

Effect of a Policy Intervention to Stimulate Lead Service Line Replacement: Evidence from Wisconsin

Chanheung Cho*

North Carolina State University

Younghyeon Jeon†

University of Florida

July 24, 2025

Abstract

Lead contamination in drinking water remains a serious public health risk in the United States, yet many households are still served by public water systems with aging lead service lines. Replacement efforts have been slow, limited by financial and logistical barriers. In 2018, the state of Wisconsin implemented a public infrastructure policy providing financial assistance for the voluntary replacement of lead service lines, aiming to improve compliance with the Safe Drinking Water Act (SDWA) and reduce associated health risks. This study estimates the causal effect of the policy using a synthetic difference-in-differences (SDID) approach, comparing SDWA violations in Wisconsin's public water systems to those in a synthetic control. The results indicate that the policy reduced SDWA violations by approximately 965, on average—implying improved regulatory compliance and safer drinking water. An event study analysis confirms the robustness of these findings. The results provide new empirical evidence on the effectiveness of state-level interventions in enhancing public health compliance and demonstrate the value of the SDID method as a robust tool for environmental policy evaluation.

JEL Codes: Q53; Q58; I18; C21

Keywords: Safe Drinking Water Act; Lead service line replacement; Drinking water violations; Synthetic difference-in-differences; Policy evaluation

* Address: Partners Building II, Raleigh, NC 27695, US. E-mail: [ccho5\[at\]ncsu.edu](mailto:ccho5[at]ncsu.edu)

† Address: 104 Gainesville FL 32611, US. E-mail: [yjeon\[at\]ufl.edu](mailto:yjeon[at]ufl.edu)

1 Introduction

Access to safe and clean drinking water is a fundamental determinant of public health, and the Safe Drinking Water Act (SDWA) has served as the cornerstone of public drinking water protection in the United States since its passage in 1974 ([U.S. Congress 2000](#)). While the regulatory framework under the SDWA has led to major improvements in drinking water quality, violations of drinking water standards remain prevalent among public water systems (PWSs), particularly those that are smaller and under-resourced ([Allaire et al. 2018](#), [Keiser et al. 2023](#)). Violations may involve health-based breaches, such as exceeding maximum contaminant levels (MCLs) for dangerous contaminants, or non-health-based failures, such as monitoring, reporting, and public notification violations ([Benneer and Olmstead 2008](#), [EPA-SDWIS 2024](#)). Persistent violations not only endanger public health but also erode trust in regulatory institutions and exacerbate environmental justice concerns, disproportionately affecting lower-income and rural areas ([Konisky and Teodoro 2016](#), [Banzhaf et al. 2019](#)).

Recent cases of lead contamination have underscored the vulnerability of U.S. drinking water systems, revealing how water insecurity can exacerbate existing weaknesses, contribute to systemic risk, and even trigger economic disruption ([Christensen et al. 2023](#), [Scanlon et al. 2023](#)). For example, the Flint water crisis and growing concerns about emerging contaminants have renewed focus on the effectiveness of regulatory oversight of water systems ([Gray et al. 2017](#), [Wang et al. 2022](#), [Christensen et al. 2023](#)). In response, several policies have been introduced to reduce regulatory violations, enhance infrastructure, and protect public health—yet their effectiveness remains empirically underexamined. In this context, evaluating the effectiveness of policy interventions aimed at reducing regulatory violations is essential not only for informing evidence-based policy design but also for identifying pathways to strengthen regulatory compliance and ensure reliable water service—ultimately improving long-term public health outcomes.

This paper examines the causal impact of a 2018 Wisconsin policy that offered public infrastructure support. Specifically, the policy provided financial assistance to property owners to identify and replace lead service lines (LSL)—pipes made of lead that carry public drinking water from the water main to a home or building’s internal plumbing system. Since

LSLs can contribute approximately 50–75% of total lead at the tap ([Camara et al. 2013](#), [Cartier et al. 2011](#), [Sandvig et al. 2009](#)), full replacement is among the most effective strategies for achieving lead standards and reducing exposure. Unlike most states, which maintained the status quo during this period, Wisconsin’s initiative provides a rare quasi-experimental setting to evaluate the effectiveness of the state-level financial assistance for LSL replacement ([Greenstone and Gayer 2009](#)).

Our primary research question is whether the 2018 Wisconsin policy intervention led to a reduction in SDWA violations among its public water systems, relative to comparison states that did not implement such interventions. To address this question, we apply a synthetic difference-in-differences (SDID) method proposed by [Arkhangelsky et al. \(2021\)](#), which combines the strengths of synthetic control (SC) and difference-in-differences (DID) approaches. The SDID estimator offers greater robustness to differential pre-trends and time-varying unobservables than conventional DID models, rendering it well-suited for policy evaluation in environmental and public health contexts where interventions are staggered, localized, and potentially confounded by dynamic factors ([Arkhangelsky et al. 2021](#)).

We construct the synthetic control group from states that did not implement significant new state-level drinking water policies between 2014 and 2023, excluding those that introduced major interventions such as expanded lead testing mandates, Per- and Polyfluoroalkyl Substances (PFAS) regulations, or major enforcement reforms. Our dataset merges annual PWS-level violation data from the Environmental Protection Agency’s (EPA) Safe Drinking Water Information System ([EPA-SDWIS 2024](#)) and EPA/State Drinking Water Dashboard ([EPA 2025a](#)) with state-level covariates including log real Gross Domestic Product (GDP), population growth rates, and the percentage of public water systems receiving one or more site visits during the review period in the selected year. These covariates proxy for economic capacity, demographic pressure, and regulatory oversight, helping to control for confounding influences and isolate the causal effect of Wisconsin’s policy.

By way of preview, our results indicate that Wisconsin experienced a substantial and sustained decline in SDWA violations following the 2018 intervention. While Wisconsin initially reported more violations than the control group, the gap narrowed after the policy was implemented. The estimated average treatment effect on the treated (ATT) suggests

that Wisconsin reduced its average violations by approximately 965 relative to the synthetic control in the post-intervention period. These findings are robust to alternative donor pool specifications, as demonstrated by placebo tests. As an additional robustness check, we limit the donor pool to states with both Senate and House League of Conservation Voters (LCV) scores below Wisconsin’s, using environmental political alignment as a proxy for baseline regulatory orientation. This restriction helps mitigate potential bias arising from the inclusion of environmentally proactive states. The larger and statistically significant treatment effect under this restriction lends credibility to the robustness of our main findings.

This study contributes to the environmental and resource economics literature, particularly in the area of environmental policy evaluation. First, we provide new empirical evidence on state-led efforts in the drinking water sector—a policy domain less studied than air pollution, energy policy, or climate regulation. Second, by applying the SDID method to the context of drinking water compliance, we demonstrate how recent advances in causal inference can be used to evaluate policies implemented under decentralized and partially observable conditions. Third, our findings enrich broader policy dialogues on federalism and cooperative enforcement agencies, where state and federal actors share oversight responsibilities ([Shimshack and Ward 2005](#), [Grant and Grooms 2017](#)). Finally, the results provide practical guidance for policymakers seeking to allocate limited regulatory resources to improve public water system performance.

The remainder of the paper proceeds as follows. Section 2 describes the background of this research. Section 3 outlines the empirical strategy and variable construction. Section 4 explains the data sources and descriptive statistics of the data. Section 5 presents the main results and robustness checks. Section 6 discusses policy implications and concludes.

2 Background

The SDWA is a federal law that regulates the safety of public drinking water supplies, covering over 148,000 PWSs as of 2024 ([EPA 2025b](#)). It authorizes the EPA to set water quality standards and to monitor their implementation by states, local jurisdictions, and water suppliers. Under the SDWA, the Lead and Copper Rule requires PWSs to control lead and

copper levels, mainly through corrosion control and public education. However, although the SDWA places responsibility on PWSs to manage lead levels, it does not authorize them to replace customer-owned portions of LSLs unless certain conditions are met.

A LSL typically consists of two portions: the public portion, owned by the utility, and the private or customer-owned portion, which connects the service line to the household. Most LSLs were installed between the late 1800s and 1940s, and the use of lead pipes in PWSs and household plumbing was banned by the 1986 SDWA amendments ([EPA 2022](#)). Despite lead being one of the most serious drinking water contaminants, identifying LSLs has been challenging due to limited records and long histories of repairs ([EPA 2022](#), [Theising 2019](#)). Therefore, many PWSs still operate with partially replaced LSLs—where only the utility-owned portion has been replaced—leaving the customer-owned segment, owned by the property owner, in place and potentially leaching lead into tap water. This split ownership creates a regulatory blind spot: utilities are responsible only for the public side, while the unregulated private side remains a recurring source of health-based SDWA violations. Moreover, utilities are often restricted from offering financial assistance for private-side infrastructure replacement, leading to inconsistent and incomplete replacements that undermine overall compliance efforts.

Building upon the SDWA framework, Wisconsin enacted 2017 Act 137 in February 2018 to address lead contamination from customer-owned LSLs and reduce the risk of drinking water contamination. ([River Network 2025](#)). Signed on February 21 and effective the next day ([Wisconsin Legislature 2018](#)), the Act establishes legal authority and financial assistance mechanisms to support full LSL replacement as a form of public infrastructure support. It authorizes municipalities and utilities—previously restricted—to assist in replacing customer-owned LSL portions through grants, loans, or both, subject to approval by the Public Service Commission of Wisconsin. The Act also allows loan repayments to be added as special charges to property tax bills, reducing implementation costs and encouraging homeowner participation. Rather than imposing a top-down mandate, it delegates authority to local governments, consistent with Wisconsin’s decentralized approach to infrastructure governance ([Wisconsin Legislature 2018](#)).

Act 137 can be evaluated from both policy design and economic perspectives. In terms of a

policy design, it reflects a coordinated governance model: local jurisdictions are authorized to implement replacement mandates, while water utilities can fund financial assistance programs using customer charges collected within their service areas ([Wisconsin Legislature 2018](#)). These programs require approval from the Public Service Commission of Wisconsin to ensure consistency, equity, and consumer protection. The Act stipulates safeguards for fairness and fiscal sustainability, such as requiring uniform grants or loans within customer classes and capping grants at 50% of customer-side replacement costs. Economically, this policy addresses a classic coordination failure by internalizing the externalities of lead exposure: property owners lacked incentives or resources to replace their lines independently, while utilities faced regulatory and financial barriers to providing assistance ([Switzer and Teodoro 2017](#), [McDonald and Jones 2018](#)). Act 137 mitigated this challenge by establishing a legal and fiscal pathway for cooperation.

