grow with L, so that only a small fraction of observation pairs are most exposed to the same shocks in g. When this condition holds, Proposition A2 shows that $\tilde{\beta}$ is consistent even when unobserved shocks affect observations jointly (through ε_ℓ), in an unspecified manner.[55] Proposition

[55] We note that the recentering of z_ℓ is key for this result: the non-recentered IV estimator may not converge to β even when z_ℓ is valid in the sense of $\mathbb{E}\left[\frac{1}{L}\sum_{\ell} z_\ell\varepsilon_\ell\right] = 0$. For instance, suppose observations with systematically high z_ℓ (i.e., high μ_ℓ) are similarly exposed to an unobserved aggregate shock in ε_ℓ: the variance in $\frac{1}{L}\sum_{\ell} z_\ell\varepsilon_\ell$ due to this shock may not vanish, even in large samples when Assumption A3 holds. This problem does not arise for recentered IV which does not systematically vary across observations. See Appendix A.11 for an example and Lee and Ogburn (2019) for a related discussion.

A2 applies to the recentered IV; Appendix A.6 extends it to the alternative μ_ℓ-controlled IV regression (see Proposition A6(v)).

Our two sufficient conditions for Assumption A3 are non-nested:

Lemma A1.

(i) *Suppose* $\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \geq 0$ *almost-surely for all* (ℓ, m) *and Assumption 1 holds. Then Assumption A3 holds if* $\text{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\right] \to 0$. *Moreover, if* $f_\ell(g; w)$ *is weakly monotone in* g *for all* ℓ *and components of* g *are jointly independent conditionally on* w*, then* $\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \geq 0$ *almost-surely.*

(ii) *Suppose* $G_\ell \subseteq \{1, \ldots, N\}$ *is such that* $f_\ell(\cdot; w)$ *does not depend on* g_n *for any* $n \notin G_\ell$ *almost-surely. Then Assumption A3 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.5. □

The first condition applies to the setting when all shocks affect all observations in the same direction, but to different extents. This holds, for example, for shift-share instruments with non-negative exposure weights. More generally nonlinear $f_\ell(\cdot)$ may also be monotone in the shock vector; for example each transportation infrastructure upgrade may weakly improve market access everywhere. In these cases, the recentered IV estimator is consistent when the first-stage covariance converges to a non-zero constant M and the average instrument $\frac{1}{L}\sum_\ell \tilde{z}_\ell$ converges to its expectation of zero in the l_2 norm. For linear shift-share IV this extra condition requires the number of shocks to grow with L with the average exposure to each individual shock becoming vanishingly small, as in Borusyak et al. (2019) and Adão et al. (2019). The assumption of independent shocks can be weakened, for instance to allow for shocks that are independent across many clusters. The second condition in Lemma A1 follows Aronow and Samii (2017) in assuming that for most pairs of observations the two instruments $\tilde{z}_\ell$ and $\tilde{z}_m$ rely on non-overlapping sets of shocks g. This would be the case, for example, when each observation receives its own random shock, and $f_\ell(\cdot)$ only depends on ℓ's shock and those of its neighbors up to a fixed network distance.

Consistency with Permutations Assumption A3 and Lemma A1 may be difficult to apply when the distribution of shocks is known conditionally on $w_c = (w, \Pi(g))$, which includes some function $\Pi(g)$ of shocks, such as a permutation class (see Section 3.3). Even if shock components g_n are *iid* conditionally on w, they can be dependent conditionally on $\Pi(g)$ (negatively correlated, in the scalar g_n case with $\Pi(g)$ denoting permutation classes). For completeness we next present a version of Proposition A3 that applies in that case, with similar conditions but some of them applied with w and others to w_c.

In this setting one may consider two expected instruments: $\mu_\ell^u = \mathbb{E}\left[f_\ell(g, w) \mid w\right]$ and $\mu_\ell^c = \mathbb{E}\left[f_\ell(g, w) \mid w, \Pi(g)\right]$, with corresponding recentered instruments $\tilde{z}_\ell^u$ and $\tilde{z}_\ell^c$. Here u and c stand for "unconditional" and "conditional" on $\Pi(g)$. Similarly, the above assumptions 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 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 A3 instead of A3c. The problem this proposition solves is that conditioning on $\Pi(g)$ creates dependencies across shock components, making $\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w_c\right]$ more difficult to bound; for example Lemma A1 is not useful conditionally on w_c. We show that when Assumption A3 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 A3. *Suppose Assumptions 1, 3c, A1c, A2, and 2 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.6. □

A.3 Asymptotic Efficiency

This appendix formalizes the efficiency result discussed in Section 3.5. We first define some useful asymptotic concepts. For a non-random sequence $r_L \to \infty$, we say that an estimator $\tilde{\beta}$ converges to β at rate r_L when $r_L\left(\tilde{\beta} - \beta\right)$ converges to a non-degenerate distribution with zero mean and variance $V > 0$ as $L \to \infty$. We refer to V as the asymptotic variance of $\tilde{\beta}$, and say that the convergence rate r_L^* is faster than r_L when $\lim_{L\to\infty} \frac{r_L}{r_L^*} = 0$.[56] We consider IV estimators of the form $\tilde{\beta} = \frac{1}{L}\tilde{z}'y / \frac{1}{L}\tilde{z}'x$ where $\tilde{z} = f\left(g, w\right)$ for an $L \times 1$ vector of functions f such that $\mathbb{E}\left[\tilde{z} \mid w\right] = 0$; the last condition requires that $\tilde{z}$ is a recentered instrument. We say that $\tilde{\beta}$ is "regular" if *(i)* it converges to β at some rate r_L, *(ii)* it has an asymptotic first stage, i.e. $\frac{1}{L}\tilde{z}'x \xrightarrow{p} M$ for some $M \neq 0$, and *(iii)* the sequences of $\frac{1}{L}\tilde{z}'x$ and $\left(r_L \frac{1}{L}\tilde{z}'\varepsilon\right)^2$ are uniformly integrable. These definitions yield the following result:

Proposition A4. *Suppose Assumption 2 holds and $\mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]$ is almost-surely invertible. Consider the recentered instrument z^* defined by equation (9). Then if the associated estimator $\beta^* = \frac{1}{L}z^{*\prime}y / \frac{1}{L}z^{*\prime}x$ is regular, it has the smallest asymptotic variance of all regular recentered IV estimators $\tilde{\beta}$: there is no $\tilde{\beta}$ that converges at a rate faster than that of β^*, and any $\tilde{\beta}$ converging at the same rate has an asymptotic variance at least as large as that of β^*.*

Proof. See Appendix C.3. □

This result characterizes the efficient instrument under weak conditions.[57]

[56] In general, the asymptotic variance concept is useful when the limiting distribution of $\tilde{\beta}$ is normal. However, it can be considered more broadly; in particular, a researcher with a quadratic loss function will generally value reductions in V outside the normal case. We therefore do not restrict the shape of the asymptotic distribution until this is required in Proposition A5.

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

We next show that under additional conditions the variance-minimizing instrument z^* also maximizes the local power of the associated RI test. The theoretical argument closely follows Lehmann and Romano (2006, Section 5.2.2), except without assuming that the distribution of shocks is the permutation distribution or that the data are *iid*. The proposition requires asymptotic normality of the estimator, for which we do not have low-level conditions because of the generality of our framework but which should hold in typical applications. For simplicity we here treat w as non-stochastic, but the argument generalizes naturally since the characterization of efficient recentered IV in Proposition A4 applies conditionally on w.

We consider a regular recentered IV estimator $\tilde{\beta} = \frac{1}{L}\tilde{z}'y / \frac{1}{L}\tilde{z}'x$, for $\tilde{z} = f(g)$ satisfying $\mathbb{E}[\tilde{z}] = 0$ (suppressing dependence on w throughout), that converges at rate r_L and has asymptotic first-stage M and asymptotic variance V. The asymptotic variance of $\frac{1}{L}\tilde{z}'\varepsilon$ is thus $\tilde{V} = M^2 V$. We consider the following condition:

Assumption A4. *Let $T(g^*, \varepsilon) = r_L \frac{1}{L} f(g^*)' \varepsilon$. For g_1^* and g_2^* distributed according to $G(\cdot)$, with g_1^*, g_2^*, and ε mutually independent, $(T(g_1^*, \varepsilon), T(g_2^*, \varepsilon)) \xrightarrow{d} \left(\sqrt{\tilde{V}} Z_1, \sqrt{\tilde{V}} Z_2\right)$, where Z_1 and Z_2 are independent standard normal variables.*

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

Proposition A5. *Suppose the assumptions of Proposition A4 hold, along with Assumption A4. Fix $\alpha \in (0,1)$ and $\delta \neq 0$. Then the limiting power of an RI test of size α based on $T(g, y - b_L x) = r_L \frac{1}{L} f(g^*)' (y - b_L x)$, against a sequence of local alternatives $b_L = \beta - \delta / r_L$, is a decreasing function of only the recentered IV estimator's asymptotic variance, V.*

Proof. See Appendix C.7. □

A.4 Potential Outcomes and Heterogeneous Treatment Effects

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

With g viewed as the source of the natural experiment, we define potential treatments and outcomes as $x_\ell = x_\ell(g, w, u)$ and $y_\ell = y_\ell(g, w, \varepsilon, u)$, where u and ε capture sources of unobserved first- and second-stage heterogeneity, respectively. We do not require that the functions $x_\ell(\cdot)$ and $y_\ell(\cdot)$ are known. We can now formalize the exclusion restriction:

Assumption A5. *(Exclusion): $y_\ell(g, w, \varepsilon, u) = y_\ell(x_\ell(g, w, u), w, \varepsilon)$ almost-surely.*

Given exclusion only, we define marginal treatment effects as $\beta_\ell(x, w, \varepsilon) = \frac{\partial}{\partial x} y_\ell(x, w, \varepsilon)$. Here for notational simplicity we assumed that x_ℓ is continuous and of full support, with $y_\ell(x, w, \varepsilon)$ differentiable in x, though below and in Appendix C.8 we show that these are both straightforward to relax. In contrast to the parametric model (2), we do not impose any restrictions on $\beta_\ell(\cdot)$, allowing for arbitrary heterogeneity (across w and ε) and nonlinearity in x.

We now have the following recentered IV identification result:

Proposition A6. *Suppose Assumptions A5 and 1 hold and $Pr(x_\ell \geq x \mid z_\ell = z, \varepsilon, w)$ is weakly increasing in z for each x almost-surely over (ε, w). Then the estimand of the recentered IV is*

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \int \beta_\ell(\gamma, \varepsilon)\omega_\ell(\gamma, \varepsilon) d\gamma\right], \tag{A5}$$

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

Proof. See Appendix C.8. □

Proposition A6 imposes a first-stage monotonicity condition: that x_ℓ is stochastically increasing in z_ℓ conditional on ε and w. This condition is substantially more general than conventional ones (e.g., Angrist et al. 2000). Conventional monotonicity specifies a causal and monotone relationship between the treatment and the instrument: i.e., $x_\ell = x_\ell(z_\ell, \eta)$ with $z \perp\!\!\!\perp \eta$ and $\frac{\partial}{\partial z} x_\ell(z, \eta) \geq 0$ almost-surely. This is sufficient for our stochastic monotonicity (with η included in the list of unobservables ε, which is without loss of generality). However, our condition also applies to settings where the shocks g affect many observations of z_ℓ and x_ℓ jointly and differentially, such that a causal first stage does not exist. For example, in the linear shift share case of $z_\ell = \sum_n w_{\ell n} g_n$, we may suppose that the shares are partially misspecified, such that $x_\ell = \sum_n \pi_{\ell n} g_n + \eta_\ell$ for unobserved $(\pi, \eta) \perp\!\!\!\perp g \mid w$. Proposition A6 shows that the recentered IV regression remains causal in this case provided x_ℓ is stochastically increasing in z_ℓ conditional on ε and w. This holds, for example, when the $w_{\ell n}$ and $\pi_{\ell n}$ are almost-surely non-negative and the g_n are mutually independent; Proposition A6 can thus be seen to generalize a monotonicity condition for shift-share IV established by Borusyak et al. (2019).

Proposition A6 shows that the recentered IV combines heterogeneous treatment effects with an intuitive convex weighting scheme under this monotonicity condition, exclusion, and shock exogeneity. Appendix C.8 shows that the weights $\omega_\ell(\gamma, \varepsilon)$ are proportional to the conditional-on-(ε, w) covariance of z_ℓ and $\mathbf{1}[x_\ell > \gamma]$. Thus, the recentered IV gives more weight to treatment effects $\beta_\ell(\gamma, \varepsilon)$ at margins γ with a larger first-stage response to the IV, given ε and w.

At the same time, Proposition A6 shows that even in conventional treatment effect settings this weighting scheme may not identify the most policy-relevant average of causal effects, such as the overall average treatment effect (ATE) or local average treatment effect (LATE). To show this simply, consider the case wherex_ℓ and z_ℓ are binary, such that there are two potential outcomes, $y_\ell(0,\varepsilon)$ and $y_\ell(1,\varepsilon)$, with each observation having a single heterogeneous treatment effect of $\beta_\ell(\varepsilon) = y_\ell(1,\varepsilon) - y_\ell(0,\varepsilon)$. We further adopt a causal first stage relationship, writing $x_\ell = x_\ell(0)(1 - z_\ell) + x_\ell(1)z_\ell$ with the potential treatments $(x_\ell(0), x_\ell(1))$ included in ε without loss. A version of Proposition A6 adapted to this setting is as follows:

Proposition A7. *Suppose Assumptions A5 and 1 holds, that* x_ℓ *and* z_ℓ *are binary, and that* $p_\ell = Pr(x_\ell(1) > x_\ell(0) \mid w)$ *is almost-surely non-negative. Then the estimand of the recentered IV is*

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\beta_\ell(\varepsilon) \mid x_\ell(1) > x_\ell(0), w\right]\tilde{\omega}_\ell\right], \tag{A6}$$

where $\tilde{\omega}_\ell$ *gives a convex weighting and is proportional to* $p_\ell \sigma_\ell^2$*, where* $\sigma_\ell^2 = \operatorname{Var}\left[z_\ell \mid w\right]$.

