BALANCING WEIGHTS FOR REGION-LEVEL ANALYSIS: THE EFFECT OF MEDICAID EXPANSION ON THE UNINSURANCE RATE AMONG STATES THAT DID NOT EXPAND MEDICAID

BY MAX RUBINSTEIN¹, AMELIA HAVILAND¹,‡, AND DAVID CHOI¹,†

¹Carnegie Mellon University, Heinz College and Department of Statistics & Data Science
mrubinst@andrew.cmu.edu; ‡amelia@andrew.cmu.edu; †davidch@andrew.cmu.edu

We predict the average effect of Medicaid expansion on the non-elderly adult uninsurance rate among states that did not expand Medicaid in 2014 as if they had expanded their Medicaid eligibility requirements. Using American Community Survey data aggregated to the region level, we estimate this effect by finding weights that approximately reweight the expansion regions to match the covariate distribution of the non-expansion regions. Existing methods to estimate balancing weights often assume that the covariates are measured without error and do not account for dependencies in the outcome model. Our covariates have random noise that is uncorrelated with the outcome errors and our assumed outcome model has state-level random effects inducing dependence between regions. To correct for the bias induced by the measurement error, we propose generating our weights on a linear approximation to the true covariates, using an idea from measurement error literature known as “regression-calibration” (see, e.g., Carroll et al. (2006)). This requires auxiliary data to estimate the variability of the measurement error. We also modify the Stable Balancing Weights objective proposed by Zubizarreta (2015)) to reduce the variance of our estimator when the model errors are homoskedastic and equicorrelated within states. We show that these approaches outperform existing methods when attempting to predict observed outcomes during the pre-treatment period. Using this method we estimate that Medicaid expansion would have caused a -2.33 (-3.54, -1.11) percentage point change in the adult uninsurance rate among states that did not expand Medicaid.

1. Introduction. We study the effect of 2014 Medicaid expansion on the non-elderly adult uninsurance rate among states that did not expand Medicaid in 2014 as if they had expanded their Medicaid eligibility requirements. We use public-use survey microdata from annual American Community Survey (ACS) aggregated to the consistent public use micro-data area (CPUMA) level, a geographic region that falls within states. We calculate weights that reweight expansion-state CPUMAs to approximately match the covariate distribution of CPUMAs in states that did not expand Medicaid in 2014. We then estimate our causal effect as the difference in means between the reweighted treated CPUMAs and the observed mean of the non-expansion CPUMAs. A key challenge is that our data consists of estimated covariates. The sampling variability in these estimates is a form of measurement error that may bias effect estimates calculated on the observed data. Additionally, CPUMAs fall within states and share a common policy-making environment. The data-generating process for the outcomes therefore may contain state-level random effects that can increase the variance of standard estimation procedures. Our study contributes to the literature on balancing weights by proposing approaches to address both of these problems. We also contribute to the literature on Medicaid expansion by estimating the foregone coverage gains of Medicaid among states that did not expand Medicaid in 2014, which to our knowledge has not yet been directly estimated.

Keywords and phrases: Medicaid expansion, balancing weights, measurement error, hierarchical data.
Approximate balancing weights are an estimation method in causal inference that grew out of the propensity score weighting literature. Rather than iteratively modeling the propensity score until the inverse probability weights achieve a desired level of balance, recent papers propose using optimization methods to generate weights that enforce covariate balance between the treated and control units (see, e.g., Hainmueller (2012), Imai and Ratkovic (2014), Zubizarreta (2015)).

From an applied perspective, there are at least four benefits of this approach: first, it does not require iterating propensity score models to generate satisfactory weights. Second, these methods (and propensity score methods generally) do not use the outcomes to determine the weights, mitigating the risk of cherry-picking an outcome model specification to obtain a desired result. Third, these methods can constrain the weights to prevent extrapolation from the data, thereby reducing model dependence (Zubizarreta (2015)). Finally, the estimates are more interpretable: by making the comparison group explicit, it is easy to communicate exactly which units contributed to the counterfactual estimate.

Most proposed methods in this literature assume that the covariates are measured without error. For our application we assume that our covariates are measured with mean-zero additive error. This error could potentially bias standard estimation procedures. As a first contribution, we therefore propose generating our weights as a function of a linear approximation to the true covariate values, using an idea from the measurement error literature known as “regression-calibration” (see, e.g., Gleser (1992), Carroll et al. (2006)). This method requires access to an estimate of the measurement error covariance matrix, which we estimate using the ACS microdata. The theoretic consistency of these estimates requires several assumptions, including that the covariate measurement errors are uncorrelated with any errors in the outcome model, the outcome model is linear in the covariates, and the covariates are i.i.d. gaussian. The first assumption is reasonable for our application since the covariates are measured on a different cross-sectional survey than our outcomes. The second is strong but somewhat relaxed because we prevent our weights from extrapolating beyond the support of the data. The third can be fully relaxed to obtain consistent estimates using ordinary least squares (OLS), but unfortunately not without additional modeling beyond our proposed method. Despite appearing costly, we show in Section 3.4.2 that this tradeoff is likely worth it in our application.

As a second contribution, we propose modifying the Stable Balancing Weights (SBW) objective (Zubizarreta (2015)) to account for possible state-level dependencies in our outcome model. We assume that the errors are homoskedastic with constant positive equicorrelation, though our general approach can accommodate other assumed correlation structures. In a setting without measurement error, we show that this modification can reduce the variance of the resulting estimates. We also connect these weights to the implied regression weights from Generalized Least Squares (GLS). Our overall approach provides a general framework that can be used by other applied researchers who wish to use balancing weights to estimate causal effects when their data are measured with error and/or the model errors are dependent.

Section 2 begins with a more detailed overview of the policy problem, and then defines the study period, covariates, outcome, and treatment. Section 3 discusses our methods, including

1These methods also use ideas from the survey literature, which had proposed similar approaches to adjust sample weights to enforce equality between sample totals and known population totals (see, e.g., Haberman (1984), Deville and Särndal (1992), Deville, Särndal and Sautory (1993), Särndal and Lundström (2005)).

2Our approach also relates to the “synthetic controls” literature (see, e.g., Abadie, Diamond and Hainmueller (2010)). Synthetic controls are a popular balancing weights approach frequently used in the applied economics literature to estimate treatment effects on the treated (ETT) for region-level policy changes when using time series cross sectional data. Our application uses a similar data structure; however, we instead consider the problem of estimating the ETC. In contrast to much of the synthetic controls literature, which assumes that the counterfactual outcomes follow a linear factor model, we instead assume that the counterfactual outcomes are linear in the observed covariates (including pre-treatment outcomes).
outlining our identification, estimation, and inferential procedures. Section 4 presents our results. Section 5 contains a discussion of our findings, and Section 6 contains a brief summary. The Appendices contain proofs, summary statistics, and additional results.

2. Policy Problem and Data.

2.1. Policy Problem Statement. Under the Affordable Care Act (ACA), states were required to expand their Medicaid eligibility requirements by 2014 to offer coverage to all adults with incomes at or below 138 percent of the federal poverty level (FPL). The United States Supreme Court ruled this requirement unconstitutional in 2012, allowing states to decide whether to expand Medicaid coverage. In 2014, twenty-six states and the District of Columbia expanded their Medicaid programs. From 2015 through 2021, an additional twelve states elected to expand their Medicaid programs. Medicaid expansion has remained a debate among the remaining twelve states that have not expanded Medicaid as of 2021.\(^3\) The effects of Medicaid expansion on various outcomes, including uninsurance rates, mortality rates, and emergency department use, have been widely studied, primarily by using the initial expansions in 2014 and 2015 to define expansion states as “treated” states and non-expansion states as “control” states (see, e.g., Courtemanche et al. (2017), Miller, Johnson and Wherry (2021), Ladhania et al. (2021)).

Medicaid enrollment is not automatic, and take-up rates have historically varied across states. This variation is partly a function of state discretion in administering programs: for example, program outreach, citizenship verification policies, and application processes differ across states (Courtemanche et al. (2017)). Estimating how Medicaid eligibility expansion actually affects the number of uninsured individuals is therefore not obvious. This is also important because many effects are mediated largely through reducing the number of uninsured individuals. Existing studies have estimated that Medicaid expansion reduced the uninsurance rate between three and six percentage points on average among states that expanded Medicaid. These estimates differed depending on the data used, specific target population, study design, and level of analysis (see, e.g., Kaestner et al. (2017), Courtemanche et al. (2017), Frean, Gruber and Sommers (2017)). However, none of these studies have directly estimated the average treatment effect on the controls (ETC).