Wisconsin has emerged as a national leader in LSL replacement ([Environmental Policy Innovation Center, 2024](#)), yet public health challenges remain. In 2016, the rate of lead poisoning among children in Wisconsin was 5%, nearly matching Flint, Michigan’s 4.9% rate during its water crisis. Between 1996 and 2016, more than 200,000 children in the state were diagnosed with lead poisoning ([Soll and Leckel, 2018](#); [Wisconsin Department of Health Services, 2025](#)). Taken together, Wisconsin’s leadership in replacement efforts and its enactment of Act 137 make the state an ideal case to evaluate the effectiveness of state-level interventions in improving public water system compliance and reducing SDWA violations.

3 Methodology

We estimate the causal effect of Wisconsin’s 2018 LSL replacement policy on PWS compliance outcomes using SDID estimator proposed by [Arkhangelsky et al. \(2021\)](#). This approach generalizes traditional DID and SC methods by flexibly reweighting both units and time periods to relax the parallel trends assumption and improve robustness to latent confounders. Let Y_{it} denote the observed outcome (number of violations) for unit $i \in 1, \dots, N$ at time $t \in 1, \dots, T$. Let $W_{it} \in 0, 1$ denote the treatment indicator, where $W_{it} = 1$ if unit i exposed to

treatment at time t and $W_{it} = 0$ otherwise.¹

Following the [Arkhangelsky et al. \(2021\)](#), SDID estimates the average treatment effect on the treated (ATT) τ by solving the weighted two-way fixed effects regression:

$$\left(\hat{\tau}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \arg \min_{\tau, \mu, \alpha, \beta} \sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau \right)^2 \hat{\omega}_i \hat{\lambda}_t, \quad (1)$$

where $\hat{\omega}_i$ are unit weights and $\hat{\lambda}_t$ are time weights optimized pre-treatment trajectories and period, respectively. The flexibility of the procedure allows for shared temporal aggregate factors given the time-fixed effects β_t and unit fixed effect α_i ([Clarke et al. 2024](#)). Compared to standard DID, which assumes equal weights across all units and time periods, SDID introduces a localization mechanism that downweights units and periods that are poorly comparable to the treated observations ([Bertrand et al. 2004](#); [Goodman-Bacon 2021](#)).

The unit weights $\hat{\omega}_i$ are chosen to minimize the discrepancy between the pre-treatment outcomes of the treated and control units:

$$\left(\hat{\omega}_0, \hat{\omega}\right) = \arg \min_{\omega_0 \in \mathbb{R}, \omega \in \Omega} \sum_{t=1}^{T_{pre}} \left(\omega_0 + \sum_{i=1}^{N_{control}} \omega_i Y_{it} - \frac{1}{N_{treated}} \sum_{i=N_{control}+1}^N Y_{it} \right)^2 + \zeta^2 T_{pre} \|\omega\|_2^2, \quad (2)$$

subject to $\omega \in \mathbb{R}_+^N$, with $\sum_{i=1}^{N_{control}} \omega_i = 1$. Here, ω_0 is an intercept term allowing for level shifts and ζ is a regularization parameter to avoid overfitting. Similarly, the time weights $\hat{\lambda}_t$ are computed by solving

$$\left(\hat{\lambda}_0, \hat{\lambda}\right) = \arg \min_{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda} \sum_{i=1}^{N_{control}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_0}^T Y_{it} \right)^2, \quad (3)$$

subject to $\lambda \in \mathbb{R}_+^T$, with $\sum_{t=1}^{T_{pre}} \lambda_t = 1$.² The objective is to select pre-treatment periods that best predict post-treatment outcomes for the control units, thereby ensuring the comparability of the pre-treatment period.

¹Technical interpretation of Synthetic Difference-in-Difference hereafter closely follows [Arkhangelsky et al. \(2021\)](#) and [Clarke et al. 2024](#).

² ζ is defined as $\zeta = (N_{treated} T_{post})^{1/4} \hat{\sigma}$, where $\hat{\sigma}^2 = (N_{control}(T_0 - 1))^{-1} \sum_{i=1}^{N_{control}} \sum_{t=1}^{T_0-1} (\Delta_{it} - \bar{\Delta})^2$ with $\Delta_{it} = Y_{it+1} - Y_{it}$ and $\bar{\Delta}$ the average first difference across control units ([Arkhangelsky et al. 2021](#)). In our study, we adopt the SDID estimation technique developed by [Clarke et al. \(2024\)](#), and it computes $\hat{\lambda}_0$ and $\hat{\lambda}$

Under the SDID, the ATT estimator $\hat{\tau}$ can be interpreted as a weighted DID estimator:

$$\hat{\tau} = \hat{\delta}_{treated} - \sum_{i=1}^{N_{control}} \hat{\omega}_i \hat{\delta}_i \quad (5)$$

where the adjusted outcome differences are

$$\hat{\delta}_{treated} = \frac{1}{N_{treated}} \sum_{i=N_{control}+1}^N \left(\frac{1}{T_{post}} \sum_{t=T_0}^T Y_{it} - \sum_{t=1}^{T_{pre}} \hat{\lambda}_t Y_{it} \right), \quad (6)$$

and similarly for control units. This framework shows that SDID combines ideas from SC (matching pre-trends) and DID (differencing out fixed effects), allowing for flexible deviation from strict parallel trends.

In this paper, we also include time-varying covariates X_{it} to improve the validity of SDID estimates by controlling for the covariates prior to estimating treatment effects. When these covariates X_{it} are included, the estimation needs to adopt a two-step procedure in which the covariates are first regressed out of the outcome variable (Clarke et al. 2024).

Therefore, we regress Y_{it} on X_{it} using the following model:

$$Y_{it} = X'_{it} \gamma + \epsilon_{it}, \quad (7)$$

where γ denotes the vector of coefficients associated with the covariates (Clarke et al. 2024). We then compute residualized outcomes:

$$\hat{Y}_{it} = Y_{it} - X'_{it} \hat{\gamma}. \quad (8)$$

The SDID estimation procedure is then applied to the residualized outcomes \hat{Y}_{it} instead of the raw outcomes Y_{it} . This procedure effectively controls variation in outcomes attributable to the covariates, allowing the SDID estimator to focus on variation associated with the

by minimizing.

$$(\hat{\lambda}_0, \hat{\lambda}) = \arg \min_{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda} \sum_{i=1}^{N_{control}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_0}^T Y_{it} \right)^2 + \zeta^2 N_{control} \|\lambda\|^2, \quad (4)$$

and assumes the very small regularization term with $\zeta = 1 \times 10^{-6} \hat{\sigma}$ to ensure the uniqueness of time weight (Clarke et al. 2024).

treatment while maintaining the method’s robustness properties. [Arkhangelsky et al. \(2021\)](#) note that this residualization step is conceptually distinct from the covariate matching typically employed in synthetic control methods, and it aligns more closely with the standard regression adjustment procedure. Further details and estimator comparisons are provided in the Appendix.

To evaluate the credibility of the estimated treatment effect and ensure that the observed results are not due to chance, we conduct a placebo-based falsification test following [Arkhangelsky et al. \(2021\)](#) and [Clarke et al. \(2024\)](#). Specifically, we iteratively reassign the treatment to each control unit (i.e., placebo units) and re-estimate the ATT under the SDID procedure for each placebo assignment. This generates a distribution of placebo estimates $\{\hat{\tau}_p^{(j)}\}_{j=1}^J$ under the null hypothesis of no treatment effect.

We then compute the placebo-based standard error and p-value as follows:

$$\hat{V}_{\text{placebo}}(\hat{\tau}) = \frac{1}{J-1} \sum_{j=1}^J (\hat{\tau}_p^{(j)} - \bar{\tau}_p)^2, \quad \text{where} \quad \bar{\tau}_p = \frac{1}{J} \sum_{j=1}^J \hat{\tau}_p^{(j)}, \quad (9)$$

$$\hat{p} = \frac{1}{J} \sum_{j=1}^J \mathbb{I}(|\hat{\tau}_p^{(j)}| \geq |\hat{\tau}|), \quad (10)$$

where $\mathbb{I}(\cdot)$ is the indicator function. The estimated p-value \hat{p} reflects the proportion of placebo estimates that are as extreme as the observed treatment effect. A small \hat{p} indicates that the observed $\hat{\tau}$ is unlikely to have arisen from the distribution of placebo effects, thereby providing strong evidence of a true causal impact.

4 Results

4.1 Data

Our analysis uses data from the Safe Drinking Water Information System ([EPA-SDWIS 2024](#)) and the EPA/State Drinking Water Dashboard ([EPA 2025a](#)), covering the period from 2014 to 2023. These sources serve as the national databases of record for monitoring compliance with the SDWA and provide detailed information on PWS activities, including inspections,

violations, and enforcement actions. Data are updated quarterly with a three-month lag, meaning that information for a given calendar year is finalized and incorporated into the database by April of the following year ([EPA 2025a](#)).

The outcome variable is the annual number of SDWA violations at the state level. Violations are categorized according to EPA standards into health-based violations, acute health-based violations, monitoring and reporting violations, and public notification violations ([EPA 2025a](#)). Health-based violations include breaches of MCLs, maximum residual disinfectant levels (MRDLs), or treatment technique (TT) requirements. Acute health-based violations are a subset of health-based violations that have the potential to cause immediate illness. Monitoring and reporting violations occur when systems fail to regularly monitor drinking water quality or fail to submit monitoring results as required. Public notification violations refer to failures to appropriately notify the public about risks to drinking water safety ([EPA 2025a](#)).

The control group consists of U.S. states that did not implement major new drinking water regulations or interventions during the study period ([Cho 2025](#)). States that undertook significant policy changes, such as the adoption of lead testing programs or PFAS regulations between 2014 and 2023, were excluded to ensure comparability. The complete list of excluded states is provided in Appendix Table C1.