Proof. See Appendix C.8. □

Proposition A7 shows that in this case the recentered IV identifies a weighted average of conditional treatment effects for "compliers" (defined by $x_\ell(1) > x_\ell(0)$), with weights given by the conditional variance of the instrument, σ_ℓ^2. In the reduced form special case, where $x_\ell = z_\ell$ and $x_\ell(1) > x_\ell(0)$ by construction, this result shows that the recentered IV identifies a variance-weighted average of conditional treatment effects $\mathbb{E}\left[\beta_\ell(\varepsilon) \mid w\right]$. Only when the σ_ℓ^2 are uncorrelated with the heterogeneous causal effects will these weighted averages identify conventional LATEs or ATEs, respectively.

A general and practical solution to obtaining more traditional causal estimands is to "reweight" the recentered IV by its conditional variance σ_ℓ^2. Given an overlap condition of $\sigma_\ell^2 > 0$ for all ℓ, the resulting instrument $(z_\ell - \mu_\ell)/\sigma_\ell^2$ is well-defined and is still a recentered IV. Since the conditional variance of this instrument equals one by construction, it is furthermore immediate from Proposition A7 that then we obtain with this instrument a conventional ATE or LATE:

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

In practice, reweighting a binary z_ℓ is no more difficult than recentering it by μ_ℓ since $\sigma_\ell^2 = \mu_\ell(1 - \mu_\ell)$. More generally σ_ℓ^2 is given by the shock assignment process (i.e., Assumption 3).[58]

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

We finally note that IV inference may be challenging when treatment effects vary. For testing the so-called "sharp null" of $\beta_\ell(x,\varepsilon) = 0$, almost surely, the randomization-based tests in Section 3.4 still apply but may reject under the "weak null" of no average effect (i.e. that the estimand in Proposition A6 is zero). Inverting RI tests to form confidence intervals for β is also no longer sensible with heterogeneous effects. This issue is not specific to RI, as asymptotic inference may also be challenging in this case. For example in the linear shift-share setting, Adão et al. (2019) derive conservative asymptotic variance estimators only for a reduced-form estimator $\tilde{z}'y/\tilde{z}'x$, under strong conditions. Aronow and Samii (2017) similarly construct conservative asymptotic variance estimators in the network interference setting. We view generalizing these approaches as a potentially fruitful area for future research.

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 estimated and used for recentering. We then adapt the Berger and Boos (1994) approach to inference with nuisance parameters to build conservative finite-sample confidence intervals.

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

Valid, but likely quite conservative confidence intervals for β in such cases can be obtained by a simple extension of the previous randomization inference procedure. Given a value of θ, the randomization test for $\beta = b$ of Proposition A1 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).[59] However, these confidence intervals are likely to be quite wide: even if the observed g is very

[59] An equivalent view on this procedure is to test joint hypotheses $\beta = b$ and $\theta = \theta_0$ using the test of Proposition A1 and then project the resulting confidence interval on the space of β.

informative about the precise value of θ, this test still searches through values very far from $\hat{\theta}$.

We propose an alternative two-step approach following Berger and Boos (1994) that is likely to be much less conservative but still valid (see Ding et al. (2016) for another application of this idea to RI). In the first step, a confidence interval CI_θ for θ with coverage $1-\gamma$ is constructed for some $\gamma \in (0, \alpha)$; Berger and Boos (1994) recommend $\gamma = 0.001$. Such tests are easy to build since the distribution of g is fully specified given θ; thus an exact RI-based confidence interval for θ can be constructed from any statistic $S = \mathcal{S}(g; w, \theta_0)$ by rerandomizing g according to $G(\cdot; w, \theta_0)$. As usual, the choice of S determines the power of the test and the length of the confidence interval. We propose a statistic that corresponds to the score test, $S = \frac{\partial}{\partial\theta} \log G(g; w, \theta_0)$, since the Hodges-Lehmann estimator induced by it is the MLE.[60] For vector-valued θ, S can be converted to a scalar LM statistic $S' \mathbb{E}_{\theta_0} \left[SS' \mid w\right]^{-1} S$; a value θ_0 is rejected if the LM statistic is in the right tail of its distribution. In the second step, the maximum p-value of the Proposition A1 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 A8.

(i) *Suppose Assumptions 1 holds, $\hat{\theta}$ is consistent for θ, and $\mu_\ell(\theta_0, w)$ is almost-surely differentiable with respect to θ_0 in a convex parameter space Θ and with a bounded gradient $\frac{\partial \mu_\ell}{\partial \theta}$. Then when Assumptions A1-A3 hold at the true value of θ, and the sequences $\frac{1}{L}\sum_\ell |x_\ell|$ and $\frac{1}{L}\sum_\ell |\varepsilon_\ell|$ are bounded in probability, the plug-in recentered IV estimator with instrument $\hat{z}_\ell$ is consistent.*

(ii) *Suppose Assumption 1 holds. Let $p_\beta(\beta; \theta_0)$ be the p-value of the randomization test of Proposition A1 for a given value of θ and let CI_θ denote a confidence interval for θ such that $Pr(\theta \in CI_\theta) \geq 1 - \gamma$ for $\gamma < \alpha$. Construct $CI_\beta = \{b \in \mathbb{R}\colon\ \max_{\theta_0 \in CI_\theta} p_\beta(\beta, \theta_0) + \gamma > \alpha\}$. Then CI_β is conservative for β, i.e. $Pr(\beta \in CI_\beta) \geq 1 - \alpha$.*

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

[60] 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.

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. [61] 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 $\text{Cov}\left[z_\ell, \mu_\ell(\theta, w)\right] = \text{Var}\left[\mu_\ell(\theta, w)\right]$ by definition of $\mu_\ell(\theta, w)$, 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).[62]

Finally, a closely related way to incorporate θ is by assuming that some one-to-one transformation of shocks $\tilde{g} = h(g; w, \theta)$ 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 $\tilde{g}_n = g_n - \rho_n(\theta, w)$ is exchangeable across n. Here the expression for the mean $\rho_n(\theta, w)$ 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 $g_n = \zeta + \rho_n(\theta, w) + u_n$, which is consistent as N grows under standard assumptions given conditionally mutually independent u_n. It is then straightforward to verify that this is the Hodges-Lehmann estimator corresponding to the RI statistic $T_\theta = \frac{1}{N}\sum_n \tilde{g}_n \frac{\partial \rho_n}{\partial \theta}$. Therefore, one may use this statistic to construct an exact confidence interval for θ. In the second step, the expected instrument given θ is constructed by the following simulation: $\tilde{g}_n$ are randomly permuted to get $\tilde{g}_n^*$ and $g_n^* = \rho_n(\theta, w) + \tilde{g}_n^*$ 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 $g_n^2 = \zeta^2\sigma_n^2(\theta, w) + u_n$.[63] With an estimate of θ,

[61] 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, $G(g \mid w)$ is the same for any binary vector g that yields the same vector $\sum_n g_n r_n$.

[62] At the same time, including $\mu_\ell(\theta, w)$ 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.

[63] 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.

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 Efficiency Controls

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

The following result extends Propositions 1, A1, 2, and A2(i) to the case of efficiency controls:

Proposition A9. *Suppose $g \perp\!\!\!\perp (a, \varepsilon) \mid w$ where a collects the $a_\ell = (a_{\ell 1}, \dots, a_{\ell r})$. Let $v_\ell^\perp$ denote the sample projection of a variable v_ℓ on a_ℓ: i.e., $v_\ell^\perp = v_\ell - \hat{\alpha}_v' a_\ell$ for $\hat{\alpha}_v = \left(\frac{1}{L}\sum_\ell a_\ell a_\ell'\right)^- \frac{1}{L}\sum_\ell a_\ell v_\ell$ and $(\cdot)^-$ denoting a generalized inverse of a matrix. Then:*

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

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

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

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

(v) *$\tilde{\beta}_\perp \xrightarrow{p} \beta$ if Assumptions 3 and A1–A3 hold, $\mathbb{E}\left[a_{\ell r}^2 \mid w\right] \leq B_a$ almost surely for all ℓ and $r = 1, \dots, R$, $\frac{1}{L}\sum_\ell a_\ell a_\ell'$ 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.10.

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

A.7 Multiple Treatments and Instruments

This appendix considers the case when the outcome equation includes several endogenous variables. For example, in network spillover regressions of Section 4.3 the researcher may specify both a direct effect of the shock to the unit and the spillover effect from other units. We show that the main results of the paper apply in that case: instrument recentering restores instrument validity, and randomization inference yields a joint confidence interval for the coefficient vector. We then discuss several special cases where separate confidence intervals may be obtained for individual coefficients, more efficiently than by projecting the joint interval. Notably, one set of shocks arising from the natural experiment is generally sufficient to identify multiple causal effects, as long as endogenous variables differ in their exposure to the same shocks. Finally, we discuss how most of the results in our framework generalize to the overidentified case, with multiple instruments.

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

$$y_\ell = \boldsymbol{\beta}'\boldsymbol{x}_\ell + \varepsilon_\ell, \tag{A7}$$

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

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

Now consider randomization inference. As before, the distribution of any scalar or vector-valued statistic $\mathcal{T}(g, y - \boldsymbol{x}\boldsymbol{b}, w)$, where $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L)'$, is known conditional on w and ε under the null of $\boldsymbol{\beta} = \boldsymbol{b}$, which can be used to construct valid tests and confidence intervals for $\boldsymbol{\beta}$ as in Proposition A1. The only complication here is that the natural choice of test statistic that extends Proposition 2, $T = \frac{1}{L}\sum_\ell \tilde{\boldsymbol{z}}_\ell'(y_\ell - \boldsymbol{b}'\boldsymbol{x}_\ell)$, is vector-valued and requires to pick a rejection region in $\mathbb{R}^M$. A natural approach is to map T into the scalar Lagrange Multiplier (LM) statistic $T_{LM} = TV(\boldsymbol{b})^{-1}T$, where $V(\boldsymbol{b})$ is the randomization variance matrix of T that imposes the null $\boldsymbol{\beta} = \boldsymbol{b}$. That is, V is computed by re-randomizing g according to $G(g \mid w)$ while holding $\varepsilon_\ell = y_\ell - \boldsymbol{b}'\boldsymbol{x}_\ell$ fixed. The null $\boldsymbol{\beta} = \boldsymbol{b}$ is then rejected when T_{LM} exceeds its $1-\alpha$ randomization quantile, which is equivalent to T being outside a particular ellipsoid in $\mathbb{R}^M$ centered at zero. This test follows the Hodges-Lehmann approach: the RI p-value is maximized at the recentered IV estimator at which $T = 0$ and thus $T_{LM} = 0$.[64] The exact joint confidence interval for $\boldsymbol{\beta}$ is constructed by inverting this test, as usual.

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

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

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

Second, if a confidence interval for one coefficient is obtained (e.g. in one of the situations discussed above), the approach of Berger and Boos (1994) and Ding et al. (2016) can be used to get conservative marginal confidence intervals for the remaining coefficient. Specifically, let CI_2 be an exact interval for β_2 with coverage $1-\gamma$ for some $\gamma \in (0, \alpha)$, e.g. $\gamma = 0.001$. Then $\beta_1 = b_1$ is rejected if $\boldsymbol{\beta} = (b_1, b_2)$ is rejected by the RI-based LM test for every $b_2 \in CI_2$ at significance level $\alpha - \gamma$. In other words, the p-value of the test for β_1 is the maximum p-value of the joint test across $b_2 \in CI_2$, plus γ. When $\gamma \to 0$, CI_2 becomes uninformative, and this procedure converges to the projection of the joint confidence interval. However, for a given γ CI_2 may be narrow in large samples, and taking the maximum across $b_2 \in CI_2$ rather than the entire real line may result in a much more powerful test for β_1.

We also conjecture that the following asymptotic approach may apply in many applications, though we leave a formal analysis of this approach to future research. One may expect under certain regularity conditions that some central limit theorem applies to $\frac{1}{\sqrt{L}} \sum_\ell \tilde{\boldsymbol{z}}_\ell \left(y_\ell - \boldsymbol{\beta}' \boldsymbol{x}_\ell\right)$, such that it converges to a jointly normal distribution. Moreover, under some conditions, this unconditional distribution may be well approximated by the randomization distribution across g only (see Lehmann (1986), Theorem

15.2.3). Furthermore, estimating this distribution at a consistent estimate $\hat{\beta}$ instead of β may be asymptotically innocuous (see Shaikh and Toulis (2019)). In such cases, $\hat{\boldsymbol{\beta}}$ is asymptotically normal and an asymptotically valid confidence interval for each coefficient separately can easily be obtained by delta method, using the randomization variance matrix of $\frac{1}{\sqrt{L}}\sum_\ell \tilde{\boldsymbol{z}}_\ell \left(y_\ell - \hat{\boldsymbol{\beta}}'\boldsymbol{x}_\ell\right)$ as an estimate of the asymptotic variance of $\frac{1}{\sqrt{L}}\sum_\ell \tilde{\boldsymbol{z}}_\ell \varepsilon_\ell$.

