cemmap
centre for microdata methods and practice

Optimal formula instruments

Kirill Borusyak
Peter Hull

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP09/25

Economic and Social Research Council

Optimal Formula Instruments

Kirill Borusyak
UC Berkeley

Peter Hull
Brown*

March 2025

Abstract

When estimating the effects of treatments defined by complex formulas, researchers often use simple functions of exogenous shocks as instruments. A leading example is "simulated instruments" for public policy eligibility, which capture variation in state-level policy generosity. We show how more powerful instruments can be constructed by incorporating heterogeneous shock exposure while using a recentering procedure to avoid bias. We characterize the asymptotically efficient instruments in this class and propose an algorithm for constructing feasible approximations to them. Compared to a simulated instrument approach, our approach yields a 44% smaller standard error on the private insurance crowd-out effect of Medicaid enrollment from the 2014 Affordable Care Act expansions.

*Contact: k.borusyak@berkeley.edu and peter_hull@brown.edu. Portions of this paper are based on material in a previously circulated working paper, *Non-Random Exposure to Exogenous Shocks: Theory and Applications*. We thank Dmitry Arkhangelsky, Tim Armstrong, Stephane Bonhomme, Kevin Chen, Tetsuya Kaji, and Michal Kolesár for helpful conversations. We thank Molly Frean, Jonathan Gruber, and Ben Sommers for sharing code and data. Elise Parrish provided outstanding research assistance.

1 Introduction

Many economic variables are given by complex formulas, incorporating multiple sources of variation. Examples include an individual's eligibility for a public program like Medicaid or their level of unemployment insurance benefits, both of which are functions of state-level policy decisions as well as various individual characteristics (e.g. family structure, household income, or work history). When estimating the causal effects of such treatments, researchers often construct instrumental variables (IVs) as simple functions of only one source of variation. For example, the influential "simulated instrument" approach of Currie and Gruber (1996a,b) leverages state-level policy shocks by constructing an index of Medicaid generosity which is then used to instrument an individual's Medicaid eligibility.[1] These instruments are valid when the policy shocks are exogenous and the chosen function of them predicts eligibility, at least somewhat.

This paper shows how more powerful instruments can be constructed and used in such settings. Intuitively, power gains can come from the instrument predicting the treatment better. This can be achieved by constructing the instrument as a treatment prediction that is a function (or "formula") of not only the exogenous shocks but also other observables capturing observations' differential shock exposure. Such treatment predictions need not be valid instruments, because of the non-random observables used in their construction. However, following the insight of Borusyak and Hull (2023), this problem can be addressed by "recentering" the treatment prediction: i.e., subtracting its expectation over the exogenous shocks, holding fixed the other observables. The class of valid formula instruments is therefore much broader than functions of exogenous shocks only.

We first characterize optimal formula instruments in a general setting. We show that the asymptotically efficient IV involves three steps: obtaining the best predictor of the treatment from both the shocks and other predetermined measures of shock exposure, recentering it to avoid bias, and adjusting for the error term's dependence on shock exposure and heteroskedasticity. This result does not require *iid* data, covering a wide range of empirical settings where both observed and unobserved shocks—potentially varying at different "levels"—affect the treatment and outcome.

We then propose an algorithm to approximate optimal IVs in practice, focusing on the first two steps: obtaining the best treatment predictor and recentering it. While implementing both steps nonparametrically may be feasible in some settings, in general they represent a high-dimensional problem that may be impractical or infeasible—especially in non-*iid* data. Instead, we propose using knowledge the researcher has on the treatment formula as well as the "design" (i.e., data-generating process) of the exogenous shocks. First, the researcher predicts the treatment from the shocks and other observables which enter the treatment formula, setting any unobserved or endogenous components of the formula to a base value (such as zero). When there are no unobserved or endogenous components, this prediction is the treatment itself. Second, the researcher recenters this prediction by drawing counterfactual sets of exogenous shocks, following Borusyak and Hull

[1] See also Cullen and Gruber (2000) and East and Kuka (2015) for simulated instruments in the unemployment insurance setting. Other simulated instrument applications include Cohodes et al. (2016), Frean et al. (2017), Brown et al. (2018), and Hackmann (2019).

(2023). Optionally residualizing the recentered prediction on covariates yields an approximation to the optimal instrument, up to the heteroskedasticity adjustment that is not popular in practice.

Specializing this algorithm to the program eligibility setting yields a likely improvement over the conventional simulated instruments approach. If program eligibility is fully determined by predetermined observables and the policy shock, our proposed approach involves instrumenting an individual's eligibility with the difference between her actual eligibility (as the best possible predictor) and her expected eligibility, where the expectation is taken for each individual across the exogenous policy shocks that could have been realized, (such as realized policies of comparable states). If program eligibility depends on other variables, such as income that can respond endogenously to policy shocks, the researcher can imperfectly predict eligibility using lagged income and recenter that prediction instead.

The proposed algorithm can also be helpful in many popular settings beyond simulated IV. Consider, for instance, Boustan et al. (2013) who construct an instrument for the change in the Herfindahl index of regional income distribution. They use exogenous national shocks to incomes of different population groups along with initial local shares of those groups. From these data, they measure and use as an IV the Herfindahl change that would follow if the exogenous income shocks were the only changes that took place, in a nonlinear version of a Bartik (1991) instrument. While this instrument is generally not valid without recentering, our results show that the recentered version of their IV is approximately optimal. Notably, conventional results on optimal instruments with *iid* data would not be applicable in their context, in which the same national shocks affect all regions simultaneously. Our results similarly justify the shift-share instrument construction proposed informally by Borusyak et al. (2025b) and instruments for changes in market access due to transportation upgrades proposed by Borusyak and Hull (2023).

We then demonstrate the power gains empirically, in an application to the partial 2014 Medicaid expansion from the Affordable Care Act (ACA). A recentered IV which incorporates variation in individuals' exposure to state expansion decisions yields a 44% smaller standard error on the private insurance crowdout effect of Medicaid enrollment, compared to a more conventional simulated IV approach leveraging expansion shocks only. These power gains are robust to different IV specifications and assumptions on the expansion shock design. Power simulations show the minimum detectable effects of recentered IV are roughly three times smaller than those of simulated IV.

Our theoretical results build on the classical literature on efficiency bounds and optimal instruments in linear and partially linear models. For linear models without functional nuisance parameters, Chamberlain (1987) characterizes the semi-parametric efficiency bound (SEB) and gives an IV estimator that achieves the bound. More closely connected to our setting is the partially linear model of, e.g., Robinson (1988): it is the special case when the outcomes, treatments, shocks, and other observed characteristics are *iid*, with the role of the nuisance function played by the expectation of the error term given the characteristics. There Chamberlain (1992) characterizes the SEB, Newey (1989) proposes an estimator that achieves this bound in the special case when the treatment is exogenous, and Ai and Chen (2003) derive an efficient sieve-based estimator in the general case. Against this backdrop, our theoretical contribution is to characterize optimal instruments in a broad class of non-*iid* settings.[2] We develop a proof technique for such settings: we characterize the estimator that minimizes an approximation to the estimator variance in finite samples and verify that the approximation is accurate in large samples under suitable regularity conditions. We further show that our optimal IV attains the Chamberlain (1992) SEB in the *iid* case.

In contemporaneous work, Coussens and Spiess (2021) characterize optimal instruments given by interactions of a single observation-specific shock with predetermined characteristics, highlighting the benefit of interacting the shock with the complier status of the individual. Our result nests theirs, with individual's exposure to the shocks generalizing the complier status beyond binary instruments and allowing multiple shocks to affect the same observation's treatment and the same shocks to affect multiple observations' treatments. While Coussens and Spiess (2021) focus on learning the compliance status nonparametrically in *iid* data, we leverage *a priori* information on the structure of the treatment to approximate the optimal IV.

We organize the rest of the paper as follows. The next section builds intuition with a simple example in the simulated instrument setting. Section 3 develops the theoretical results while Section 4 demonstrates them in an application. Section 5 concludes. Additional theoretical results are given in Appendix A; all proofs are given in Appendix B.

2 Motivating Example

Consider estimating the causal effect of eligibility for a public program like Medicaid on an outcome like program takeup or later health.[3] Formally, we consider a simple causal model of

$$y_i = \beta x_i + \varepsilon_i,$$

relating outcome y_i for individual i to her Medicaid eligibility x_i. Here ε_i denotes the potential outcome that individual i would see when ineligible for Medicaid (i.e., when $x_i = 0$) and β is the causal parameter of interest. We wish to estimate β while allowing for the possibility of endogenous eligibility: i.e., that x_i and ε_i are correlated. To focus on estimation efficiency, we assume here that the causal effect β is homogeneous.

Medicaid eligibility can be represented as a formula which incorporates state-level government policy and individual characteristics that determine an individual's exposure to different polices. To formalize this, let c_i be a vector of characteristics for individual i (e.g., family structure and income), let $s(i) \in \{1, \ldots, 50\}$ index i's state of residence, and let g_k be the Medicaid policy in state k formalized as the set of family types and income combinations that make one eligible for Medicaid

[2] Among non-*iid* settings, optimal instruments have been most thoroughly studied in time-series data (e.g., Hansen (1985); see Anatolyev (2007) for a review).

[3] This example builds on one which Borusyak and Hull (2023, Section 2) use to motivate instrument recentering. Here we use it to motivate the new theoretical results in Sections 3 and the empirical analysis in Section 4.

in that state. Eligibility is then given by:

$$x_i = \mathbf{1}\left[c_i \in g_{s(i)}\right].$$

Consider estimation of β in an idealized scenario where Medicaid policies are drawn in a natural experiment: i.e., randomly from some pool of potential policies.[4] Formally we suppose the g_k are drawn from some distribution, independently of all c_i, $s(i)$, and ε_i. We do not assume the other determinants of eligibility are exogenous (i.e., unrelated to ε_i) : individuals with certain characteristics or living in particular states may have systematically higher or lower potential outcomes. Hence, despite the exogeneity of state policies, ordinary least squares (OLS) estimation of β is likely biased. We need an instrument for eligibility, which is uncorrelated with ε_i but correlated with x_i.

Simulated instruments leverage the exogenous policy shocks by constructing an IV as a function of state policy only: $z_i = f\left(g_{s(i)}\right)$. The function $f(\cdot)$ is chosen to make the instrument powerful—specifically, to make z_i a strong predictor of the eligibility treatment. The strongest predictor that only varies through state policy is $\mathbb{E}\left[x_i \mid g_{s(i)}\right]$, which can be interpreted as the average generosity of i's state policy. Currie and Gruber (1996a,b) propose a simple approximation to this predictor. They build a large and nationally-representative group of individuals $j = 1, \ldots, J$, simulate individuals' eligibility under each state policy $\bar{g}$, and define $f(\bar{g})$ as the fraction of individuals who would be eligible under that policy:

$$f(\bar{g}) = \frac{1}{J}\sum_{j=1}^{J} \mathbf{1}\left[c_j \in \bar{g}\right].$$

The simulated instrument $z_i = f\left(g_{s(i)}\right)$ is a fixed function of the exogenous policy in an individual's state and is therefore uncorrelated with ε_i.[5] It is nevertheless correlated with x_i because we expect the eligibility of any given individual to be higher in states where the policy is more generous.

A drawback of such instruments, which limits their power, is that they discard all within-state variation in eligibility due to c_i. This may seem unavoidable, as such variation is non-random and using it may introduce bias. For example, one might consider constructing an instrument to approximate $\mathbb{E}\left[x_i \mid g_{s(i)}, c_i\right]$ instead of $\mathbb{E}\left[x_i \mid g_{s(i)}\right]$; the former has a stronger first-stage correlation with x_i, so using this IV would likely produce a smaller standard error than z_i. However, this instrument may be correlated with ε_i through some characteristics in c_i, making the IV estimates biased. Indeed, here $\mathbb{E}\left[x_i \mid g_{s(i)}, c_i\right] = x_i$, since eligibility is fully determined by $g_{s(i)}$ and c_i. Using it as an instrument is thus equivalent to OLS estimation, with the same bias concerns as before.

The main practical insight of this paper is that improved predictions of formula treatments,

[4]This does not presume that any policy could arise: for instance, the pool may only include potential policies that prioritize the poor.

[5]Formally, for any fixed $f(\cdot)$, $\mathbb{E}\left[z_i \mid s(i), \varepsilon_i\right] = \mathbb{E}\left[f\left(g_{s(i)}\right) \mid s(i), \varepsilon_i\right] = \int f(\bar{g})dG(\bar{g}) = \mathbb{E}\left[z_i\right]$ where $G(\cdot)$ is the distribution of potential Medicaid policies. Thus $\mathrm{Cov}\left[z_i, \varepsilon_i\right] = \mathbb{E}\left[\left(z_i - \mathbb{E}\left[z_i\right]\right)\varepsilon_i\right] = \mathbb{E}\left[\left(\mathbb{E}\left[z_i \mid s_i, \varepsilon_i\right] - \mathbb{E}\left[z_i\right]\right)\varepsilon_i\right] = 0$.

including x_i itself, can be used to construct valid and more powerful instruments. Consider:

$$\tilde{z}_i = \underbrace{\mathbf{1}[c_i \in g_{s(i)}]}_{=x_i} - \underbrace{\frac{1}{50}\sum_{k=1}^{50}\mathbf{1}[c_i \in g_k]}_{\equiv \mu_i}.$$

The first term of $\tilde{z}_i$ is individual i's eligibility, determined by both her characteristics c_i and the realized policy draw in her state $g_{s(i)}$. The second term is her *expected* eligibility over policy draws. This μ_i is derived from a thought experiment in which the realized policies are randomly reshuffled across the 50 states, since each permutation of $(g_1, \ldots, g_{50})$ is as likely to have been realized when policies are drawn randomly from some set. On average, an individual's eligibility across such permutations equals the share of states in which her characteristics would make her eligible. Borusyak and Hull (2023) show that the *recentering* of x_i by μ_i makes $\tilde{z}_i$ a valid instrument. Intuitively, IV regressions that use $\tilde{z}_i$ compare individuals who have more Medicaid eligibility than expected given the policy experiment (i.e., those with $\tilde{z}_i > 0$) to those with less-than-expected eligibility (with $\tilde{z}_i < 0$). Since this delineation is by chance, driven only by the random policy shocks, such IV regressions are free from bias.[6] In the next section we discuss different strategies for recentering that follow from different assumptions of how the shocks are drawn.