To adjust for confounding factors that could simultaneously influence both violation outcomes and policy adoption, the analysis incorporates several time-varying covariates. These are the log of real GDP, which captures state-level economic conditions; the population growth rate, which reflects demographic pressures that could affect water system demand and operational strain; and the percentage of public water systems that received a site visit during the year, which serves as a proxy for regulatory oversight intensity. These covariates were selected based on their theoretical and empirical relevance to water system compliance dynamics.³

Table 1 Panel A presents descriptive statistics for the pooled sample, Wisconsin as the

³Real GDP data are from the U.S. Bureau of Economic Analysis ([U.S. Bureau of Economic Analysis 2024](#)); population growth data are from the Federal Reserve Bank of St. Louis ([Federal Reserve Bank of St. Louis 2024](#)); and site visit data are obtained from the U.S. Environmental Protection Agency’s State Drinking Water Dashboard [EPA 2025a](#).

treated unit, and the control group of states. Wisconsin exhibits lower economic output compared to the control group, with an average log real GDP of 10.90 relative to 12.05. Its population growth is also slower, at 0.331% compared to 0.710% in the control group. In contrast, Wisconsin shows a much higher rate of site visits, with 67.11% of its public water systems receiving at least one site visit during the study period, compared to 39.20% in the control states.

Regarding violation outcomes, Wisconsin initially reported more violations than the control group. In 2014, the number of violations in Wisconsin was 5,305, while the control group’s average was 3,044.48. By 2018, the year of policy intervention, Wisconsin’s violations declined to 4,216, whereas the control group averaged 3,051.32 violations. By 2022, Wisconsin further reduced its violations to 2,753, while the control group’s violations increased to 4,069.56. These descriptive patterns suggest a relative improvement in Wisconsin’s compliance over the study period relative to other states, supporting the need for a formal causal analysis to quantify the policy’s effect.

4.2 Unit Weight Analysis

Figure 1 presents the estimated unit-specific weights assigned to each control state under both the SC and SDID estimators. These weights reflect how much each untreated state contributes to constructing the counterfactual trend for Wisconsin’s compliance outcomes in the absence of the policy intervention. Consistent with the properties of the SC method, the SC weights are highly sparse, with only a few control states receiving non-zero weight. In particular, Texas and Utah emerge as the dominant contributors, with Texas receiving a weight of close to 0.5. This sparsity is a defining feature of SC, which seeks to match the pre-treatment outcome path of the treated unit as closely as possible, typically relying on a small number of donor units that best replicate the treated unit’s pre-policy trajectory.

In contrast, the SDID estimator yields a markedly different weighting pattern. Under SDID, the weights are distributed more evenly across a broader set of control units, with no single state receiving a disproportionate share. All individual weights under SDID remain below 0.1, and most weights are positive but small. This more diffuse weighting structure is the result of two key modifications introduced by SDID relative to traditional SC. First,

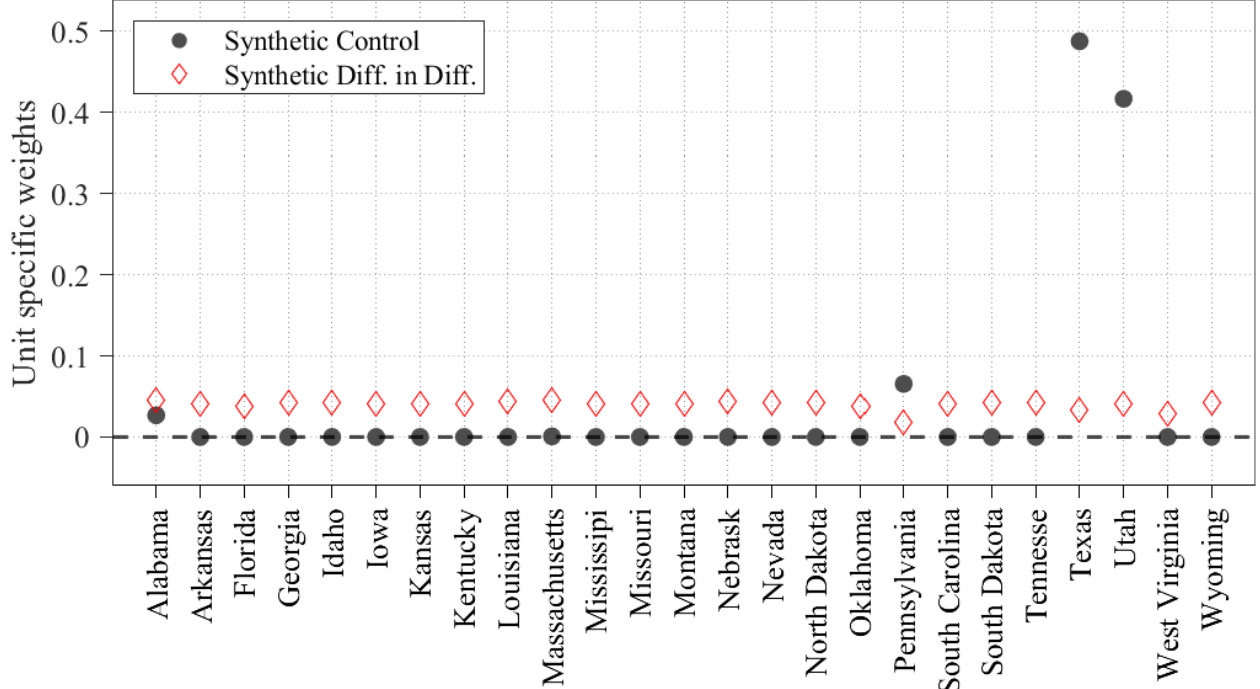


Figure 1: Estimated Unit Specific Weights. The weights are computed using the SDID and SC estimator with covariate adjustment, where outcome variables were first residualized on observed covariates prior to weight estimation, following the two-step procedure described in [Clarke et al. \(2024\)](#).

SDID incorporates regularization when solving for unit weights, discouraging sparsity and preventing the over-reliance on a few units ([Arkhangelsky et al. 2021](#)). Second, by integrating a two-way fixed effects structure (including unit fixed effects) into the outcome model, SDID partially absorbs permanent differences across states, reducing the burden on the weighting scheme to exactly replicate pre-treatment outcome levels ([Arkhangelsky et al. 2021](#)).

The differences in weighting behavior between SC and SDID are especially important in the context of this study, where we now observe the extended pre-treatment period from 2014 to 2017. While traditional SC relies on matching the pre-treatment levels of the outcome variable, it may still overfit short-term fluctuations, especially when a small number of donor units receive disproportionately large weights. In contrast, SDID mitigates this risk by placing greater emphasis on matching pre-treatment trends and not just levels. By incorporating both unit and time weights, SDID ensures that donor units with more stable and parallel trajectories to Wisconsin’s pre-policy trend are given priority, while regularization avoids overreliance on any single state.

4.3 Synthetic Difference-in-Differences Estimation and Validation

Table 1 Panel B presents the estimated ATT for Wisconsin using three methods—SDID, SC, and traditional DID—under both the covariate-adjusted and unadjusted specifications. The estimates provide several important insights into the performance and robustness of each approach.

First, the SDID estimates consistently show a significant reduction in violations, with an ATT of -1200.107 without the covariates and -965.063 with the covariates.⁴ This robustness across each specification illustrates a key strength of the SDID estimator: by combining unit and time weighting with two-way fixed effects, SDID reduces bias from unobserved confounders and achieves greater stability in settings with differential pre-treatment trends (Arkhangelsky et al. 2021; Clarke et al. 2024). Moreover, the covariates are incorporated through a two-step residualization procedure (Equation 8), which allows the treatment effect estimation to focus on variation not explained by the observed covariates.

In contrast, the SC estimates vary widely depending on whether the covariates are included.⁵ Without the covariates, the SC estimator yields an ATT of -733.17, but this flips to a large and imprecise positive value (333.405) with the covariate adjustment, accompanied by a standard error nearly four times that of the SDID estimate. This instability reflects known limitations of SC when few donor units dominate the synthetic match, especially under noisy or idiosyncratic pre-treatment trajectories (Abadie et al. 2010, Arkhangelsky et al. 2021). Because SC solves a constrained optimization problem that selects a sparse set of weights to closely match outcome levels, it is sensitive to overfitting and poorly suited for controlling pre-trend divergence unless very close matches exist.

DID estimates remain negative and statistically significant across both specifications (ATTs of -1513.85 and -958.25), but they come with higher standard errors than SDID. DID’s reliability is contingent on the validity of the parallel trends assumption, which is questionable here based on visual evidence from Appendix Figure C1. Violations of this assumption can bias estimates when treated and control units exhibit divergent trends prior to intervention

⁴The covariate coefficients are reported in Appendix Table A1.

⁵SC does not include unit fixed effects because it aims to replicate the treated unit’s level and trend using a weighted combination of control units. Including unit fixed effects would absorb level differences, undermining this matching objective (Abadie et al. 2010).

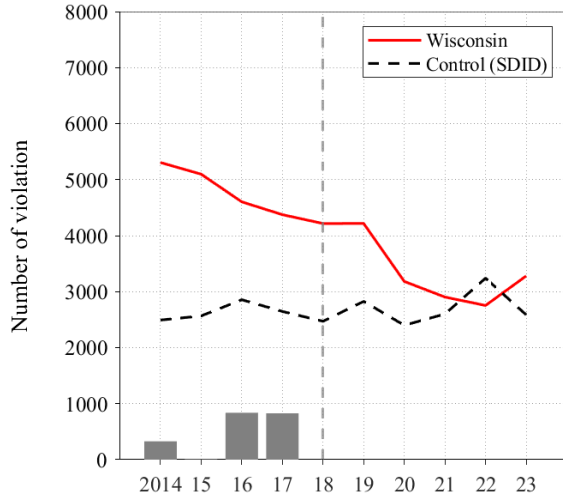
	Pooled sample	Wisconsin	Control group
log Real GDP	12.005	10.900	12.050
Population growth (%)	0.696	0.331	0.710
Site Visit (%)	40.270	67.110	39.204
Number of Violations			
2014 ($T_0 - 4$)	3131.423	5305	3044.480
2018 (T_0)	3096.115	4216	3051.320
2022 ($T_0 + 4$)	4018.923	2753	4069.560