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

A.8 Nonlinear Outcome Models

This appendix considers settings where the parameter of interest is specified in terms of a nonlinear model:

$$y_\ell = m_\ell(x;\beta) + \varepsilon_\ell, \tag{A8}$$

where $\{m_\ell(\cdot)\}_{\ell=1}^L$ is a set of known functions and x includes an unrestricted set of observables. We show that our results on identification, inference, and asymptotic efficiency generalize naturally to this setup.

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

$$\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \left(y_\ell - m_\ell(x;\beta)\right)\right] = 0, \tag{A9}$$

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

Valid finite-sample inference on β is similarly obtained as in the linear case. The test statistic which induces as a Hodges-Lehmann estimator the solution to the sample analog of (A9) is

$$T = \frac{1}{L}\sum_{\ell}\left(f_\ell(g,w)-\mu_\ell\right)\left(y_\ell - m_\ell(x;b)\right), \tag{A10}$$

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

Derivation of the efficient instrument also follows similarly. As in Proposition A4, we consider the class of recentered instruments yielding IV estimators that converge at some rate. Given analogous regularity conditions, it is straightforward to verify that the asymptotically variance-minimizing instrument in this class is

$$z^* = \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1}\left(\mathbb{E}\left[\frac{\partial}{\partial\beta}m(x;\beta)\mid g,w\right] - \mathbb{E}\left[\frac{\partial}{\partial\beta}m(x;\beta)\mid w\right]\right), \tag{A11}$$

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

A.9 Identification with Exogenous Exposure

In this section we discuss identification when shock exposure is exogenous but recentering of z_ℓ is not possible either because shocks are endogenous or because the shock assignment process is not known. We first present a high-level condition and then discuss a more intuitive condition for the linear shift-share setting. We connect both conditions to identification in market access regressions with endogenous transportation upgrades.

We formalize exogeneity of shock exposure by $\mathbb{E}\left[\varepsilon_\ell \mid w\right] = 0$, which makes any function of w a valid instrument for estimating β.[65] The instrument is however constructed as $z_\ell = f_\ell(g,w)$ which may depend on endogenous shocks. We allow for such endogeneity by letting $\mathbb{E}\left[\varepsilon_\ell \mid g,w\right] = \xi_\ell$ be non-zero, in violation of Assumption 1 (and, in particular, Assumption 2(i)).

We first consider two high-level assumptions:

Assumption A6. ***(i)*** $\mathbb{E}\left[\frac{1}{L}\sum_\ell \xi_\ell^2\right] \to 0$; *and* ***(ii)*** $\frac{1}{L}\sum_\ell z_\ell^2 \le B_z$ *almost-surely for fixed* B_z.

The first part of Assumption A6 requires that the endogeneity of shocks is asymptotically weak, in the mean-squared sense. Intuitively, this condition is satisfied when the number of shocks is small and the residuals are sufficiently mutually independent, such that g cannot have strong dependence

[65] A more general formulation is that $\mathbb{E}\left[\varepsilon_\ell \mid w\right]$ is known and can be used to recenter y_ℓ, as in the control functions approach. One might further assume $\mathbb{E}\left[\varepsilon_\ell \mid w\right] = w_\ell'\alpha$ for unknown α and some observed exposure variables w_ℓ, such that controlling for w_ℓ identifies β.

with many ε_ℓ at once. The second part of Assumption A6 is a regularity condition on z_ℓ that can be straightforwardly relaxed.

We then have the following result:

Lemma A2. *If Assumption A6 holds,*

$$\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] \to 0, \tag{A12}$$

such that β is identified by the instrument z_ℓ in large samples, provided $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell\right] \to M \neq 0$.

Proof. See Appendix C.11. □

Here large-sample identification follows in the sense of the IV moment restriction $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell (y_\ell - b x_\ell)\right] = 0$ having a limiting solution of β as $L \to \infty$.

Since the key Assumption A6(i) is high-level, we complement it with lower-level conditions sufficient for exposure share exogeneity to identify β with linear shift-share instruments. The result generalizes that of Goldsmith-Pinkham et al. (2020), which is obtained when the shocks are non-stochastic and their number does not grow with the sample size.

Consider a shift-share instrument $z_\ell = \sum_n w_{\ell n} g_n$ with $w_{\ell n} \geq 0$. Let $w_n = \frac{1}{L}\sum_\ell w_{\ell n}$ denote the average exposure to the nth shock and $\bar{\varepsilon}_n = \frac{\sum_\ell w_{\ell n}\varepsilon_\ell}{\sum_\ell w_{\ell n}}$ be the average residual of observations exposed to that shock. Borusyak et al. (2019) establish a simple representation for the sample covariance between the instrument and the residual at the shock level:

$$\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell = \sum_n w_n g_n \bar{\varepsilon}_n. \tag{A13}$$

We impose the following lower-level analog of Assumption A6:

Assumption A7. ***(i)*** $\mathbb{E}\left[\sum_n w_n \bar{\varepsilon}_n^2 / \sum_n w_n\right] \to 0$ *and* ***(ii)*** $\sum_n w_n g_n^2 \leq B_g$ *and* $\sum_n w_n \leq B_w$ *almost-surely for fixed* B_g *and* B_w.

The substantive first part of this assumption requires the $\bar{\varepsilon}_n$ to be sufficiently close to zero in the mean-squared sense. Since $\mathbb{E}\left[\bar{\varepsilon}_n\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_\ell w_{\ell n}\varepsilon_\ell}{\sum_\ell w_{\ell n}} \mid w\right]\right] = 0$ for all n, this assumption (like Assumption A6(i)) requires that the law of large numbers applies to $\bar{\varepsilon}_n$, i.e. that there are sufficiently many observations exposed to each given shock and their residuals are sufficiently independent. This condition holds, for example, when the number of shocks does not grow with L, and exposure is well-balanced across them. Under this assumption we show that the shift-share IV estimator identifies β without strong substantive assumptions on the shocks (specifically, with a regularity condition in Assumption A6(ii) only that can be further relaxed):

Lemma A3. *If Assumption A7 holds, then* $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] \to 0$.

Proof. See Appendix C.11. □

In the market access setting, ξ_ℓ may be non-zero when line placement g is strategic (i.e., dependent on the productivity shocks ε_ℓ) even when geography w is exogenous in the sense of $\mathbb{E}[\varepsilon_\ell \mid w] = 0$. Lemma A2 shows how conventional market access regressions can be nevertheless identified in large samples provided this endogeneity is not too strong. Lemma A3 can be used to build intuition for this condition by taking a first-order approximation of the nonlinear market access measure in the set of potential upgrades: $\Delta \log MA_\ell \approx \sum_n w_{\ell n} g_n$ where $w_{\ell n} \geq 0$ captures the predicted increase in market access in region ℓ when only line n is constructed. Assumption A7 may then hold when each line affects market access of many regions with sufficiently uncorrelated residuals, such that the average productivity shock of lines exposed to each given line (or at least most of them) is close to zero. This assumption is likely to hold when there is a small number of lines which are long and cross diverse regions. It is more restrictive when lines are shorter and regions with similar unobservables.

A.10 Recentering in General Simulated Instrument Settings

This appendix discusses extensions to our baseline approach to simulated eligibility instruments, presented in Section 4.2. We further discuss advantages of our recentered IV relative to a controlling strategy used in the literature estimating eligibility effects of unemployment insurance.

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

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

Kuka (2015) use a similar approach to augment simulated instrument construction in evaluating the effects of unemployment insurance eligibility.

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

A.11 Recentering Helps with Consistency: An Example

We provide a minimal statistical example to illustrate the general idea that recentering an instrument substantially weakens the conditions for estimator consistency.

Suppose a sample of observations ℓ is available, and a reduced-form effect β of treatment $x_\ell = z_\ell$ on outcome $y_\ell = \beta z_\ell + \varepsilon_\ell$ is of interest (ignoring the constant for simplicity). Let

$$z_\ell = w_\ell + g_\ell \tag{A14}$$

with w_ℓ and g_ℓ are independent $N(0,1)$ variables. Suppose there is a macroeconomic shock $\nu = \pm 1$ (with equal probabilities) that determines whether high-μ_ℓ observations have higher or lower residuals:

$$\varepsilon_\ell = \nu w_\ell. \tag{A15}$$

All of ν, w_ℓ, and g_ℓ are jointly independent.

Then the intuitive definition of validity of the non-recentered IV z_ℓ holds:

$$\mathbb{E}\left[\varepsilon_\ell \mid z_\ell\right] = \mathbb{E}\left[w_\ell \nu \mid z_\ell\right] = \mathbb{E}\left[\nu \mid z_\ell\right] \mathbb{E}\left[w_\ell \mid z_\ell\right] = 0. \tag{A16}$$

Yet, the corresponding OLS estimator is inconsistent (although unbiased) because the impact of the ν shock does not vanish as $L \to \infty$:

$$\frac{\sum_\ell y_\ell z_\ell}{\sum_\ell z_\ell^2} = \beta + \frac{\frac{1}{L}\sum_\ell \left(w_\ell + g_\ell\right) w_\ell \cdot \nu}{\frac{1}{L}\sum_\ell z_\ell^2} = \beta + \frac{1}{2}\nu + o_p(1). \tag{A17}$$

On the other hand, the recentered IV estimator does not suffer from this problem. The expected instrument here $\mu_\ell = \mathbb{E}[z_\ell \mid w_\ell] = w_\ell$, this the recentered IV is $\tilde{z}_\ell = z_\ell - \mu_\ell = g_\ell$. The corresponding IV estimator is now consistent, as guaranteed by Proposition A2:

$$\frac{\sum_\ell y_\ell \tilde{z}_\ell}{\sum_\ell z_\ell \tilde{z}_\ell} = \beta + \frac{\frac{1}{L}\sum_\ell g_\ell \varepsilon_\ell \cdot \nu}{\frac{1}{L}\sum_\ell g_\ell (w_\ell + g_\ell)} = \beta + o_p(1) \cdot \nu = \beta + o_p(1). \tag{A18}$$

B Empirical Appendix

B.1 Data for Section 4.1

Our analysis of market access effects uses data on 340 prefectures of mainland China. This excludes the islands of Hainan and Taiwan and the special administrative regions of Hong Kong and Macau. At the same time this includes six sub-prefecture-level cities (e.g. Shihezi) that do not belong to any prefecture. We use United Nations shapefiles to geocode each prefecture by the location of its main city (or, in a few cases, by the prefecture centroid).[66]

We use a variety of sources to assemble a comprehensive database of the HSR network in each year between 2003 (when the first new line was completed) and 2016. Our starting points are Map 1.2 of Lawrence et al. (2019), China Railway Yearbooks, and the replication files of Lin (2017). We cross-check network links across these sources and use Internet resources such as Wikipedia and Baidu Baike to confirm and fill in missing information. Our database includes various types of HSR lines, including the National HSR Grid (4+4 and 8+8) and high-speed intercity railways. However, we only consider newly built HSR lines, excluding traditional lines upgraded to higher speeds. We do not put further restrictions on the class of trains (e.g. to G- and D-classes only) or specify an explicit minimum speed. The operating speed therefore ranges between 160 and 380kph, although the majority of lines are at 250kph. For each line we collect the date of its official opening, the date when construction began, its operating speed, and the list of stops (attributed to prefectures). When different sections of the same line opened in a staggered way, we classify each section as a separate line for the purposes of constructing our 999 counterfactuals, following the definition of a line in footnote 24. We include only one contiguous stop per prefecture and drop lines that do not cross prefecture borders.

To measure market access according to the formula given in the text, we compute travel time between all pairs of cities k and ℓ as of the end of each year $t = 2003, \dots, 2016$ for both the actual and counterfactual networks. Travel time combines traditional modes of transportation (car or low-speed train) with HSR, where available. We allow for unlimited changes between different HSR lines and between HSR and traditional modes without a layover penalty, as HSR trains tend to operate frequently and traditional modes also involve downtime. Following the existing literature, we proxy for travel time by traditional modes by the straight-line distance, and specify the speed

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

of $100 = 120/1.2$kph, where 120kph is their typical speed and the 1.2 adjustment for actual routes that are longer than a straight line. For two prefectures connected by an HSR line, we compute the distance along the line as the sum of straight-line distances between adjacent prefectures on the line. We use the operating speed of each line divided by an adjustment factor of 1.3 to capture the fact that the average speed is lower than the nominal speed we record. Computing market access further requires the population of each of the 340 prefectures from the 2000 population Census, which we obtain from the CityPopulation.de website.[67]

We measure prefecture employment in the 2000–2017 China City Yearbooks.[68] Each yearbook covers the previous year (so our data cover 1999–2016). The yearbooks provide most variables for two spatial definitions: the entire prefecture and the "urban district" (Shixiaqu), which is the main urban area of the prefecture; we use the former in the main analysis but also collect the latter for the robustness analysis. The employment variables we describe below measure urban employment, but they are still measured both for the main urban district as well as the for the entire prefecture which may include other urban areas. The main data in the yearbooks are reported at the prefecture level but some urban district variables are also provided for county-level cities—a finer administrative division. We use county-level city data to complete some missing data in the prefecture-level variables where possible; this however does not apply to our main variable as it is not for urban districts.