The recentered instrument $\tilde{z}_i$ is likely more powerful than the simulated instrument z_i, as it is more predictive of the treatment. By construction, the policy generosity measure $f\left(g_{s(i)}\right)$ is not tailored to an individual's exposure to the policies, limiting z_i's correlation with x_i. In contrast, $\tilde{z}_i$ is perfectly correlated with x_i conditional on c_i (i.e., among individuals with the same family structure and income but who reside in different states). Only individuals for whom the policy variation is relevant are in the effective sample with $\tilde{z}_i$, since $\tilde{z}_i = 0$ for any individual who is, e.g., so rich that they are not eligible under any policy or so poor that they are eligible under all policies. The next section formalizes the sense in which such instruments likely yield precise estimates of β, and proposes a general algorithm for producing them from formula treatments like x_i.

Power when using $\tilde{z}_i$ may be further increased by including functions of c_i (e.g. family size or some income bins) as controls. Such controls soak up residual variation in ε_i while not affecting the instrument's first stage (because $\tilde{z}_i$ is uncorrelated with all predetermined characteristics as well as with ε_i), typically increasing estimation efficiency.[7] The next section formalizes this logic by showing how the theoretically most efficient IV estimation of β involves such residual adjustment in addition to forming the recentered best predictor $\tilde{z}_i$; we add an optional step to the algorithm that involves such adjustment. Although extra covariates can increase the power of simulated IV estimation too, the two approaches do not coincide even when controlling for c_i flexibly.[8]

[6] More formally, Borusyak and Hull (2023) show μ_i and x_i have the same covariance with ε_i so $\mathbb{E}\left[(x_i - \mu_i)\,\varepsilon_i\right] = 0$.

[7] If the included controls linearly span expected eligibility μ_i, using $\tilde{z}_i$ as the instrument is numerically equivalent to using OLS estimation on x_i with the same controls—slightly simplifying implementation (Borusyak and Hull, 2023).

[8] The two approaches would coincide if policy generosity was measured for each combination of characteristics in c_i separately. This strategy, however, is only feasible if c_i is discrete with a sufficiently small number of distinct values, which is not the case in most applications of simulated IV (see., e.g., Gruber (2003, p.47)).

Before proceeding, we note that while x_i in this example is fully determined by the exogenous policy shocks $g_{s(i)}$ and predetermined individual characteristics c_i, similar recentered IVs may be constructed for treatments with endogenous components. For example, suppose an individual's income u_i is relevant to her Medicaid eligibility, $c_i = (u_i, \tilde{c}_i)$, but it is not predetermined: income may respond to the policy shocks as individuals change employment. In this case $\tilde{z}_i$ will not be a valid instrument, as recentering by μ_i will not account for the endogenous response of c_i to $g_{s(i)}$. Still, a strong predictor of x_i can be formed from its formula: one can compute the predicted eligibility based on the realized policy shocks, other characteristics $\tilde{c}_i$, and an earlier measure of income u_{0i} that replaces u_i. This approximation of $\mathbb{E}\left[x_i \mid g_{s(i)}, u_{0i}, \tilde{c}_i\right]$ can then be recentered, as before, to obtain a valid and likely powerful instrument. Cases where some characteristics relevant for eligibility are unobserved by the researcher can be handled similarly.[9]

3 Theory

We now develop general theory for optimal formula instruments. Section 3.1 introduces the setting and the class of valid recentered instruments. Section 3.2 derives the recentered instrument that is asymptotically most efficient, while Section 3.3 develops an algorithm for obtaining feasible approximations to the optimal IV.

3.1 Setting

An outcome y_i and treatment $x_i = h_i(g, w, u)$ are observed for a set of units $i = 1, \dots, N$. Here $h_1(\cdot), \dots, h_N(\cdot)$ is a set of known functions, $g = (g_1, \dots, g_K)$ is a set of observed shocks (potentially varying at a different level than i), w is a set of observed predetermined variables, and u is another set of variables (potentially unobserved and also potentially varying at different levels). This formulation of x_i is so far without loss of generality; below we consider assumptions that make it restrictive and introduce substantive distinctions between g, w, and u. Since multiple observations may be exposed to the same observed and potentially unobserved shocks, we do not make any assumptions of independently or identically distributed (*iid*) data and instead work with a finite population of size N; the results apply to data randomly drawn from some population as well.

A causal effect or structural parameter β relates the outcome to treatment by

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

where ε_i is an unobserved error. Here we assume y_i and x_i are scalar and demeaned (such that no constant is required in estimation), and that the outcome model is linear with a constant effect.[10]

[9] For example, we might be interested in the effects of Medicaid *enrollment* rather than eligibility and not observe an individual's compliance status (i.e. whether they would take up Medicaid when eligible). This u_i can be ignored when predicting x_i from the shocks and predetermined observables, e.g. by presuming that all individuals are compliers. If compliance status can be partially predicted, incorporating this may yield further precision gains.

[10] The constant effects assumption follows the optimal IV literature, facilitating an analysis of relative efficiency across different IV estimators since different instrument constructions will generally identify different weighted averages

To estimate β, we look for an instrument $z = (z_1, \ldots, z_N)'$ that satisfies an exogeneity condition:

$$\mathbb{E}\left[\frac{1}{N}\sum_i z_i \varepsilon_i\right] = 0 \tag{2}$$

as well as a relevance condition $\mathbb{E}\left[\frac{1}{N}\sum_i z_i x_i\right] \neq 0$, such that $\beta = \mathbb{E}\left[\sum_i z_i y_i\right] / \mathbb{E}\left[\sum_i z_i x_i\right]$.[11] The IV estimator corresponding to z is then given by $\hat{\beta}[z] = \left(\sum_i z_i y_i\right) / \left(\sum_i z_i x_i\right)$.

While the class of IV estimators is restrictive relative to the more general class considered in, for instance, Chamberlain (1992), our non-*iid* analysis makes it much less restrictive than it may seem. In particular, by the Frisch-Waugh-Lovell theorem, it includes IV estimators with controls since z_i can be chosen to be an in-sample residualization of some instrument on a set of covariates. Indeed, as we show below, our optimal instrument involves such adjustment.

We form instruments by assuming the shocks g are exogenous and relevant to the treatment. Specifically, we suppose that, conditional on the predetermined variables w, the outcome errors are mean-independent of g and the treatment is affected by g (or is otherwise dependent on it):

Assumption 1. *(Shock exogeneity):* $\mathbb{E}\left[\varepsilon_i \mid g, w\right] = \mathbb{E}\left[\varepsilon_i \mid w\right]$ *a.s. for all* i*.*

Assumption 2. *(Relevance):* $\mathbb{E}\left[x_i \mid g, w\right] \neq \mathbb{E}\left[x_i \mid w\right]$ *with positive probability for some* i*.*

In the basic Medicaid example, g is the vector of state eligibility policies and w contains the relevant characteristics of all individuals along with their states of residence. Then $h_i(\cdot)$ is the known algorithm which checks the eligibility of individual i using these inputs, and u is empty. If income changes in response to the policy, we include it in u while adding pre-period income to w.

Two remarks about Assumption 1 are due here. First, in some applications, the shocks are randomized after the realization of w; this makes them fully independent from w as well as from the errors under an exclusion restriction (that shocks only affect outcomes through the treatment), satisfying Assumption 1. Under full independence, the class of valid moment conditions is wider (Poirier, 2017); by making the weaker and more conventional Assumption 1, we limit ourselves to IV estimators. Second, Assumption 1 allows ε_i to be arbitrarily correlated with the variables in w such that OLS estimation is generally biased even if u is empty.

The class of instruments satisfying exogeneity under Assumption 1 can be sharply characterized. We refer to instruments constructed as $z_i = f_i(g, w)$ for a set of non-stochastic functions $\{f_i(\cdot)\}_{i=1}^N$ as *formula instruments*. We further call them recentered formula instruments, or just *recentered instruments*, if they are mean-zero given w:

$$\mathbb{E}\left[f_i(g, w) \mid w\right] = 0 \text{ a.s. for all } i. \tag{3}$$

Let $\mathfrak{R}$ denote the class of recentered instruments.[12] We then have the following result:

of heterogeneous effects. See Borusyak and Hull (2021, appendix C.1) for a characterization of these averages.

[11] We assume throughout that relevant expectations and other moments are well defined.

[12] $\mathfrak{R}$ clearly includes instruments of the form $z_i = p_i(g, w) - \mathbb{E}\left[p_i(g, w) \mid w\right]$ for any $(p_i(\cdot))_{i=1}^N$. Moreover, for any

Proposition 1. *Under Assumption 1, all recentered instruments satisfy the exogeneity condition (2). Moreover, only recentered instruments satisfy (2): for any* $z = (z_i)_{i=1}^N$ *there exists a conditional distribution of* $\varepsilon \mid (z, g, w)$ *such that Assumption 1 holds but (2) fails unless* z *is a deterministic function of* (g, w) *satisfying (3).*

The first part of the proposition follows Borusyak and Hull (2023) to show that recentered instruments are valid. The second part is new, and highlights two ideas. First, formula instruments that are not recentered include some variation from w and are thus prone to exogeneity failures without further restrictions on the error term. Second, Assumption 1 would not justify the validity of instruments constructed from any other data besides the shocks and predetermined variables in w. With this characterization, we next look for the most efficient recentered instrument.

3.2 Optimal IV

We take a non-standard approach to deriving the efficient instrument, given the non-*iid* setup. We first give an approximation to the finite-population variance of any recentered IV estimator which we show is accurate in large samples under appropriate regularity conditions. We then find the recentered instrument that minimizes this approximation in finite samples.

The approximate variance of the IV estimator using recentered instrument $z \in \mathfrak{R}$ is defined as:

$$\mathcal{V}[z] = \frac{\mathrm{Var}[z'\varepsilon]}{\mathbb{E}[z'x]^2},$$

for $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$ and $x = (x_1, \ldots, x_N)'$, assuming the relevant variance and expectation exist. This expression represents the variance of $\frac{1}{N}z'\varepsilon / \mathbb{E}\left[\frac{1}{N}z'x\right]$; if the first-stage covariance $\frac{1}{N}z'x$ converges to a non-zero constant as $N \to \infty$, we expect $\mathrm{Var}\left[\hat{\beta}[z]\right] \approx \mathcal{V}[z]$ when N is large and appropriate regularity conditions hold.

We define some additional concepts to formalize this asymptotic approximation. First, for a non-random sequence $r_N \to \infty$, we say that an estimator $\tilde{\beta}$ converges to β at rate r_N when $r_N\left(\tilde{\beta} - \beta\right)$ converges to a non-degenerate distribution with zero mean and variance $0 < V < \infty$ as $N \to \infty$.[13] We refer to V as the asymptotic variance of $\tilde{\beta}$.[14] Second, we say that a recentered IV estimator $\hat{\beta}[z]$ for $z \in \mathfrak{R}$ is *regular* if it converges to β at some rate r_N, if it has an asymptotic first stage (i.e. $\frac{1}{N}z'x \xrightarrow{p} M$ for some $M \neq 0$), and if the sequences of $\frac{1}{N}z'x$ and $\left(r_N \frac{1}{N}z'\varepsilon\right)^2$ are uniformly integrable. We then have the following result:

Proposition 2. *For any regular recentered IV estimator* $\hat{\beta}[z]$ *that converges to* β *at rate* r_N *with*

$z \in \mathfrak{R}$, in-sample residuals of z on any set of functions $r(w)$ are also in $\mathfrak{R}$; by the Frisch–Waugh–Lovell theorem, these correspond to estimators that use z as an IV while controlling for $r(w)$. This means estimators using any $p_i(g, w)$ as an IV while controlling for $\mathbb{E}[p_i(g, w) \mid w]$ can be represented as recentered IV estimators because, with this control, the estimator is numerically equivalent to the one which instruments with $p_i(g, w) - \mathbb{E}[p_i(g, w) \mid w]$.

[13] Borusyak and Hull (2023) provide sufficient conditions for consistency of recentered IV estimators.

[14] The asymptotic variance concept is most 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/r_N^2 outside the normal case.

variance V, $\mathcal{V}[z]$ provides a good large-sample approximation to the variance:

$$\lim_{N\to\infty} \frac{\mathcal{V}[z]}{V/r_N^2} = 1.$$

This result justifies looking for the recentered IV estimator with the smallest approximate variance $\mathcal{V}[z]$. The following theorem characterizes the solution:

Theorem 1. *Suppose Assumption 1 holds, $\mathbb{E}[\varepsilon\varepsilon' \mid g, w] = \mathbb{E}[\varepsilon\varepsilon' \mid w]$ a.s., and this matrix is a.s. invertible. Consider the recentered instrument*

$$z^* = \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1} \tilde{z} \quad \textit{for } \tilde{z} = \mathbb{E}[x \mid g, w] - \mathbb{E}[x \mid w]. \tag{4}$$

The associated $\hat{\beta}[z^*]$ has the smallest approximate variance of all recentered IV estimators:

$$z^* \in \arg\min_{z\in\mathfrak{R}} \mathcal{V}[z],$$

with

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

Three comments are due here. First, note that we impose weak conditions on the data-generating process: only that the shocks are exogenous in the sense of Assumption 1 and that the errors are not perfectly collinear.[15] While Theorem 1 also requires the second moments of ε to be independent of g given w (which holds when $g \perp\!\!\!\perp (\varepsilon, w)$, as when shocks are fully randomized), Appendix Theorem A1 provides a more cluttered expression for the optimal IV without that assumption. Second, we note that the approximate variance is minimized among all recentered IV estimators, regardless of whether they are regular, although $\mathcal{V}[z]$ need not be a useful object otherwise. Third, note that by Assumption 2 the instruments $\tilde{z}$ and z^* are non-trivial and $\mathcal{V}[z^*]$ is finite.

Theorem 1 builds on classic results on efficient estimation in *iid* data. In particular, Chamberlain (1987) characterizes optimal instruments in the *iid* linear model (see also Newey and McFadden (1994) for a different characterization that our proof leverages). Newey (1989, Section 5) derives an optimal estimator for the partially linear model with conditionally exogenous treatment, in which the functional nuisance parameter corresponds to $\mathbb{E}[\varepsilon \mid w]$ in our notation.[16] Ai and Chen (2003) propose a sieve-based efficient estimator in the general *iid* case. While the key advantage of Theorem 1 is that it applies in non-*iid* data, it is also instructive to specialize it to the *iid* case in which the semi-parametric efficiency bound (SEB) of Chamberlain (1992) applies. Appendix Proposition A1 shows that z^* attains this SEB asymptotically, suggesting that our limiting of the estimator class to recentered IV does not carry an asymptotic efficiency cost.[17]

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

[16] Conditional treatment exogeneity means $\mathbb{E}[\varepsilon - \mathbb{E}[\varepsilon \mid w] \mid x] = 0$, which here holds when $u = \emptyset$.