We believe that the ETC may differ from the ETT. Every state had different coverage policies prior to 2014, and non-expansion states tended to have less generous policies than expansion states. “Medicaid expansion” therefore represents a set of treatments of varying intensities that are distributed unevenly across expansion and non-expansion states. Averaged over the non-expansion states, which tended to have less generous policies and higher uninsurance rates prior to Medicaid expansion, we might expect the average effect to be larger in absolute magnitude than among the expansion states, where “Medicaid expansion” on average reflected smaller policy changes.\(^4\) Even limited to states with equivalent coverage policies prior to 2014 we still might expect the ETT to differ from the ETC. For example, all states that were entirely controlled by the Democratic Party at the executive and legislative levels expanded their Medicaid programs, while only states where the Republican Party controlled at least part of the state government failed to expand their programs. Prior to the 2014 Medicaid expansion, Sommers et al. (2012) found that conservative governance was

\(^3\)https://www.nbcnews.com/politics/politics-news/changed-hearts-minds-biden-s-funding-offer-shifts-medicaid-expansion-n1262229

\(^4\)As a part of our analysis strategy, we limit our pool of expansion states to those where the policy changes were comparable to the non-expansion states. We also control for pre-treatment uninsurance rates (see Section 2.3).
associated with lower Medicaid take-up rates. This might reflect differences in program implementation, which could serve as effect modifiers for comparable policy changes.\footnote{Interestingly, Sommers et al. (2012) also find that the association between conservative governance and lower take-up rates prior to 2014 existed even after controlling for a variety of factors pertaining to state-level policy administration decisions. They conjecture that this may reflect cultural conservatism: people in conservative states are more likely to view enrollment in social welfare programs negatively, and therefore be less likely to enroll.} These factors may then attenuate the effects of Medicaid expansion averaged over non-expansion states relative to expansion states.

As a more general causal quantity, the ETC is also interesting in its own right: to the extent that the goal of studying Medicaid expansion is to understand the foregone benefits - or potential harms - of Medicaid expansion among non-expansion states, the ETC is the relevant quantity of interest. Authors have previously made claims about the ETC without directly estimating it. For example, Miller, Johnson and Wherry (2021) use their estimates of the ETT to predict that had non-expansion states expanded Medicaid, they would have seen 15,600 fewer deaths from 2014 through 2017. However, as these authors note, this estimate assumes that the ETT provides a comparable estimate to the ETC. From a policy analysis perspective, we recommend that researchers estimate the ETC directly when it answers a substantive question of interest. We therefore contribute to the literature on Medicaid expansion by directly estimating this quantity.

2.2. 

\textit{Data Source and Study Period.} Our primary data source is the annual household and person public use microdata files from the ACS from 2011 through 2014. The ACS is an annual cross-sectional survey of approximately three million individuals across the United States. The public use microdata files include information on individuals in geographic areas greater than 65,000 people. The smallest geographic unit contained in these data are public-use microdata areas (PUMAs), arbitrary boundaries that nest within states but not within counties or other more commonly used geographic units. One limitation of these data is a 2012 change in the PUMA boundaries, which do not overlap well with the previous boundaries. As a result, the smallest possible geographic areas that nest both PUMA coding systems are known as consistent PUMAs (CPUMAs). The United States contains 1,075 total CPUMAs, with states ranging from having one CPUMA (South Dakota, Montana, and Idaho) to 123 CPUMAs (New York). Our primary dataset contains 925 CPUMAs among 45 states (see also Section 2.3). The average total number of sampled individuals per CPUMA across the four years is 1,001; the minimum number of people sampled was 334 and the maximum is 23,990. We aggregate the microdata to the CPUMA level using the survey weights.

This aggregation naturally raises concerns about measurement error and hierarchy. Any CPUMA-level variable is an estimate, leading to concerns about measurement error. The hierarchical nature of the dataset – CPUMAs within states – raises concerns about geographic dependence.

Our study period begins in 2011, following Courtemanche et al. (2017), who note that several other aspects of the ACA were implemented in 2010 – including the provision allowing for dependent coverage until age 26 and the elimination of co-payments for preventative care – and likely induced differential shocks across states. We also restrict our post-treatment period to 2014. We therefore avoid additional assumptions required for identification given that several states expanded Medicaid in 2015, including Indiana, Pennsylvania, and Alaska.

2.3. 

\textit{Treatment assignment.} Reducing the concept of “Medicaid expansion” to a binary treatment simplifies a more complex reality. There are at least three reasons to be cautious about this simplification. First, states differed substantially in their Medicaid coverage policies prior to 2014. Given perfect data we might ideally consider Medicaid expansion as a
continuous treatment with values proportional to the number of newly eligible individuals. The challenge is correctly identifying newly eligible individuals in the data (though see Frean, Gruber and Sommers (2017) and Miller, Johnson and Wherry (2021), who attempt to address this). Second, Frean, Gruber and Sommers (2017) note that five states (California, Connecticut, Minnesota, New Jersey, and Washington) and the District of Columbia adopted partial limited Medicaid expansions prior to 2014. The “2014 expansion” therefore actually occurred in part prior to 2014 for several states.\(^6\) Finally, timing is an issue: among the states that expanded Medicaid in 2014, Michigan’s expansion did not go into effect until April 2014, while New Hampshire’s expansion did not occur until September 2014.

Our primary analysis excludes New York, Vermont, Massachusetts, Delaware, and the District of Columbia from our pool of expansion states. These states had comparable Medicaid coverage policies prior to 2014 and therefore reflect invalid comparisons (Kaestner et al. (2017)). We also exclude New Hampshire because it did not expand Medicaid until September 2014. While Michigan expanded Medicaid in April 2014, we leave this state in our pool of “treated” states. We consider the remaining expansion states, including those with early expansions, as “treated” and the non-expansion states, including those that later expanded Medicaid, as “control” states. We later consider the sensitivity of our results to these classifications by removing the early expansion states indicated by Frean, Gruber and Sommers (2017). Our final dataset contains data for 925 CPUMAs, with 414 CPUMAs among 24 non-expansion states and 511 CPUMAs among 21 expansion states.\(^7\) When we exclude the early expansion states, we are left with 292 CPUMAs across 16 expansion states. We provide a complete list of states by Medicaid expansion classification in Appendix C.

2.4. Outcome. Our outcome is the non-elderly (individuals aged 19-64) adult uninsurance rate in 2014. While take-up among the Medicaid-eligible population is a more natural outcome, we choose the non-elderly adult uninsurance rate for two reasons, one theoretic and one practical. First, Medicaid eligibility post-expansion is likely endogenous: Medicaid expansion may affect an individual’s income and poverty levels, which often define Medicaid eligibility. Second, we can better compare our results with the existing literature, including Courtemanche et al. (2017), who also use this outcome. One drawback is that the simultaneous adoption of other ACA provisions by all states in 2014 also affects this outcome. As a result, we only attempt to estimate the effect of Medicaid expansion in 2014 in the context of this changing policy environment. We discuss this further in Sections 3.2 and 3.3.

2.5. Covariates. We choose our covariates to approximately align with those considered in Courtemanche et al. (2017) and that are likely to be potential confounders. Specifically, using the ACS public use microdata, we calculate the unemployment and uninsurance rates for each CPUMA from 2011 through 2013. We also estimate a variety of demographic characteristics averaged over this same time period, including percent female, white, married, Hispanic ethnicity, foreign-born, disabled, students, and citizens. We estimate the percent in discrete age categories, education attainment categories, income-to-poverty ratio categories, and categories of number of children. Finally, we calculate the average population growth and number of households to adults. We provide a more extensive description of our calculation of these variables in Appendix B.

\(^6\)Kaestner et al. (2017) and Courtemanche et al. (2017) also consider Arizona, Colorado, Hawaii, Illinois, Iowa, Maryland, and Oregon to have had early expansions.

\(^7\)We additionally include the 4 CPUMAs from New Hampshire in the covariate adjustment procedure described in Section 3.
In addition to the ACS microdata we use 2010 Census data to estimate the percentage living in an urban area for each CPUMA. Lastly, we include three state-level covariates reflecting the partisan composition of each state’s government in 2013. These include an indicator for states with a Republican governor, Republican control over the lower legislative chamber, and Republican control over both legislative chambers and the governorship. ⁸

3. Methods. In this section we present our causal estimand, identifying assumptions, estimation strategy, and inferential procedure. Our primary methodological contributions are contained in the subsection on estimation. We begin by outlining notation.

3.1. Notation. We let \( s \) index states, and \( c \) index CPUMAs within states. Let \( m \) denote the number of states, \( p_s \) the number of CPUMAs in state \( s \), and \( n = \sum_{s=1}^{m} p_s \) the total number of CPUMAs. For each state \( s \), let \( A_s \) denote its treatment assignment according to the discussion given in Section 2.3, with \( A_s = 1 \) indicating treatment and \( A_s = 0 \) indicating control. For each CPUMA \( c \) in state \( s \), let \( Y_{sc} \) denote its uninsurance rate in 2014; let \( X_{sc} \) denote a \( q \)-dimensional covariate vector; and let \( A_{sc} = A_s \) denote its treatment status. We assume potential outcomes (Rubin (2005)), defining a CPUMA’s potential uninsurance under treatment by \( Y_{sc}^1 \), and under control by \( Y_{sc}^0 \). Finally, let \( n_1 \) and \( n_0 \) denote the number of treated and control CPUMAs, and define \( m_1 \) and \( m_0 \) analogously for states.