Total urban employment data come from two chapters of the Yearbooks: "People's Living Conditions and Social Security" and "Population, Labor Force, and Land Resources." The economic difference between them is not entirely clear. We use the former one, labeled "The Average Number of Staff and Workers", as its whole-prefecture version has by far the lowest number of strong year-to-year deviations which may indicate data quality issues. The other variable, "Persons Employed in Various Units at Year End", is used for robustness checks in Appendix Table A1, together with the urban district versions of both variables. In that table we further use a measure of total rail ridership originating in the prefecture, which is only available until 2014; for this analysis we thus use the 2007–2014 change in ridership instead of 2007–2016 as elsewhere.

We finally apply a data cleaning procedure to all outcome variables used in the analysis. We first mark a prefecture-year observation as a one-off jump, and replace it with a missing value, if (*i*) the variable changes by more than twice in either direction relative to the previous non-missing value for the prefecture, (*ii*) it is followed by a change in the opposite direction that is at least 75% as large (in terms of log-changes), and (*iii*) the previous value has not been marked as a jump. We then mark an observation as a sustained change if condition (*i*) is satisfied but (*ii*) is not. We view the outcome change between 2007 and 2016 as valid only if neither 2007 nor 2006 are marked as jumps and there

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

[68] Data for 2000-2015, excluding 2009 and 2011, are from http://oversea.cnki.net.proxy.uchicago.edu/kns55/default.aspx (accessed on January 23, 2019 via a University of Chicago portal). Data from 2009, 2011, 2016, and 2017, are from http://tongji.oversea.cnki.net/chn/navi/HomePage.aspx?id=N2018050234&name=YZGCA (accessed on January 23, 2019). We checked that these two sources agreed in years where both were available.

are no sustained changes in any year in between. For the main outcome variable this reduces the sample from 282 to the final set of 274 prefectures, but for other outcomes the sample reduction is more substantial.

B.2 Data for Section 4.2

Our application to simulated eligibility instruments uses a repeated cross-section of annual data from the American Community Survey (ACS; Ruggles et al. 2020). Our baseline estimation uses a representative 1% sample of individuals from 2013 and 2014 and we use the analogous 1% sample from 2012 to explore pre-trends. We restrict the sample to non-disabled adults (aged 21-64) residing in one of the 43 states eligible for Medicaid expansion under the ACA. To define this sample of states we follow Frean et al. (2017) in excluding "early expansion" states which had previously expanded Medicaid before 2013, as well as Massachusetts and Vermont who had previously made all adults with household income less than 138% FPL eligible. We also follow Frean et al. (2017) in designating 19 of these states as having expanded under the ACA in 2014, with 24 not expanding.[69]

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

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

Our recentered IV is constructed by predicting the actual Medicaid eligibility of each individual. In 2013 we use actual 2013 eligibility policies, again following Frean et al. (2017). In 2014 we predict eligibility by combining information on the 2013 policies and a state's decision to expand. An individual is eligible for Medicaid in 2014 if either she was eligible under the 2013 policies of her state (whether or not the state expanded eligibility) or if her household income is below 138% FPL and her

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

state expanded eligibility under the ACA. To compute the expected instrument we identify individuals who would have been eligible in 2014 if their state expanded but not otherwise (the "Exposed Sample"). Outside of this sample the expected instrument in 2014 is simply the individual's actual 2014 eligibility, while inside this sample the expected instrument is the fraction of states which expanded conditional on the governor's party. The 2013 expected eligibility IV is actual 2013 eligibility. Political party affiliation of state governors is determined as of December 2013[70], and in all regressions we control for an indicator for state party affiliation (interacted with year indicators). In robustness checks we control for other time-interacted state characteristics, such as a state's 2012 median income or share insured under Medicaid (both obtained from the ACS).

B.3 Robustness Checks for Section 4.2

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

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

Appendix Table A2 shows that we obtain relatively small pre-trend estimates across all specifications, with similar coefficients obtained by conventional simulated IV (odd columns) and recentered IV (even columns). The same efficiency gains we document in Table 4 are found here, with significantly smaller 95% confidence intervals for the recentered IV (again obtained by a wild score bootstrap) which exclude zero for the take-up and crowd-out outcomes. We find no significant pre-trends in the employer-sponsored insurance outcome in columns 5 and 6.

Alternative Assignment Processes Second, we explore the robustness of our estimates to alternative assumptions on the shock assignment process. Specifically we allow a state's decision to expand Medicaid coverage to depend not only on the party of its governor (as in our baseline specification) but additionally on the state's 2012 median income and 2012 level of Medicaid coverage. We accomplish this by including a quadratic in these three state characteristics (including their interaction), interacted with year indicators, in the control vector $c_{\ell t}$. This allows the expected instrument to depend flexibly on these characteristics in the exposed sample. Appendix Table A3 shows that we obtain virtually identical estimates, standard errors, and 95% confidence intervals.

Alternative IV Implementations Third, we apply alternative IV estimators implied by our framework. Recall that in the even-numbered columns of Table 4 we restrict the sample to indi-

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

viduals whose individual characteristics make them exposed to the expansion natural experiment in 2014. A different approach is to recenter the IV $z_{\ell t}$ by (or control for) the expected instrument $\mu_{\ell t}$, while keeping the full sample of individuals. Appendix Table A4 reports estimates from this approach for the three outcomes of interest. Panel B, which includes demographic controls, again finds much narrower confidence intervals relative to the simulated eligibility instrument. However, excluding these controls in Panel A yields an intriguing pattern: confidence intervals for the recentered IV are much wider than those of the simulated instrument.

In this section we explain how a combination of two factors generates the discrepancy between panels A and B of Appendix Table A4. First, the regression residuals are strongly correlated with the indicator for an individual being exposed to the expansion experiment, which is not controlled for in this regression. Second, exogenous shocks are assigned at the level of states, which include both exposed and non-exposed individuals. This discussion reveals why the problem does not arise when focusing on the non-exposed sample or when appropriate controls are included. We further relate this problem to Step 3 of the optimal instrument construction in Section 3.5.

For clarity of the theoretical discussion, we simplify the setup. First, we suppose that a single 2014 cross-section is available and state fixed effects are not included; we correspondingly drop the t subscript throughout. Second, we assume states only change eligibility as prescribed by their expansion decision, i.e. $x_{\ell t} = z_{\ell t}$. Finally, we assume that state decisions to expand are independent with a known propensity $\mathbb{E}[g_n \mid w]$ (e.g., as a function of the state governor's party). Thus, the recentered expansion indicator $\tilde{g}_n = g_n - \mathbb{E}[g_n \mid w]$ can be computed without permutations.[71]

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

We now consider the variances of the two estimators, approximated as in the proof of Proposition A4: $\text{Var}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}^{CG}\varepsilon_{\ell}^{\perp}\right]/\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}^{CG}x_{\ell}^{\perp}\right]^2$ and $\text{Var}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\varepsilon_{\ell}^{\perp}\right]/\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}x_{\ell}^{\perp}\right]^2$, respectively, where $\perp$ denotes the in-sample projection residual on the control variables (including a constant). We focus our attention on the numerators of these expressions because the first-stage covariances in the denominator are asymptotically equivalent (and equal in finite samples without controls).[72] For simplicity of exposition we also consider an individual's state of residence s_{ℓ} as fixed. Letting $L_n =$

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

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

$\sum_\ell \mathbf{1}[s_\ell = n]$ denotes the (fixed) number of individuals in each state n, it can then be shown that

$$\frac{\text{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell^\perp\right]}{\text{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell^{CG} \varepsilon_\ell^\perp\right]} = \frac{\sum_n \left(\frac{L_n}{L}\right)^2 \text{Var}\left[\tilde{g}_n\right] \mathbb{E}\left[e_{SEIV,n}^2\right]}{\sum_n \left(\frac{L_n}{L}\right)^2 \text{Var}\left[\tilde{g}_n\right] \mathbb{E}\left[e_{CG,n}^2\right]}, \tag{A19}$$

where $e_{SEIV,n} = \frac{1}{L_n}\sum_{\ell:\ s_\ell = n} \varepsilon_\ell^\perp f_\ell$ denotes the sum of residuals of all *exposed* individuals in state n (normalized by L_n), while $e_{CG,n} = \sum_{\ell:\ s_\ell = n} \varepsilon_\ell^\perp$ averages over *all* observations in the state.[73]

Equation (A19) shows that the recentered IV delivers power gains relative to the simulated instrument approach whenever the normalized sum of residuals is closer to zero for a typical state, in the mean-squared sense, when restricting to exposed individuals. The restricted sum has fewer summands, working in favor of the recentered IV. If the expansion shocks were assigned at the individual level, without state clustering, this would guarantee that the recentered IV is more efficient (since $e_{SEIV,n} = e_{CG,n} = \varepsilon_\ell^\perp$ in that case).

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

$$\begin{aligned}
\mathbb{E}\left[e_{SEIV,n}\right] &= \mathbb{E}\left[\mathbb{E}\left[e_{SEIV,n} \mid w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L_n}\sum_{\ell:\ s_\ell = n} \mathbb{E}\left[\varepsilon_\ell^\perp f_\ell \mid w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L_n}\sum_{\ell:\ s_\ell = n} \mathbb{E}\left[\varepsilon_\ell^\perp \mid f_\ell = 1, w\right] f_\ell\right] \\
&= \alpha \cdot \mathbb{E}\left[\frac{\sum_{\ell:\ s_\ell = n} f_\ell}{L_n}\right] \neq 0.
\end{aligned} \tag{A20}$$

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

The predictions of this discussion are borne out in the data. In Panel C of Table A4 we verify that the confidence interval of recentered IV become dramatically narrowed with a single control of f_ℓ (interacted with the 2014 dummy appropriately for difference-in-differences).[74] Moreover, demographic controls in Panel B of Table A4 capture most of the variation in f_ℓ, delivering similar results. Our

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

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

recentered IV specifications in the main text, by restricting the sample to the exposed individuals, effectively control for state dummies interacted with f_ℓ and achieve the best efficiency properties.

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

Monte Carlo Simulation Finally, we verify large and pervasive power gains from using the recentered IV in a Monte Carlo study, in which the true causal effect and the shock assignment process are known. We draw 999 counterfactual state expansion decisions by choosing random sets of 8 Republican- and 11 Democratic-controlled states as expansion states and use these shocks to compute counterfactual instruments $\tilde{z}_{\ell t}^{CG}$ and $\tilde{z}_{\ell t}$. We do not specify a model of the first stage (i.e., which exact policies states would have implemented if they randomly changed their decision to adopt the ACA Medicaid expansion), instead imagining that states either expand to 138% FPL or keep their 2013 policy. We therefore use $\tilde{x}_{\ell t} = \tilde{z}_{\ell t}$ as the endogenous variable. Finally, for the Medicaid take-up and ESI crowd-out outcomes we take the second-stage residuals $\varepsilon_{\ell t}^*$ from columns 2 and 6 of Table 4, panel A. These outcomes are unrelated to the endogenous variable by design, corresponding to the true causal effect of zero for all individuals, while keeping the correlation structure from the actual data. With these generated data, we re-estimate equation (11) with the fixed effects and controls as in our baseline implementation in Panel A of Table 4. By design, both sets of estimates should be centered at the true effects of zero, while we expect the recentered IV procedure to systematically reject a larger set of alternative hypotheses.

Figure A5 first shows the simulated distribution of simulated and recentered IV estimates from this exercise. Both estimators are approximately unbiased, with both distributions in both panels centered around the true effects of zero. However, consistent with the dramatically shorter confidence intervals in Table 4, the distribution of recentered IV coefficients is dramatically tighter around this mean. The estimate standard deviation falls from 0.014 to 0.006 as we move from the simulated IV to recentered IV in Panel A, with a larger decline from 0.020 to 0.007 in Panel B. With minimal bias, these correspond to simulated root mean-squared error reductions of 58.5% and 66.5% with the recentered IV, respectively.

Figure A6 shows that these reductions in estimator variance translate to increased rejection rates of false null hypotheses for both outcomes, while also suggesting the wild bootstrap 95% confidence intervals in Table 4 have approximately correct size. Away from the true null hypothesis of zero the recentered IV power curve is much more steeply sloping, with uniformly higher rejection rates. With the Medicaid take-up outcome, for example, the recentered IV is found to reject coefficients outside the range of $[-0.018, 0.017]$ with probability of at least 0.8, while the simulated IV only has such high power outside a nearly three times as long range, of $[-0.042, 0.056]$. For the ESI crowd-out outcome this contrast in minimum detectable effects is even starker, at $[-0.022, 0.018]$ for the recentered IV versus $[-0.073, 0.051]$ for the simulated IV.

B.4 Data for Section 4.4

Our Monte Carlo simulations of linear SSIV size and power are based on a data-generating process that Borusyak et al. (2019) develop for the setting of Autor et al. (2013). The baseline process, used in Panel A of Figure 4, is calibrated to the IV estimates in column 3 of Table 4 in Borusyak et al. (2019) with a second and first stage of

$$y_{\ell t} = \beta x_{\ell t} + \gamma' r_{\ell t} + \varepsilon_{\ell t}, \tag{A21}$$

$$x_{\ell t} = \pi z_{\ell t} + \rho' r_{\ell t} + u_{\ell t}. \tag{A22}$$

The outcome $y_{\ell t}$ corresponds to the change in manufacturing employment as a fraction of the working-age population in U.S. commuting zone ℓ in decade t (either 1990-2000 or 2000-2007), the treatment $x_{\ell t}$ is a measure of regional import competition with China, and the shift-share instrument $z_{\ell t} = \sum_n s_{\ell nt} g_{nt}$ is constructed by combining the industry-level growth of China imports in eight developed economies g_{nt} with lagged regional employment weights of different industries $s_{\ell nt}$. The vector $r_{\ell t}$ includes the sum of lagged employment shares, interacted with period indicators, and other pre-treatment controls as described in Borusyak et al. (2019). The sum-of-share controls linearly span the expected instrument when the industry shocks have a common mean in each period, and without loss we demean g_{nt} by period. There are a total of 1,444 observations (722 commuting zones in two periods) and estimation is weighted by the start-of-period population of the commuting zone.