[17] Despite this, the estimators in Newey (1989) and Ai and Chen (2003) do not coincide with the one in Theorem 1 in the *iid* case: both of them are weighted versions of the Robinson (1988) estimator, which involves non-parametric

Equation (4) reveals the structure of the optimal instrument z^*. It is based on the best predictor of treatment given the exogenous shocks and predetermined variables, $\mathbb{E}[x \mid g, w]$, recentered by its expectation over the shocks $\mathbb{E}[x \mid w] = \mathbb{E}[\mathbb{E}[x \mid g, w] \mid w]$. This recentered best predictor $\tilde{z}$ is then adjusted by $\mathbb{E}[\varepsilon\varepsilon' \mid w]^{-1}$. The following proposition unpacks this last step:

Proposition 3. *Let $\psi = \mathbb{E}[\varepsilon \mid w]$ and $\Omega = \operatorname{Var}[\varepsilon \mid w]$ be the conditional mean and variance of the errors. Then z^* from Theorem 1 can be written as*

$$z^* = \Omega^{-1}(\tilde{z} - \nu\rho\psi), \tag{5}$$

where $\rho\psi = \frac{\psi'\Omega^{-1}\tilde{z}}{\psi'\Omega^{-1}\psi}\psi$ is the Ω^{-1}-weighted projection of $\tilde{z}$ on ψ, and $\nu = \frac{\psi'\Omega^{-1}\psi}{1+\psi'\Omega^{-1}\psi} \in [0, 1)$.

This result shows that z^* is obtained by two sequential adjustments to the recentered best predictor $\tilde{z}$. First, $\tilde{z}$ is partially residualized on ψ. In many cases, $\nu \to 1$ when $N \to \infty$; by the Frisch-Waugh-Lovell theorem, the limit case of $\nu = 1$ amounts to controlling for ψ while instrumenting with $\tilde{z}$.[18] Second, there is an adjustment for the inverse conditional variance of the errors Ω^{-1}, as in conventional generalized least squares estimation.

Appendix Proposition A2 specializes Theorem 1 to the class of "shift-share" IVs: instruments that are linear in the shocks and for which (as discussed below) feasible versions may be easier to obtain. The optimal instrument in this class resembles z^*, except with $\tilde{z}$ replaced by an optimal linear prediction of the treatment—i.e., the fitted value from a linear regression of x_i on the (recentered) vector of shocks, with coefficients that can vary across i and depend on w. This linear regression parallels the non-parametric regression of x_i on the shocks that yields $\mathbb{E}[x_i \mid g, w]$ in Theorem 1. Like the non-parametric analog, the linear regression is generally infeasible as it is fit separately for each observation, over the distribution of the shock vector (of which there is only one observation). However, it can guide the choice of feasible instruments.

3.3 Feasible Approximations

Drawing on the above results, we propose an approach (Algorithm 1) for obtaining feasible approximations to z^* in three steps: approximating the best predictor $\mathbb{E}[x \mid g, w]$, recentering it, and adjusting for $\mathbb{E}[\varepsilon\varepsilon' \mid w]^{-1}$. The algorithm leverages the researcher's knowledge of the formula of the treatment and the "design" (i.e. data-generating process) of the exogenous shocks. Consequently, it can be used even in situations when the conditional means and variances underlying z^* cannot be consistently estimated because the data are not *iid* (i.e., we allow the entire sample of (x, g, w, ε) to be one draw from some joint data-generating process) or otherwise high-dimensional.

residualization of y and x on w, while $\hat{\beta}[z^*]$ involves partially controlling for $\mathbb{E}[\varepsilon \mid w]$ (see Proposition 3 and footnote 18 below).

[18] In general the residualization is partial for the same reason why, in conventional panel data settings, the efficient random effects estimator partially demeans the data by unit (e.g. Wooldridge (2002, p.286)). As with the unit-specific residual means in the random effects case, ψ is orthogonal to $\tilde{z}$ in expectation but not in the observed realization. Full residualization which imposes in-sample orthogonality is thus generally inefficient, as with the fixed effect estimator.

Algorithm 1 Feasible Approximation to the Optimal IV

0. Represent the treatment as a formula $x_i = h_i(g, w, u)$ for known functions $h_i(\cdot)$, where the inputs are exogenous shocks g, predetermined variables w, and possibly other variables u that are either unobserved or may respond to the shocks.

1. Construct a treatment prediction $p_i(g, w)$. If the treatment is a function of g and w only, set $p_i = x_i$. Otherwise, set $p_i = h_i(g, w, \bar{u})$ by replacing u with some base value $\bar{u}$.
   - *Optional:* Replace $p_i(g, w)$ with its linear approximation around some base value $g = \bar{g}$.

2. Form the instrument by recentering the prediction: $\tilde{z}_i = p_i(g, w) - \mu_i$ for $\mu_i = \mathbb{E}[p_i(g, w) \mid w]$.
   - To obtain μ_i, draw some (preferably large) number J of counterfactual shock vectors $g^{(j)}$ from the data-generating process of g (e.g., permutations of g when the shocks are exchangeable). Set $\mu_i = \frac{1}{J} \sum_{j=1}^{J} p_i(g^{(j)}, w)$.
   - If $p_i(g, w)$ is linear in the shocks, μ_i can be computed analytically from a specification for $\mathbb{E}[g \mid w]$ without a full specification of the shock data-generating process.

3. *Optional:* Residualize $\tilde{z}$ on predetermined variables (i.e., functions of w) that may predict the error ε, and reweight it by an estimate of $\text{Var}[\varepsilon \mid w]^{-1}$.

The first step is to approximate the best predictor $\mathbb{E}[x_i \mid g, w]$ with some $p_i(g, w)$ using the treatment formula $x_i = h_i(g, w, u)$. This is trivial when u is empty, as in the initial Section 2 example, since then $\mathbb{E}[x_i \mid g, w] = x_i \equiv p_i(g, w)$. Otherwise, we propose forming a predictor by integrating $h_i(g, w, u)$ over some hypothetical distribution of $u \mid g, w$. In the simplest case, if there is a typical or default value $\bar{u}$ of u, it can be plugged in to form a prediction of $p_i(g, w) = h_i(g, w, \bar{u})$. This is the approach suggested at the end of Section 2, where lagged income u_{0i} substitutes for the potentially endogenous contemporaneous income u_i.[19] Importantly, per Proposition 1, a misspecified distribution for $u \mid g, w$ may lead to an imperfect approximation of $\mathbb{E}[x_i \mid g, w]$ but this will not bias estimation so long as the approximation is recentered.

The second step is to recenter the predictor $p_i(g, w)$ using knowledge of the shock assignment process: formally, the distribution of $g \mid w$. Here we summarize several approaches, following Borusyak and Hull (2023) and Borusyak et al. (2025a). Recentering is straightforward when the shock assignment process is known, such as when g is generated by a randomized control trial with some experimental protocol (conditional on w, or more typically independent from w). The researcher can draw a large number of counterfactual shock vectors $g^{(j)}$ from this protocol, recompute the predictor $p_i(g^{(j)}, w)$ for each $j = 1, \ldots, J$, and average them: $\frac{1}{J} \sum_{j=1}^{J} p_i(g^{(j)}, w)$. This approximation to $\mathbb{E}[p_i(g, w) \mid w]$ can then be subtracted from $p_i(g, w)$, or controlled for in estimation.[20] The same

[19] Berry et al. (1999) take a similar approach in the context of estimating demand for differentiated products, setting unobserved product quality shifters to zero when constructing a (non-recentered) instrument. Conlon and Gortmaker (2020) propose a refinement to this approach, integrating over the empirical distribution of estimated quality shifters rather than using a single default value.

[20] While a large J reduces noise in the instrument $p(g, w) - \frac{1}{J} \sum_{j=1}^{J} p(g^{(j)}, w)$, approximating μ_i with any finite

steps can be followed in observational data when the elements of g are assumed to be exchangeable conditional on w, as with the exchangeable policy shocks in Section 2. The counterfactual $g^{(j)}$ can then be generated as permutations of g, bypassing the challenge of specifying the distribution of the shocks.[21] More generally, $g^{(j)}$ could be drawn from conditional permutations of shocks within—but not across—groups with the same shock distribution; we illustrate this approach below.[22]

As Borusyak et al. (2025a) note, recentering requires weaker yet assumptions when $p_i(\cdot)$ is linear in the shocks—as in "shift-share" instrument constructions. In that case, only the expectation of the shocks matters for $\mathbb{E}[p_i(g,w) \mid w]$ and thus needs to be modeled. In cases where that approach to recentering is more tenable, Borusyak et al. (2025a, Section 3.3) propose approximating a nonlinear $p_i(g,w)$ with its linear approximation around some fixed $g = \bar{g}$. This approach yields coefficients approximating those in the infeasible regressions of Appendix Proposition A2, discussed above. As with other imperfect predictions of $\mathbb{E}[x_i \mid g, w]$, the cost of such simpler recentering is a likely power loss from the linear approximation.

The third step is to adjust for $\mathbb{E}[\varepsilon\varepsilon' \mid w]^{-1}$. Per Proposition 3, this involves two feasible adjustments: residualizing on (i.e., controlling for) functions of w that predict ε, and weighting by an estimate of $\text{Var}[\varepsilon \mid w]^{-1}$ (akin to feasible generalized least squares). In non-*iid* or high-dimensional settings, both steps may be challenging without *a priori* restrictions on how the unobserved ε relates to w. Controlling for arbitrary functions of w may improve efficiency, though this is not guaranteed (see Appendix C.9 of Borusyak and Hull (2021) for a counterexample). Weighting by a flexible but noisy estimate of $\text{Var}[\varepsilon \mid w]^{-1}$ may introduce bias, as in conventional settings (e.g., Angrist and Pischke 2008, p. 69).

A practical alternative is to disregard the third step of Algorithm 1 and just instrument with the recentered $p_i(g,w)$ as an approximation to $\tilde{z}$. This approach has two formal justifications. First, it is straightforward to show that the unadjusted $\tilde{z}$ has the highest correlation with the treatment among all recentered IVs:

Lemma 1. $\tilde{z} \in \arg\max_{z \in \mathfrak{R}} \frac{\mathbb{E}[z'x]^2}{\mathbb{E}[z'z]\mathbb{E}[x'x]}$.

This result immediately implies that $\tilde{z}$ produces the highest first-stage R^2, and so approximations to it should have favorable power. Second, because of this property, $\tilde{z}$ minimizes the worst-case estimator variance absent sharp information on the conditional error distribution:

Lemma 2. *Let $\bar{\lambda} \geq 0$ be some constant and let $\mathcal{E}$ be the class of random vectors $e = (e_1, \ldots, e_N)'$ such that the maximum eigenvalue of $\mathbb{E}[ee' \mid g, w]$ is bounded by $\bar{\lambda}$ uniformly over (g, w). Suppose*

number of draws $J \geq 1$ is sufficient for identification. To see this, redefine the shocks as $\tilde{g} = \left(g, g^{(1)}, \ldots, g^{(J)}\right)$. The original Assumption 1 implies Assumption 1 with respect to $\tilde{g}$. Moreover, $p(g,w) - \frac{1}{J}\sum_{j=1}^J p(g^{(j)}, w)$ is in $\mathfrak{R}$ redefined with respect to $\tilde{g}$ because $\mathbb{E}\left[p(g,w) - \frac{1}{J}\sum_{j=1}^J p(g^{(j)}, w) \mid w\right] = 0$ when $g \mid w$ and $g^{(j)} \mid w$ have the same distribution.

[21] Formally, this requires including the permutation class of g in w.

[22] See Borusyak and Hull (2023) for a discussion of other approaches to recentering, such as using a theory-based approximation to $\mu_i = \mathbb{E}[p_i(g,w) \mid w]$ (as in Abdulkadiroglu et al. (2017)) or using an estimated μ_i based on first-step estimates of the shock assignment process.

the only knowledge a researcher has about ε is $\varepsilon \in \mathcal{E}$. Then $\tilde{z}$ is approximately minimax, in that:

$$\tilde{z} \in \arg\min_{z \in \mathfrak{R}} \max_{e \in \mathcal{E}} \frac{\text{Var}\left[z'e\right]}{\mathbb{E}\left[z'x\right]^2}.$$

This result captures the intuition that maximizing the first-stage R^2 is optimal when one has weak priors on the distribution of ε, and worries about making further incorrect adjustments.

4 Application

We illustrate our approach by estimating the private insurance crowdout effects of Medicaid eligibility, using the partial Affordable Care Act (ACA) expansion in 2014. Section 4.1 describes the setting and adapts our theoretical results to this setting while Section 4.2 presents the results.

4.1 Setting and Estimators

In January 2014, many US states expanded Medicaid eligibility under the ACA to cover non-elderly adults with incomes up to 138% of the federal poverty level (FPL). This expansion did not cover all states since a 2012 Supreme Court decision (NFIB v. Sebelius, 567 U.S. 519) let individual state governors opt out of the more generous ACA coverage level. In practice, expansion decisions partially followed partisan lines: among the 43 states with less generous Medicaid policies in 2013, only 8 out of 30 states with Republican governors expanded while 11 out of 13 states with Democratic governors did.[23] Non-expansion states mostly kept their 2013 eligibility rules in place, though some increased coverage slightly. Expansion states mostly adopted the ACA's 138% FPL threshold, though some extended coverage further.

We use state expansion decisions as policy shocks for estimating Medicaid eligibility effects. To formalize this approach, consider a repeated cross section of individuals i observed in years $t(i) \in \{2013, 2014\}$ with states of residence $s(i)$. We write i's Medicaid eligibility, $x_i \in \{0, 1\}$, as:

$$x_i = h^{t(i)}(c_i, e_{s(i)}^{2013}, g_{s(i)}, e_{s(i)}^{\Delta}) \tag{6}$$

where c_i collects relevant individual characteristics (income, work status, and parental status), e_k^{2013} is state k's Medicaid eligibility policy in 2013, $g_k \in \{0, 1\}$ indicates whether or not state k expanded coverage under the ACA in 2014, and e_k^{Δ} includes other 2014 changes to Medicaid coverage (i.e. non-ACA coverage increases or ACA expansions beyond the 138% FPL threshold). These inputs are sufficient to determine individual i's eligibility, as formalized by the $h^{2013}(\cdot)$ and $h^{2014}(\cdot)$ functions.[24]

We assume the expansion shocks are exogenous, conditional on the political party of states' governors, when estimating eligibility effects in a difference-in-differences setup. Formally, we relate

[23] See Frean et al. (2017) for more background on the partial Medicaid expansions and related ACA policy changes.