Panel A: Covariate and outcome means

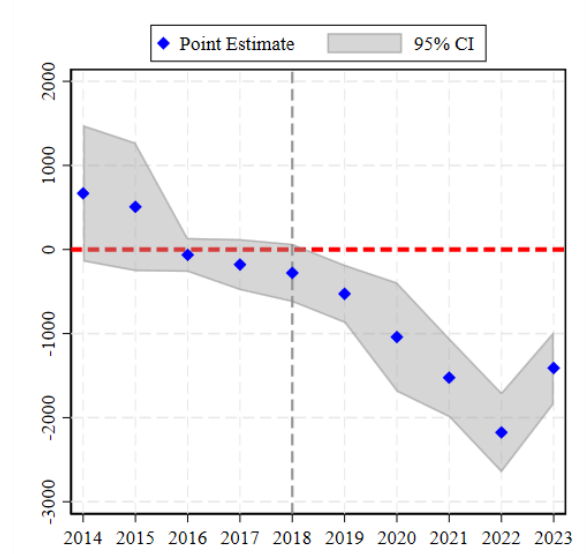
	SDID		SC		DID	
ATT	-1200.107**	-965.063*	-733.17	333.405	-1513.85***	-958.246*
Standard error	(537.793)	(501.927)	(671.171)	(1276.643)	(548.544)	(535.209)
Covariates		✓		✓		✓
Time FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓			✓	✓

Panel B: Estimates for average treatment effect on the treated (ATT) on Wisconsin

Table 1: Descriptive Statistics and Estimated Treatment Effects.. We employ the placebo-based standard error estimator. Placebo treatments in estimation is to control units and compute the distribution of placebo estimates $\hat{\tau}_p$ to approximate the sampling variability of the estimator. The variance estimate is given by $\hat{V}_{placebo}(\hat{\tau}) = \text{Var}(\hat{\tau}_p)$, and a $(1 - \alpha)$ level confidence interval is constructed as $\hat{\tau} \pm z_{\alpha/2} \sqrt{\hat{V}_{placebo}(\hat{\tau})}$, where $z_{\alpha/2}$ denotes the standard normal critical value (Arkhangelsky et al. 2021, Clarke et al. 2024). Methodological details of SC and DID are in the Appendix. The standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.



Panel A: Outcome trends and time-specific weights



Panel B: Event Study

Figure 2: Outcome Trends and Time-Specific Weights and Event Study The dashed red line represents the observed compliance trend in Wisconsin between 2014 and 2023. The solid black line traces the SDID estimate. The gray bars at the bottom indicate the time weights used in SDID, indicating the relative importance assigned to each pre-treatment year in constructing the counterfactual trajectory.

(Bertrand et al. 2004; Goodman-Bacon 2021).

Interestingly, the covariate-adjusted SDID and DID estimates are relatively close in magnitude (-965 vs. -958), but the underlying assumptions differ substantially. SDID does not assume parallel trends; instead, it optimally reweights both control units and time periods to match the pre-treatment trajectory of the treated unit. This design allows SDID to construct a more credible counterfactual in the presence of latent confounders and imperfect trend alignment (Arkhangelsky et al. 2021).

Figure 2 Panel A displays trends in SDWA violations for Wisconsin and the covariate-adjusted synthetic control group from 2014 to 2023. The vertical dashed line indicates the 2018 policy intervention year. The red dashed line shows the observed violations in Wisconsin, the black solid line traces the SC estimate, and the black dashed line shows the SDID counterfactual. The gray bars at the bottom depict the time weights λ_t used in the SDID estimator, illustrating the relative importance of each pre-treatment year in constructing the counterfactual.

These figures complement the estimates in Table 1 Panel B, where the SDID method

consistently yields a statistically significant and negative treatment effect across specifications. The SDID estimator’s relative robustness and precision stem from its ability to combine unit and time weighting with two-way fixed effects ([Arkhangelsky et al. 2021](#); [Clarke et al. 2024](#)). Unlike DID, which assumes parallel trends, and SC, which emphasizes level matching, SDID explicitly focuses on matching pre-treatment trends. This feature is critical in this application. Time weights play an important role in improving SDID’s credibility. As shown in Figure 2 Panel A, 2016 and 2017 receive the greatest weight, suggesting that these years were the most informative for approximating post-treatment outcomes. Earlier years like 2014–2015 received smaller but non-zero weights. This selective emphasis allows the estimator to focus on periods with the highest predictive power while down-weighting noisy or structurally atypical years ([Arkhangelsky et al. 2021](#)).

Notably, the SDID control trajectory does not exactly match Wisconsin’s outcome levels before treatment. This is by design: rather than overfitting to level fluctuations, SDID prioritizes capturing trend direction and slope—features more predictive of post-intervention outcomes. As such, the gap between Wisconsin and its SDID control even before 2018 does not undermine the estimator’s credibility. Instead, it reflects SDID’s emphasis on trend alignment over exact matching, and this design ensures that the post-treatment counterfactual is valid even when treated and control units differ in levels.

Overall, Figure 2 Panel A provides compelling visual evidence that violations in Wisconsin declined substantially following the 2018 policy intervention, and that SDID accurately captures this effect. Unlike SC, which can be unstable, or DID, which relies on strong identifying assumptions, SDID offers a robust and flexible framework for causal inference under imperfect pre-treatment comparability.

To evaluate the credibility of the estimated treatment effect in Wisconsin, we examine the dynamic effects of Wisconsin’s 2018 policy intervention, we conduct an event study analysis following the method proposed by [Clarke et al. \(2024\)](#). Figure 2 Panel B presents the estimated treatment effects for each year from 2014 to 2023, using the SDID framework. Each point estimate reflects the annual difference in SDWA violations between Wisconsin and its synthetic control, controlling for the observed covariates. The blue diamonds indicate the point estimates, and the shaded band represents the 95% confidence interval. The vertical

dashed line at 2018 marks the policy intervention year, while the red dashed horizontal line at zero provides a reference for evaluating the null of no effect. The estimates prior to 2018 are close to zero, supporting the credibility of the identifying assumption that the treated and control units followed parallel trends in the absence of the policy change. After 2018, the estimates become increasingly negative, indicating a substantial and statistically significant reduction in SDWA violations relative to the synthetic control.

This event study is constructed using a two-step procedure. First, the outcomes are residualized by regressing the outcome variable on the covariates to obtain by Equation (8). Second, the year-specific treatment effects are estimated by comparing the residualized outcomes of the treated unit to a weighted average of the control units. The treatment effect at time t is given by:

$$\hat{\tau}_t = \frac{1}{N_{\text{treated}}} \sum_{i=1}^{N_{\text{treated}}} \left(\hat{Y}_{it} - \sum_{t=1}^{T_0-1} \hat{\lambda}_t \hat{Y}_{it} \right) - \sum_{i=1}^{N_{\text{control}}} \hat{\omega}_i \left(\hat{Y}_{it} - \sum_{t=1}^{T_0-1} \hat{\lambda}_t \hat{Y}_{it} \right), \quad (11)$$

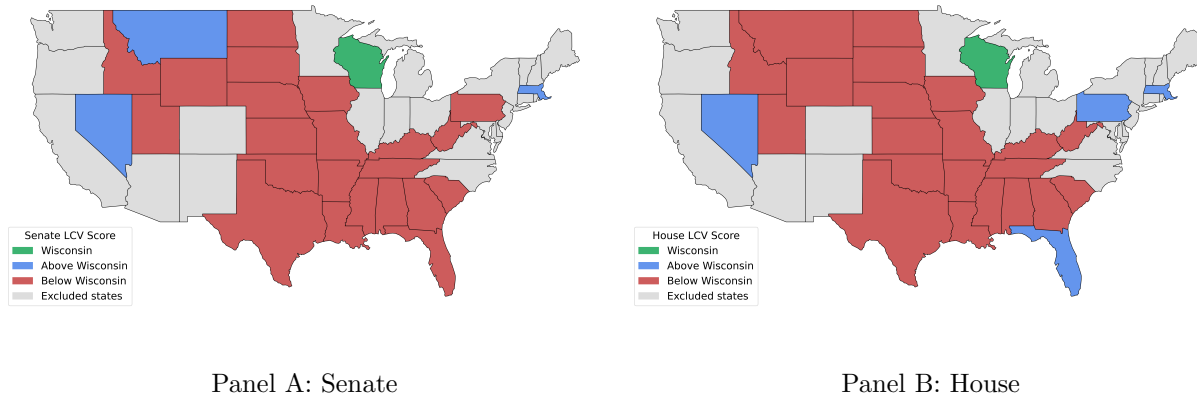
where $\hat{\omega}_i$ and $\hat{\lambda}_t$ are unit and time weights estimated from the pre-treatment period, T_0 is the intervention year, and N_{treated} is the number of treated units (Clarke et al. 2024). In this setting, Wisconsin is the only treated unit, and the time weights are optimized to reweight the pre-treatment path of the treated unit.⁶

This dynamic pattern of treatment effects corroborates our main results and provides further evidence that the 2018 LSL replacement policy had a persistent and growing impact on regulatory compliance in Wisconsin. Further consideration of robustness, we conduct the predictive error-based placebo analysis in the Appendix.

4.4 Policy Evaluation Against Environmentally Unaligned States

Political orientation can influence not only the adoption of environmental regulations but also the strength and consistency with which they are implemented. States with stronger environmental voting records may possess greater institutional capacity, administrative will, or citizen-driven oversight, all of which can enhance eco-consciousness (Konisky 2007). To examine whether the effectiveness of Wisconsin’s policy intervention depends on this political

⁶Confidence intervals are computed using a cluster bootstrap.