Given a collection of objects indexed over CPUMAs, we denote the complete set by removing the subscript. For example, \( X \) denotes \( \{ X_{sc} : s \in C \} \), where \( C \) is the set of all CPUMAs. Subscripting by the labels \( A = 1 \) or \( A = 0 \) denotes the subset corresponding to the treated or control units; for example, \( X_{A=0} \) denotes \( \{ X_{sc} : A_{sc} = 0 \} \), the covariates of the control units. To denote averaging, we will use an overbar, while also abbreviating \( A = 0 \) and \( A = 1 \) respectively by 0 and 1. For example, \( \bar{X}_0 \) (which abbreviates \( X_{A=0} \)) denotes the average covariate vector for the control units, and \( \bar{Y}_1 \) denotes the average potential outcome under treatment for the control units.

3.2. Estimand. Letting \( X, Y^1, \) and \( Y^0 \) be random, we define the causal estimand \( \psi_0 \)

\[
\psi_0 = \mathbb{E}[Y_0^1 - Y_0^0 \mid X_{A=0}]
\]

\[
= \psi_0^1 - \psi_0^0
\]

where \( \psi_0^0 \) denotes the expectation \( \mathbb{E}[Y_0^0 \mid X_{A=0}] \). The estimand \( \psi_0 \) represents the expected treatment effect on non-expansion states conditioning on \( X_{A=0} \), the observed covariate values of the non-expansion states (see, e.g., Imbens (2004)). The challenge is that we do not observe \( Y^1_{A=0} \), the counterfactual outcomes for non-expansion CPUMAs had their states expanded their Medicaid programs, nor their average \( \bar{Y}_0^1 \). We therefore require causal assumptions to identify this counterfactual quantity using our observed data. ⁹

---

⁸ Nebraska is the only state with a unicameral legislature and the legislature is technically non-partisan. We nevertheless classified them as having Republican control of the legislature for this analysis.

⁹ As noted previously, the 2014 Medicaid expansion occurred simultaneously with the implementation of several other major ACA provisions, including (but not limited to) the creation of the ACA-marketplace exchanges, the individual mandate, health insurance subsidies, and community-rating and guaranteed issue of insurance plans (Courtemanche et al. (2017)). Almost all states broadly implemented these reforms beginning January 2014. Conceptually we think of the other ACA components as a state-level treatment \( R \) separate from Medicaid expansion \( A \). Our total estimated effect may include interactions between these policy changes; however, we do not attempt to separately identify these effects. Without further assumptions, we therefore cannot generalize these results beyond 2014.
3.3. Identification. We appeal to the following causal assumptions to identify $\psi_0$ from our observed data: the stable unit treatment value assumption (SUTVA), no unmeasured confounding given the true covariates, and no anticipatory treatment effects. We also invoke parametric assumptions to model the measurement error and to express our estimand in terms of parameters from a linear model. We conclude by using ideas from the “regression-calibration” literature (see, e.g., Gleser (1992)) to ensure that identifying our target estimand is possible given auxiliary data on the measurement error covariance matrix.

We first assume the SUTVA at the CPUMA level. Assuming the SUTVA has two implications for our analysis: first, that there is only one version of treatment; second, that each unit’s potential outcome only depends on its treatment assignment. We discussed potential violations of the first implication previously when considering how to reduce Medicaid expansion to a binary treatment. The second implication could be violated if one CPUMA’s expansion decision affected uninsurance rates in another CPUMA (see, e.g., Frean, Gruber and Sommers (2017)). On the other hand, our assumption allows for interference among individuals living within CPUMAs and is therefore weaker than assuming no interference among any individuals. Further addressing this is beyond the scope of this paper.

Second, we assume no effects of treatment on the observed covariates. This includes assuming no anticipatory effects on pre-2014 uninsurance rates. This is violated in our study, as some treated states allowed specific counties to expand their Medicaid programs prior to 2014, thereby affecting their pre-2014 uninsurance rates. We later test the sensitivity of our results to the exclusion of these states.

Third, we assume no unmeasured confounding. Specifically, we posit that in 2014 the potential outcomes for each CPUMA are jointly independent of the state-level treatment assignment conditional on CPUMA and state-level covariates $X_{sc}$:

\[(Y^1_{sc}, Y^0_{sc}) \perp A_s \mid X_{sc}\]

The covariate vector $X_{sc}$ includes both time-varying pre-treatment covariates, including pre-treatment outcomes, and covariates averaged across 2011-2013, such as demographic characteristics, and the state-level governance indicators discussed in Section 2.2. We believe this assumption is reasonable given our rich covariate set.

Fourth, we assume that the outcomes for each treatment group are linear in the true covariates:

\[Y^a_{sc} = \alpha_a + X^\top_{sc} \beta_a + \epsilon_{sc} + \varepsilon_s \quad a = 0, 1\]

where $^\top$ denotes vector transpose, and the errors $\epsilon_{sc}$ and $\varepsilon_s$ are mean-zero; independent from the covariates, treatment assignment, and each other; and have finite variances $\sigma^2_\epsilon$ and $\sigma^2_\varepsilon$, respectively. This implies that the errors for each CPUMA within a given state have a constant within-state correlation $\sigma^2_\varepsilon/(\sigma^2_\epsilon + \sigma^2_\varepsilon)$, which we denote as $\rho$. To fix ideas, $\epsilon_{sc}$ may capture time-specific idiosyncracies at the local level, possibly due to the local policy or economic conditions. By contrast $\varepsilon_s$ captures time-specific idiosyncracies at the state-level that are common across CPUMAs within a state due to the shared policy and economic environment.

Fifth, we assume that the covariates $X$ and outcomes $Y$ are not observed directly. Instead, survey sampled versions $W$ and $J$ are available, with additive gaussian measurement error arising due to sample variability.

\footnote{Because our covariates include pre-treatment outcomes, this assumption also implies that $\epsilon_{sc}$ and $\varepsilon_{sc}$ are uncorrelated with pre-treatment outcomes, including any error terms that might appear in their generative models.}
\( J_{sc} = Y_{sc} + \xi_{sc} \)

where \((\xi_{sc}, \nu_{sc})\) is independent of \((X, Y)\) and has distribution

\[
(\xi_{sc}, \nu_{sc}) \overset{\text{ind}}{\sim} \text{MVN} \left( 0, \begin{bmatrix} \sigma^2_{\xi,sc} & 0 \\ 0 & \Sigma_{\nu,sc} \end{bmatrix} \right)
\]

We believe equations (5) and (6) are reasonable because measurement error in our context is sampling variability. While equation (6) further implies the measurement errors in our covariates and outcomes are uncorrelated, this is also reasonable because our outcomes are measured on a different cross-sectional survey than our covariates.\(^{11}\)

Sixth, we assume that the covariates for the treated units \(X_{sc}\) are drawn i.i.d. multivariate normal conditional on treatment:

\[
X_{sc} | A_{sc} = 1 \overset{\text{iid}}{\sim} \text{MVN} (\nu_1, \Sigma_{X|1})
\]

Under equations (6)-(7), the conditional expectation of \(X_{sc}\) given noisy observation \(W\) among the treated units can be seen to equal

\[
\mathbb{E}[X_{sc}|W, A] = \nu_1 + \Sigma_{X|1} (\Sigma_{X|1} + \Sigma_{\nu,sc})^{-1} (W_{sc} - \nu_1), \quad \forall \, sc : A_{sc} = 1
\]

Equation (7) is a convenient simplification to motivate equation (8). For example, \(X_{sc}\) includes state-level covariates, so the covariates cannot be independent. More generally, many of the covariates are bounded, and therefore cannot be gaussian. In fact, assuming (7) is not strictly necessary for consistent estimation of \(\psi_0^1\) (see, e.g., Gleser (1992)); however, it is required by the weighting approaches that we consider here. In our validation experiments described in Section 4.3, we find that our approaches which assume (7) outperform those that do not, as the latter group evidently relies more heavily on the linearity assumption (4).\(^{12}\)

Regardless, to use equation (8) to estimate \(\mathbb{E}[X_{sc}|W_{sc}, A_{sc} = 1]\), we require \(\nu_1, \Sigma_{\nu,sc},\) and \(\Sigma_{X|1}\). \(\bar{W}_1\) provides a consistent estimate of \(\nu_1\). However, the data does not identify \(\Sigma_{\nu,sc}\) and \(\Sigma_{X|1}\). Our final assumption is that we can consistently estimate the covariance matrices \(\Sigma_{\nu,sc}\) and \(\Sigma_{X|1}\) using auxiliary data, so that we can use (8) to estimate the conditional mean. The ACS microdata serves as our auxiliary data; further details are discussed in Section 3.4.2.