[24] Here, as in the initial Section 2 motivating example, we assume income (and other characteristics) do not respond to expansion decisions. In our repeated cross-section, we do not observe pre-period income so cannot apply the extension at the end of that section.

individual outcomes to eligibility in the repeated cross section by:

$$y_i = \beta x_i + \alpha_{s(i)} + \tau_{t(i)} + \varepsilon_i, \tag{7}$$

where α_k and τ_t are state and time fixed effects. Let q_k^{2013} be an indicator for state k having a Republican governor in 2013. Following Assumption 1, we formalize shock exogeneity as $\mathbb{E}\left[\varepsilon_i \mid g, w\right] = \mathbb{E}\left[\varepsilon_i \mid w\right]$ where g collects the 43 expansion dummies and w collects all c_i, $s(i)$, $t(i)$, e_k^{2013}, and q_k^{2013}. This assumption is consistent with earlier difference-in-differences analyses that compare outcome trends of expansion and non-expansion states before and after 2014, adjusting for state party or other observables (e.g. Averett et al. (2019); Miller and Wherry (2017)). We impose no assumptions on other coverage changes e_k^{Δ}, which we collect in u.

Our baseline design assumption is that the shocks are drawn from the same (unknown) distribution among states with the same-party governor. This view of the 2014 expansion decisions, as arising from a natural experiment, conforms with some earlier analyses (e.g. Black et al., 2019) and allows us to construct counterfactual $g^{(j)}$ vectors by permuting the shocks conditional on state party. That is, all $g^{(j)}$ with expansions in some 8 Republican-governor states and some 11 Democratic-governor states are valid shock counterfactuals. Below we check sensitivity to richer designs, which allow expansion probabilities to vary by additional state observables.

Following Section 2, we consider two IV specifications using these assumptions. First, we construct a simulated instrument $z_i = f^{t(i)}\left(g_{s(i)}\right)$ which only leverages the shock variation. Since the policy only changes in 2014, we set z_i to zero for all individuals in 2013 as well as those in non-expansion states. For the expansion states in 2014, the instrument measures the change in the policy generosity. Specifically, it equals the difference in the fraction of the 2014 nationally representative sample who would be eligible for Medicaid between two situations: had all states adopted the ACA's 138% FLP threshold vs. had all of them kept their 2013 policies intact. Exogeneity of the expansion shocks and the design assumption make this z_i uncorrelated with ε_i controlling for state fixed effects, year fixed effects, and the interaction of the state party indicator $q_{s(i)}$ and year. These controls furthermore make estimation with z_i equivalent to instrumenting by the interaction of $g_{s(i)}$ and a 2014 indicator, as in an instrumented difference-in-differences specification.[25]

Next, we construct an approximation to the recentered best predictor $\tilde{z}_i$ by applying the first two steps of Algorithm 1 to the formula for Medicaid eligibility. In the first step, we predict eligibility from the exogenous policy shocks g and predetermined variables w, ignoring the non-ACA eligibility changes in u. In our notation, this involves replacing e_k^{Δ} with $\emptyset$ in equation (6) which yields a predicted eligibility of $p_i(g, w) = h^{t(i)}(c_i, e_{s(i)}^{2013}, g_{s(i)}, \emptyset)$. Second, we recenter this prediction by subtracting its expectation μ_i with respect to the expansion shocks, conditional on individual characteristics and 2013 policies. Consistent with our design assumption, we account for

[25] Indeed, given state and year fixed effects, the specific values of the instrument in the four cells (2013 vs. 2014 and treated vs. untreated states) are immaterial as long as z_i increases more in 2014 in the treated states.

differences in expansion probabilities by state party when taking the expectation:

$$\mu_i \equiv \mathbb{E}\left[p_i(g,w) \mid w\right] = \pi(q_{s(i)}) \cdot h^{t(i)}\left(c_i, e^{2013}_{s(i)}, 1, \emptyset\right) + (1-\pi(q_{s(i)})) \cdot h^{t(i)}\left(c_i, e^{2013}_{s(i)}, 0, \emptyset\right),$$

where $\pi(0) = 11/13$ and $\pi(1) = 8/30$ correspond to the fractions of expansion states among those with Democratic and Republican governors, respectively.

The recentered instrument $p_i(g,w) - \mu_i$ equals zero for all individuals in 2013, as well as individuals in 2014 whose characteristics make them eligible (or ineligible) with or without an ACA expansion in their state. It only varies among individuals in 2014 whose eligibility is affected by expansion decisions. We thus restrict estimation to these "exposed" individuals in 2014 and, in keeping with the difference-in-differences structure, the corresponding group of individuals in 2013 (i.e., those whose characteristics and state of residence would make them exposed to the expansion shocks in 2014)—what we call the exposed sample. We control for state fixed effects, year fixed effects, and the interaction of the state party indicator $q_{s(i)}$ and year, as in the simulated IV specification. With these controls and the restriction to the exposed sample, recentered IV estimation is again equivalent to instrumenting by the interaction of $g_{s(i)}$ and a 2014 indicator.[26]

We apply both IV strategies using data from the 2013 and 2014 American Community Surveys, using representative 1% samples of non-disabled U.S. adults (ages 21-64) residing in the 43 states eligible for ACA expansion in 2014. Appendix C.1 details the sample construction.

Our primary outcomes are indicators for Medicaid enrollment and for private insurance coverage. Effects on the former outcome capture takeup of the expanded Medicaid coverage; effects on the latter outcome capture how Medicaid eligibility crowds out other forms of insurance—an important policy parameter in the literature (e.g. Frean et al., 2017; Leung and Mas, 2018). We also look at effects on an indicator for employer-sponsored insurance coverage, which is the most common type of private insurance and is the focus of the classic literature on Medicaid crowdout (e.g. Cutler and Gruber (1996)). In our setting, Medicaid may also crowdout private insurance obtained directly on ACA state health exchanges (as documented by Frean et al. (2017)). This would have different economic implications, as such crowdout would not typically have employment effects.

4.2 Results

Our recentered IV is much more predictive of actual Medicaid eligibility than the simulated IV. First-stage estimates in Table 1 show that the simulated IV predicts eligibility with a coefficient of 0.85 and a standard error of 0.11 (column 1) while the recentered IV in the exposed sample has a higher coefficient of 0.97 and a smaller standard error of 0.02 (column 2); the latter coefficient is statistically indistinguishable from one, as should be the case with the recentered best predictor.[27]

[26] This is because the recentered instrument can be written in the exposed sample as $g_{s(i)} \times \mathbf{1}[t(i) = 2014] - \pi(0) \times \mathbf{1}[t(i) = 2014] - (\pi(1) - \pi(0)) \times q_{s(i)} \times \mathbf{1}[t(i) = 2014]$ with the latter two terms absorbed by the controls.

[27] Throughout we report state-clustered standard errors. To address finite-sample concerns with a relatively small number of states, we report confidence intervals by a wild score bootstrap as suggested by Kline and Santos (2012) and use them for hypothesis testing. This computationally efficient approach requires inverting bootstrapped test statistics, which generally makes confidence intervals asymmetric around the IV point estimate.

Table 1: Medicaid Application: First-Stage Effects on Eligibility

	(1)	(2)	(3)
Simulated IV	0.851		0.032
	(0.113)		(0.140)
	[0.585,1.108]		[-0.252,0.479]
Recentered IV		0.972	0.817
		(0.015)	(0.171)
		[0.941,1.014]	[0.394,1.161]
Partial R^2	0.012	0.799	0.103
Exposed Sample	N	Y	N
States	43	43	43
Individuals	2,397,313	421,042	2,397,313

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

The partial R^2 for the recentered IV is also dramatically higher (0.80, vs. 0.01 for the simulated IV). Column 3 of the table shows that just adding the recentered IV to the simulated IV specification increases the partial R^2 meaningfully (to 0.10) and renders the simulated IV coefficient small and statistically insignificant. All of these results are consistent with the recentered IV giving a better approximation to the recentered best predictor $\tilde{z}_i$.

This improved first-stage prediction translates to meaningful precision gains for the effects of eligibility on Medicaid and private insurance coverage. Columns 1 and 2 of Table 2, Panel A, show the standard error is 64% smaller with recentered IV vs. simulated IV (0.010 vs. 0.028) when estimating the effects of Medicaid eligibility on Medicaid enrollment. For crowdout effects, which take private insurance coverage as the outcome, standard errors are reduced by 70% (0.007 vs. 0.023; columns 3 and 4) from an insignificant simulated IV estimate to a significant recentered IV estimate. These estimates incorporate effects from both the conventional crowdout margin of employer-sponsored insurance as well as crowdout from ACA state health exchanges; in columns 5 and 6 we isolate crowd-out of employer-sponsored plans. Here neither simulated nor recentered IV yields significant estimates, though the latter is again much more precise.

In economic terms, the recentered IV estimates suggest a total private insurance crowdout rate of 32.1%, with a 7.2 percentage point increase in Medicaid coverage offset by a 2.3 percentage point decrease in private insurance coverage. This relative effect, reported in Panel B column 4 as the coefficient from an IV regression of private insurance coverage on Medicaid enrollment (instead of eligibility), is similar to the 42% crowd-out that Leung and Mas (2018) find in applying a difference-

Table 2: Medicaid Application: IV Estimates

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Simulated IV	Recentered IV	Simulated IV	Recentered IV	Simulated IV	Recentered IV
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Medicaid Eligibility Effects						
Eligible	0.132	0.072	−0.048	−0.023	0.009	−0.009
	(0.028)	(0.010)	(0.023)	(0.007)	(0.014)	(0.005)
	[0.082,0.198]	[0.050,0.094]	[-0.127,0.037]	[-0.040,-0.008]	[-0.057,0.070]	[-0.020,0.003]
Panel B. Medicaid Enrollment Effects						
Has Medicaid			−0.361	−0.321	0.068	−0.125
			(0.165)	(0.092)	(0.111)	(0.061)
			[-1.156,0.382]	[-0.549,-0.119]	[-0.494,0.734]	[-0.252,0.056]
Exposed Sample	N	Y	N	Y	N	Y
States	43	43	43	43	43	43
Individuals	2,397,313	421,042	2,397,313	421,042	2,397,313	421,042

Notes: Panel A of this table reports second-stage coefficients from the two IV regressions described in the text: one using a conventional simulated instrument and the other using as an instrument a recentered prediction of Medicaid eligibility. Columns 1, 3, and 5 estimate regressions in the full sample of individuals in 2013–2014, while Columns 2, 4, and 6 restrict to the sample of individuals whose characteristics and state of residence make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year. Panel B shows estimates from IV regressions which use an indicator for Medicaid enrollment as the endogenous variable, instead of an indicator for Medicaid eligibility. State-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

in-differences specification to the 2014 Medicaid expansion.[28] However, we find minimal evidence for crowdout from employer-sponsored insurance plans even with our more powerful recentered IV. Instead, our estimates suggest crowdout arises from direct-purchase private insurance via state health exchanges (implying minimal employment effects). This aligns with the findings of Frean et al. (2017), who exploit multiple sources of ACA-induced policy variation with a simulated instrument (see also Courtemanche et al. (2017), Kaestner et al. (2017), and Maclean and Saloner (2019)). For both private insurance and employer-sponsored insurance, the standard errors of the recentered IV crowdout parameter estimates in Panel B are around 44% smaller than those of simulated IV.

Importantly, the recentered IV power gains are not a product of the relatively simple simulated IV specification. Appendix Table A1 shows that adding flexible controls for the individual characteristics which drive exposure to Medicaid expansions (specifically: full interactions of household income deciles, parental status, work status, and year) has little effect on the point estimates and standard errors for both estimators. Here the similar power of recentered IV with and without flexible controls suggests little added benefit from trying to approximate the adjustment for ψ in the optimal IV after approximating the recentered best predictor.

Three additional checks are presented in the Appendix. First, we check the assumption of expansion exogeneity with a pre-trend test that replaces the 2013-2014 difference-in-differences IVs with a comparable 2012-2013 analysis.[29] Although with the increased precision of recentered IV we are able to reject the null hypothesis of no pre-trends, Appendix Table A2 shows that the magnitude of the placebo coefficient is small (around 0.01–0.02) regardless of the outcome and the instrument. Second, we check sensitivity to our baseline design assumption by allowing a state's decision to expand to depend not only on the governor's political party but also on the state's median household income and the 2012 rate of Medicaid coverage (specifically, by adding as controls a quadratic in these three state characteristics, interacted with year dummies). Appendix Table A3 shows that estimated effects of eligibility remain very similar across these specifications. Third, we explore robustness to using the recentered IV without restricting to the exposed sample. Appendix Table A4 shows that this approach only yields power gains when the additional demographic controls (those from Appendix Table A1) or an indicator for being in the exposed sample interacted with year are included as covariates. We discuss the reason for this in Appendix C.3 by relating it to our general efficiency theory of Section 3.2.

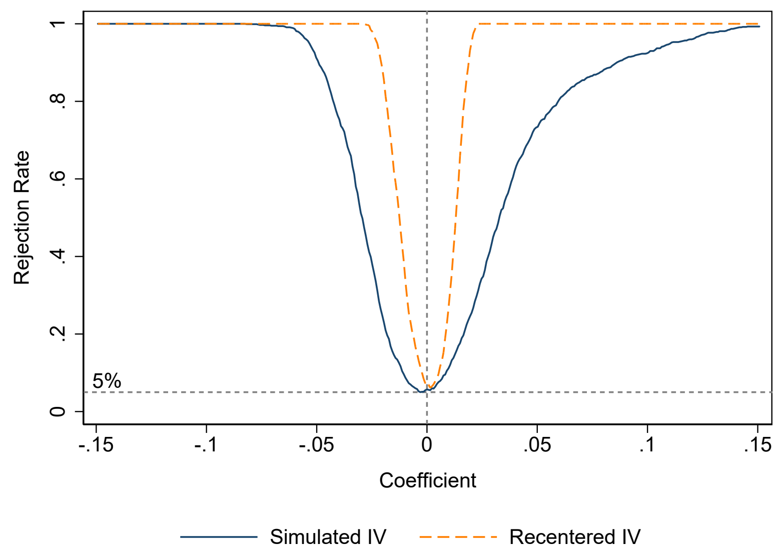
Large power gains from recentered IV are confirmed in a Monte Carlo study based on our baseline estimates. Figure 1 plots simulated power curves for the primary Medicaid enrollment

[28]The corresponding simulated IV specification yields a private insurance crowd-out rate of 36.1%, reported in Panel B column 3, which is not statistically distinguishable from the recentered IV estimate ($p = 0.719$). Recentered and simulated IV also yield statistically indistinguishable estimates for the employer-sponsored insurance outcome ($p = 0.104$), reported in Panel B columns 5-6. In contrast, the recentered and simulated IV estimates which use Medicaid eligibility as the endogenous variable are statistically distinguishable, while they should have the same estimating according to the theory presented thus far. Appendix C.2 discusses how this pattern can be explained by measurement error in self-reported income and demographics entering the eligibility calculation. The specifications which use Medicaid enrollment as the endogenous variable are free from such bias.