Each draw of the baseline simulation generates 1,444 new observations of $(y_{\ell t}, x_{\ell t}, z_{\ell t})$ by holding fixed the employment shares, pre-treatment controls, and estimated coefficients and residuals of equations (A21) and (A22) but redrawing the industry shocks g_{nt}. We generate new shocks from a wild bootstrap of $g^*_{nt} = g_{nt}\nu^*_{nt}$ by multiplying the original year-demeaned shocks by a standard normal ν^*_{nt}. This process preserves the heteroskedasticity of the shocks, and corresponds to the process in row (b) of Table C6 in Borusyak et al. (2019).

In Panel B of Figure 4 we modify the baseline process to reduce the number of shocks in each period, from $N = 397$ manufacturing SIC industries to 20 two-digit industries. This modification corresponds to the process in row (g) of Table C6 in Borusyak et al. (2019). We aggregate imports from China to the U.S. and either developed economies as well as the number of U.S. workers by manufacturing industry to construct the new g_{nt}, $z_{\ell t}$, and $x_{\ell t}$, as described in Appendix A.10 of Borusyak et al. (2019), holding fixed other variables. We then redraw shocks again by a wild bootstrap, In Panels C and D of Figure 4 we modify the baseline process to add treatment effect heterogeneity, by period and Census division. We use the original number of shocks but instead estimate versions of equations (A21) and (A22) which interact both $x_{\ell t}$ and $z_{\ell t}$ with period or division fixed effects. In Panel C the estimated second-stage effects are -0.491 for the 1990s and -0.225 for the 2000s, replicating Table C3 of Borusyak et al. (2019). In Panel D the estimated effects vary between -0.609 for the East North Central Census division and -0.135 for the West North Central division. We then generate data as before, with a wild bootstrap for shocks.

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

C Proofs of Propositions

C.1 Proof of Proposition 1

For the recentered IV regression,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{L}\sum_{\ell} \tilde{z}_{\ell}\varepsilon_{\ell}\right] &= \mathbb{E}\left[\frac{1}{L}\sum_{\ell} \tilde{z}_{\ell}\mathbb{E}\left[\varepsilon_{\ell} \mid g, w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_{\ell} \tilde{z}_{\ell}\mathbb{E}\left[\varepsilon_{\ell} \mid w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_{\ell} \mathbb{E}\left[\tilde{z}_{\ell} \mid w\right]\mathbb{E}\left[\varepsilon_{\ell} \mid w\right]\right] \\
&= 0. \qquad \text{(A23)}
\end{aligned}$$

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

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell} z_{\ell}\varepsilon_{\ell}^{\perp}\right] = \mathbb{E}\left[\frac{1}{L}\sum_{\ell} z_{\ell}y_{\ell}^{\perp}\right] - \beta \cdot \mathbb{E}\left[\frac{1}{L}\sum_{\ell} z_{\ell}x_{\ell}^{\perp}\right] = 0, \qquad \text{(A24)}$$

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,

$$\begin{aligned}\mathbb{E}\left[\varepsilon^\perp \mid g, w\right] &= (I - P_\mu)\,\mathbb{E}\left[\varepsilon \mid g, w\right] \\ &= (I - P_\mu)\,\mathbb{E}\left[\varepsilon \mid w\right] \\ &= \mathbb{E}\left[\varepsilon^\perp \mid w\right], \end{aligned} \tag{A25}$$

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

$$\begin{aligned}\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell^\perp\right] &= \mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell)\mathbb{E}\left[\varepsilon_\ell^\perp \mid g, w\right]\right] \\ &= \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\tilde{z}_\ell \mid w\right]\mathbb{E}\left[\varepsilon_\ell^\perp \mid w\right]\right] \\ &= 0, \end{aligned} \tag{A26}$$

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

C.2 Proof of Proposition 2

The Hodges-Lehmann estimator of interest solves:

$$\begin{aligned}\frac{1}{L}\sum_\ell f_\ell\left(g, w\right)\left(y_\ell - b x_\ell\right) &= \mathbb{E}\left[\frac{1}{L}\sum_\ell f_\ell\left(g^*, w\right)\left(y_\ell - b x_\ell\right) \mid w, y, x\right] \\ &= \frac{1}{L}\sum_\ell \mu_\ell\left(y_\ell - b x_\ell\right), \end{aligned} \tag{A27}$$

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

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \left(f_\ell(g, w) - \mu_\ell\right) y_\ell}{\frac{1}{L}\sum_\ell \left(f_\ell(g, w) - \mu_\ell\right) x_\ell}, \tag{A28}$$

which coincides with the recentered IV estimator.

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

$$\begin{aligned}\frac{1}{L}\sum_\ell f_\ell\left(g, w\right)\left(y_\ell^\perp - b x_\ell^\perp\right) &= \mathbb{E}\left[\frac{1}{L}\sum_\ell f_\ell\left(g^*, w\right)\left(y_\ell^\perp - b x_\ell^\perp\right) \mid w, y, x\right] \\ &= \frac{1}{L}\sum_\ell \mu_\ell\left(y_\ell^\perp - b x_\ell^\perp\right). \end{aligned} \tag{A29}$$

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

C.3 Proof of Proposition A4 and Lemma 3

Proposition A4 Consider some recentered IV $\tilde{z}$ associated with a regular estimator $\tilde{\beta}$ that converges at rate $\tilde{r}_L$ to an asymptotic distribution $\tilde{\mathcal{D}}$ with variance $\tilde{V}$. Uniform integrability of $\frac{1}{L}\tilde{z}'x$ implies that $\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right] \to M$. Then, by the continuous mapping theorem,

$$\tilde{r}_L \frac{\frac{1}{L}\tilde{z}'\varepsilon}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]} = \tilde{r}_L\left(\tilde{\beta}-\beta\right)\cdot\frac{\frac{1}{L}\tilde{z}'x}{M}\cdot\frac{M}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]} \Rightarrow \tilde{\mathcal{D}}, \tag{A30}$$

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

$$\mathrm{Var}\left[\tilde{r}_L\frac{\frac{1}{L}\tilde{z}'\varepsilon}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]}\right] = \tilde{r}_L^2\frac{\mathrm{Var}\left[\frac{1}{L}\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]^2} \to \tilde{V}. \tag{A31}$$

The same argument applies to β^*:

$$\mathrm{Var}\left[r_L^*\frac{\frac{1}{L}z^{*\prime}\varepsilon}{\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]}\right] = {r_L^*}^2\frac{\mathrm{Var}\left[\frac{1}{L}z^{*\prime}\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]^2} \to V^*, \tag{A32}$$

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

$$\frac{\tilde{r}_L^2/\tilde{V}}{{r_L^*}^2/V^*}\cdot\frac{\mathrm{Var}\left[\frac{1}{L}\tilde{z}'\varepsilon\right]/\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]^2}{\mathrm{Var}\left[\frac{1}{L}z^{*\prime}\varepsilon\right]/\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]^2} \to 1. \tag{A33}$$

We prove below that

$$\frac{\mathrm{Var}\left[\frac{1}{L}\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]^2} \geq \frac{\mathrm{Var}\left[\frac{1}{L}z^{*\prime}\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]^2}. \tag{A34}$$

whenever the denominators on both sides are not equal to zero (which holds for large enough L, since both $\tilde{z}$ and z^* have asymptotic first-stages). This concludes the proof, since (A33) and (A34) jointly imply that

$$\limsup_{L\to\infty}\frac{\tilde{r}_L^2/\tilde{V}}{{r_L^*}^2/V^*} \leq 1. \tag{A35}$$

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

To establish (A34), first note that by the law of iterated expectations and Assumption 2,

$$
\begin{aligned}
\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon' z^*\right] &= \mathbb{E}\left[\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon' z^* \mid g, w\right]\right] \\
&= \mathbb{E}\left[\tilde{z}'\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)\right] \\
&= \mathbb{E}\left[\tilde{z}'\mathbb{E}\left[x \mid g, w\right]\right] \\
&= \mathbb{E}\left[\tilde{z}'x\right], \qquad \text{(A36)}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\text{Var}\left[(z^*)'\varepsilon\right]}{\mathbb{E}\left[(z^*)'x\right]^2} &= \text{Var}\left[(z^*)'\varepsilon\right]^{-1} = \mathbb{E}\left[(z^*)'x\right]^{-1} \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1}\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)\right]^{-1}. \qquad \text{(A37)}
\end{aligned}
$$

It also shows that with

$$
U = \frac{\tilde{z}'\varepsilon}{\mathbb{E}\left[\tilde{z}'x\right]} - \frac{(z^*)'\varepsilon.}{\mathbb{E}\left[(z^*)'x\right]} \qquad \text{(A38)}
$$

we have

$$
\begin{aligned}
\frac{\text{Var}\left[\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} - \frac{\text{Var}\left[(z^*)'\varepsilon\right]}{\mathbb{E}\left[(z^*)'x\right]^2} &= \frac{\text{Var}\left[\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} - 2\frac{\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon' z^*\right]}{\mathbb{E}\left[\tilde{z}'x\right]\mathbb{E}\left[(z^*)'x\right]} + \frac{\text{Var}\left[(z^*)'\varepsilon\right]}{\mathbb{E}\left[(z^*)'x\right]^2} \\
&= \mathbb{E}\left[U^2\right] \\
&\geq 0, \qquad \text{(A39)}
\end{aligned}
$$

implying equation (A34).

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

$$
\left(\Omega + \psi\psi'\right)^{-1} = \Omega^{-1} - \Omega^{-1}\psi\frac{\psi'\Omega^{-1}}{1 + \psi'\Omega^{-1}\psi}. \qquad \text{(A40)}
$$

Thus,

$$
\begin{aligned}
z^* &= \left(\Omega + \psi\psi'\right)^{-1} \tilde{z} \\
&= \Omega^{-1}\left(\tilde{z} - \frac{\psi'\Omega^{-1}\tilde{z}}{1+\psi'\Omega^{-1}\psi}\psi\right) \\
&= \Omega^{-1}\left(\tilde{z} - \hat{\rho}\hat{\nu}\psi\right), \qquad \text{(A41)}
\end{aligned}
$$

as required.

C.4 Proof of Proposition A1

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

$$
Pr\left(T^* \in R \mid \varepsilon, w\right) = Pr\left(T^* \in R \mid y, x, w\right) \geq 1 - \alpha \qquad \text{(A42)}
$$

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

$$
Pr\left(T \in R \mid \varepsilon, w\right) = Pr\left(T^* \in R \mid \varepsilon, w\right) \geq 1 - \alpha. \qquad \text{(A43)}
$$

C.5 Proof of Proposition A2

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

$$
\begin{aligned}
\tilde{\beta} - \beta &= \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell} \\
&= \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{M}\left(1 + o_p(1)\right) \qquad \text{(A44)}
\end{aligned}
$$

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

$$\begin{aligned}
\text{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right] &= \mathbb{E}\left[\left(\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right)^2\right] \\
&= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \varepsilon_\ell \varepsilon_m\right] \\
&= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid w\right]\mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid w\right]\right] \\
&\le \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\left|\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid w\right]\right| \sqrt{\mathbb{E}\left[\varepsilon_\ell^2 \mid w\right]\mathbb{E}\left[\varepsilon_m^2 \mid w\right]}\right] \\
&\le B\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m} \left|\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0 \qquad \text{(A45)}
\end{aligned}$$

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

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

$$\begin{aligned}
T = \frac{1}{L}\sum_\ell \tilde{z}_\ell (y_\ell - bx_\ell) &= \frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell + (\beta - b)\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell \\
&\xrightarrow{p} (\beta - b)\, M \neq 0. \qquad \text{(A46)}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}^*_\ell (y_\ell - bx_\ell)\right] &= \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\tilde{z}^*_\ell (y_\ell - bx_\ell) \mid w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\tilde{z}^*_\ell \mid w\right]\mathbb{E}\left[y_\ell - bx_\ell \mid w_\ell\right]\right] \\
&= 0 \qquad \text{(A47)}
\end{aligned}$$

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

$$\begin{aligned}
\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}^*_\ell(y_\ell - bx_\ell)\right] &= \mathbb{E}\left[\left(\frac{1}{L}\sum_\ell \tilde{z}^*_\ell\left(y_\ell - bx_\ell\right)\right)^2\right] \\
&= \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\mathbb{E}\left[\tilde{z}^*_\ell \tilde{z}^*_m \mid w\right]\mathbb{E}\left[\left(y_\ell - bx_\ell\right)\left(y_m - bx_m\right) \mid w\right]\right] \\
&\leq \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\left|\mathbb{E}\left[\tilde{z}^*_\ell \tilde{z}^*_m \mid w\right]\right|\sqrt{\mathbb{E}\left[\left(y_\ell - bx_\ell\right)^2 \mid w\right]\mathbb{E}\left[\left(y_m - bx_m\right)^2 \mid w\right]}\right] \\
&\leq C(b)\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0, \qquad \text{(A48)}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[\left(y_\ell - bx_\ell\right)^2 \mid w\right] &= \mathbb{E}\left[\varepsilon^2_\ell + 2\left(\beta - b\right)x_\ell\varepsilon_\ell + \left(\beta - b\right)^2 x^2_\ell \mid w\right] \\
&\leq B + 2\left|\beta - b\right| \cdot \left|\mathbb{E}\left[x_\ell\varepsilon_\ell \mid w\right]\right| + \left(\beta - b\right)^2 \mathbb{E}\left[x^2_\ell \mid w\right]. \qquad \text{(A49)}
\end{aligned}$$

Proof of Lemma A1(i). For the first statement of the lemma, we have

$$\begin{aligned}
\frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] &= \sum_{\ell,m}\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right] \\
&= \mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \mid w\right]\right] \\
&= \operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\right] \to 0, \qquad \text{(A50)}
\end{aligned}$$

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 C1. *If $h \colon \mathbb{R}^N \to \mathbb{R}$ is monotone and random variables $g_1, \dots, g_N$ are independent, then for any $k \in \{1, \dots, N-1\}$ the conditional expectation $\mathbb{E}\left[h\left(g_1, \dots, g_N \mid g_1, \dots g_k\right)\right]$ is monotone.*

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

$$\begin{aligned}\mathbb{E}\left[h\left(g' \mid g_1, \ldots g_k\right)\right] &= \int \cdots \int h\left(g'\right) dG_{k+1}(g_{k+1}) \ldots dG_N(g_N) \\ &\geq \int \cdots \int h\left(g\right) dG_{k+1}(g_{k+1}) \ldots dG_N(g_N) \\ &= \mathbb{E}\left[h(g \mid g_1, \ldots g_k)\right], \end{aligned} \tag{A51}$$

as required.