Under these assumptions we can rewrite our causal estimand in terms of the model parameters. As \(\epsilon_{sc}\) and \(\varepsilon_{s}\) are zero-mean and independent of covariates, we can rewrite

\[
\psi_0^a = \mathbb{E}[Y_0^a | X_{A=0}] \quad \text{by applying expectations to (4), yielding}
\]

\[
\psi_0^a = \alpha_a + \bar{X}_0^T \beta_a
\]

\(\text{and}

\[
W_{sc} = X_{sc} + \nu_{sc}
\]
If we observed \((A, Y, X)\), we would have \(Y_{A=a} \text{ and } X_{A=a}\) for \(a = \{0, 1\}\). Therefore, by (4) the data would identify \((\alpha_a, \beta_a)\), which identifies \(\psi_0^a\) since \(\bar{X}_0\) is observed. However, we only observe the noisy measurements \(J\) and \(W\). As \(\xi_{sc}\) is zero-mean, equation (5) implies that \(J_0\) estimates \(Y_0^1\) and therefore \(\psi_0^1\). Estimating \(\psi_0^1\) remains challenging: as \(Y_{sc}^1 \bot A_{sc} | X_{sc} \equiv J_{sc} \bot A_{sc} | W_{sc}\), it is well-known that noisy measurements will bias standard estimation procedures, such as linear regression, that naively use them without adjustment (see also Appendix A.1).

Let \(\bar{X}_{sc} = \mathbb{E}[X_{sc}|W, A]\) abbreviate the conditional expectation of the covariates given the noisy observations, as given by equation (8). Substituting \(X_{sc} = \bar{X}_{sc} + \tilde{X}_{sc} - \bar{X}_{sc}\) into the outcome model given by (4) and then (4) into (5) yields

\[
J_{sc} = \alpha_1 + \tilde{X}_{sc}^\top \beta_1 + (X_{sc} - \bar{X}_{sc})^\top \beta_1 + \xi_{sc} + \epsilon_{sc} + \bar{e}_s \quad \forall sc: A_{sc} = 1
\]

As \(X_{sc} - \tilde{X}_{sc}\) equals \(X_{sc} - \mathbb{E}[X_{sc}|W, A]\), this quantity is zero-mean conditioned on \((W, A)\). The term \((X_{sc} - \bar{X}_{sc})^\top \beta_1\) appearing in (10) is therefore also conditionally zero-mean. Finally, the outcome noise \(\xi_{sc}\) as well as the noise terms \(\epsilon_{sc}\) and \(\bar{e}_s\) appearing in this equation are also zero-mean and independent of \((W, A)\). It follows that if we observed \((\bar{X}, J, A)\), we would have \(\bar{X}_{A=1}\) and \(J_{A=1}\); therefore, by (10) the data would identify \((\alpha_1, \beta_1)\). In turn, equation (9) implies that \(\alpha_1, \beta_1\) and \(\bar{X}_{A=0}\) (which can be estimated without bias from \(W_{A=0}\)) identifies \(\psi_0^1\). Since have assumed that \(\bar{X}\) follows equation (8), and that we have auxiliary data available to estimate this equation, we therefore have sufficient data to estimate \(\psi_0\) under our models and assumptions. We now discuss estimation.

3.4. Estimation. We propose to use approximate balancing weights to estimate \(\psi_0^1\). We first review approximate balancing weights and the SBW objective proposed by Zubizarreta (2015). These methods typically assume that the covariates are measured without error. We will show in Proposition 1 that under the classical-errors-in-variables model, the SBW estimate of \(\psi_0^1\) has the same bias as the OLS estimate.

We first attempt to remove this bias by estimating equation (8), leveraging the ACS microdata replicate survey weights to estimate this model. We consider two adjustments: (a) a homogeneous adjustment that assumes the noise covariance \(\Sigma_{\nu,sc}\) is constant across all CPUMAs; and (b) a heterogeneous adjustment that allows \(\Sigma_{\nu,sc}\) to vary according to the sample sizes associated with each CPUMA. We next propose a modification to SBW that we call H-SBW, which accounts for the state-level random effects \(\bar{e}_s\). Using SBW and H-SBW we generate weights that balance the adjusted data to the mean covariate values of the non-expansion states. To further reduce imbalances that remain after weighting, we consider bias-corrections using ridge-regression augmentation, following Ben-Michael, Feller and Rothstein (2021).

3.4.1. Stable balancing weights. Zubizarreta (2015) proposes the Stable Balancing Weights (SBW) algorithm to generate a set of weights \(\gamma\) that reweight a set of covariates \(Z = \{Z_{sc}\}\) to a target covariate vector \(\nu\) within a tolerance parameter \(\delta\) by solving the optimization problem:

\[
\min_{\gamma} \sum_{sc} \gamma_{sc}^2 \quad \text{such that} \quad \gamma \in \Gamma(Z, \nu, \delta)
\]

where the constraint set \(\Gamma(Z, \nu, \delta)\) is given by

\[
\Gamma(Z, \nu, \delta) = \left\{ \gamma : \sum_{sc} \gamma_{sc} Z_{sc} - \nu \leq \delta, \gamma \geq 0, \sum_{sc} \gamma_{sc} = 1 \right\}
\]
\[ \hat{\psi}_0^1 = \sum_{A,\nu=1} \hat{\gamma}_{sc} Y_{sc} \quad \hat{\psi}_0^0 = \bar{Y}_0^0 \quad \hat{\psi} = \hat{\psi}_0^1 - \hat{\psi}_0^0 \]

In the case where the potential outcomes follow the linear model specified in equation (4), the bias of \( \hat{\psi}_0^1 \) is less than or equal to \( |\beta_1|^T \delta \), and therefore equal to zero if \( \delta = 0 \) (Zubizarreta, 2015). Moreover, \( \hat{\psi}_0^0 \) produces the minimum variance estimator within the constraint set - conditional on \( X \) - assuming that the errors in the outcome model are independent and identically distributed.

### 3.4.2. Measurement error

In the presence of measurement error the estimation procedure described in Section 3.4.1 will be biased. We show in Appendix A.1, Proposition 1, that under the classical errors-in-variables model where \( \Sigma_{\nu,sc} = \Sigma_{\nu} \) for all units, if \( \hat{\gamma} \) is found by solving the SBW objective (11) with \( Z = X_{A=1} \), \( v = 0 \) equal to the estimated mean \( \bar{W}_0 \) of the control units, and \( \delta = 0 \), the estimator \( \hat{\psi}_0^1 \) in equation (12) has bias

\[ \mathbb{E}[\hat{\psi}_0^1] - \psi_0^1 = (\bar{X}_0 - v_1)^T (\kappa_1 - \delta) \beta_1 \]

where \( \kappa_1 = (\Sigma_{X|1} + \Sigma_{\nu})^{-1} \Sigma_{X|1} \). This is equivalent to the bias for an OLS estimator of \( \psi_0^1 \), where \( (\alpha_1, \beta_1) \) are estimated by regression of \( Y_{A=1} \) on \( W_{A=1} \).

We mitigate this bias by setting \( Z = \bar{X}_{A=1} \), where \( \bar{X}_{A=1} \) is an estimate of \( \bar{X}_{A=1} \) given by equation (8). This requires estimating \( v_1, \Sigma_{\nu,sc} \) and \( \Sigma_{X|1} \). To estimate \( v_1 \) we simply use \( \bar{W}_1 \). To estimate \( \Sigma_{X|1} \) and \( \Sigma_{\nu,sc} \) we use the ACS microdata’s set of 80 replicate survey weights to construct 80 additional CPUMA-level datasets. For each CPUMA among the treated states, we take the empirical covariance matrix of its covariates over the datasets to derive unpooled estimates \( \Sigma_{\nu,sc} \), which we average over CPUMAs to create \( \hat{\Sigma}_\nu \). We then estimate \( \Sigma_{X|1} \) by subtracting \( \hat{\Sigma}_\nu \) from the empirical covariance matrix of \( W_{A=1} \),

\[ \hat{\Sigma}_{X|1} = \frac{1}{n_1} \sum_{sc:A,sc=1} (W_{sc} - \bar{W}_1)(W_{sc} - \bar{W}_1)^\top - \hat{\Sigma}_\nu \]

We consider two estimates of \( \Sigma_{\nu,sc} \): first, where we let \( \hat{\Sigma}_{\nu,sc} = \hat{\Sigma}_\nu \) for all units, which we call the homogeneous adjustment; second, where each \( \hat{\Sigma}_{\nu,sc} \) equals \( \hat{\Sigma}_\nu \) rescaled according to the sample size of the estimate \( W_{sc} \), which we call the heterogeneous adjustment. We describe these adjustments fully in Appendix B. Using \( \hat{\Sigma}_{X|1} \) and \( \hat{\Sigma}_{\nu,sc} \), we estimate \( \bar{X}_{A=1} \) using the empirical version of equation (8), inducing estimates \( \bar{X}_{A=1} \) given by

\[ \bar{X}_{sc} = \bar{W}_1 + \hat{\Sigma}_{X|1}(\hat{\Sigma}_{X|1} + \hat{\Sigma}_{\nu,sc})^{-1}(W_{sc} - \bar{W}_1), \quad \forall sc: A_{sc} = 1 \]

We then compute debiased balancing weights \( \hat{\gamma} \) by solving equation (11) with \( Z = \bar{X}_{A=1}, v = \bar{W}_0 \), and tuning parameter \( \delta \) chosen as described in Section 3.4.4. Given \( \hat{\gamma} \), we find \( \hat{\psi}_0^1 \) and \( \hat{\psi} \) again using equation (12).