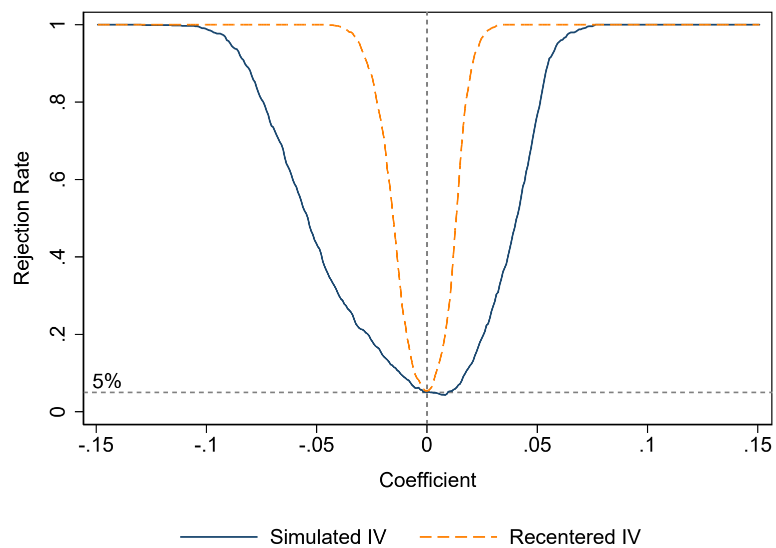
[29]Specifically, we replace 2013 individuals with 2012 individuals and replace 2014 individuals with 2013 individuals. We continue to construct the endogenous variable and instrument as an individual's Medicaid eligibility in 2013 and 2014 for comparability, and also keep all controls unchanged.

Figure 1: Medicaid Application: Simulated Power Curves

A. Has Medicaid



B. Has Employer-Sponsored Insurance



Notes: This figure plots the simulated rejection rates of the two IV regressions discussed in the text: one using a conventional simulated instrument and the other using as an instrument a recentered prediction of Medicaid eligibility. See Appendix C.3 for a description of the simulation procedure. Rejection rates are for nominal 5%-level tests of each coefficient based on wild score bootstraps, clustered by state. The true effect of zero is indicated by the dashed vertical line. The nominal 5% level of the tests is indicated by the dashed horizontal lines.

outcome and the employer-sponsored insurance coverage outcome, while Appendix Figure A1 plots the simulated estimator distributions (Appendix C.4 details the simulation procedure). In this controlled environment the true causal effect and the shock assignment process are known, allowing us to verify that the recentered IV estimator is both close to unbiased and substantially more efficient than the simulated IV estimator. We find, for example, that the minimum detectable effects of recentered IV (i.e., the smallest null hypotheses which are rejected by a 0.05-size test with probability 0.8) are roughly three times smaller than those of simulated IV.

5 Conclusion

Many economic treatments are given by formulas which incorporate multiple sources of variation, only some of which are exogenous. Rather than discarding the other non-random variation, we show how it can be used to construct powerful formula instruments. By combining the known treatment structure with knowledge of the design of exogenous shocks, researchers can construct recentered instruments which strongly predict the treatment. We show how such recentered best predictors are formally justified, and how they can be further adjusted to approximate the asymptotically optimal formula instrument. Empirically, we show substantial power gains from using such instruments to estimate the takeup and crowdout effects of Medicaid eligibility—with standard errors around half the size of those from a conventional simulated instrument approach.

Importantly, while we have focused on the popular setting of simulated instruments, the insights of this paper may apply to a large class of formula treatments and instruments from a variety of fields. These include linear and nonlinear shift-share instruments, treatments capturing spillovers across social networks or geography, instruments based on centralized school assignment mechanisms, "free-space" instruments capturing access to mass media, and treatments or instruments leveraging weather shocks (Borusyak and Hull, 2021). We expect this class to only increase as researchers find new and creative ways to exploit exogenous variation in complex treatments.

References

Abdulkadiroglu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 2017, *85* (5), 1373–1432.

Ai, Chunrong and Xiaohong Chen, "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 2003, *71* (6), 1795–1843.

Anatolyev, Stanislav, "Optimal instruments in time series: A survey," *Journal of Economic Surveys*, 2007, *21* (1), 143–173.

Angrist, Joshua D. and JS Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press, 2008.

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

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

Berry, Steven, James Levinsohn, and Ariel Pakes, "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *American Economic Review*, 1999, *89* (3), 400–430.

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

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

_ and _ , "Non-Random Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

_ , _ , and Xavier Jaravel, "Design-based identification with formula instruments: A review," *Econometrics Journal*, 2025.

_ , _ , and _ , "A practical guide to shift-share instruments," *Journal of Economic Perspectives*, 2025, *39* (1), 181–204.

Boustan, Leah, Fernando Ferreira, Hernan Winkler, and Eric M. Zolt, "The effect of rising income inequality on taxation and public expenditures: Evidence from U.S. Municipalities and school districts, 1970-2000," *Review of Economics and Statistics*, 2013, *95* (4), 1291–1302.

Brooks, Tricia, "What Some Researchers Get Wrong About Medicaid's Income Eligibility Requirements," *Health Affairs Blog*, 2019.

Brown, David W., Amanda E. Kowalski, and Ithai Z. Lurie, "Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood," *NBER Working Paper No. 20835*, 2018.

Chamberlain, Gary, "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 1987, *34* (3), 305–334.

— , "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 1992, *60* (3), 567–596.

Cohodes, Sarah R., Daniel S. Grossman, Samuel A. Kleiner, and Michael F. Lovenheim, "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions," *Journal of Human Resources*, 2016, *51* (3), 727–759.

Conlon, Christopher and Jeff Gortmaker, "Best practices for differentiated products demand estimation with PyBLP," *RAND Journal of Economics*, 2020, *51* (4), 1108–1161.

Courtemanche, Charles, James Marton, Benjamin Ukert, Aaron Yelowitz, and Daniela Zapata, "Early impacts of the Affordable Care Act on health insurance coverage in Medicaid expansion and non-expansion states," *Journal of Policy Analysis and Management*, 2017, *36* (1), 178–210.

Coussens, Stephen and Jann Spiess, "Improving Inference from Simple Instruments Through Compliance Estimation," *Mimeo*, 2021.

Cullen, Julie Berry and Jonathan Gruber, "Does Unemployment Insurance Crowd out Spousal Labor Supply?," *Journal of Labor Economics*, 2000, *18* (3), 546–572.

Currie, Janet and Jonathan Gruber, "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *The Quarterly Journal of Economics*, 1996, *111* (2), 431–466.

— **and** — , "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," *Journal of Political Economy*, 1996, *104* (6), 1263–1296.

Cutler, David M. and Jonathan Gruber, "Does Public Insurance Crowd out Private Insurance?," *The Quarterly Journal of Economics*, 1996, *111* (2), 391–430.

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

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers, "Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act," *Journal of Health Economics*, 2017, *53* (September 2010), 72–86.

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

Hackmann, Martin B., "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry," *American Economic Review*, 2019, *109* (5), 1684–1716.

Hansen, Lars Peter, "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators," *Journal of Econometrics*, 1985, *30* (1-2), 203–238.

Kaestner, Robert, Bowen Garrett, Jiajia Chen, Anuj Gangopadhyaya, and Caitlyn Fleming, "Effects of ACA Medicaid Expansions on Health Insurance Coverage and Labor Supply," *Journal of Policy Analysis and Management*, 2017, *36* (3), 608–642.

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

Leung, Pauline and Alexandre Mas, "Employment Effects of the ACA Medicaid Expansions," *Industrial Relations*, 2018, *57* (2), 206–234.

Maclean, Johanna Catherine and Brendan Saloner, "The Effect of Public Insurance Expansions on Substance Use Disorder Treatment: Evidence from the Affordable Care Act," *Journal of Policy Analysis and Management*, 2019, *38* (2), 366–393.

Miller, Sarah and Laura R Wherry, "Health and access to care during the first 2 years of the ACA Medicaid expansions," *New England Journal of Medicine*, 2017, *376* (10), 947–956.

Newey, Whitney K., "The Asymptotic Variance of Semiparametric Estimators," *Unpublished manuscript*, 1989.

Newey, Whitney K and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 1994, *4*, 2111–2245.

Poirier, Alexandre, "Efficient estimation in models with independence restrictions," *Journal of Econometrics*, 2017, *196* (1), 1–22.

Robinson, P.M., "Root-N-Consistent Semiparametric Regression," *Econometrica*, 1988, *56* (4), 931–954.

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

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

A Additional Theoretical Results

Theorem A1. *Suppose Assumption 1 holds and* $\mathbb{E}[\varepsilon\varepsilon' \mid g, w]$ *is a.s. invertible. Then the recentered instrument*

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

solves

$$z^* \in \arg\min_{z \in \mathfrak{R}} \mathcal{V}\left[z\right].$$

Moreover, for $b = \mathbb{E}[x \mid g, w]$ *and* $A = \mathbb{E}[\varepsilon\varepsilon' \mid g, w]$,

$$\mathcal{V}\left[z^*\right] = \mathbb{E}\left[b'A^{-1}b - \mathbb{E}\left[b'A^{-1} \mid w\right] \mathbb{E}\left[A^{-1} \mid w\right]^{-1} \mathbb{E}\left[A^{-1}b \mid w\right]\right]^{-1}.$$

Proposition A1. *Suppose the observations of* $(y_i, x_i, \varepsilon_i, g_i, w_i)$ *are* iid *across* i *(with distributions not changing with* N*) and that Assumptions 1 and 2 hold with* $w = (w_i)_{i=1}^N$ *and* $g = (g_i)_{i=1}^N$. *Suppose further* $\mathbb{E}\left[\varepsilon_i^2 \mid g_i, w_i\right] = \mathbb{E}\left[\varepsilon_i^2 \mid w_i\right]$ *as in Theorem 1. Then if* $\hat{\beta}\left[z^*\right]$ *is regular and the sequence* $\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}$ *is uniformly integrable,* $\hat{\beta}\left[z^*\right]$ *asymptotically achieves the Chamberlain (1992) semi-parametric efficiency bound.*

For the next result, let $\mathfrak{R}^S \subset \mathfrak{R}$ denote the class of recentered shift-share instruments $z = S(w) \cdot (g - \mathbb{E}[g \mid w])$ that are characterized by an $N \times K$ matrix $S(w)$ that is measurable with respect to w.

Proposition A2. *Suppose assumptions of Theorem 1 hold and* $\text{Var}[g \mid w]$ *is a.s. invertible. Let* $\tilde{g} = g - \mathbb{E}[g \mid w]$. *Consider the recentered shift-share instrument*

$$z^S = \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1} \tilde{z}^S \quad \text{for } \tilde{z}^S = \mathbb{E}\left[x\tilde{g}' \mid w\right] \text{Var}\left[\tilde{g} \mid w\right]^{-1} \tilde{g}.$$

The associated $\hat{\beta}\left[z^S\right]$ *has the smallest approximate variance of all recentered shift-share IV estimators*

$$z^S \in \arg\min_{z \in \mathfrak{R}^S} \mathcal{V}\left[z\right].$$

B Proofs

B.1 Proof of Proposition 1

The first statement of the proposition has been established by Borusyak and Hull (2023). We now prove the second statement, focusing for clarity on the case of discrete g, w, and z. The continuously-distributed case follows similarly, under appropriate regularity conditions.

To show that z has to be a formula instrument, i.e. measurable with respect to g, w, suppose by contradiction there is an observation ι and values $(\bar{g}, \bar{w})$ in the support of (g, w) such that z_ι

takes distinct values $\bar{z}_1$ and $\bar{z}_2$ with probabilities $\pi_{\bar{z}_j|\bar{g},\bar{w}} > 0$, $j = 1, 2$. Consider the following data-generating process for ε: $\varepsilon_i = 0$ for $i \neq \iota$ and

$$\varepsilon_\iota = \begin{cases} (\bar{z}_1 - \bar{z}_2)/\pi_{\bar{z}_1|\bar{g},\bar{w}} & g = \bar{g}, w = \bar{w}, z_\iota = \bar{z}_1 \\ -(\bar{z}_1 - \bar{z}_2)/\pi_{\bar{z}_2|\bar{g},\bar{w}} & g = \bar{g}, w = \bar{w}, z_\iota = \bar{z}_2 \\ 0, & \text{otherwise.} \end{cases}$$

It is straightforward to verify that this ε satisfies $0 = \mathbb{E}[\varepsilon \mid g, w] = \mathbb{E}[\varepsilon \mid w]$. Yet,

$$\begin{aligned} \mathbb{E}\left[z'\varepsilon\right] &= \left(\bar{z}_1 \cdot (\bar{z}_1 - \bar{z}_2) - \bar{z}_2 \cdot (\bar{z}_1 - \bar{z}_2)\right) \cdot Pr\left(g = \bar{g}, w = \bar{w}\right) \\ &= (\bar{z}_1 - \bar{z}_2)^2 \cdot Pr\left(g = \bar{g}, w = \bar{w}\right) \neq 0. \end{aligned}$$

violating (2).