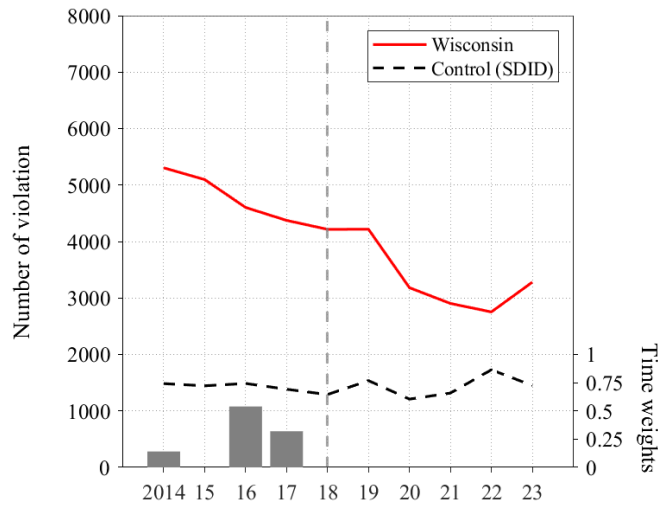


Panel B: House

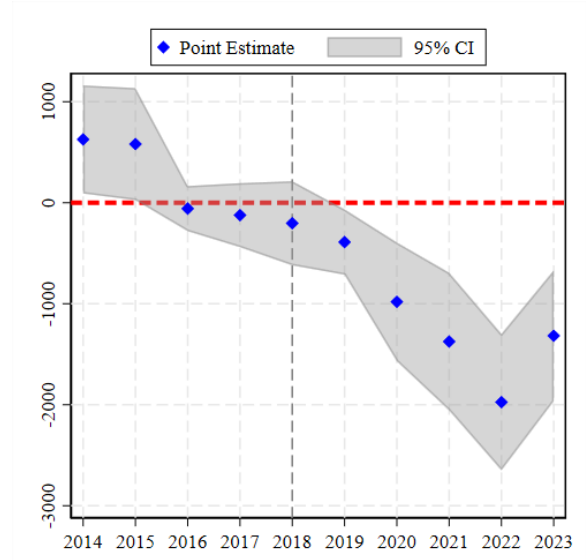
alignment, we use the LCV score—a widely used measure that quantifies how members of Congress vote on environmental legislation—as a proxy for each state’s environmental orientation.⁷ Higher LCV scores indicate greater political support for environmental protection, while lower scores suggest weaker alignment with environmental goals (League of Conservation Voters 2025).

To isolate the role of political context more precisely, we restrict the donor pool to a conservative subset of states with both Senate and House LCV scores below those of Wisconsin. These environmentally unaligned states are less likely to engage in robust

⁸The actual LCV scores are reported in the Online Appendix Table C3.



Panel A: Outcome trends and time-specific weights



Panel B: Event Study

Figure 4: Outcome trends and event study with environmentally unaligned states

environmental regulations or possess the administrative capacity to support proactive and ambitious environmental action. Comparing Wisconsin to this group allows us to examine whether the policy impact is especially pronounced in less supportive political environments, where interventions may face more institutional frictions but yield greater marginal returns.

While we already omitted states that enacted major drinking water reforms during the study period, environmentally proactive states may still have undertaken unobserved or informal efforts to improve compliance. Including such states could bias the estimated treatment effects downward, as observed improvements may reflect underlying political will or institutional quality rather than the absence of formal regulation. By limiting the donor pool to politically unaligned states, we construct a more conservative and credible counterfactual to assess the causal impact of Act 137.

Using this environmentally unaligned comparison group, the estimated ATT is -1,035.33 in the covariate-adjusted model, with a statistically significant at the 5% level.⁹ Compared to the baseline SDID estimate in Table 1 (-965.21), the adjusted values indicate that Wisconsin's intervention appears even more effective when benchmarked against politically unaligned

⁹Estimation results and estimated unit specific weights are depicted in Appendix Table C4 and Figure C3, respectively.

states. These findings strengthen the interpretation that Act 137 led to the actual improvement in compliance behavior, particularly in institutional settings where such outcomes are typically less likely.

Figure 4 provides additional support for these conclusions. Panel A shows a clear post-intervention decline in Wisconsin’s trajectory relative to its synthetic control group, which consists solely of environmentally unaligned states. Panel B presents the dynamic event study estimates, which remain close to zero during the pre-treatment period, indicating a good pre-treatment fit and suggesting that the post-treatment effects are unlikely to be driven by pre-existing trends. Following the 2018 intervention, the treatment effects become increasingly negative, with their magnitude growing over time. This temporal pattern suggests a sustained and compounding policy impact.

Taken together, these findings indicate that the effectiveness of Wisconsin’s policy intervention was not dependent on political alignment. Rather, the institutional and financial mechanisms introduced by Act 137—such as enabling full service line replacement and offering local financing tools—proved effective in overcoming structural and behavioral barriers to compliance. The policy delivered significant and measurable improvements, even when compared to states with weaker environmental orientation. This offers a broader insight: well-designed, locally implemented environmental policy can achieve substantial compliance gains, even in politically challenging environments.

5 Discussion and Conclusion

This study provides evidence that targeted policy interventions can lead to significant improvements in the compliance outcomes among public utilities, even in the presence of imperfect pre-treatment trends. By applying the SDID estimator, we address concerns about latent confounding and deviations from parallel trends—common challenges in causal inference with observational settings. These findings contribute to both the methodological literature on causal inference and ongoing efforts to design and evaluate effective environmental and public health regulations.

Methodologically, this study provides empirical support for SDID as a credible alternative

to traditional DID and SC approaches, particularly in settings where pre-treatment comparability between treated and control units is imperfect. As emphasized by [Arkhangelsky et al. \(2021\)](#) and [Clarke et al. \(2024\)](#), SDID enhances robustness by reweighting both units and time periods while incorporating fixed effects, thereby effectively controlling for both observable and latent sources of bias. The empirical application presented here demonstrates that these theoretical advantages translate into practical gains: the resulting estimates are more precise than those generated by DID or SC alone. This supports the broader view that there is a need for causal inference frameworks to accommodate flexible counterfactual construction, rather than relying on rigid assumptions such as strict parallel trends or convex hull conditions.

The findings also contribute to the literature on regulatory compliance and policy evaluation. While previous studies have documented the challenges in enforcing environmental and public health regulations, particularly in decentralized settings, this study shows that carefully designed interventions could lead to sustained improvements in compliance. Notably, the substantial reduction in violations observed in Wisconsin suggests that proactive policy interventions—such as those providing public infrastructure support and financial assistance—can yield immediate and meaningful improvements in regulatory outcomes. This reinforces arguments in the policy literature advocating for proactive, rather than purely punitive, regulatory strategies ([Coglianese and Kagan 2007](#)).

Moreover, these results have important implications for the design and evaluation of future policy initiatives. First, the success of Wisconsin’s program highlights the importance of targeted, state-level policies tailored to local community conditions, rather than one-size-fits-all federal mandates. Second, the methodology employed in this study demonstrates the advantages of adopting more flexible econometric techniques in policy evaluation, particularly when dealing with complex real-world data that may violate traditional modeling assumptions. It is critical for both policymakers and researchers to recognize that the parallel-trends assumption is often implausible and that methodological innovations such as SDID can substantially improve the credibility of causal impact evaluations.

Several limitations warrant discussion. While SDID substantially relaxes the assumptions underlying DID and SC, it still relies on the availability of sufficient pre-treatment periods

and appropriate donor units to construct reliable weights. If unobserved shocks occurring contemporaneously with the intervention differentially affect the treated unit, even SDID estimates may be biased. Although the placebo tests in this study help mitigate concerns about such threats, future work could benefit from explicitly modeling potential confounders or exploring extensions ([Athey and Imbens 2017](#)). Furthermore, this analysis focuses on the aggregate outcomes at the state level; disaggregated analyses by PWS size, ownership type, or violation category could provide additional insights into the heterogeneity of treatment effects. Future research may also examine the persistence of the observed compliance improvements over longer horizons, assessing whether the policy intervention impacts diminish, stabilize, or even amplify over time.

Nevertheless, this study highlights the value of innovative causal inference methods such as SDID and demonstrates how well-designed policy interventions can ensure regulatory compliance and thereby address potential public health concerns. As policymakers grapple with increasingly complex environmental and public health challenges and with questions about the effectiveness of the policy initiatives proposed in response, evidence-based and methodologically rigorous evaluations will be essential for guiding effective governance.

References

- ABADIE, A., A. DIAMOND, AND J. HAINMUELLER (2010): “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program,” *Journal of the American Statistical Association*, 105, 493–505.
- ALLAIRE, M., H. WU, AND U. LALL (2018): “National trends in drinking water quality violations,” *Proceedings of the National Academy of Sciences*, 115, 2078–2083.
- ANGRIST, J. D. AND J.-S. PISCHKE (2009): *Mostly harmless econometrics: An empiricist’s companion*, Princeton university press.
- ARKHANGELSKY, D., S. ATHEY, D. A. HIRSHBERG, G. W. IMBENS, AND S. WAGER (2021): “Synthetic difference-in-differences,” *American Economic Review*, 111, 4088–4118.
- ATHEY, S. AND G. IMBENS (2017): “The state of applied econometrics: Causality and policy evaluation,” *Journal of Economic Perspectives*, 31, 3–32.
- BANZHAF, S., L. MA, AND C. TIMMINS (2019): “Environmental justice: The economics of race, place, and pollution,” *Journal of Economic Perspectives*, 33, 185–208.
- BENNEAR, L. S. AND S. M. OLMSTEAD (2008): “The impacts of the “right to know”: Information disclosure and the violation of drinking water standards,” *Journal of Environmental Economics and Management*, 56, 117–130.
- BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): “How much should we trust differences-in-differences estimates?” *The Quarterly journal of economics*, 119, 249–275.
- CAMARA, E., K. R. MONTREUIL, A. K. KNOWLES, AND G. A. GAGNON (2013): “Role of the water main in lead service line replacement: A utility case study,” *Journal-American Water Works Association*, 105, E423–E431.
- CARTIER, C., L. LAROCHE, E. DESHOMMES, S. NOUR, G. RICHARD, M. EDWARDS, AND M. PRÉVOST (2011): “Investigating dissolved lead at the tap using various sampling protocols,” *Journal-American Water Works Association*, 103, 55–67.
- CHO, C. (2025): “Human Right to Water Act and drinking water compliance: A synthetic control approach,” *Economics Letters*, 112361.
- CHRISTENSEN, P., D. A. KEISER, AND G. E. LADE (2023): “Economic effects of environmental crises: Evidence from Flint, Michigan,” *American Economic Journal: Economic Policy*, 15, 196–232.
- CLARKE, D., D. PAILAÑIR, S. ATHEY, AND G. IMBENS (2024): “On synthetic difference-in-differences and related estimation methods in Stata,” *The Stata Journal*, 24, 557–598.
- COGLIANESE, C. AND R. A. KAGAN (2007): *Regulation and regulatory processes*, Oxford University Press.