The homogeneous adjustment approximately aligns with the adjustments suggested by Carroll et al. (2006) and Gleser (1992). In Appendix A.1, Propositions 2-4, we show that this procedure returns consistent estimates of \( \psi_0^1 \) and \( \psi_0 \) under the identifying assumptions discussed and assuming \( \delta = 0 \) is feasible. This is the first application we are aware of to
apply regression calibration in the context of balancing weights to address measurement error. However, this method requires access to knowledge about \( \Sigma_\nu \). We use survey microdata to identify this parameter for our application. Alternatively, region-level datasets often contain region-level variance estimates. If a researcher is willing to assume \( \Sigma_\nu \) is diagonal, she could leverage this information to use this approach. If no auxiliary data is available, she could also consider \( \Sigma_\nu \) to be a sensitivity parameter and conduct estimates over a range of possible values (see, e.g., Huque, Bondell and Ryan (2014), Illenberger, Small and Shaw (2020)).

3.4.3. H-SBW criterion. Unlike the setting outlined in Zubizarreta (2015), our application likely has state-level dependencies in the error terms which may increase the variance of the SBW estimator. We therefore add a tuning parameter \( \rho \in [0, 1) \) to penalize the within-state cross product of the weights, as detailed in equation (14), representing a constant within-state correlation of the errors.

\[
\min_{\gamma} \sum_{s=1}^{m} \left( \sum_{c=1}^{p_s} \gamma_{sc}^2 + \sum_{c \neq d} \rho \gamma_{sc} \gamma_{sd} \right) \quad \text{such that} \quad \gamma \in \Gamma(Z, \nu, \delta)
\]

To build intuition about this objective, for \( \delta \to \infty \), the following solution is attained:

\[
\hat{\gamma}_{sc} \propto \frac{1}{(p_s - 1)\rho + 1}
\]

Setting \( \rho = 0 \) returns the SBW solution: \( \hat{\gamma}_{sc} \propto 1 \). When setting \( \rho \approx 1 \), we get \( \hat{\gamma}_{sc} \propto \frac{1}{p_s} \). In other words, as we increase \( \rho \), this objective downweights CPUMAs in states with large numbers of CPUMAs and upweights CPUMAs in states with small numbers of CPUMAs (assigning each CPUMA within a state equal weight). As we increase \( \rho \), the objective will therefore more uniformly disperse weights across states. We show in Appendix A.3 that solving the H-SBW produces the minimum conditional variance estimator of \( \psi_0 \) within the constraint set assuming homoskedasticity and equicorrelated errors. We also highlight the connection between the H-SBW solution and the implied regression weights from GLS.

An important caveat emerges in the context of measurement error. In settings where the covariates are dependent, the conditional expectation \( \mathbb{E}[X_{sc}| W, A] \) is no longer given by equation (8), and therefore \( \hat{X}_{sc} \) given by equation (13) is no longer consistent for \( \hat{X}_{sc} \). As a result, finding weights solving H-SBW equation (14) with \( Z = \hat{X}_{A=1} \) is not unbiased in these settings. A simulation study in Appendix F also shows that this bias increases with \( \rho \), and that the SBW solution remains approximately unbiased. To regain unbiasedness in general, \( \hat{X}_{sc} \) must be modified from equation (13) to account for dependencies, requiring new modeling assumptions. We demonstrate this more formally in Appendix A.3 and propose an adjustment to account for dependent gaussian covariates in Appendix B.

3.4.4. Hyperparameter selection. Practical guidance in the literature is that \( \delta \) should reduce the standardized mean differences to be less than 0.1 (see, e.g., Zhang et al. (2019)). In our application, all of our covariates are measured on the same scale. Additionally, because some of these covariates have very small variances (for example, percent female), we instead target the percentage point differences. We can then estimate \( \psi_0 \) using equation (12), substituting \( J_{sc} \) for \( Y_{sc} \) and using the weights \( \hat{\gamma} \).

We choose the vector \( \delta \) using domain knowledge about which covariates are most likely to be important predictors of the potential outcomes under treatment, again recalling that the bias of our estimate is bounded by \( \sum_{d=1}^{q} \delta_d |\beta_{1,d}| \) (where we index the covariates from \( d = 1, ..., q \)). Specifically, we know that pre-treatment outcomes are often strong predictors
of post-treatment outcomes, so we constrain the imbalances to fall within 0.05 percentage points (out of 100) for pre-treatment outcomes. Because health insurance is often tied to employment, we also prioritize balancing pre-treatment uninsurance rates, seeking to reduce imbalances below 0.15 percentage points. On the opposite side of the spectrum, we constrain the Republican governance indicators to fall within 25 percentage points. While we believe that these covariates are important to balance, given the data we are unable to reduce the constraints further without generating extreme weights. We detail the remaining constraints in Appendix D.

We consider \( \rho \in \{0, 1/6\} \). The first choice is equivalent to the SBW objective, while the second assumes constant equicorrelation of 1/6. We choose \( \rho \) to be small to limit additional bias induced by H-SBW in the context of dependent data and measurement error.

Data-driven approaches to select these parameters could also be used. For example, absent measurement error if pre-treatment outcomes and covariates were available one could use the residuals from GLS to estimate \( \rho \). Data-driven procedures for \( \delta \) are also possible. Wang and Zubizarreta (2020) propose a data-driven approach that only uses the covariate information. When data exists for a long pre-treatment period, Abadie, Diamond and Hainmueller (2015) propose tuning their weights with respect to covariate balance using a “training” and “validation” period, an idea that could be adapted to choose \( \delta \). Expanding these ideas to this setting would be an interesting area for future work.

3.5. Sensitivity to covariate imbalance. Our initial set of weights allow for some large covariate imbalances. We follow the proposal of Ben-Michael, Feller and Rothstein (2021) and use ridge-regression augmentation to reduce these imbalances. While these weights achieve better covariate balance, this comes at the cost of extrapolating beyond the support of the data. Letting \( \hat{\gamma}^{H-SBW} \) be our H-SBW weights, and \( \hat{X}_1 \) denote the matrix whose columns are the members of \( \hat{X}_{A=1} \), we define these weights as:

\[
\hat{\gamma}_{BC-HSBW} = \hat{\gamma}^{H-SBW} + (\hat{X}_1 \hat{\gamma}^{H-SBW} - \bar{W}_0)^\top (\hat{X}_1 \Omega^{-1} \hat{X}_1^\top + \lambda I_q)^{-1} \hat{X}_1 \Omega^{-1}
\]

where \( \Omega \) is a block diagonal matrix with diagonal entries equal to one and the within-group off diagonals equal to \( \rho \). We choose \( \lambda \) so that all imbalances fall within 0.5 percentage points.

We refer to Ben-Michael, Feller and Rothstein (2021) for more details about this procedure. For our results we consider estimators using SBW weights (\( \rho = 0 \)), H-SBW weights (\( \rho = 1/6 \)), and their ridge-augmented versions that we respectively call BC-SBW and BC-HSBW.

---

13 While both our approach and the synthetic controls literature prioritizes balancing pre-treatment outcomes, the motivation is somewhat different. The synthetic controls literature frequently assumes that the potential outcomes absent treatment follow a linear factor model. Under this model a heuristic is that the effects of covariates must show up via the pre-treatment outcomes, so that balancing pre-treatment outcomes is sufficient to balance any other relevant covariates (see Botosaru and Ferman (2019), who formalize this idea). However, both the theory and common practice recommends balancing a large vector of pre-treatment outcomes. We instead have very limited pre-treatment data. We therefore do not assume a factor model, and instead treat the pre-treatment outcomes simply as covariates. We then justify prioritizing balancing these pre-treatment outcomes assuming they have large coefficients relative to other covariates in the assumed model for the (post-treatment) potential outcomes under treatment.

14 Our key estimand - \( \psi_1^1 \) - differs from the traditional target of the synthetic controls literature - \( \psi_0^1 \). If there exist covariates that are not relevant for producing \( Y_0^1 \) but are relevant for \( Y_1^1 \), then balancing pre-treatment outcomes alone will not in general balance these covariates (Botosaru and Ferman (2019)). Therefore, even with access to a large vector of pre-treatment outcomes and assuming that \( Y_1^1 \) follows a factor model, explicitly balancing such covariates in addition to pre-treatment outcomes may be necessary to estimate the ETC well.
3.6. Model validation. To check model validity, we rerun our procedures on pre-treatment data to compare the performance of our models for a fixed \( \delta \). In particular, we train our model on 2009-2011 data to predict 2012 outcomes, and 2010-2012 data to predict 2013 outcomes. We limit to one-year prediction error since our estimand is only one-year forward. We examine the performance of SBW against H-SBW and their bias-corrected versions BC-SBW and BC-HSBW, using the covariate adjustment methods described in Section 3.4.2 to account for measurement error. In Appendix E, we additionally compare to “Oaxaca-Blinder” OLS and GLS weights (see, e.g. Kline (2011)). The OLS weights do not require the gaussian assumption of (7) for consistency, but rely more heavily on the linear model (4).