Lemma C2. *For any monotone* $h_1, h_2 \colon \mathbb{R}^N \to \mathbb{R}$*,* $\text{Cov}\left[h_1(g), h_2(g)\right] \geq 0$ *for* $g = (g_1, \ldots, g_n)$ *with independent components.*

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

$$\text{Cov}\left[h_1(g), h_2(g)\right] = \mathbb{E}\left[\text{Cov}\left[h_1(g), h_2(g) \mid g_1\right]\right] + \text{Cov}\left[\mathbb{E}\left[h_1(g) \mid g_1\right], \mathbb{E}\left[h_2(g) \mid g_1\right]\right]. \tag{A52}$$

The first term is the expectation of a covariance between two monotone functions of $N-1$ variables, where monotonicity follows by Lemma C1. The second term, again by Lemma C1, is a covariance of two monotone functions of random scalars. Thus both of the terms are non-negative.

Applying Lemma C2 to $\tilde{z}_\ell = f_\ell(g, w) - \mu_\ell(w)$ and $\tilde{z}_m = f_m(g, w) - \mu_m(w)$ and conditioning on w everywhere, we obtain the second result of Lemma A1(i).

Proof of Lemma A1(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 $\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] = 0$ a.s. for such (ℓ, m) pairs. We therefore obtain

$$\begin{aligned}\frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] &= \frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \emptyset\right]\mathbb{E}\left[\sum_{\ell,m}\left|\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \\ &\leq \frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \emptyset\right]\mathbb{E}\left[\sqrt{\text{Var}\left[\tilde{z}_\ell \mid w\right]\text{Var}\left[\tilde{z}_m \mid w\right]}\right] \\ &\leq B_Z \cdot \frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \emptyset\right] \to 0. \end{aligned} \tag{A53}$$

C.6 Proof of Proposition A3

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 A1c, 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 $\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell \mid w_c\right] = 0$ by the law of iterated expectations. Consider the variance now, conditionally on w:

$$\begin{aligned}
\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell \mid w\right] &= \mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&\overset{a.s.}{\leq} \mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell + \frac{1}{L}\sum_\ell (\mu_\ell^c - \mu_\ell^u)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&= \mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^u)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&\overset{a.s.}{\leq} \operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^u)\,\varepsilon_\ell \mid w\right] \\
&\overset{a.s.}{\leq} B \cdot \mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m} \left|\operatorname{Cov}\left[z_\ell, z_m \mid w\right]\right| \mid w\right] \xrightarrow{p} 0 \text{ a.s.}
\end{aligned} \tag{A54}$$

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

$$\begin{aligned}
\operatorname{Cov}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell, \frac{1}{L}\sum_m (\mu_m^c - \mu_m^u)\,\varepsilon_\ell \mid w_c\right] &= \mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell \cdot \frac{1}{L}\sum_m (\mu_m^c - \mu_m^u)\,\varepsilon_\ell \mid w_c\right] \\
&= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[(z_\ell - \mu_\ell^c)\,\varepsilon_\ell \varepsilon_m \mid w_c\right] \cdot (\mu_m^u - \mu_m^c) \\
&= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[z_\ell - \mu_\ell^c \mid w_c\right] \cdot \mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid w_c\right] (\mu_m^c - \mu_m^u) \\
&= 0.
\end{aligned} \tag{A55}$$

When two random variables are uncorrelated, the variance of the sum exceeds the variance of one. The fourth line of (A54) again follows by the law of total variance, specifically that

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^u)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
=&\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^u)\,\varepsilon_\ell \mid w\right] - \operatorname{Var}\left[\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^u)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
\leq&\operatorname{Var}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^u)\,\varepsilon_\ell \mid w\right].
\end{aligned} \tag{A56}$$

Finally, the last line of (A54) directly follows from the proof of Proposition A2 (equation (A45), conditionally on w) using Assumptions 1, A2, and A3.

Since $\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\,\varepsilon_\ell \mid w\right] = 0$, (A54) 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.7 Proof of Proposition A5

Let $\hat{R}(t,e) = \int \mathbf{1}[T(\gamma,e) \leq t]\, dG(\gamma)$ denote the re-randomization distribution of the normalized RI test statistic. We first prove that when testing the correct null, i.e. for $e = \varepsilon$, this cdf converges in probability to $\Phi\left(t/\sqrt{\tilde{V}}\right)$ for each t, where $\Phi(\cdot)$ is the cdf of the standard normal distribution. By Assumption A4 and the Law of Iterated Expectations

$$\mathbb{E}\left[\hat{R}(t,\varepsilon)\right] = Pr\left(T(g^*,\varepsilon) \leq t\right) \rightarrow \Phi\left(t/\sqrt{\tilde{V}}\right). \tag{A57}$$

Similarly,

$$\begin{aligned}\mathbb{E}\left[\hat{R}(t,\varepsilon)^2\right] &= \mathbb{E}\left[\int\int \mathbf{1}[T(\gamma_1,\varepsilon) \leq t]\, \mathbf{1}[T(\gamma_2,\varepsilon) \leq t]\, dG(\gamma_1)\, dG(\gamma_2)\right] \\ &= Pr\left(T(g_1^*,\varepsilon) \leq t,\ T(g_2^*,\varepsilon) \leq t\right) \\ &\rightarrow \Phi^2\left(t/\sqrt{\tilde{V}}\right),\end{aligned} \tag{A58}$$

where the last line again uses Assumption A4. Thus $\text{Var}\left[\hat{R}(t,\varepsilon)\right] = \mathbb{E}\left[\hat{R}^2(t,\varepsilon)\right] - \mathbb{E}\left[\hat{R}(t,\varepsilon)\right]^2 \rightarrow 0$, showing that $\hat{R}(t,\varepsilon) \xrightarrow{p} \Phi\left(t/\sqrt{\tilde{V}}\right)$.

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

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

$$\begin{aligned}T(g^*, y - b_L x) &= r_L \frac{1}{L} f(g^*)' (\varepsilon + x \cdot \delta/r_L) \\ &= r_L T(g^*,\varepsilon) + \delta \frac{1}{L} f(g^*)' x.\end{aligned} \tag{A59}$$

While the first term converges to a distribution as before, the second term converges to zero in probability under the assumptions of Proposition A1(ii) (see equation (A48)). Thus, by contiguity, the RI critical values are asymptotically the same and converge in probability to $\sqrt{\tilde{V}}\Phi^{-1}(\alpha/2)$ and $\sqrt{\tilde{V}}\Phi^{-1}(1-\alpha/2)$. In contrast, the asymptotic distribution of the test statistic is shifted by δM:

$$\begin{aligned}T(g, y - b_L x) &= r_L \frac{1}{L} z' (\varepsilon + x \cdot \delta/r_L) \\ &= r_L T(g,\varepsilon) + \delta \frac{1}{L} f(g)' x \\ &\xrightarrow{d} \mathcal{N}\left(\delta M, \sqrt{\tilde{V}}\right).\end{aligned} \tag{A60}$$

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

$$\begin{aligned}
&Pr\left(\delta M+\sqrt{\tilde{V}}\cdot Z<\sqrt{\tilde{V}}\Phi^{-1}(\alpha/2)\right)+Pr\left(\delta M+\sqrt{\tilde{V}}\cdot Z>\sqrt{\tilde{V}}\Phi^{-1}(1-\alpha/2)\right)\\
&=Pr\left(Z<\Phi^{-1}(\alpha/2)-\delta M/\sqrt{\tilde{V}}\right)+Pr\left(-Z<\delta M/\sqrt{\tilde{V}}-\Phi^{-1}(1-\alpha/2)\right)\\
&=\Phi\left(\Phi^{-1}(\alpha/2)-\delta/\sqrt{V}\right)+\Phi\left(\Phi^{-1}(\alpha/2)+\delta/\sqrt{V}\right), \qquad \text{(A61)}
\end{aligned}$$

by symmetry of $\Phi(\cdot)$. Differentiating (A61) by V yields

$$-\frac{1}{2}V^{-3/2}\cdot\delta\left[\Phi'\left(\Phi^{-1}(\alpha/2)+\delta/\sqrt{V}\right)-\Phi'\left(\Phi^{-1}(\alpha/2)-\delta/\sqrt{V}\right)\right]. \qquad \text{(A62)}$$

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

C.8 Proof of Propositions A6 and A7

Proposition A6 Letting $\kappa_\ell(\varepsilon)=\lim_{x\to-\infty}y_\ell(x,\varepsilon)$, we have $y_\ell=\kappa_\ell(\varepsilon)+\int_{-\infty}^{x_\ell}\beta_\ell(\gamma,\varepsilon)d\gamma$. 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,

$$\begin{aligned}
\mathbb{E}\left[\tilde{z}_\ell y_\ell\right]&=\mathbb{E}\left[\tilde{z}_\ell\int_{-\infty}^{x_\ell}\beta_\ell(\gamma,\varepsilon)d\gamma\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{x_\ell}\beta_\ell(\gamma,\varepsilon)\tilde{z}_\ell d\gamma\mid\varepsilon,w\right]\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{\infty}\beta_\ell(\gamma,\varepsilon)\tilde{z}_\ell\mathbf{1}\left[x_\ell\geq\gamma\right]d\gamma\mid\varepsilon,w\right]\right]\\
&=\mathbb{E}\left[\int_{-\infty}^{\infty}\beta_\ell\left(x,\varepsilon\right)\phi_\ell(\gamma,\varepsilon)d\gamma\right] \qquad \text{(A63)}
\end{aligned}$$

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

$$\begin{aligned}
\phi_\ell(x,\varepsilon)&=\mathbb{E}\left[\tilde{z}_\ell\mathbf{1}\left[x_\ell\geq x\right]\mid\varepsilon,w\right]\\
&=\mathrm{Cov}\left[\tilde{z}_\ell,\mathbf{1}\left[x_\ell\geq x\right]\mid\varepsilon,w\right]. \qquad \text{(A64)}
\end{aligned}$$

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

$$\begin{aligned}
\phi_\ell(x,\varepsilon)&=\mathrm{Cov}\left[\tilde{z}_\ell,Pr\left(x_\ell\geq x\mid z_\ell,\varepsilon,w\right)\mid\varepsilon,w\right]\\
&=\mathrm{Cov}\left[z_\ell,Pr\left(x_\ell\geq x\mid z_\ell,\varepsilon,w\right)\mid\varepsilon,w\right], \qquad \text{(A65)}
\end{aligned}$$

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

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \beta_\ell\left(\gamma,\varepsilon\right)\omega_\ell(\gamma,\varepsilon)d\gamma\right], \tag{A66}$$

where

$$\omega_\ell(\gamma,\varepsilon) = \frac{\phi_\ell(\gamma,\varepsilon)}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \phi_\ell(\tau,\varepsilon)d\tau\right]} \tag{A67}$$

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

Proposition A7 Here

$$\begin{aligned} y_\ell &= y_\ell(0,\varepsilon) + \beta_\ell(\varepsilon)x_\ell \\ &= y_\ell(0,\varepsilon) + \beta_\ell(\varepsilon)x_\ell(0) + \beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right)z_\ell \\ &= y_\ell(0,\varepsilon) + \beta_\ell(\varepsilon)x_\ell(0) + \beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right)\left(\tilde{z}_\ell + \mu_\ell\right) \end{aligned} \tag{A68}$$

and

$$\begin{aligned} &\mathbb{E}\left[\tilde{z}_\ell\left(y_\ell(0,\varepsilon) + \beta_\ell(\varepsilon)x_\ell(0) + \beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right)\mu_\ell\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell\left(y_\ell(0,\varepsilon) + \beta_\ell(\varepsilon)x_\ell(0) + \beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right)\mu_\ell\right) \mid w\right]\right] \\ &= 0, \end{aligned} \tag{A69}$$