- ENVIRONMENTAL POLICY INNOVATION CENTER (2024): “Here’s why Milwaukee and Wisconsin are models for lead pipe replacement—and why President Joe Biden is there delivering the Biden-Harris swan song on lead in drinking water,” Accessed: May 21, 2025.
- EPA (2022): “EPA Researchers Share Approaches to Identify Lead Service Lines,” <https://www.epa.gov/sciencematters/epa-researchers-share-approaches-identify-lead-service-lines>, accessed: 2025-05-25.
- (2025a): “Enforcement and compliance history online: EPA/State Drinking Water Dashboard,” <https://echo.epa.gov/trends/comparative-maps-dashboards/drinking-water-dashboard?state=National>, accessed: 2025-02-25.
- (2025b): “Information about Public Water Systems,” <https://www.epa.gov/dwreginfo/information-about-public-water-systems>, accessed: 2025-05-25.
- EPA-SDWIS (2024): “Safe Drinking Water Information System (SDWIS) - Federal Reporting Services,” Accessed: 2025-04-28.
- FEDERAL RESERVE BANK OF ST. LOUIS (2024): “Annual Population Estimates by State,” Accessed April 2025.
- GOODMAN-BACON, A. (2021): “Difference-in-differences with variation in treatment timing,” *Journal of econometrics*, 225, 254–277.
- GRANT, L. E. AND K. K. GROOMS (2017): “Do nonprofits encourage environmental compliance?” *Journal of the Association of Environmental and Resource Economists*, 4, S261–S288.
- GRAY, S., A. SINGER, L. SCHMITT-OLABISI, J. INTRONE, AND J. HENDERSON (2017): “Identifying the causes, consequences, and solutions to the Flint Water Crisis through collaborative modeling,” *Environmental Justice*, 10, 154–161.
- GREENSTONE, M. AND T. GAYER (2009): “Quasi-experimental and experimental approaches to environmental economics,” *Journal of Environmental Economics and Management*, 57, 21–44.
- KEISER, D. A., B. MAZUMDER, D. MOLITOR, AND J. S. SHAPIRO (2023): “Water works: Causes and consequences of safe drinking water in America,” in *National Center for Environmental Economics Seminar, Washington, DC, April*, vol. 13.
- KONISKY, D. M. (2007): “Regulatory competition and environmental enforcement: Is there a race to the bottom?” *American Journal of Political Science*, 51, 853–872.
- KONISKY, D. M. AND M. P. TEODORO (2016): “When governments regulate governments,” *American Journal of Political Science*, 60, 559–574.
- KRANZ, S. (2022): “Synthetic difference-in-differences with time-varying covariates,” .

- LEAGUE OF CONSERVATION VOTERS (2025): “National Environmental Scorecard,” <https://www.lcv.org/congressional-scorecard/>, accessed: 2025-05-26.
- MCDONALD, Y. J. AND N. E. JONES (2018): “Drinking water violations and environmental justice in the United States, 2011–2015,” *American Journal of Public Health*, 108, 1401–1407.
- RIVER NETWORK (2025): “State Policy Hub: Drinking Water,” Accessed: 2025-02-25.
- SANDVIG, A., P. KWAN, G. KIRMEYER, B. MAYNARD, D. MAST, R. R. TRUSSELL, S. TRUSSELL, A. CANTOR, AND A. PRESCOTT (2009): “Contribution of Service Line and Plumbing Fixtures to Lead and Copper Rule Compliance Issues,” Tech. rep., Awwa Research Foundation, Denver, CO.
- SCANLON, B. R., R. C. REEDY, S. FAKHREDDINE, Q. YANG, AND G. PIERCE (2023): “Drinking water quality and social vulnerability linkages at the system level in the United States,” *Environmental Research Letters*, 18, 094039.
- SHIMSHACK, J. P. AND M. B. WARD (2005): “Regulator reputation, enforcement, and environmental compliance,” *Journal of Environmental Economics and Management*, 50, 519–540.
- SOLL, D. AND R. LECKEL (2018): “Guidebook for Wisconsin Municipalities and Citizens on Replacement of Lead Service Lines Using Wisconsin Act 137: A Step-by-Step Guide with Links to Related Tools and Information,” Developed by University of Wisconsin–Eau Claire and University of Wisconsin–Stout.
- SWITZER, D. AND M. P. TEODORO (2017): “The color of drinking water,” *Journal (American Water Works Association)*, 109, 40–45.
- THEISING, A. (2019): “Lead pipes, prescriptive policy and property values,” *Environmental and Resource Economics*, 74, 1355–1382.
- U.S. BUREAU OF ECONOMIC ANALYSIS (2024): “Gross Domestic Product (GDP) by State,” Accessed April 2025.
- U.S. CONGRESS (2000): *The Safe Drinking Water Act as Amended by the Safe Drinking Water Act Amendments of 1996: Public Law 104-182, August 6, 1996*, S. Prt. 106-59, Washington, DC: U.S. Government Printing Office, printed for the use of the Committee on Environment and Public Works, 106th Congress, 2nd Session.
- WANG, R., X. CHEN, AND X. LI (2022): “Something in the pipe: the Flint water crisis and health at birth,” *Journal of Population Economics*, 35, 1723–1749.
- WISCONSIN DEPARTMENT OF HEALTH SERVICES (2025): “Environmental Public Health Tracking: Lead Poisoning Data,” <https://www.dhs.wisconsin.gov/epht/lead.htm>, accessed: 2025-05-28.

WISCONSIN LEGISLATURE (2018): “2017 Wisconsin Act 137,” Enacted on February 21, 2018. Effective on February 22, 2018. Relates to financial assistance for the replacement of lead-containing customer-side water service lines.

Appendix

A Comparison Estimators

For reference, we compare our primary estimates from the synthetic difference-in-differences (SDID) method with those obtained from conventional difference-in-differences (DID) and synthetic control (SC) estimators. Each method relies on different identifying assumptions and weighting structures. For technological and methodological explanations, we closely follow the [Arkhangelsky et al. \(2021\)](#),

The DID estimator is based on a two-way fixed effects model:

$$(\hat{\tau}^{\text{DID}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau)^2 \right\} \quad (\text{A1})$$

where Y_{it} is the observed outcome, W_{it} is an indicator equal to one if unit i is treated at time t , and α_i and β_t are unit and time fixed effects. The parameter τ captures the average treatment effect on the treated (ATT). This specification assumes that, in the absence of treatment, treated and control units would have followed parallel trends in outcomes ([Angrist and Pischke, 2009](#)). However, this assumption is violated in our observational policy evaluations, particularly when treated units exhibit different pre-treatment dynamics in Figure C1.

The SC method constructs a weighted average of control units to approximate the pre-treatment trajectory of the treated unit. Formally, it solves:

$$(\hat{\tau}^{\text{SC}}, \hat{\mu}, \hat{\beta}) = \arg \min_{\tau, \mu, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \beta_t - W_{it}\tau)^2 \hat{\omega}_i^{\text{SC}} \right\}, \quad (\text{A2})$$

Here, α_i and β_t are unit and time fixed effects, and W_{it} is a binary treatment indicator. The SC method constructs a synthetic counterfactual by assigning weights to control units to match the treated unit's pre-treatment outcome path subject to $\omega_j \geq 0$ and $\sum_j \omega_j = 1$. The estimated treatment effect is then computed as the post-treatment difference between the treated unit and its synthetic counterpart. Unlike DID, SC does not rely on the parallel trends assumption ([Abadie et al., 2010](#)), but it assumes that a convex combination of control units can reproduce the treated unit's counterfactual. The method is sensitive to the quality

	SDID	SC	DID
ATT	-965.063	333.405	-958.248
log Real GDP	-47.512	3240.072	-59.079
Population growth (%)	188.778	-1061.727	290.847
Site Visit (%)	-351.430	-1223.387	-395.817
Time FE	✓	✓	✓
State FE	✓		✓

Table A1: Estimated Covariate Coefficients from SDID, SC, and DID Models.

of pre-treatment fit and may perform poorly when the treated unit lies far outside the convex hull of the donor pool.

The SDID estimator integrates the strengths of DID and SC by incorporating both unit and time fixed effects and applying data-driven weights to units and periods. The estimator solves the Equation (1), where $\hat{\omega}_i^{\text{SDID}}$ and $\hat{\lambda}_t^{\text{SDID}}$ are unit and time weights estimated to balance treated and control units on pre-treatment trends and to emphasize informative time periods. As described by [Arkhangelsky et al. \(2021\)](#), SDID improves robustness to violations of the parallel trends assumption and achieves double robustness: the ATT can be consistently estimated if either the regression model or the weighting procedure is correctly specified ([Clarke et al., 2024](#)).

Table A1 reports the estimated coefficients on covariates included in the SDID, SC, and DID models. These covariates—log real GDP, population growth, and site visit rate—are incorporated to adjust for systematic differences between treated and control units. In the SDID model, covariate adjustment is implemented through residualization, whereby the outcome is regressed on covariates, time fixed effects, and unit fixed effects to obtain a residualized outcome:

$$\hat{Y}_{it} = Y_{it} - X'_{it}\hat{\gamma}. \quad (8)$$

Following [Kranz \(2022\)](#), the covariate coefficients $\hat{\gamma}$ are estimated in a two-way fixed

effects regression using only untreated units (i.e., those with $W_{it} = 0$) to avoid contamination from treatment effects:

$$Y_{it} = X'_{it}\gamma + \alpha_i + \beta_t + \epsilon_{it} \quad \text{for units with } W_{it} = 0. \quad (8')$$

The resulting residuals \hat{Y}_{it} are then used in the SDID estimation step, which minimizes the following weighted least squares objective in Equation 1.

For SC, covariates are incorporated as additional matching targets in the optimization of unit weights ω_j , solving:

$$\min_{\omega} \left\| \bar{X}_{\text{treated}} - \sum_j \omega_j \bar{X}_j^{\text{control}} \right\|^2. \quad (A3)$$

In the DID specification, covariates are included directly in the regression equation along with time and unit fixed effects:

$$Y_{it} = \tau \cdot W_{it} + X'_{it}\gamma + \alpha_i + \beta_t + \varepsilon_{it}. \quad (A4)$$

While ATT remains the primary quantity of interest, reporting the covariate coefficients in Table A1 provides insight into the adjustment behavior across estimation methods.

B Placebo Analysis Based on Predictive Error Ratios

To further evaluate the credibility of the estimated treatment effect in Wisconsin, we conduct a placebo-based falsification test using the Ratio of Mean Squared Prediction Error (RMSPE) (Abadie et al. 2010). Figure B1 presents the results of the placebo test using the absolute value of the log Ratio of Mean Squared Prediction Error ($|\log \text{RMSPE}|$) for Wisconsin compared to the control states.¹⁰ Each black dot represents a control state's $|\log \text{RMSPE}|$, while the red dashed line indicates Wisconsin's value. The $|\log \text{RMSPE}|$ metric captures the relative change in prediction error between the pre- and post-treatment periods. Specifically, a $|\log \text{RMSPE}|$ value close to zero indicates that the model's predictive accuracy remained

¹⁰Figure C2 in Appendix shows the raw differences between observed and synthetic control outcomes for Wisconsin and placebo states.