It remains to be shown that z must be a recentered instrument. Consider a different data-generating process for ε, in which $\mathbb{E}[\varepsilon \mid w] = \mathbb{E}[z \mid w]$. Then

$$\mathbb{E}\left[z'\varepsilon\right] = \mathbb{E}\left[z'\mathbb{E}[\varepsilon \mid g, w]\right] = \mathbb{E}\left[z'\mathbb{E}[\varepsilon \mid w]\right] = \mathbb{E}\left[\mathbb{E}[z \mid w]'\,\mathbb{E}[\varepsilon \mid w]\right] = \mathbb{E}\left[\mathbb{E}[z \mid w]'\,\mathbb{E}[z \mid w]\right] \neq 0,$$

unless $\mathbb{E}[z \mid w] = 0$ a.s., violating (2) again.

B.2 Proof of Proposition 2

Uniform integrability of $\frac{1}{N}z'x$ implies $\mathbb{E}\left[\frac{1}{N}z'x\right] \to M$. Then, denoting the asymptotic distribution of $r_N\left(\hat{\beta}[z] - \beta\right)$ by $\tilde{\mathcal{D}}$ and using the continuous mapping theorem,

$$r_N \frac{\frac{1}{N}z'\varepsilon}{\mathbb{E}\left[\frac{1}{N}z'x\right]} = r_N\left(\hat{\beta}[z] - \beta\right) \cdot \frac{\frac{1}{N}z'x}{M} \cdot \frac{M}{\mathbb{E}\left[\frac{1}{N}z'x\right]} \Rightarrow \tilde{\mathcal{D}}, \tag{A1}$$

as $\frac{\frac{1}{N}z'x}{M} \xrightarrow{p} 1$, and $\frac{M}{\mathbb{E}\left[\frac{1}{N}z'x\right]} \to 1$. Furthermore, uniform integrability of $\left(r_N \frac{1}{N}z'\varepsilon\right)^2$ implies

$$\text{Var}\left[r_N \frac{\frac{1}{N}z'\varepsilon}{\mathbb{E}\left[\frac{1}{N}z'x\right]}\right] = r_N^2 \mathcal{V}[z] \to V, \tag{A2}$$

which is equivalent to the Proposition's claim.

B.3 Proof of Theorem 1 and Appendix Theorem A1

Since Theorem 1 is a special case of Appendix Theorem A1 when $\mathbb{E}[\varepsilon\varepsilon' \mid g, w] = \mathbb{E}[\varepsilon\varepsilon' \mid w]$ a.s., we focus on the latter.

First, we confirm that $z^* \in \mathfrak{R}$:

$$\begin{aligned}\mathbb{E}\left[z^* \mid w\right] &= \mathbb{E}\left[\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1} \mathbb{E}\left[x \mid g, w\right] \mid w\right] \\ &\quad - \mathbb{E}\left[\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1} \mid w\right] \mathbb{E}\left[\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1} \mid w\right]^{-1} \mathbb{E}\left[\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1} \mathbb{E}\left[x \mid g, w\right] \mid w\right] \\ &= 0.\end{aligned}$$

Second, we show that $\mathbb{E}\left[z'\varepsilon\varepsilon' z^*\right] = \mathbb{E}\left[z'x\right]$ for any $z \in \mathfrak{R}$:

$$\begin{aligned}\mathbb{E}\left[z'\varepsilon\varepsilon' z^*\right] &= \mathbb{E}\left[z'\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right] z^*\right] \\ &= \mathbb{E}\left[z'\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1} \mid w\right]^{-1} \mathbb{E}\left[\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1} \mathbb{E}\left[x \mid g, w\right] \mid w\right]\right)\right] \\ &= \mathbb{E}\left[z'x\right] - 0,\end{aligned}$$

where the last line follows because z is mean-zero conditional on any function of w. Note that $\mathbb{E}\left[z'\varepsilon\varepsilon' z^*\right] = \mathbb{E}\left[z'x\right]$ also implies $\mathcal{V}\left[z^*\right] = \mathbb{E}\left[z^{*\prime}x\right]^{-1}$.

Following the proof of Theorem 5.3 in Newey and McFadden (1994), let

$$U = \frac{z'\varepsilon}{\mathbb{E}\left[z'x\right]} - \frac{z^{*\prime}\varepsilon}{\mathbb{E}\left[z^{*\prime}x\right]}.$$

Then

$$\begin{aligned}\mathbb{E}\left[U^2\right] &= \frac{\operatorname{Var}\left[z'\varepsilon\right]}{\mathbb{E}\left[z'x\right]^2} - 2\frac{\mathbb{E}\left[z'\varepsilon\varepsilon' z^*\right]}{\mathbb{E}\left[z'x\right]\mathbb{E}\left[z^{*\prime}x\right]} + \frac{\mathbb{E}\left[z^{*\prime}\varepsilon\varepsilon' z^*\right]}{\mathbb{E}\left[z^{*\prime}x\right]^2} \\ &= \frac{\operatorname{Var}\left[z'\varepsilon\right]}{\mathbb{E}\left[z'x\right]^2} - \frac{1}{\mathbb{E}\left[z^{*\prime}x\right]} \\ &= \mathcal{V}\left[z\right] - \mathcal{V}\left[z^*\right] \geq 0,\end{aligned}$$

showing that z^* minimizes the approximate estimator variance.

It remains to obtain the expression for $\mathcal{V}\left[z^*\right]$. We have:

$$\begin{aligned}\mathcal{V}\left[z^*\right] &= \mathbb{E}\left[z^{*\prime}x\right]^{-1} \\ &= \mathbb{E}\left[\left(b - \mathbb{E}\left[A^{-1} \mid w\right]^{-1} \mathbb{E}\left[A^{-1}b \mid w\right]\right)' A^{-1}x\right] \\ &= \mathbb{E}\left[\left(b - \mathbb{E}\left[A^{-1} \mid w\right]^{-1} \mathbb{E}\left[A^{-1}b \mid w\right]\right)' A^{-1}b\right] \\ &= \mathbb{E}\left[b'A^{-1}b - \mathbb{E}\left[b'A^{-1} \mid w\right] \mathbb{E}\left[A^{-1} \mid w\right]^{-1} A^{-1}b\right] \\ &= \mathbb{E}\left[b'A^{-1}b - \mathbb{E}\left[b'A^{-1} \mid w\right] \mathbb{E}\left[A^{-1} \mid w\right]^{-1} \mathbb{E}\left[A^{-1}b \mid w\right]\right].\end{aligned}$$

B.4 Proof of Proposition 3

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

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

Thus:

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

B.5 Proof of Lemma 1

We have $\mathbb{E}[z'x] = \mathbb{E}[z'\mathbb{E}[x \mid g, w]] = \mathbb{E}[z'\tilde{z}]$, where the first equality follows from the law of iterated expectations and the second equality holds because $\mathbb{E}[z'\mathbb{E}[x \mid w]] = 0$ for any $z \in \mathfrak{R}$. Thus, maximizing $\frac{\mathbb{E}[z'x]^2}{\mathbb{E}[z'z]\mathbb{E}[x'x]}$ is equivalent to maximizing $\frac{\mathbb{E}[z'\tilde{z}]^2}{\mathbb{E}[z'z]\mathbb{E}[\tilde{z}'\tilde{z}]}$. By the Cauchy-Schwarz inequality, this ratio attains its maximum of one at $z = \tilde{z}$.

B.6 Proof of Lemma 2

Since $\mathbb{E}[z] = 0$,

$$\operatorname{Var}\left[z'\varepsilon\right] = \mathbb{E}\left[z'\varepsilon\varepsilon'z\right] = \mathbb{E}\left[z'\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right] z\right].$$

By standard results in linear algebra,

$$\max_{e \in \mathcal{E}} z'\mathbb{E}\left[ee' \mid g, w\right] z = \bar{\lambda}z'z,$$

which is achieved when $\mathbb{E}[ee' \mid g, w] = \bar{\lambda}\frac{zz'}{z'z}$ a.s. Thus, the minimax problem simplifies to

$$\min_{z \in \mathfrak{R}} \frac{\mathbb{E}\left[z'z\right]}{\mathbb{E}\left[z'x\right]^2},$$

which is equivalent to $\max_{z \in \mathfrak{R}} \frac{\mathbb{E}[z'x]^2}{\mathbb{E}[z'z]}$ and thus to $\max_{z \in \mathfrak{R}} \frac{\mathbb{E}[z'x]^2}{\mathbb{E}[z'z]\mathbb{E}[x'x]}$. By Lemma 1, $\tilde{z}$ is a solution.

B.7 Proof of Appendix Proposition A1

By Proposition 2, $\mathcal{V}[z^*]$ is the (rescaled) asymptotic variance of the estimator. We have:

$$\begin{aligned}\mathcal{V}[z^*] &= \mathbb{E}\left[\tilde{z}'\left(\Omega+\psi\psi'\right)^{-1}\tilde{z}\right]^{-1} \\ &= \mathbb{E}\left[\tilde{z}'\left(\Omega^{-1}-\Omega^{-1}\psi\frac{\psi'\Omega^{-1}}{1+\psi'\Omega^{-1}\psi}\right)\tilde{z}\right]^{-1} \\ &= \mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}-\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}\right]^{-1},\end{aligned}$$

where the first line follows from Theorem 1, the second line plugs in the result of Proposition 3, and the third line rearranges terms.

We next show that this expression asymptotically coincides with $\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right]^{-1}$. Note that

$$\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}-\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}\right]^{-1}-\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right]^{-1}=\frac{\mathbb{E}\left[\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}\right]}{\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right]\left(\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right]-\mathbb{E}\left[\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}\right]\right)}$$

and, since $\mathrm{Var}\left[\tilde{z}_i \mid w_i\right]\neq 0$ with positive probability by Assumption 2,

$$\begin{aligned}\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right] &= \sum_i \mathbb{E}\left[\frac{\tilde{z}_i^2}{\mathrm{Var}\left[\varepsilon_i \mid w_i\right]}\right] \\ &= \mathbb{E}\left[\frac{\mathrm{Var}\left[\tilde{z}_i \mid w_i\right]}{\mathrm{Var}\left[\varepsilon_i \mid w_i\right]}\right]\times N \\ &\to \infty,\end{aligned}$$

so long as $\mathbb{E}\left[\frac{\mathrm{Var}[\tilde{z}_i|w_i]}{\mathrm{Var}[\varepsilon_i|w_i]}\right]$ exists. To establish the result, it remain to show that $\mathbb{E}\left[\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}\right]$ does not diverge to infinity. This is obvious if $\psi_i=0$ a.s. Otherwise, $\frac{1}{\sqrt{N}}\tilde{z}'\Omega^{-1}\psi=O_p(1)$ by the central limit theorem while $\frac{1}{N}\left(1+\psi'\Omega^{-1}\psi\right)\xrightarrow{p}\mathbb{E}\left[\psi_i^2/\mathrm{Var}\left[\varepsilon_i \mid w_i\right]\right]>0$. Thus $\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}=O_p(1)$. By uniform integrability of $\frac{\left(\tilde{z}'\Omega^{-1}\psi\right)^2}{1+\psi'\Omega^{-1}\psi}$, its expectation converges to the finite expectation of the limit distribution: specifically, to

$$\frac{\mathbb{E}\left[\mathrm{Var}\left[\tilde{z}_i \mid w_i\right]\psi_i^2/\mathrm{Var}\left[\varepsilon_i \mid w_i\right]\right]}{\mathbb{E}\left[\psi_i^2/\mathrm{Var}\left[\varepsilon_i \mid w\right]\right]}.$$

Finally, we show that $\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right]^{-1}$ coincides with the semi-parametric efficiency bound in this setting. Following the notation of Chamberlain (1992), we have a model that satisfies the moment condition $E[\rho(y_i,x_i,g_i,w_i,\beta,h_0(w_i)) \mid g_i,w_i]=0$ where $\rho(y,x,g,w,\beta,\tau)=y-\beta x-\tau$ and $h_0(\bar{w})=\mathbb{E}\left[\varepsilon_i \mid w_i=\bar{w}\right]$. Chamberlain (1992) shows the efficiency bound is:

$$J_0^{-1}=\mathbb{E}\left[\mathbb{E}\left[D_0'\Sigma_0^{-1}D_0 \mid w_i\right]-\mathbb{E}\left[D_0'\Sigma_0^{-1}H_0 \mid w_i\right]\mathbb{E}\left[H_0'\Sigma_0^{-1}H_0 \mid w_i\right]^{-1}\mathbb{E}\left[H_0'\Sigma_0^{-1}D_0 \mid w_i\right]\right]^{-1}$$

for

$$
\begin{aligned}
D_0 &= \mathbb{E}\left[\frac{\partial}{\partial \beta}\rho(y_i, x_i, g_i, w_i, \beta, h_0(w_i)) \mid g_i, w_i\right] \\
\Sigma_0 &= \mathbb{E}\left[\rho(y_i, x_i, g_i, w_i, \beta, h_0(w_i))\rho(y_i, x_i, g_i, w_i, \beta, h_0(w_i))' \mid g_i, w_i\right] \\
H_0 &= \mathbb{E}\left[\frac{\partial}{\partial \tau}\rho(y_i, x_i, g_i, w_i, \beta, h_0(w_i))' \mid g_i, w_i\right].
\end{aligned}
$$

In our model, $D_0 = -\mathbb{E}\left[x_i \mid g_i, w_i\right]$ and $H_0 = -1$. Furthermore $\Sigma_0 = \mathbb{E}\left[(\varepsilon_i - E[\varepsilon_i \mid w_i])^2 \mid g_i, w_i\right] = \mathrm{Var}\left[\varepsilon_i \mid w_i\right]$ since $\mathbb{E}\left[\varepsilon_i^2 \mid g_i, w_i\right] = \mathbb{E}\left[\varepsilon_i^2 \mid w_i\right]$. Hence

$$
\begin{aligned}
J_0 =& \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[x_i \mid g_i, w_i\right]' \mathrm{Var}\left[\varepsilon_i \mid w_i\right]^{-1} \mathbb{E}\left[x_i \mid g_i, w_i\right] \mid w_i\right]\right] \\
&- \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[x_i \mid g_i, w_i\right]' \mathrm{Var}\left[\varepsilon_i \mid w_i\right]^{-1} \mid w_i\right] \mathbb{E}\left[\mathrm{Var}\left[\varepsilon_i \mid w_i\right]^{-1} \mid w_i\right]^{-1} \mathbb{E}\left[\mathrm{Var}\left[\varepsilon_i \mid w_i\right]^{-1} \mathbb{E}\left[x_i \mid g_i, w_i\right] \mid w_i\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\left(\mathbb{E}\left[x_i \mid g_i, w_i\right] - \mathbb{E}\left[x \mid w_i\right]\right)' \mathrm{Var}\left[\varepsilon_i \mid w_i\right]^{-1} \left(\mathbb{E}\left[x_i \mid g_i, w_i\right] - \mathbb{E}\left[x_i \mid w_i\right]\right)' \mid w_i\right]\right] \\
=& \mathbb{E}\left[\frac{\mathrm{Var}\left[\tilde{z}_i \mid w_i\right]}{\mathrm{Var}\left[\varepsilon_i \mid w_i\right]}\right],
\end{aligned}
$$

while as shown above the variance of the optimal IV asymptotically coincides with $\mathbb{E}\left[\tilde{z}'\Omega^{-1}\tilde{z}\right]^{-1} = \frac{1}{N}\mathbb{E}\left[\frac{\mathrm{Var}\left[z_i^2 \mid w_i\right]}{\mathrm{Var}\left[\varepsilon_i \mid w_i\right]}\right]^{-1}$, where $\frac{1}{N}$ reflects the notational difference between the asymptotic variance (scaled by N) and the approximate variance (not scaled).