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

$$\begin{aligned} \mathbb{E}\left[\tilde{z}_\ell y_\ell\right] &= \mathbb{E}\left[\beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right)\tilde{z}_\ell^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right)\tilde{z}_\ell^2 \mid w\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\beta_\ell(\varepsilon)\left(x_\ell(1) - x_\ell(0)\right) \mid w\right]\mathbb{E}\left[\tilde{z}_\ell^2 \mid w\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\beta_\ell(\varepsilon) \mid x_\ell(1) > x_\ell(0), w\right]p_\ell\sigma_\ell^2\right] \end{aligned} \tag{A70}$$

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

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell)y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell)x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\beta_\ell(\varepsilon) \mid x_\ell(1) > x_\ell(0), w\right]\tilde{\omega}_\ell\right] \tag{A71}$$

where

$$\tilde{\omega}_\ell = \frac{p_\ell \sigma_\ell^2}{\mathbb{E}\left[p_\ell \sigma_\ell^2\right]}. \tag{A72}$$

C.9 Proof of Proposition A8

By the mean value theorem, $\mu_\ell\left(\hat{\theta}, w\right) - \mu_\ell(\theta, w) = \frac{\partial \mu_\ell}{\partial \theta}(\theta^*, w)'\left(\hat{\theta} - \theta\right)$ for some $\theta^* \in \Theta$ and with $\frac{\partial \mu_\ell}{\partial \theta}$ component-wise bounded by a scalar B_μ. Thus, for any variable v_ℓ satisfying $\frac{1}{L}\sum_\ell |v_\ell| = O_p(1)$,

$$\begin{aligned}
\left|\frac{1}{L}\sum_\ell v_\ell\left(\mu_\ell\left(\hat{\theta}, w\right) - \mu_\ell(\theta, w)\right)\right| &\leq \frac{1}{L}\sum_\ell \left|v_\ell\left(\mu_\ell\left(\hat{\theta}, w\right) - \mu_\ell(\theta, w)\right)\right| \\
&= \frac{1}{L}\sum_\ell \left|v_\ell \frac{\partial \mu_\ell}{\partial \theta}(\theta^*, w)'\left(\hat{\theta} - \theta\right)\right| \\
&\leq \left(\frac{1}{L}\sum_\ell |v_\ell|\right) B_\mu \left\|\hat{\theta} - \theta\right\|_1 \xrightarrow{p} 0.
\end{aligned} \tag{A73}$$

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

$$\frac{1}{L}\sum_\ell \hat{z}_\ell x_\ell = \frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell - \frac{1}{L}\sum_\ell x_\ell\left(\mu_\ell\left(\hat{\theta}, w\right) - \mu_\ell(\theta, w)\right) \xrightarrow{p} M \neq 0 \tag{A74}$$

and

$$\frac{1}{L}\sum_\ell \hat{z}_\ell \varepsilon_\ell = \frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell - \frac{1}{L}\sum_\ell \varepsilon_\ell\left(\mu_\ell\left(\hat{\theta}, w\right) - \mu_\ell(\theta, w)\right) \xrightarrow{p} 0, \tag{A75}$$

where the first line uses Assumption A1 and stochastic boundness of $\frac{1}{L}\sum_\ell |x_\ell|$, and the second line follows from Proposition A2 and stochastic boundness of $\frac{1}{L}\sum_\ell |\varepsilon_\ell|$. Together equations (A74) and (A75) show consistency of the plug-in recentered estimator $\sum_\ell \hat{z}_\ell y_\ell / \sum_\ell \hat{z}_\ell x_\ell$.

C.10 Proof of Proposition A9

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 $\mathbb{E}[\varepsilon_\ell \mid g, a, w] = \mathbb{E}[\varepsilon_\ell \mid a, w]$ can be constructed along the lines of Proposition 1, see equation (A25)). Part (ii) follows because under the null the distribution of $g \mid \varepsilon^\perp, w$ is the same as $g \mid w$, by independence established in part (i). Part (iii) is analogous to the proof of Proposition 2 for the μ_ℓ-controlled regression (Appendix C.2). Part (iv) follows from the fact that for any variable v_ℓ, $\frac{1}{L}\sum_\ell z_\ell v_\ell^\perp = \frac{1}{L}\sum_\ell \tilde{z}_\ell v_\ell^\perp$ because $\frac{1}{L}\sum_\ell \mu_\ell v_\ell^\perp = 0$ by the properties of projection. Finally, for part (v) we write $\tilde{\beta}_\perp - \beta = \frac{1}{L}\sum_\ell \varepsilon_\ell^\perp \tilde{z}_\ell / \frac{1}{L}\sum_\ell x_\ell^\perp \tilde{z}_\ell$. We first show that the numerator converges to zero in probability. We have:

$$\frac{1}{L}\sum_{\ell}\varepsilon_{\ell}^{\perp}\tilde{z}_{\ell}=\frac{1}{L}\sum_{\ell}\varepsilon_{\ell}\tilde{z}_{\ell}-\hat{\alpha}_{\varepsilon}'\left(\frac{1}{L}\sum_{\ell}a_{\ell}\tilde{z}_{\ell}\right). \tag{A76}$$

By Proposition A2(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 A3 and applying the proof of Proposition A2(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 $r=1,\dots,R$. Since $\hat{\alpha}_{\varepsilon}=O_p(1)$, we have $\frac{1}{L}\sum_{\ell}\varepsilon_{\ell}^{\perp}\tilde{z}_{\ell}=o_p(1)$.

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

$$\frac{1}{L}\sum_{\ell}x_{\ell}^{\perp}\tilde{z}_{\ell}=\frac{1}{L}\sum_{\ell}x_{\ell}\tilde{z}_{\ell}-\hat{\alpha}_{x}'\left(\frac{1}{L}\sum_{\ell}a_{\ell}\tilde{z}_{\ell}\right), \tag{A77}$$

where $\frac{1}{L}\sum_{\ell}x_{\ell}\tilde{z}_{\ell}=M+o_p(1)$ by Assumption A1 and $\hat{\alpha}_x=O_p(1)$. Therefore, $\tilde{\beta}_{\perp}\xrightarrow{p}\beta$.

C.11 Proofs of Lemmas A2 and A3

Proof of Lemma A2. With $\frac{1}{L}\sum_{\ell}z_{\ell}^2\leq B_z$ almost surely, we have

$$\begin{aligned}\left|\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}\varepsilon_{\ell}\right]\right|&=\left|\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}\mathbb{E}\left[\varepsilon_{\ell}\mid g,w\right]\right]\right|\\&\leq\mathbb{E}\left[\sqrt{\frac{1}{L}\sum z_{\ell}^2}\cdot\sqrt{\frac{1}{L}\sum_{\ell}\xi_{\ell}^2}\right]\\&\leq B_z\cdot\sqrt{\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\xi_{\ell}^2\right]}\to 0,\end{aligned}$$

where the first line follows by the Law of Iterated Expectations, the second line follows by the Cauchy-Schwartz inequality, and the third line follows by Jensen's inequality.

Proof of Lemma A3. We have:

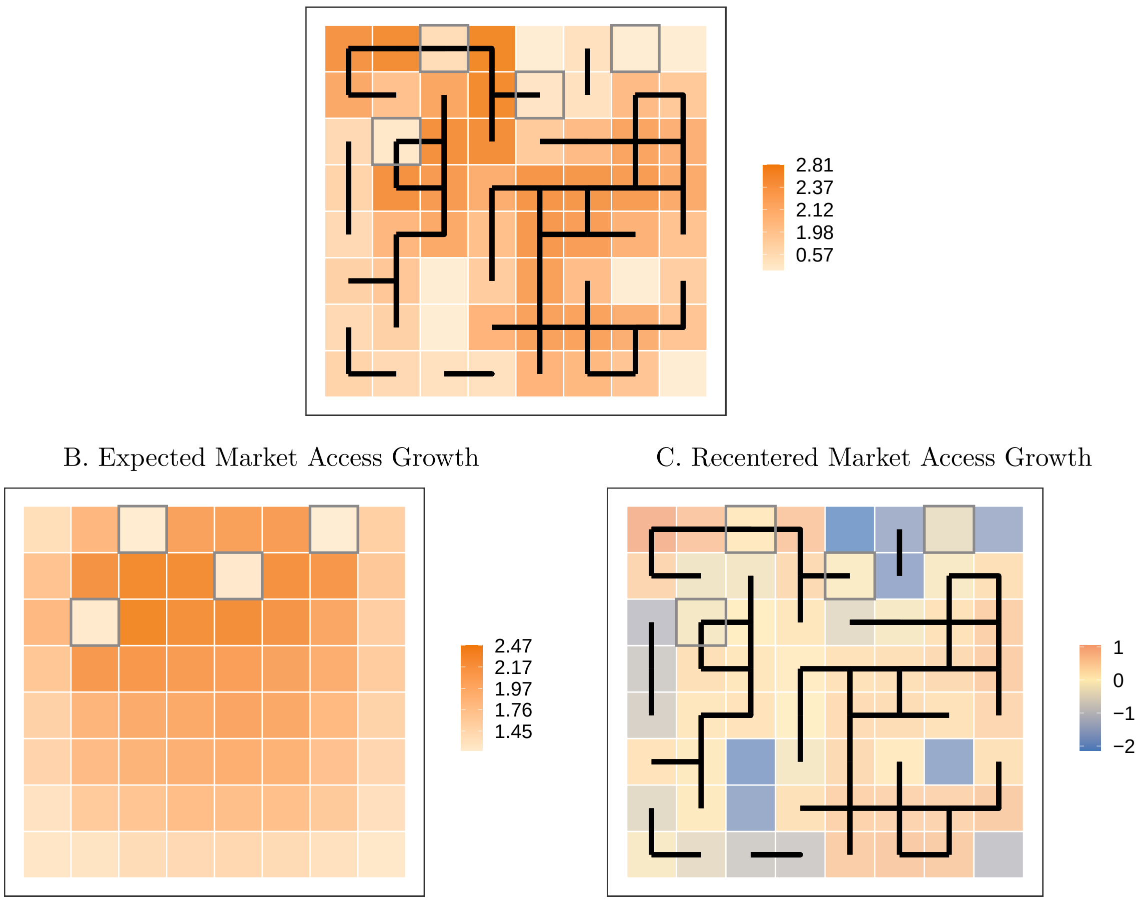
$$\begin{aligned}\left|\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}\varepsilon_{\ell}\right]\right|&=\left|\mathbb{E}\left[\sum_{n}w_n g_n\bar{\varepsilon}_n\right]\right|\\&\leq\mathbb{E}\left[\sqrt{\sum_{n}w_n g_n^2}\cdot\sqrt{\sum_{n}w_n\bar{\varepsilon}_n^2}\right]\\&\leq\sqrt{B_gB_w}\cdot\mathbb{E}\left[\frac{\sum_n w_n\bar{\varepsilon}_n^2}{\sum_n w_n}\right]\to 0,\end{aligned}$$

where the first line uses the Borusyak et al. 2019 representation, the second line uses the Cauchy-Schwartz inequality, and the third line uses Assumption A7 and Jensen's inequality.

Appendix Figures and Tables

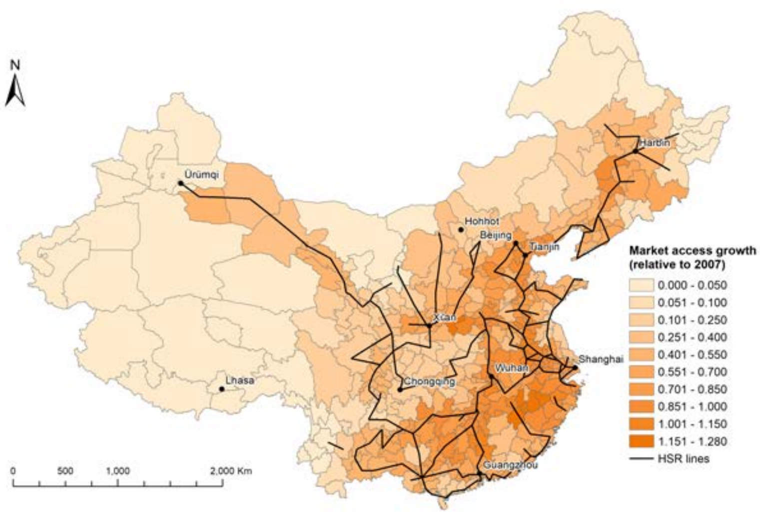
Figure A1: Market Access Growth with Unequal Population

A. Line Construction and Market Access Growth



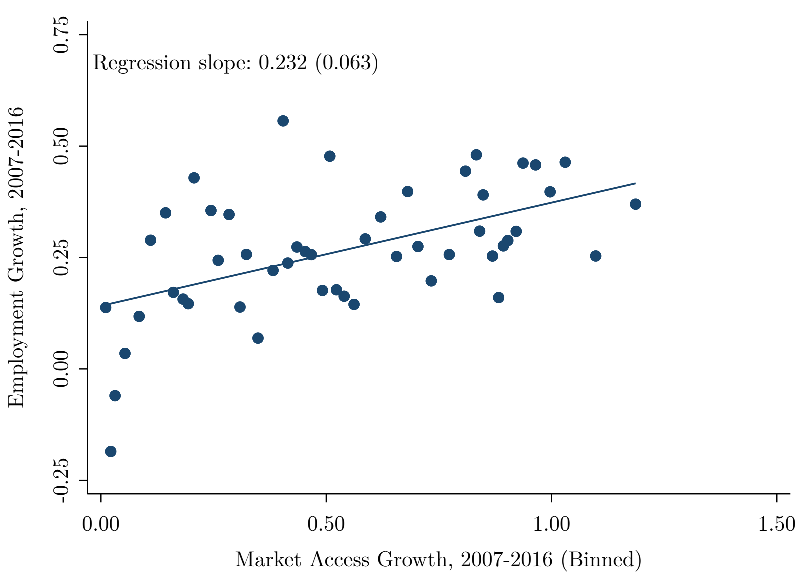
Notes: This figure parallels Figure 1, except assuming that four highlighted regions have 10 times larger market size than all other regions.