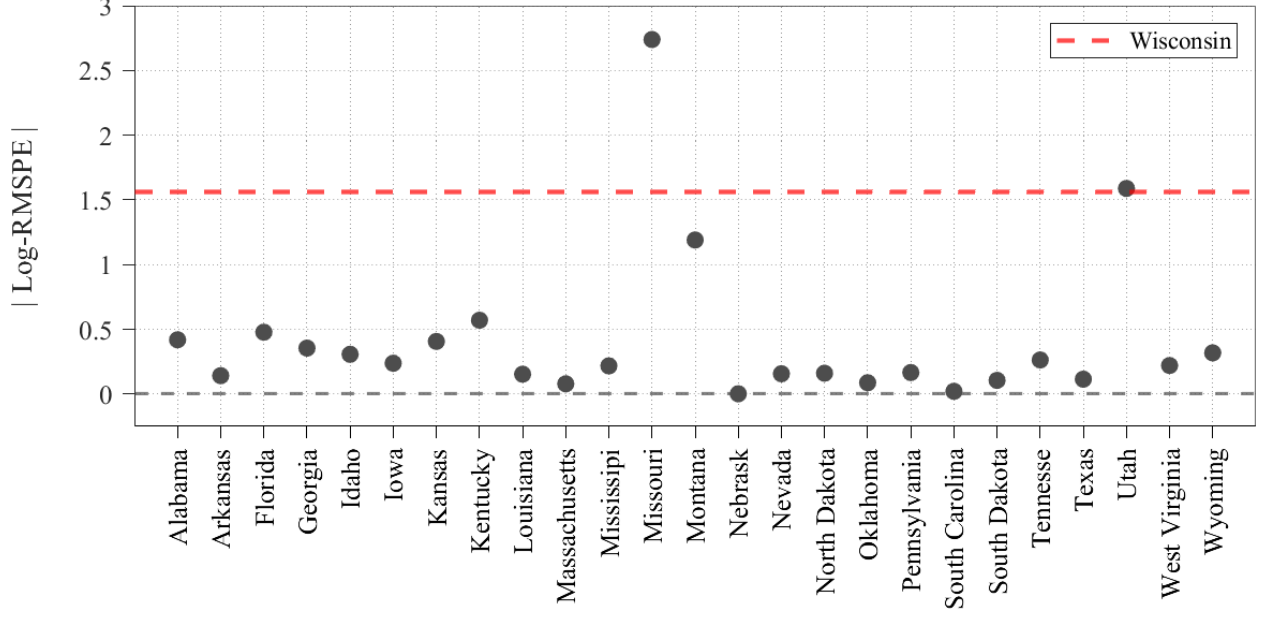


Figure B1: Log-RMSPE ratios across states. Each black dot represents a control state, and the red dashed line denotes Wisconsin. The values are computed using the SDID estimator with the covariate adjustment.

stable before and after the intervention, implying no substantial structural change. In contrast, a large $|\log \text{RMSPE}|$ indicates a substantial increase in prediction error after the intervention, which is consistent with a real underlying shift in outcomes due to the treatment. Thus, higher $|\log \text{RMSPE}|$ values are interpreted as evidence of a treatment effect relative to the counterfactual trajectory.

The measure used in this analysis for each state j is as follows:

$$|\log \text{RMSPE}_j| = \left| \log \left(\frac{\frac{1}{T-(T_0-1)} \sum_{t=T_0}^T (Y_{jt} - \hat{Y}_{jt})^2}{\frac{1}{T_0-1} \sum_{t=1}^{T_0-1} (Y_{jt} - \hat{Y}_{jt})^2} \right) \right| \quad (\text{B1})$$

where Y_{jt} represents the observed outcome for state j in year t , specifically the number of violations. \hat{Y}_{jt} is the counterfactual prediction for state j generated by the SDID estimator. T_0 denotes the treatment year, and T is the last observed year. The numerator captures the mean squared prediction error in the post-treatment period, while the denominator captures the mean squared prediction error in the pre-treatment period.

The placebo test is designed to assess the validity of the SDID estimates by examining whether untreated states exhibit similar post-treatment deviations. If large treatment effects

were common among untreated units, it would cast doubt on the causal interpretation of the Wisconsin estimates. However, the figure clearly shows that Wisconsin's $|\log \text{RMSPE}|$ is substantially higher than that of any control state. The values of most control states are located near zero, indicating minimal post-treatment deviations and reinforcing the stability of their pre- and post-treatment trajectories. Although a few control states, such as Missouri, Montana, and Utah, exhibit increased deviations, none show changes comparable to those observed in Wisconsin.

C Additional Figures and Tables

State	Policy	Date	Description
Alaska	AK H 209	07.28.2016	committee studies rural water and sewer needs
Arizona	HB 2049	04.28.2017	expands grant eligibility for small water systems
	SB 1459	05.12.2016	assist low-income homeowners with well improvements
California	HR2W	09.25.2012	ensuring affordable, accessible, acceptable and safe water
Colorado	HB 1306	06.08.2017	funds lead testing in public schools
	HB 20-1119	06.29.2020	regulates PFAS storage, disposal, and firefighting foam
	SB 20-2018	06.29.2020	establishes PFAS fund for grants, takeback, and assistance
	HB 22-1358	06.07.2022	law mandates lead testing in schools, childcare
Connecticut	HB 5509	06.14.2018	protects vulnerable groups from sewer foreclosures
Delaware	HB 200	07.22.2021	funds clean water projects, prioritizing equity
Illinois	SB 550	01.17.2017	mandates lead testing, inventory, and notification
	SB 2146	08.23.2019	invests in clean water infrastructure and workforce training
	HB 0414	08.06.2021	creates low-income water and sewer assistance program
	HB 3739	01.01.2022	mandates full lead pipe replacement and assistance
Indiana	HB 1138	05.01.2023	preschools and childcare must test for lead
Maine	S.P. 64	06.21.2021	mandates PFAS monitoring, notification, and mitigation
	HP 113	07.15.2021	nation's first comprehensive PFAS product ban enacted
Maryland	SB 96	04.30.2019	prohibits tax sales for water bill liens
Michigan	HB 4342	10.24.2023	child care centers must label water safety
	SB 88	10.24.2023	child care centers must manage lead exposure
Minnesota	HF 1	10.21.2020	funds water infrastructure upgrades and protection
	HF 2310	05.24.2023	funds PFAS mitigation, bans, and regulations
New Hampshire	SB 309	07.10.2018	sets PFAS water standards, adds toxicologist
	HB 1264	07.23.2020	sets PFAS MCLs, funds programs, expands standards
New Jersey	SB 968/A2863	05.11.2021	law mandates lead level notifications quickly
	SB 994	09.13.2022	mandates utility affordability
New Mexico	SB 1	03.13.2023	facilitates regionalization of water utilities
New York	SB S8158	09.06.2016	schools must test for lead, provide aid
	VolA-5-5-1	08.26.2020	sets maximum contaminant levels for contaminants

State	Policy	Date	Description
North Carolina	HB 1087	07.01.2020	funds utilities, reviews, mergers, and projects
Ohio	HB 512	09.09.2016	strengthens Lead and copper testing requirement
	3745-81-84	05.01.2018	revised Lead and Copper Rule
	HB 166	11.01.2019	H2Ohio fund for water quality projects
Oregon	Water Vision	2019	improvements to our infrastructure and ecosystems
Rhode Island	SB 2298	06.24.2022	mandates PFAS testing, standards, and monitoring
	SB 0724	06.22.2023	revises PFAS contamination response
Vermont	Act 21	05.15.2019	regulation of poly-fluoroalkyl substances
	Act 139	07.06.2020	construction grants for public water improvement
Virginia	HJ538	02.24.2021	access to clean, potable, and affordable water
	HB 1257	01.01.2022	sets maximum contaminant levels
Washington	SB 6413	06.07.2018	bans PFAS firefighting foam, mandates disclosure
	SB 5135	07.28.2019	regulates priority toxic chemicals in products

Table C1: States removed from analysis because of policy interventions (2014–2023). HF: House File, HB: House Bill, SB: Senate Bill, PFAS: Perfluoroalkyl and Polyfluoroalkyl Substances. Data Source: [River Network 2025](#), retrieved on April 29, 2025. Table information are from [Cho \(2025\)](#). We also excluded D.C. and Hawaii from the control group due to their structural dissimilarity to Wisconsin. D.C., as a city-state, lacks rural drinking water systems and exhibits administrative characteristics fundamentally distinct from continental states. Hawaii, being a geographically isolated island state, operates under water supply and enforcement systems that differ markedly from those on the mainland. Including these units would violate the synthetic control method’s requirement for comparable untreated units and risk undermining the credibility of our causal estimates.