B.8 Proof of Appendix Proposition A2

Let $\Omega = \mathrm{Var}\left[\tilde{g} \mid w\right]$ be the covariance matrix of the shocks and let $S^* = \mathbb{E}\left[x\tilde{g}' \mid w\right]\Omega^{-1}$. For brevity, we suppress the dependence of S, S^*, and Ω on w.

As with Theorem 1, we follow the proof of Theorem 5.3 in Newey and McFadden (1994). We first show that $\mathbb{E}\left[z'\varepsilon\varepsilon' z^S\right] = \mathbb{E}\left[z'x\right]$ for any $z \in \mathfrak{R}^S$: repeatedly using the law of iterated expectations,

$$
\begin{aligned}
\mathbb{E}\left[z'\varepsilon\varepsilon' z^S\right] &= \mathbb{E}\left[\tilde{g}' S' \varepsilon\varepsilon' S^* \tilde{g}\right] \\
&= \mathbb{E}\left[\tilde{g}' S' \mathbb{E}\left[\varepsilon\varepsilon' \mid h, w\right] S^* \tilde{g}\right] \\
&= \mathbb{E}\left[\tilde{g}' S' \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right] \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1} \mathbb{E}\left[x\tilde{g}' \mid w\right] \Omega^{-1} \tilde{g}\right] \\
&= \mathrm{tr}\,\mathbb{E}\left[S' \mathbb{E}\left[x\tilde{g}' \mid w\right] \Omega^{-1} \tilde{g}\tilde{g}'\right] \\
&= \mathrm{tr}\,\mathbb{E}\left[S' \mathbb{E}\left[x\tilde{g}' \mid w\right]\right] \\
&= \mathrm{tr}\,\mathbb{E}\left[S' x g'\right] \\
&= \mathbb{E}\left[g' S' x\right] = \mathbb{E}\left[z' x\right],
\end{aligned}
$$

where the third line imposed the condition $\mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right] = \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]$ from Theorem 1. The rest follows identically to Theorem 1, as $\mathcal{V}\left[z\right] - \mathcal{V}\left[z^S\right] = \mathbb{E}\left[U^2\right] \geq 0$ for $U = \frac{z'\varepsilon}{\mathbb{E}[z'x]} - \frac{z^{S\prime}\varepsilon}{\mathbb{E}[z^{S\prime}x]}$.

C Empirical Appendix

C.1 Sample Construction

Our application uses a repeated cross-section of annual data from the American Community Survey (ACS; Ruggles et al., 2020). Our main estimates use representative 1% samples from 2013 and 2014; we use an analogous sample from 2012 to estimate pre-trends. We restrict these samples 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 expanded Medicaid after the ACA but before 2013, as well as Massachusetts and Vermont which had made all adults with household income less than 138% FPL eligible prior to the ACA. 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.[30]

In each year, we classify an individual as insured under Medicaid when they report being covered by Medicaid or an equivalent government-assistance program, excluding Medicare and Veterans Affairs 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 by employer-sponsored insurance.

We use the formulas in Frean et al. (2017) to define Medicaid eligibility. Income is given by a household's total pre-tax personal income or losses (*inctot*), adjusted for inflation. Other inputs to the eligibility calculation include an indicator for whether an individual is a parent (i.e. an adult with children in the household) and an indicator for whether an individual is in the labor force (*labforce*). We note that these may be noisy proxies for the characteristics actual used to assign Medicaid eligibility.

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 (24.5%), and that of individuals who would be eligible if their state kept 2013 policy intact (11.6%). We assign the difference in these shares to all individuals in 2014 residing in expansion states and zero to all other individuals.

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

[30] 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 Massachusetts and Vermont.

the individual's actual 2014 eligibility, while inside this sample the expected instrument in 2014 is the fraction of states which expanded conditional on the governor's party. The 2013 expected instrument is actual 2013 eligibility for all individuals. Political party affiliation of state governors is determined as of December 2013. In all regressions we control for a Republican governor indicator interacted with year. In robustness checks we control for other time-interacted state characteristics: a state's 2012 median income or share insured under Medicaid (both from the ACS).

C.2 Reconciling Simulated and Recentered IV Estimates

This appendix first discusses how measurement error in Medicaid eligibility calculations can bias both simulated and recentered IV estimates of eligibility effects. We then show how such bias is avoided in IV regressions which use Medicaid enrollment as the endogenous variable instead.

For clarity of the theoretical discussion, we simplify the setup. First, we suppose that a single *iid* cross-section of 2014 data is available and state fixed effects are not included; we correspondingly drop the t subscript throughout. Second, we assume that state decisions to expand Medicaid coverage are unconditionally as-good-as-random and mutually independent with the same probability . We then consider the simple causal model of $y_i = \beta x_i^* + \varepsilon_i$ where y_i is a measure of insurance coverage for individual i and $x_i^* \in \{0,1\}$ is i's true Medicaid eligibility. We suppose changes in x_i^* come from the exogenous expansion of eligibility policy; formally, we write $x_i^* = x_{i0}^* + e_i^* g_i$ where $x_{i0}^*, e_i^*, g_i \in \{0,1\}$ with $x_{i0}^* e_i^* = 0$. The x_{i0}^* indicator identifies individuals who would be eligible for Medicaid regardless of the binary expansion shock g_i, while e_i^* indicates those who become eligible when $g_i = 1$ (i.e. exposed individuals). Individuals who are never eligible regardless of g_i are identified by $x_{i0}^* = e_i^* = 0$. Rather than observing x_{i0}^* and e_i^* directly, we assume the researcher computes eligibility x_i from a mismeasured x_{i0} and e_{i0}: i.e., $x_i = x_{i0} + e_i g_i$ where again $x_{i0}, e_i \in \{0,1\}$ and $x_{i0} e_i = 0$. Such measurement error could reflect error in self-reported household income or demographics (Brooks, 2019). We assume the as-good-as-random expansion shocks only affect outcomes by changing eligibility, making them exogenous: $g_i \perp\!\!\!\perp (\varepsilon_i, x_{i0}^*, e_i^*, x_{i0}, e_i)$.

We first show how simulated and IV estimates of β, which use measured eligibility as the right-hand side variable, are biased by this measurement error. The simulated IV estimate uses the expansion shock to instrument measured eligibility in the full sample of individuals. This IV regression identifies

$$\begin{aligned}\beta^{SIV} &= \frac{\text{Cov}\left[y_i, g_i\right]}{\text{Cov}\left[x_i, g_i\right]} \\ &= \frac{\text{Cov}\left[\beta(x_{i0}^* + e_i^* g_i) + \varepsilon_i, g_i\right]}{\text{Cov}\left[x_{i0} + e_i g_i, g_i\right]} \\ &= \beta \frac{\mathbb{E}\left[e_i^*\right]}{\mathbb{E}\left[e_i\right]}, \end{aligned} \tag{A3}$$

where we use the independence of g_i in the third line. The recentered IV estimate of β, implemented as in Panel A of Table 2, regresses the outcome on measured eligibility in the sample of individuals who are exposed according to the observed e_i. This estimator identifies

$$\begin{aligned}\beta^{RIV} &= \frac{\text{Cov}\left[y_i, x_i \mid e_i = 1\right]}{\text{Var}\left[x_i \mid e_i = 1\right]} \\ &= \frac{\text{Cov}\left[\beta(x_{i0}^* + e_i^* g_i) + \varepsilon_i, g_i \mid e_i = 1\right]}{\text{Var}\left[g_i \mid e_i = 1\right]} \\ &= \beta\mathbb{E}\left[e_i^* \mid e_i = 1\right], \end{aligned} \tag{A4}$$

where we again use the independence of g_i in the third line. Since $\mathbb{E}\left[e_i^* \mid e_i = 1\right] \in [0, 1]$ and $\mathbb{E}\left[e_i^* \mid e_i = 1\right] = \mathbb{E}\left[e_i^* e_i\right] / \mathbb{E}\left[e_i^*\right] < \mathbb{E}\left[e_i\right] / \mathbb{E}\left[e_i^*\right]$ whenever $Pr\left(e_i \neq e_i^*\right) > 0$, these expressions show that the recentered IV estimand is attenuated relative to both the causal parameter of interest and the simulated IV estimand: $|\beta^{RIV}| \leq |\beta|$ and $|\beta^{RIV}| \leq |\beta^{SIV}|$. The simulated IV estimand can either be larger or smaller than the causal parameter of interest, depending on the relative shares of true and computed exposure $\mathbb{E}\left[e_i^*\right]$ and $\mathbb{E}\left[e_i\right]$.

Such bias, however, does not arise when estimating the effects of Medicaid enrollment using either of the two instruments. This follows because in such IV regressions both the first stage regression (of enrollment on eligibility) and the reduced form (some outcome, such as private insurance coverage, on eligibility) have the same proportionate bias given by equations (A3) and (A4), depending on the instrument.

C.3 Efficiency of Full-Sample Recentered IV Estimates

Appendix Table A4 reports recentered IV estimates of Medicaid takeup and crowdout effects in the full sample of 2014 and 2013 individuals, not restricting to the exposed sample as in our baseline specification. 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.

Here we explain how a combination of two factors generates the discrepancy between panels A and B of the table. First, the regression residuals are strongly correlated with the indicator for an individual being exposed to the ACA 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 exposed sample or when appropriate controls are included. We further relate this problem to the third step of the instrument construction in Section 3.3.

For clarity of the theoretical discussion, we simplify the setup. First, we suppose that a single 2014 cross-section is available and thus state fixed effects are not included; we correspondingly drop the t subscript throughout. We allow for other controls to be included. Second, we assume states only change eligibility as prescribed by their expansion decision, i.e. $x_i = z_i$. Finally, we assume that state decisions to expand are independent with a known propensity $\mathbb{E}\left[g_k \mid w\right]$ (e.g., as a function of the state governor's party). Thus, the recentered expansion indicator $\tilde{g}_k = g_k - \mathbb{E}\left[g_k \mid w\right]$ can be computed without permutations.[31] We compare the recentered simulated instrument z_i^{SI} to the recentered best predictor instrument $\tilde{z}_i$, ignoring controls that span $\mathbb{E}[g_k \mid w]$ (e.g. the state party indicator).

Under these additional assumptions, using the recentered simulated instrument is equivalent to using the recentered expansion indicator: $z_i^{SI} = \tilde{g}_{s(i)}$. The recentered best predictor instrument only differs by setting z_i^{SI} to zero for the non-exposed sample: $\tilde{z}_i = z_i - \mathbb{E}[z_i \mid w] = f_i \tilde{g}_{s(i)}$, where f_i is an indicator for individual i being in the exposed group. With $x_i = z_i$, the first stage can be written $x_i = \mu_i + f_i \tilde{g}_{s(i)}$, where the expected instrument μ_i equals 0 for individuals who are not eligible regardless of $g_{s(i)}$, 1 for those always eligible, and $\mathbb{E}\left[g_{s(i)} \mid w\right]$ otherwise.

We now consider the approximate variances of the two estimators, $\operatorname{Var}\left[\frac{1}{N}\sum_i z_i^{SI}\varepsilon_i^{\perp}\right] / \mathbb{E}\left[\frac{1}{N}\sum_i z_i^{SI} x_i^{\perp}\right]^2$ and $\operatorname{Var}\left[\frac{1}{N}\sum_i \tilde{z}_i\varepsilon_i^{\perp}\right] / \mathbb{E}\left[\frac{1}{N}\sum_i \tilde{z}_i x_i^{\perp}\right]^2$, 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).[32] For simplicity of exposition we also consider an individual's state of residence $s(i)$ as fixed. Letting $N_k = \sum_i \mathbf{1}[s(i) = k]$ denote the (fixed) number of individuals in each state k, it can then be shown that

$$\frac{\operatorname{Var}\left[\frac{1}{N}\sum_i \tilde{z}_i\varepsilon_i^{\perp}\right]}{\operatorname{Var}\left[\frac{1}{N}\sum_i z_i^{SI}\varepsilon_i^{\perp}\right]} = \frac{\sum_k \left(\frac{N_k}{N}\right)^2 \operatorname{Var}[\tilde{g}_k]\, \mathbb{E}\left[e_{RI,k}^2\right]}{\sum_k \left(\frac{N_k}{N}\right)^2 \operatorname{Var}[\tilde{g}_k]\, \mathbb{E}\left[e_{SI,k}^2\right]}, \tag{A5}$$

where $e_{RI,k} = \frac{1}{N_k}\sum_{i:\, s(i)=k} \varepsilon_i^{\perp} f_i$ is the sum of residuals of *exposed* individuals in state k (normalized by N_k), while $e_{SI,k} = \frac{1}{N_k}\sum_{i:\, s(i)=k} \varepsilon_i^{\perp}$ averages over *all* observations in the state.[33]

Equation (A5) 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 $e_{RI,k} = e_{SI,k}$ for exposed individuals in that case).

However, this simplified example shows that the recentered IV is likely to deliver a power loss when the shocks g_k are clustered and $\varepsilon_i^{\perp}$ is strongly correlated with the indicator of exposed sample f_i (i.e., exposed individuals have systematically different residuals, and f_i is not controlled for). To see this simply, suppose $\mathbb{E}\left[\varepsilon_i^{\perp} \mid f_i = 1, w\right] = \alpha \neq 0$ for all i. In this scenario $e_{RI,k}$ is not mean-zero,

[31] Formally, we assume that w does not include the permutation class of g. Under this assumption, $\tilde{g}_k$ is independent across states conditionally on w, simplifying the analysis.