Figure A2: Simulated HSR Lines and Market Access Growth



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

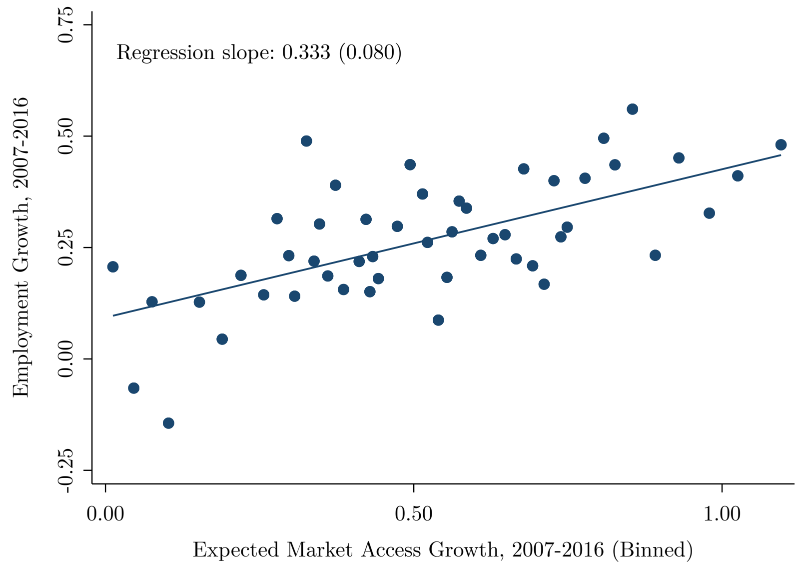
Figure A3: Employment Growth and Market Access Growth



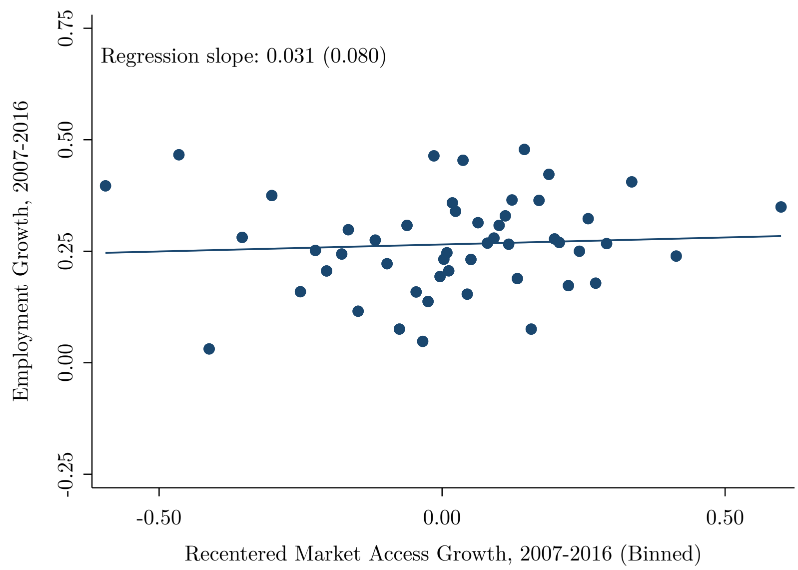
Notes: This figure shows a binned scatterplot of employment growth across 274 prefectures in China, from 2007 to 2016, against market access growth in the same period. Fifty bins of approximately equal size are shown. The regression line of best fit is also indicated, along with the coefficient and spatial-clustered standard error.

Figure A4: Employment Growth and Expected/Recentered Market Access Growth

A. Expected Market Access Growth



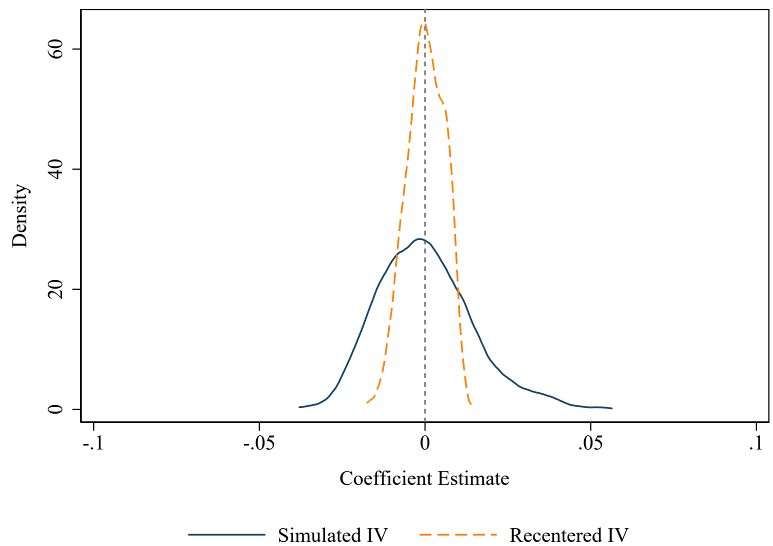
B. Recentered Market Access Growth



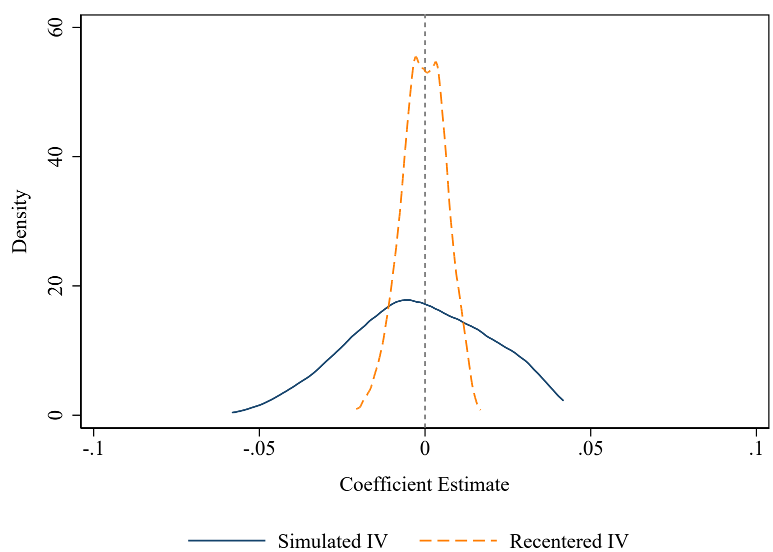
Notes: These figures show binned scatterplots of employment growth across 274 prefectures in China, from 2007 to 2016, against the expected and recentered market access growth in the same period. Expected and recentered market access is constructed by permuting opening dates of lines with the same number of links as described in Section 4.1. Fifty bins of approximately equal size are shown. Regression lines of best fit are also indicated, along with coefficients and spatial-clustered standard errors.

Figure A5: Medicaid Eligibility Effects: Simulated Distributions of Simulated and Recentered IVs

A. Has Medicaid

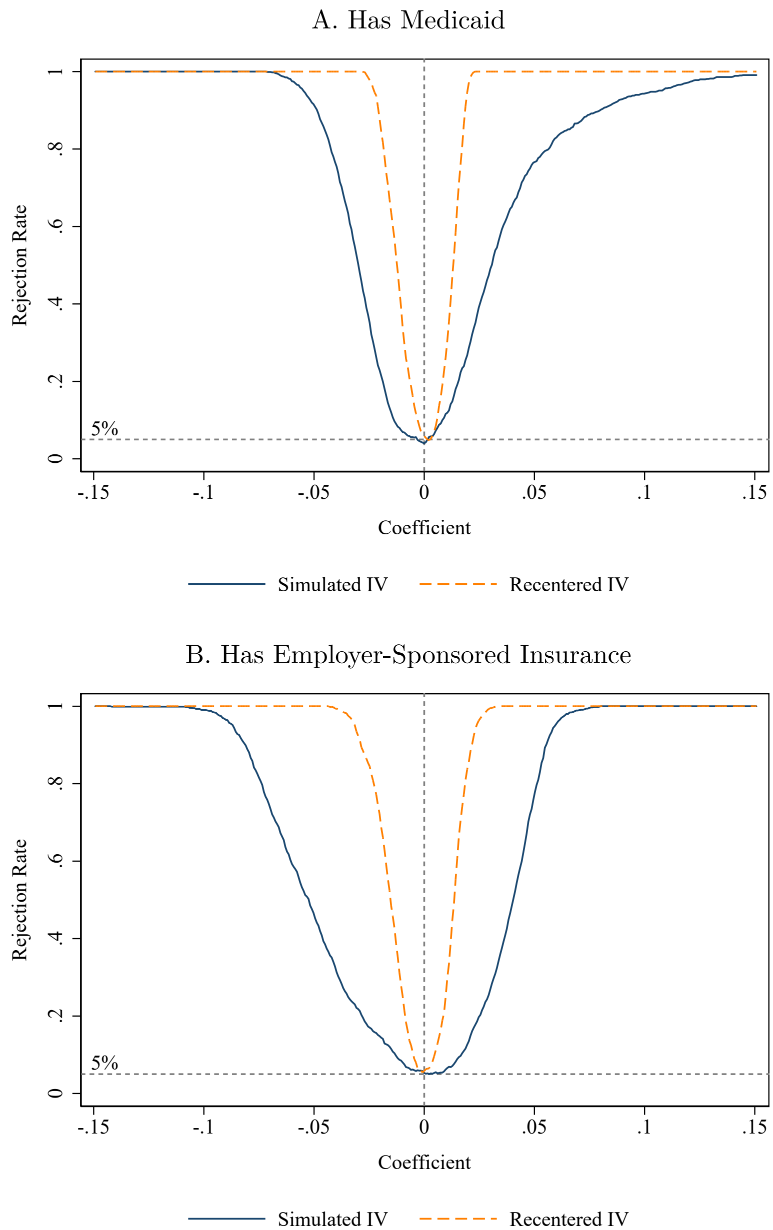


B. Has Employer-Sponsored Insurance



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

Figure A6: Medicaid Eligibility Effects: Simulated Size and Power of Simulated and Recentered IVs



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

Table A1: Effects of Market Access on Additional Outcomes

	Unadjusted OLS (1)	Recentered IV (2)	Controlled OLS (3)
	A. Average Number of Employed Staff and Workers (Urban District)		
Market Access Growth	0.179	0.025	0.053
	(0.080)	(0.118)	(0.095)
		[-0.201, 0.205]	[-0.151, 0.234]
Expected Market Access Growth			0.275
			(0.098)
Prefectures	262	262	262
	B. Persons Employed in Various Units at Year End (Whole City)		
Market Access Growth	0.198	0.047	0.079
	(0.096)	(0.116)	(0.094)
		[-0.175, 0.217]	[-0.112, 0.249]
Expected Market Access Growth			0.283
			(0.123)
Prefectures	267	267	267
	C. Persons Employed in Various Units at Year End (Urban District)		
Market Access Growth	0.169	0.027	0.055
	(0.084)	(0.109)	(0.089)
		[-0.216, 0.215]	[-0.169, 0.243]
Expected Market Access Growth			0.256
			(0.107)
Prefectures	263	263	263
	D. Railway Passenger Traffic (Whole City)		
Market Access Growth	0.366	0.231	0.253
	(0.104)	(0.178)	(0.147)
		[-0.126, 0.571]	[-0.057, 0.608]
Expected Market Access Growth			0.455
			(0.132)
Prefectures	191	191	191
Recentered	No	Yes	Yes

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

Table A2: Simulated and Recentered Medicaid Eligibility Pre-Trends

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

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

Table A3: Recentered IV Estimates of Medicaid Eligibility Effects, Alternative Designs

	Has Medicaid	Has Private Insurance	Has Employer-Sponsored Insurance
	(1)	(2)	(3)
	A. Republican Governor and 2012 Median Income		
Eligibility	0.077	−0.018	−0.005
	(0.011)	(0.008)	(0.006)
	[0.053,0.092]	[-0.042,0.002]	[-0.019,0.011]
	B. Republican Governor, 2012 Median Income and 2012 Medicaid Coverage		
Eligibility	0.076	−0.023	−0.009
	(0.011)	(0.007)	(0.005)
	[0.054,0.102]	[-0.040,-0.008]	[-0.020,0.003]
Exposed Sample	Y	Y	Y
States	43	43	43
Individuals	421,042	421,042	421,042

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

Table A4: Alternative Estimates of Medicaid Eligibility Effects

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

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

References

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

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

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

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

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

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2019. "Quasi-Experimental Shift-Share Research Designs." *NBER Working Paper 24997.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Southworth, Lucinda K., Stuart K. Kim, and Art B. Owen. 2009. "Properties of balanced permutations." *Journal of Computational Biology* 16:625–638.