3.7. Inference. We use the leave-one-state-out jackknife to estimate the variance of \( \hat{\psi}_0^1 \) (see, e.g., Cameron and Miller (2015)). Specifically, we take the pool of treated states and generate a list of datasets that exclude each state. For each dataset in this list we calculate the weights and the leave-one-state-out estimate of \( \psi_0^1 \). Throughout all iterations we hold our targeted mean fixed at \( \hat{W}_0 \). When generating these estimates, if our preferred initial choice of \( \delta \) does not converge, we gradually reduce the constraints (increase \( \delta \)) until we can obtain a solution. For each dataset we also re-estimate \( \hat{X}_{A=1} \) before estimating the weights to account for the variability in the covariate adjustment procedure. We then estimate the variance:

\[
\text{Var}(\hat{\psi}_0^1) = \frac{m_1 - 1}{m_1} \sum_{s:A_s=1} \left( S(s) - S(\cdot) \right)^2
\]

where \( S(s) \) is the estimator of \( \psi_0^1 \) with treated state \( s \) removed, and \( S(\cdot) = \frac{1}{m_1} \sum_{s:A_s=1} S(s) \). In other settings the jackknife has been shown to be a conservative approximation of the bootstrap, such as in Efron and Stein (1981), which we apply in Appendix A.1, Proposition 5 to give a partial result for our application. In a simulation study mirroring our setting with \( m_1 = 25 \) units (available in Appendix F), we obtain close to nominal coverage rates using these variance estimates.

To estimate the variance of \( \hat{\psi}_0^0 \) we run an auxiliary regression model on the non-expansion states and estimate the variance of the linear combination \( \hat{W}_0^T \hat{\beta}_0 \) using cluster-robust standard errors. We do not need to adjust the non-expansion state data to estimate this quantity: a linear regression line always contains the point \( (\hat{W}_0, J_0) \), which are unbiased estimates of \( (\hat{X}_0, \psi_0^0) \). Therefore, \( \mathbb{E}(\hat{W}_0^T \hat{\beta}_0 | X) = \psi_0^0 \). Our final variance estimate \( \text{Var}(\hat{\psi}_0^0) \) is the sum of \( \text{Var}(\hat{\psi}_0^1) \) and \( \text{Var}(\hat{\psi}_0^0) \). We use the t-distribution with \( m_1 - 1 \) degrees of freedom to generate 95 percent confidence intervals.

4. Results. We first present summary statistics regarding the variability of the pre-treatment outcomes on our adjusted and unadjusted datasets. The second sub-section contains covariate balance diagnostics. The third sub-section contains a validation study, and the final sub-section contains our ETC estimates.

---

15 The jackknife approximates the bootstrap, which is sometimes used to estimate the variance of the OLS-based estimates using regression-calibration in the standard setting with i.i.d. data (Carroll et al. (2006)).

16 That is, we treat \( \hat{W}_0 \) as identical to \( X_0 \) and ignore the variability of the estimate. This variability is of smaller order than the variability in \( \hat{\psi}_0^1 \); the former does not depend on the number of states but instead decreases with the number of CPUMAs among the control states and the sample sizes used to estimate each CPUMA-level covariate. Recall that \( \hat{X}_0 \) is fixed because our estimand is conditional on the observed covariate values of the non-expansion states.

17 When more substantial undercoverage occurs it is likely due to bias.

18 The latter is much smaller than the former – specifically, we estimate \( \text{Var}(\hat{\psi}_0^0) = 0.017 \).
4.1. **Covariate adjustment.** Table 1 displays the sample variance of our pre-treatment outcomes among the expansion states using each covariate adjustment strategy (see Section 3.4.2 for definitions of these adjustments). Although we most heavily prioritize balancing these covariates, they are also among the least precisely estimated, as most of our other covariates average over multiple years of data. Both the homogeneous and heterogeneous adjustments reduce the variability in the data by comparable amounts. Intuitively, these adjustment reduce the likelihood that our balancing weights will fit to noise in the covariate measurements. These results are consistent across most of our other covariates. Tables containing distributional information for each covariate are available in Appendix C.

<table>
<thead>
<tr>
<th>Variable Unadjusted</th>
<th>Heterogeneous</th>
<th>Homogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninsured Pct 2011</td>
<td>8.35</td>
<td>8.04</td>
</tr>
<tr>
<td>Uninsured Pct 2012</td>
<td>8.20</td>
<td>7.89</td>
</tr>
<tr>
<td>Uninsured Pct 2013</td>
<td>8.09</td>
<td>7.78</td>
</tr>
</tbody>
</table>

4.2. **Covariate balance.** Figure 1 displays the weighted and unweighted imbalances in our adjusted covariate set (using the homogeneous adjustment) using the H-SBW weights. Our unweighted data shows substantial imbalances in the Republican governance indicators as well as pre-treatment uninsurance rates. H-SBW reduces these differences; however, some remain, particularly among the Republican governance indicators. All other imbalances are relatively small, both on the absolute difference and standardized mean difference scale. In fact, despite not targeting the SMD, all but five remaining covariate imbalances fall within 0.1 SMD, compared to 23 covariates prior to reweighting. The largest remaining imbalance on the SMD scale is among percent female (-0.3), though, as noted previously, this variable has low variance in our dataset. A complete balance table is in Appendix D.

**Fig 1: Balance plot, homogeneous adjustment (primary dataset)**
To further reduce these imbalances we augment these weights using ridge-regression. Figure 2 shows the total weights summed across states (and standardized to sum to 100) for three estimators: H-SBW, BC-HSBW, and SBW. For BC-HSBW we also display the sum of the negative weights separately from the sum of the positive weights to highlight the extrapolation. This figure illustrates two key points: first, that H-SBW more evenly disperses the weights across states relative to SBW; second, that BC-HSBW extrapolates somewhat heavily in order to achieve balance, particularly from CPUMAs in California. This is likely in part because California has the most CPUMAs of any state in the dataset.

4.3. Model validation. We compare the performance of our models by repeating the covariate adjustments and calculating our procedure on 2009-2011 ACS data to predict 2012 outcomes, and similarly for 2010-2012 data to predict 2013 outcomes for the non-expansion states. We run this procedure using the primary dataset and excluding the early expansion states. Table 2 displays the mean error and RMSE of these results, with the rows ordered by RMSE for each dataset. We see that the estimators trained on the covariate adjusted data perform substantially better than the unadjusted data and the estimators trained on the homogeneous adjustment outperform their counterparts on the heterogeneous adjustment. We therefore prioritize the results using the homogeneous adjustment. We also observe that H-SBW has comparable performance to SBW throughout.

Interestingly, the bias-corrected estimators perform relatively poorly on the primary dataset, but are the best performing estimators when excluding early expansion states – with RMSEs lower than any other estimator on the primary dataset. As the early expansion states include California, and Figure 2 shows that the bias-corrected estimators extrapolate heavily from this state, the differences in these results suggest that preventing this extrapolation may improve the model’s performance. While these results do not imply that the bias-corrected models will perform poorly when predicting $\psi_{10}$ in the post-treatment period on the primary dataset (or that they will perform especially well when excluding the early expansion states),
it does highlight the dangers of extrapolation: linearity may approximately hold on the support of the data where we have sufficient covariate overlap, but beyond this region this may be a more costly assumption.\footnote{In Table 13 in Appendix \textit{E}, we also compare the performance against the implied regression weights from OLS and GLS. These weights exactly balance the observed covariates, but are almost always the worst performing estimators in the validation study on either dataset. This also illustrates the benefits of the regularization inherent in the bias-corrected weights.}

Finally, we observe that the mean errors for the estimators generated on the unadjusted dataset are all negative, reflecting that these estimates under-predict the true uninsurance rate (see also Table 13 in Appendix \textit{E}, which shows similar results for each year individually). These results likely reflect a form of regression-to-the-mean caused by overfitting our weights to noisy covariate measurements. More formally, we can think of the uninsurance rates in time period $t$ in expansion and non-expansion regions as being drawn from separate distributions with means $(\upsilon_1, \upsilon_0)$, respectively, where $\upsilon_1 < \upsilon_0$. For simplicity assume that the only covariate in the outcome model at time $t$ is $Y_{sct-1}$. Under equation (4), we obtain $Y_{sc} = Y_{sct} = \alpha_1 + \beta_1 Y_{sct-1} + \epsilon_{sc} + \epsilon_{st}$. The pre-treatment outcomes are likely positively correlated with the post-treatment outcomes, implying that $\beta_1 > 0$. Because $\upsilon_1 < \upsilon_0$, when reweighting the vector of noisy pre-treatment outcomes $J_{A=1}$ to $\upsilon_0$ the expected value of the weighted measurement error $E[\sum_{A=1} \hat{\gamma}_{sc} Y_{sct-1}]$ should be positive. In other words, our weights are likely to favor units with covariate measurements that are greater than their true covariate values. The expected value of the weighted pre-treatment outcome will then be less than the target $\upsilon_0$. This implies that our estimates will be negatively biased, since $E[\beta_1 \sum_{A=1} \hat{\gamma}_{sc} Y_{sct-1} - \upsilon_0] \leq 0$.\footnote{This phenomenon has also been discussed in the difference-in-differences and synthetic controls literature (see, e.g., \textit{Daw and Hatfield (2018)}).}

While our covariate adjustments are meant to eliminate this bias, in practice they appear more likely only to reduce it. Assuming these errors reflect a (slight) negative bias that will also hold for our estimates of $\psi_0$, these results suggest that the true treatment effect may be closer to zero than our estimates.