Treatment Group	Control Group (25)
Wisconsin	Alabama, Arkansas, Florida, Georgia, Idaho, Iowa, Kansas, Kentucky, Louisiana, Massachusetts, Mississippi, Missouri, Montana, Nebraska, Nevada, North Dakota, Oklahoma, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, West Virginia, Wyoming

Table C2: States assigned to treatment and control groups

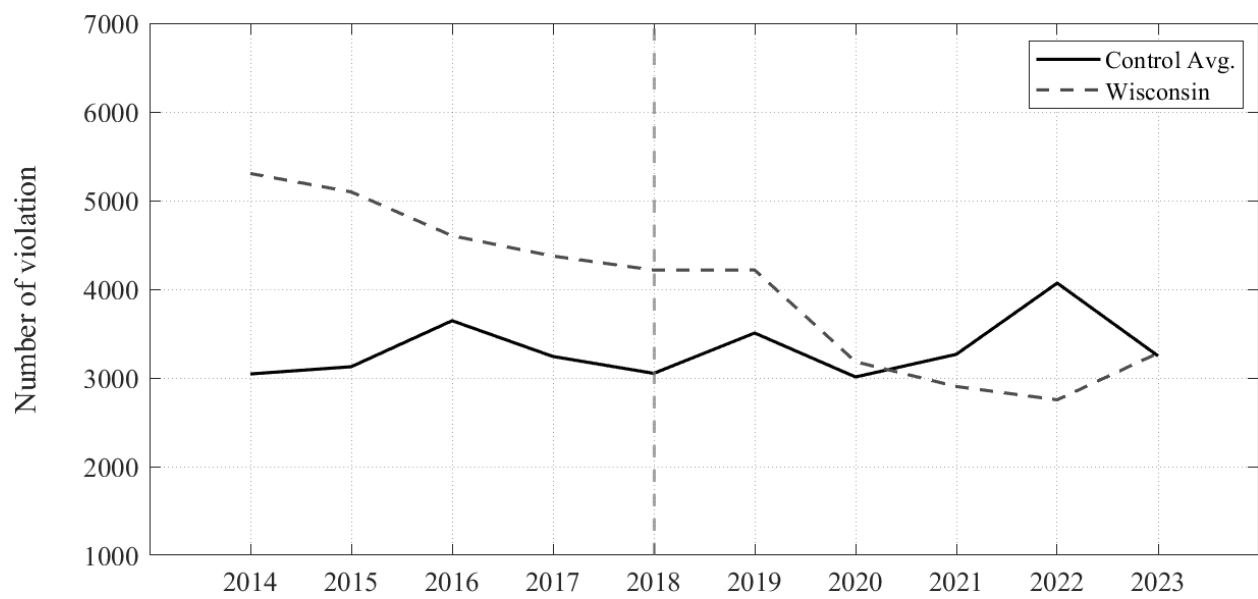


Figure C1: Trends in the Number of Violations for Wisconsin and Control States (2014–2023)

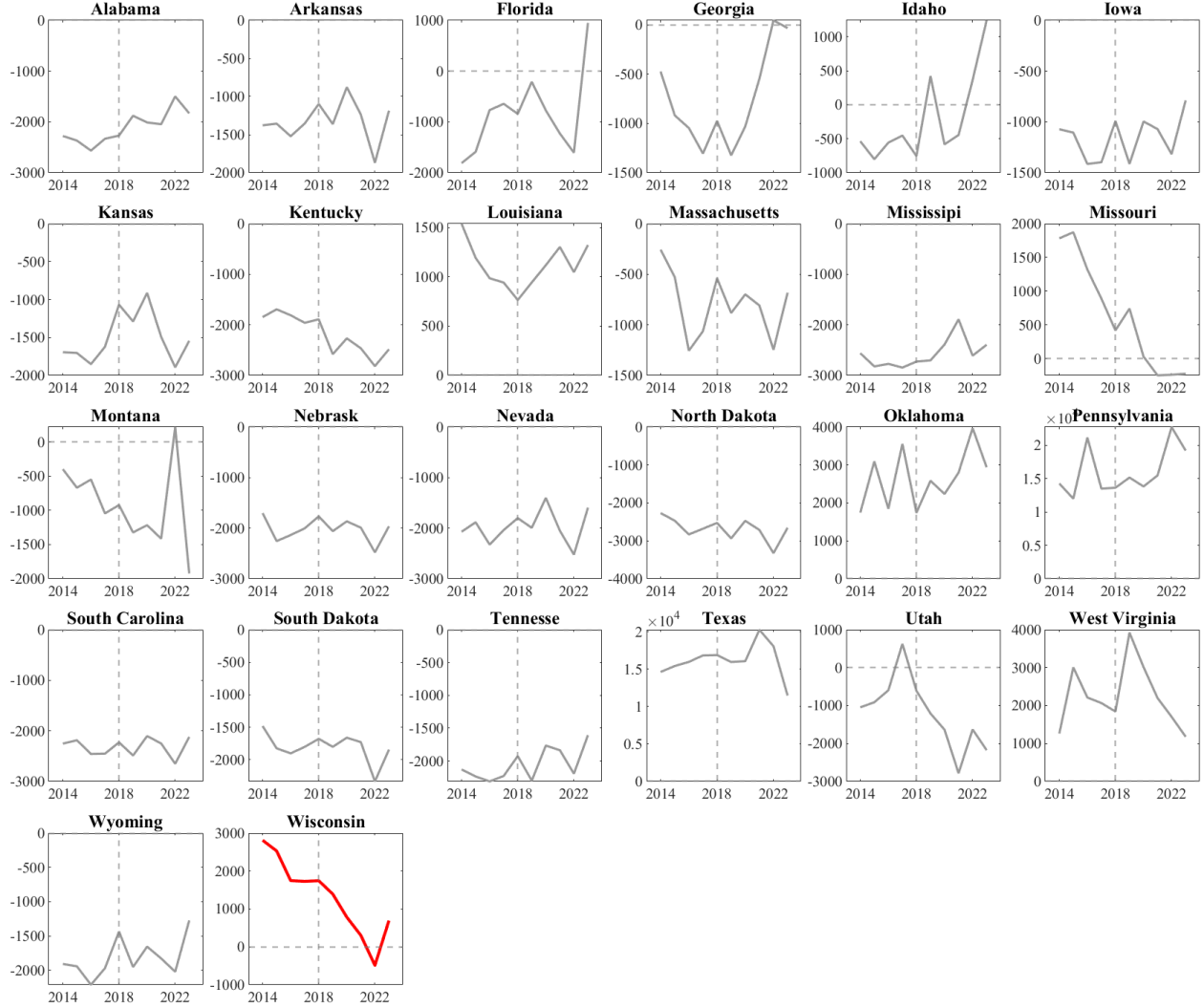


Figure C2: Estimated treatment-control gaps by state. This figure presents the difference between observed outcomes and synthetic control outcomes for Wisconsin and each placebo control state from 2014 to 2023. Each subplot corresponds to a different state, showing the trend of the gap between the state’s actual violations and the synthetic counterfactual, with the vertical dashed line indicating the intervention year (2018). The trajectory for Wisconsin is highlighted in red.

	LCV Score (2018)		
	Senate	House	Above WI (Senate > or House >)
Alabama	43	14	
Arkansas	4	4	
Florida	36	48	✓
Georgia	7	26	
Idaho	7	10	
Iowa	4	27	
Kansas	11	2	
Kentucky	11	19	
Louisiana	7	9	
Massachusetts	100	90	✓
Mississippi	5	22	
Missouri	43	24	
Montana	54	3	✓
Nebraska	11	10	
Nevada	54	74	✓
North Dakota	29	3	
Oklahoma	11	4	
Pennsylvania	46	45	✓
South Carolina	7	20	
South Dakota	11	3	
Tennessee	7	23	
Texas	11	28	
Utah	7	6	
West Virginia	25	7	
Wyoming	7	0	
Wisconsin	50	36	

Table C3: League of conservation voters scores by state in 2018. This table presents the 2018 LCV scores for both Senate and House delegations by state. The LCV score measures the percentage of pro-environment votes cast by each member of Congress based on key environmental legislation identified annually by the League of Conservation Voters. Scores range from 0 to 100, where 100 indicates full alignment with environmental protection priorities.

SDID		
ATT	-1096.43**	-1035.326**
Standard error	(474.694)	(431.965)
Covariates		✓
Time FE	✓	✓
State FE	✓	✓

Table C4: Estimates for average treatment effect on the treated (ATT) with environmentally unaligned states. We employ the placebo-based standard error estimator. The standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

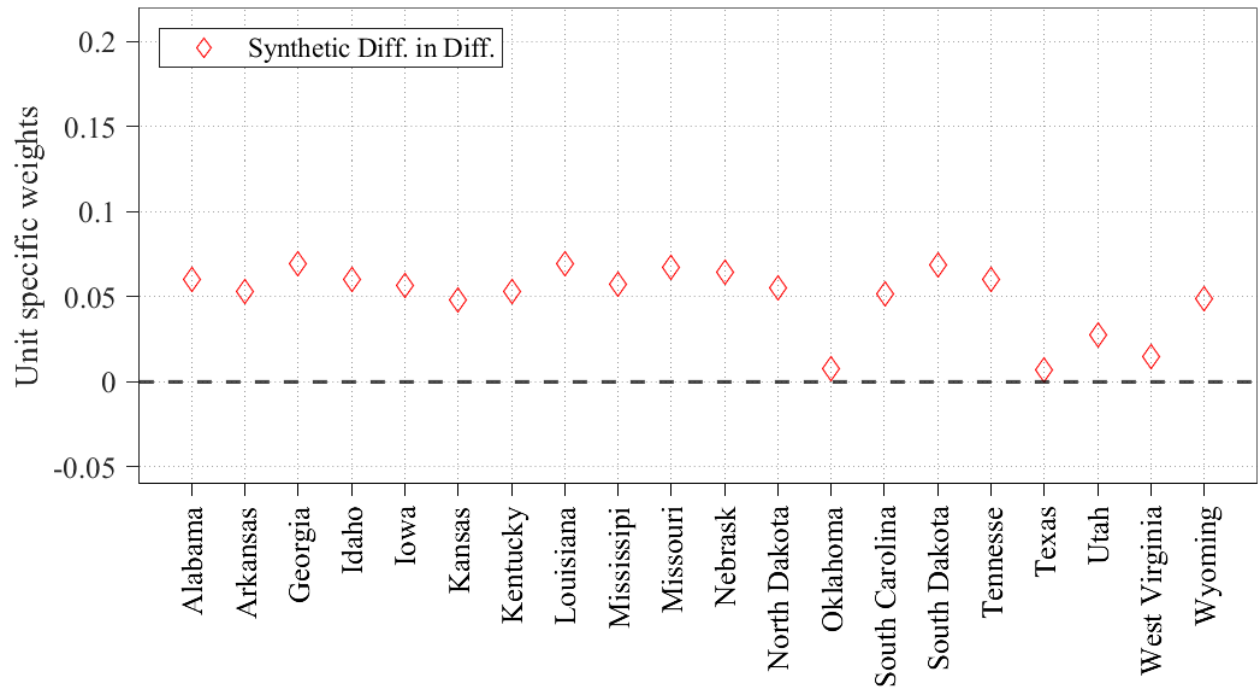


Figure C3: Unit weights from SDID estimation using environmentally unaligned states. This figure displays the unit-specific weights assigned by the SDID estimator when the donor pool is restricted to states with lower Senate or House LCV scores than Wisconsin. The weights are relatively evenly distributed across units, indicating that no single state dominates the construction of the synthetic control. This diffuse weighting improves robustness and helps mitigate the risk of over-reliance on a small subset of donor states.