[32] Namely, since f_i is binary, $\mathbb{E}\left[\frac{1}{N}\sum_i z_i^{SI} x_i\right] = \mathbb{E}\left[\frac{1}{N}\sum_i \tilde{g}_{s(i)}\left(\mu_i + f_i\tilde{g}_{s(i)}\right)\right] = \mathbb{E}\left[\frac{1}{N}\sum_i f_i\tilde{g}_{s(i)}^2\right] = \mathbb{E}\left[\frac{1}{N}\sum_i f_i\tilde{g}_{s(i)}\left(\mu_i + f_i\tilde{g}_{s(i)}\right)\right] = \mathbb{E}\left[\frac{1}{N}\sum_i \tilde{z}_i x_i\right]$. With controls this equality holds asymptotically, since the difference between x_i and $x_i^{\perp}$ is uncorrelated with $z_i^{SI} - \tilde{z}_i = (1 - f_i)\,\tilde{g}_{s(i)}$.

[33] Namely, $\operatorname{Var}\left[\frac{1}{N}\sum_i \tilde{z}_i\varepsilon_i^{\perp}\right] = \sum_k \left(\frac{N_k}{N}\right)^2 \mathbb{E}\left[\left(\frac{1}{N_k}\sum_{i:\, s(i)=k}\tilde{z}_i\varepsilon_i^{\perp}\right)^2\right] = \sum_k \left(\frac{N_k}{N}\right)^2 \mathbb{E}\left[\tilde{g}_k^2 \cdot \left(\frac{1}{N_k}\sum_{i:\, s(i)=k} f_i\varepsilon_i^{\perp}\right)^2\right] = \sum_k \left(\frac{N_k}{N}\right)^2 \operatorname{Var}[\tilde{g}_k]\,\mathbb{E}\left[e_{RI,k}^2\right]$, since $\mathbb{E}\left[\frac{1}{N}\sum_i \tilde{z}_i\varepsilon_i^{\perp}\right] = 0$, and similarly for $\operatorname{Var}\left[\frac{1}{N}\sum_i z_i^{SI}\varepsilon_i^{\perp}\right]$.

even on average across states, which potentially yields a high mean-squared residual:

$$\mathbb{E}\left[e_{RI,k}\right] = \mathbb{E}\left[\mathbb{E}\left[e_{RI,k} \mid w\right]\right] = \mathbb{E}\left[\frac{1}{N_k}\sum_{i:\, s(i)=k} \mathbb{E}\left[\varepsilon_i^{\perp} f_i \mid w\right]\right]$$

$$= \mathbb{E}\left[\frac{1}{N_k}\sum_{i:\, s(i)=k} \mathbb{E}\left[\varepsilon_i^{\perp} \mid f_i = 1, w\right] f_i\right] = \alpha \cdot \mathbb{E}\left[\frac{\sum_{i:\, s(i)=k} f_i}{N_k}\right] \neq 0.$$

The simulated instrument, which does not condition on $f_i = 1$, does not suffer from this problem since $\varepsilon_i^{\perp}$ 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 simulated instrument that uses both subsamples.

The predictions of this discussion are borne out in the data. In Panel C of Appendix Table A4 we verify that the confidence interval of recentered IV becomes dramatically narrowed with a single control of f_i (interacted with the 2014 dummy appropriately for the difference-in-differences setting).[34] Moreover, demographic controls in Panel B of Appendix Table A4 capture most of the variation in f_i, delivering similar results. Our recentered IV specifications in the main text, by restricting the sample to the exposed individuals, effectively control for year interacted with f_i.

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 3.2: this control happens to play the role of the predetermined predictors of the residual, ψ. Our application therefore highlights that in general there is no guarantee of an efficiency gain from improving the first stage with a recentered IV (i.e., approximating the recentered best predictor) if adjustment for ψ is not feasible.

C.4 Power Simulations

We verify large and robust power gains from 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 a counterfactual simulated instrument and a counterfactual recentered best predictor instrument. We do not specify a model for e_k^{Δ}, and instead take the counterfactual best predictor as the endogenous variable. Finally, for the Medicaid take-up and employer-sponsored insurance crowd-out outcomes we take the second-stage residuals from Columns 2 and 6 of Table 2, 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 the simulated and recentered IV specifications as in our baseline implementation in Panel A of Table 2. 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.

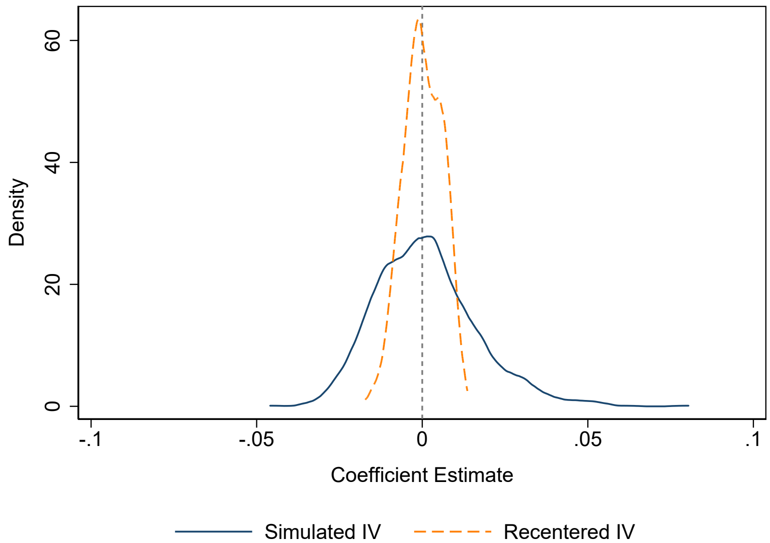
[34]The efficiency of IV specifications that only control for the expected eligibility prediction in columns 2, 4, and 6 of Panel A is lower because this control does not span f_i.

Appendix Figure A1 shows the distribution of simulated and recentered IV estimates for the two outcomes across Monte Carlo draws. Both estimators are approximately unbiased, with both distributions in both panels centered around the true effect of zero. However, consistent with the much shorter confidence intervals in Table 2, the distribution of recentered IV coefficients is dramatically tighter around this mean. The estimator 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.

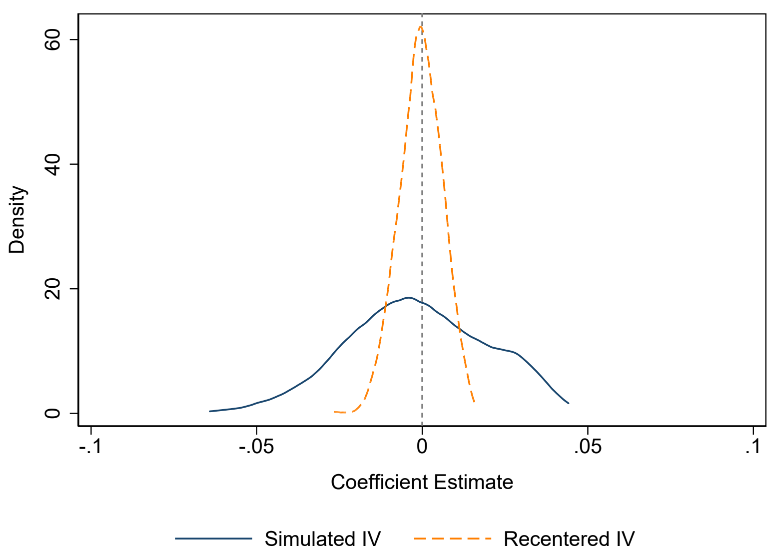
Figure 1 shows that these reductions in estimator variance translate to increased rejection rates of false null hypotheses for both outcomes, while also suggesting the wild bootstrap 95% confidence intervals in Table 2 have approximately correct size. Away from the true null hypothesis of zero the recentered 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 employer-sponsored insurance 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.

Figure A1: Medicaid Application: Simulated Estimator Distributions

A. Has Medicaid



B. Has Employer-Sponsored Insurance



Notes: This figure plots the simulated distributions of the two IV regression discussed in the text: one using a conventional simulated instrument and the other using as an instrument a recentered prediction of Medicaid eligibility. See Appendix C.3 for a description of the simulation procedure. The true effect of zero in both panels is indicated by the dashed vertical line.

Table A1: Medicaid Application: Second Stage-Estimates with Additional Controls

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Simulated IV (1)	Recentered IV (2)	Simulated IV (3)	Recentered IV (4)	Simulated IV (5)	Recentered IV (6)
Panel A. Medicaid Eligibility Effects						
Eligible	0.135	0.073	−0.050	−0.024	0.003	−0.008
	(0.029)	(0.010)	(0.022)	(0.007)	(0.013)	(0.005)
	[0.082,0.224]	[0.050,0.095]	[-0.112,0.001]	[-0.042,-0.008]	[-0.039,0.037]	[-0.019,0.005]
Panel B. Medicaid Enrollment Effects						
Has Medicaid			−0.372	−0.334	0.023	−0.108
			(0.146)	(0.091)	(0.097)	(0.060)
			[-0.767,0.008]	[-0.566,-0.121]	[-0.238,0.322]	[-0.235,0.081]
Exposed Sample	N	Y	N	Y	N	Y
States	43	43	43	43	43	43
Individuals	2,397,313	421,042	2,397,313	421,042	2,397,313	421,042

Notes: Panel A of this table reports second-stage coefficients from the two IV regressions described in the text: one using a conventional simulated instrument and the other using as an instrument a recentered prediction of Medicaid eligibility. Columns 1, 3, and 5 estimate regressions in the full sample of individuals in 2013–2014, while Columns 2, 4, and 6 restrict to the sample of individuals whose characteristics and state of residence make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects, an indicator for Republican-governed states interacted with year, and full interactions of deciles of household income, parental status, work status, and year. Panel B shows estimates from IV regressions which use an indicator for Medicaid enrollment as the endogenous variable, instead of an indicator for Medicaid eligibility. State-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

Table A2: Medicaid Application: Pre-Trend Estimates

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Simulated IV (1)	Recentered IV (2)	Simulated IV (3)	Recentered IV (4)	Simulated IV (5)	Recentered IV (6)
Panel A. Baseline Specification						
Eligibile	0.022	0.020	−0.015	−0.011	−0.011	−0.007
	(0.009)	(0.004)	(0.017)	(0.004)	(0.017)	(0.005)
	[-0.002,0.042]	[0.009,0.029]	[-0.069,0.026]	[-0.020,-0.003]	[-0.058,0.028]	[-0.019,0.004]
Panel B. With Additional Controls						
Eligibile	0.023	0.020	−0.019	−0.014	−0.016	−0.011
	(0.010)	(0.004)	(0.014)	(0.004)	(0.016)	(0.005)
	[-0.012,0.040]	[0.009,0.027]	[-0.056,0.020]	[-0.022,-0.006]	[-0.051,0.031]	[-0.023,0.001]
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 pre-trend estimates for the two IV regressions described in the text: one using a conventional simulated instrument and the other using as an instrument a recentered prediction of Medicaid eligibility. Pre-trend estimates come from the IV specifications described in the text replacing 2013 individuals with 2012 individuals and replacing 2014 individuals with 2013 individuals. Columns 1, 3, and 5 estimate regressions in the full sample of individuals, while Columns 2, 4, and 6 restrict to the sample of individuals whose characteristics and state of residence make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year. The regressions in Panel B additionally control for deciles of household income, interacted with indicators for parental and work status and year. Conventional state-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

Table A3: Medicaid Application: Alternative Designs

	Has Medicaid	Has Private Insurance	Has Employer-Sponsored Insurance
	(1)	(2)	(3)
Panel A. Republican Governor and 2012 Median Income			
Eligible	0.077	−0.018	−0.005
	(0.011)	(0.008)	(0.006)
	[0.053,0.094]	[-0.040,0.002]	[-0.018,0.011]
Has Medicaid		−0.231	−0.067
		(0.101)	(0.070)
		[-0.541,0.027]	[-0.216,0.176]
Panel B. Republican Governor, 2012 Median Income, 2012 Medicaid Coverage			
Eligible	0.076	−0.023	−0.009
	(0.011)	(0.007)	(0.005)
	[0.053,0.102]	[-0.040,-0.007]	[-0.021,0.003]
Has Medicaid		−0.304	−0.125
		(0.096)	(0.058)
		[-0.546,-0.099]	[-0.256,0.048]
Exposed Sample	Y	Y	Y
States	43	43	43
Individuals	421,042	421,042	421,042

Notes: This table adds additional controls to the recentered IV regressions in Table 2. The regressions in both panels add 2012 state median income interacted with year and the Republican governor control. The regressions in Panel B additionally control for 2012 state Medicaid coverage rates interacted with year, the Republican governor control, and 2012 state median income. State-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

Table A4: Medicaid Application: Recentered IV Including Non-Exposed Individuals

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Recentered (1)	Controlled (2)	Recentered (3)	Controlled (4)	Recentered (5)	Controlled (6)
Panel A. Baseline Controls						
Eligible	0.032	0.072	0.193	0.061	0.208	0.071
	(0.085)	(0.038)	(0.290)	(0.123)	(0.301)	(0.127)
	[-0.433,0.148]	[-0.072,0.128]	[-0.214,1.796]	[-0.123,0.526]	[-0.204,1.907]	[-0.118,0.565]
Panel B. With Demographics × Post						
Eligible	0.116	0.116	−0.029	−0.030	−0.018	−0.019
	(0.012)	(0.011)	(0.013)	(0.012)	(0.012)	(0.012)
	[0.092,0.149]	[0.093,0.148]	[-0.052,0.003]	[-0.051,0.002]	[-0.040,0.012]	[-0.040,0.011]
Panel C. With Exposed Sample × Post						
Eligible	0.094	0.093	−0.012	−0.013	−0.005	−0.006
	(0.011)	(0.011)	(0.015)	(0.016)	(0.017)	(0.018)
	[0.064,0.118]	[0.059,0.118]	[-0.039,0.034]	[-0.041,0.038]	[-0.034,0.047]	[-0.037,0.052]
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: Panel A of this table reports second-stage coefficients from versions of the recentered IV regressions in Table 2, estimated in the full sample of individuals in 2013–14. Columns 1, 3, and 5 use as an instrument a recentered prediction of Medicaid eligibility while Columns 2, 4, and 6 do not recenter but control for the expected prediction. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year. The regressions in Panel B additionally control for deciles of household income, interacted with indicators for parental and work status and year. The regressions in Panel C instead add the interaction of year and an indicator for an individual having characteristics that make them exposed to the partial ACA Medicaid expansion in 2014. State-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.