4.4. Primary Results. Table 3 presents all of our estimates of the ETC. Using H-SBW we estimate an effect of $-2.33$ (-3.54, -1.11) percentage points on our primary dataset. The SBW results are almost identical with $-2.35$ (-3.76, -0.95) percentage points. Compared to the unadjusted data we find very similar point estimates at $-2.34$ (-2.88, -1.79) percentage points for H-SBW and $-2.39$ (-2.99, -1.79) percentage points for SBW. We see that H-SBW

\begin{table}
\centering
\caption{Estimator pre-treatment outcome mean prediction error and RMSE (in \% pts)}
\begin{tabular}{llll|llll}
\hline
\textbf{Sigma estimate} & \textbf{Estimator} & \textbf{Mean Error} & \textbf{RMSE} & \textbf{Sigma estimate} & \textbf{Estimator} & \textbf{Mean Error} & \textbf{RMSE} \\
\hline
Homogeneous & SBW & -0.20 & 0.20 & Homogeneous & BC-HSBW & -0.02 & 0.07 \\
Homogeneous & H-SBW & -0.23 & 0.23 & Homogeneous & BC-SBW & -0.03 & 0.12 \\
Homogeneous & SBW & -0.27 & 0.27 & Heterogeneous & BC-HSBW & -0.08 & 0.15 \\
Homogeneous & H-SBW & -0.39 & 0.36 & Homogeneous & BC-SBW & -0.03 & 0.25 \\
Heterogeneous & BC-SBW & -0.39 & 0.39 & Homogeneous & SBW & 0.07 & 0.26 \\
Homogeneous & H-SBW & -0.42 & 0.42 & Heterogeneous & BC-SBW & -0.07 & 0.28 \\
Unadjusted & SBW & -0.56 & 0.56 & Unadjusted & H-SBW & -0.04 & 0.29 \\
Unadjusted & H-SBW & -0.57 & 0.57 & Heterogeneous & BC-HSBW & 0.01 & 0.25 \\
Homogeneous & BC-SBW & -0.58 & 0.58 & Unadjusted & SBW & -0.37 & 0.42 \\
Heterogeneous & BC-SBW & -0.63 & 0.63 & Heterogeneous & BC-HSBW & -0.43 & 0.46 \\
Unadjusted & BC-SBW & -0.88 & 0.88 & Unadjusted & BC-HSBW & -0.60 & 0.60 \\
Unadjusted & BC-HSBW & -0.96 & 0.96 & Unadjusted & BC-HSBW & -0.70 & 0.71 \\
\hline
\end{tabular}
\end{table}
reduces the confidence interval length relative to SBW on our primary dataset, though the lengths are nearly identical when excluding early expansion states. This suggests that H-SBW had at best only modest variance improvements relative to SBW in this setting. Using the adjusted covariate set also increases the length of the confidence intervals relative to the unadjusted data. This increase in the estimated variance is expected in part because the adjustment procedure generally reduces the variability in the data, as we saw in Table 1, requiring that the variance of the weights to increase to achieve approximate balance. More generally this increase also reflects the additional uncertainty due to the measurement error.

Adding the bias-correction decreases the absolute magnitude of the estimates: we estimate effects of -2.05 (-3.30, -0.80) percentage points for BC-HSBW and -2.07 (-3.14, -1.00) percentage points for BC-SBW. This contrasts to our validation study, where the bias-corrected estimators tended to predict lower uninsurance rates than the other estimators (implying we might see larger absolute magnitude effect estimates).

<table>
<thead>
<tr>
<th>Weights</th>
<th>Adjustment</th>
<th>Estimate (95% CI)</th>
<th>Difference</th>
<th>Estimate (95% CI)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-SBW</td>
<td>Homogeneous</td>
<td>-2.33 (-3.54, -1.11)</td>
<td>0.01</td>
<td>-2.09 (-3.24, -0.94)</td>
<td>0.19</td>
</tr>
<tr>
<td>H-SBW</td>
<td>Unadjusted</td>
<td>-2.34 (-2.88, -1.79)</td>
<td>-</td>
<td>-2.28 (-2.87, -1.70)</td>
<td>-</td>
</tr>
<tr>
<td>BC-HSBW</td>
<td>Homogeneous</td>
<td>-2.05 (-3.30, -0.80)</td>
<td>0.17</td>
<td>-1.94 (-3.27, -0.61)</td>
<td>0.28</td>
</tr>
<tr>
<td>BC-HSBW</td>
<td>Unadjusted</td>
<td>-2.22 (-2.91, -1.52)</td>
<td>-</td>
<td>-2.22 (-3.14, -1.31)</td>
<td>-</td>
</tr>
<tr>
<td>SBW</td>
<td>Homogeneous</td>
<td>-2.35 (-3.76, -0.95)</td>
<td>0.04</td>
<td>-2.05 (-3.19, -0.91)</td>
<td>0.16</td>
</tr>
<tr>
<td>SBW</td>
<td>Unadjusted</td>
<td>-2.39 (-2.99, -1.79)</td>
<td>-</td>
<td>-2.21 (-2.75, -1.68)</td>
<td>-</td>
</tr>
<tr>
<td>BC-SBW</td>
<td>Homogeneous</td>
<td>-2.07 (-3.14, -1.00)</td>
<td>0.13</td>
<td>-1.99 (-3.33, -0.66)</td>
<td>0.23</td>
</tr>
<tr>
<td>BC-SBW</td>
<td>Unadjusted</td>
<td>-2.19 (-2.94, -1.45)</td>
<td>-</td>
<td>-2.23 (-3.12, -1.33)</td>
<td>-</td>
</tr>
</tbody>
</table>

(a) “Difference” column reflects difference between adjusted and unadjusted estimators

All adjusted estimators were closer to zero than the corresponding unadjusted estimators. This includes estimates using the heterogeneous adjustment (see Appendix E). However, the difference between the adjusted SBW and H-SBW and the unadjusted versions is close to zero. This contrasts with our validation study where the unadjusted SBW and H-SBW estimators ranged from about 0.3 to 0.4 percentage points lower than the adjusted estimators. We interpret this difference as due to chance: while theory and our validation study suggests that our unadjusted estimators are biased, bias only exists in expectation. Our primary dataset is a random draw where our unadjusted and adjusted estimators give similar estimates.

We next consider the sensitivity of our analysis with respect to the no anticipatory treatment effects assumption by excluding the early expansion states (California, Connecticut, Minnesota, New Jersey, and Washington) and re-running our analyses. The columns under “Early expansion excluded” in Table 3 reflects these results. The overall patterns of the results are consistent with our primary estimates; however, our point estimates almost all move somewhat closer to zero. This may indicate that either the primary estimates have a slight negative bias, or that these estimates have a slight positive bias. Given our analysis of the validation study we view the first case as more likely. This would imply that our primary estimators reflect a lower bound on the true treatment effect. Regardless, the differences are small relative to our uncertainty estimates. Overall we view our primary results as relatively robust to the exclusion of these states. Additional results are available in Appendix E.

21 When excluding the early expansion states, the difference between the estimates on the adjusted and unadjusted data persist but are also smaller in magnitude.
5. Discussion. We divide our discussion into two sections: methodological considerations and policy considerations.

5.1. Methodological considerations and limitations. We make multiple contributions to the literature on balancing weights. First, our estimation procedure accounts for mean-zero random noise in our covariates that is uncorrelated with the outcome errors. We modify the SBW constraint set to balance on a linear approximation to the true covariate values, applying the idea of regression calibration from the measurement error literature to the context of balancing weights. Our results illustrate the benefits of this procedure: using observed pre-treatment outcomes generated by an unknown data generating mechanism, Table 2 demonstrates that our proposed estimators have better predictive performance when balancing on the adjusted covariates. This finding is consistent with concerns about overfitting to noisy covariate measurements and subsequent regression-to-the-mean.

This approach has several limitations: first, it requires access to auxiliary data with which to estimate the measurement error covariance matrix $\Sigma_\nu$. Many applications may not have access to such information. Even without such data, $\Sigma_\nu$ could also be considered a sensitivity parameter to evaluate the robustness of results to measurement error (see, e.g., Huque, Bondell and Ryan (2014), Illenberger, Small and Shaw (2020)). Second, from a theoretic perspective, we require strong distributional assumptions on the covariates to consistently estimate $\psi_0$ using convex balancing weights. This contrasts to Gleser (1992), who shows that the OLS estimates are consistent with only very weak distributional assumptions on the data (see also Propositions 6 and 7 in Appendix A.1). This relates to a third limitation: we require strong outcome modeling assumptions. Yet by preventing extrapolation, SBW and H-SBW estimates may be less sensitive than OLS estimates to these assumptions. Our validation results support this: the standard regression calibration adjustment using OLS and GLS weights performs the worst out of any methods we consider (see Table 13 in Appendix E). In contrast, using regression-calibration with balancing weights – even when allowing for limited extrapolation using ridge-augmentation – performs better. Developing methods to relax these assumptions further would be a valuable area for future work.

A final concern is that this procedure may be sub-optimal with respect to the mean-square error of our estimator. In particular, the bias induced by the measurement error decreases with the sample size used to calculate each CPUMA’s covariate values, the minimum of which were over three hundred. Yet the variance of our counterfactual estimate decreases with the number of treated states. From a theoretic perspective, the variance is of a larger order than the bias, so perhaps the bias from measurement error should not be a primary concern. Our final results support this: the changes in our results on the adjusted versus unadjusted data are of smaller magnitude than the associated uncertainty estimates. Other studies have proposed tests of whether the measurement error corrections are “worth it,” though we do not do this here (see, e.g., Gleser (1992)). However, as an informal observation, our simulation study in Appendix F shows that the MSE of the SBW estimator that naively balances on the noisy covariates $W$ is comparable to the MSE of the SBW estimator that balances on the adjusted covariates $\hat{X}$ when the ratio of the variance of $X$ to the variance of $W$ is 0.95. However, even in this setting we find that confidence interval coverage can fall below nominal rates when balancing on $W$, and that the measurement error correction can improve our ability to construct valid confidence intervals.

Our second contribution is to introduce the H-SBW objective. This objective can improve upon the SBW objective assuming that the errors in the outcome model are homoskedastic.

\footnote{We also provide a suggestion of how to adapt our procedure to accommodate a basis expansion of gaussian covariates in Appendix A.1 Remark 5}
with constant positive equicorrelation $\rho$ within known groups of units. Assuming no measurement error and that $\rho$ is known, we show that H-SBW returns the minimum variance estimator within the constraint set by more evenly dispersing weights across the groups. We also demonstrate the connection between these weights and the implied weights from GLS (see Propositions 9 and 10 in Appendix A.2). While studies have considered balancing weights in settings with hierarchical data (see, e.g., Keele et al. (2020), Ben-Michael, Feller and Hartman (2021)), we are the first to our knowledge to propose changing the criterion to account for correlated outcomes.

This estimation procedure has at least three potential drawbacks. First, we make a very specific assumption on the covariance structure of the error terms that is useful for our application. For applications where a different structure $\Omega$ is more appropriate one can still follow our approach and minimize the more general criterion $f(\gamma) = \gamma^\top \Omega \gamma$. Second, we require specifying the parameters $(\rho, \delta)$ in advance. Choosing $\delta$ is a challenging problem shared with SBW and we do not offer any new suggestions (though see Wang and Zubizarreta (2020), who offer an interesting approach to this problem.) Choosing $\rho$ is a new problem of this estimation procedure. Encouragingly, our simulation study shows the H-SBW estimator for any chosen $\rho$ almost always has lower variance than SBW in the presence of state-level random effects (see Appendix F).\footnote{We caution that this finding solely reflects the simulation space we examined.} Even so, identifying a principled approach to choosing this parameter would be a useful future contribution. Third, in the presence of both measurement error and dependent data, using H-SBW in combination with the standard regression-calibration adjustment may be biased. This bias arises because the standard adjustment assumes independent data. Our simulations show that this bias can increase with $\rho$ and that SBW remains approximately unbiased if the covariates are gaussian - though even in this setting H-SBW may still yield modest MSE improvements. We also show a theoretical modification to the adjustment where H-SBW would return unbiased estimates in this setting (see Propositions 8 and 11).

5.2. Policy considerations and limitations. We estimate that had states that did not expand Medicaid in 2014 instead expanded their programs, they would have seen a -2.33 (-3.54, -1.11) percentage point change in the average adult uninsurance rate. Our validation study and robustness checks indicate that this estimate may be biased downwards (away from zero), in which case we can interpret this estimate as a lower bound on the ETC. Existing estimates place the ETT between -3 and -6 percentage points. These estimates vary depending on the targeted sub-population of interest, the data used, the level of modeling (individuals or regions), and the modeling approach (see, e.g., Courtemanche et al. (2017), Kaestner et al. (2017), Frean, Gruber and Sommers (2017)). Our estimate of the ETC are closer to zero than these estimates. When we attempt to estimate the ETT using our proposed method, the resulting estimates have high uncertainty due to limited covariate overlap.\footnote{In particular, there were no states entirely controlled by Democrats that did not expand Medicaid. Even allowing for large imbalances, our standard error estimates were approximately three percentage points and our confidence intervals all contained zero. When estimating a simple difference-in-differences model on the unadjusted dataset we estimate that the ETT is -2.05 (-3.23, -0.87), where the standard errors account for clustering at the state level.} The differences may reflect different modeling strategies and data, or it may suggest that the ETC is smaller in absolute magnitude than the ETT.

We are ultimately unable to draw any statistical conclusions about these differences; nevertheless, we continue to emphasize caution about assuming these estimands are equal. Due in part to the different covariate distributions of expansion versus non-expansion states, we
may still suspect that the ETC differs from the ETT with respect to the uninsurance rates. Moreover, because almost every outcome of interest is mediated through increasing the number of insured individuals, such a difference may be important. For example, Miller, Johnson and Wherry (2021) study the effect of Medicaid expansion on mortality. Using their estimate of the ETT with respect to mortality they project that had all states expanded Medicaid, 15,600 deaths would have been avoided from 2014 through 2017. If we believe that this number increases monotonically with the number of uninsured individuals, this estimate may be an overestimate if the ETC with respect to the uninsurance rate is less than the ETT, or an underestimate if the ETC is greater than the ETT. Obtaining more precise inferences on the ETC with respect to uninsurance rates, if possible, would therefore be valuable future work.

Our analytic approach is not without limitations. Specifically, we require strong assumptions, including SUTVA, no anticipatory treatment effects, no unmeasured confounding conditional on the true covariates, and several parametric assumptions regarding the outcome and measurement error models. We address some concerns about possible violations of these assumptions. For example, our results were qualitatively similar whether we excluded possible “early expansion states,” or used different weighting strategies (including relaxing the positivity restrictions and changing the tuning parameter $\rho$). However, we do not attempt to address concerns about the impact of spillovers across regions. And while we believe that no unmeasured confounding is reasonable for this problem, we did not conduct a sensitivity analysis (see, e.g., Bonvini and Kennedy (2021)) with respect to this assumption.

Medicaid expansion remains an ongoing policy debate in the United States, where as of 2022 twelve states have not expanded their eligibility requirements. Our study estimates the effect of Medicaid expansion on adult uninsurance rates; however, the primary reason this effect interesting is that Medicaid enrollment is not automatic for eligible individuals. If the goal of Medicaid expansion is to increase insurance access for low-income adults, state policy-makers also may wish to make it easier or even automatic to enroll in Medicaid.

6. Conclusion. We predict the average change in the non-elderly adult uninsurance rate in 2014 among states that did not expand their Medicaid eligibility thresholds as if they had. We use survey data aggregated to the CPUMA-level to estimate this quantity. The resulting dataset has both measurement error in the covariates that may bias standard estimation procedures, and a hierarchical structure that may increase the variance of these same approaches. We therefore propose an estimation procedure that uses balancing weights that accounts for these problems. We demonstrate that our bias-reduction approach improves on existing methods when predicting observed outcomes from an unknown data generating mechanism. Applying this method to our problem, we estimate that states that did not expand Medicaid in 2014 would have seen a -2.33 (-3.54, -1.11) percentage point change in their adult uninsurance rates had they done so. This is the first study we are aware of that directly estimates the treatment effect on the controls with respect to Medicaid expansion. From a methodological perspective, we demonstrate the value of our proposed method relative to existing methods. From a policy-analysis perspective, we emphasize the importance of directly estimating the relevant causal quantity of interest. More generally if the goal of Medicaid expansion is to improve access to insurance, state and federal policy-makers should consider policies that make Medicaid enrollment easier if not automatic.

Acknowledgements. We gratefully acknowledge invaluable advice and comments from Zachary Branson, Riccardo Fogliato, Edward Kennedy, Brian Kovak, Akshaya Jha, Lowell Taylor, Jose Zubizaretta. We also thank the two anonymous reviewers and the associate editor for their feedback which has greatly improved the paper.
SUPPLEMENTARY MATERIAL

Analyses conducted using R Version 4.0.2 (R Core Team (2020)), and the optweight (Greifer (2021)) and tidyverse (Wickham et al. (2019)) packages. Programs and supporting materials are available at github.com/mrubinst757/medicaid-expansion. Proofs and additional results are available in the Appendix.

REFERENCES


