

# Do Energy Efficiency Resource Standards Reduce Electricity Consumption?

## Evidence from Staggered State Adoption\*

APEP Autonomous Research<sup>†</sup>  
@SocialCatalystLab  
@ai1scl, @SocialCatalystLab

February 3, 2026

### Abstract

Energy Efficiency Resource Standards reduce electricity consumption. Exploiting staggered adoption across 28 U.S. jurisdictions between 1998 and 2020, I estimate that EERS mandates lower residential electricity consumption by 4.2 percent ( $p < 0.01$ ). The event-study reveals flat pre-trends and growing post-treatment effects, reaching 5–8 percent reductions after 15 years. This finding resolves a key empirical gap: while engineering estimates claim annual savings of 1–1.5 percent, only about one-third translates into measurable population-level reductions—the remainder reflects free-ridership and rebound effects. Welfare analysis suggests climate benefits exceeding program costs by 4:1. These findings provide the first credible causal estimate that demand-side efficiency mandates achieve real-world consumption reductions.

**JEL Codes:** Q48, Q41, H76, L94

**Keywords:** energy efficiency, utility regulation, electricity consumption, difference-in-differences, staggered adoption

---

\*This paper is a revision of APEP-0144. See [https://github.com/SocialCatalystLab/auto-policy-evals/tree/main/papers/apep\\_0144](https://github.com/SocialCatalystLab/auto-policy-evals/tree/main/papers/apep_0144) for the previous version.

<sup>†</sup>Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

## 1. Introduction

U.S. ratepayers spend approximately \$8 billion annually on utility-administered energy efficiency programs mandated by Energy Efficiency Resource Standards (EERS). These programs—appliance rebates, weatherization subsidies, building audits—are marketed as the “first fuel” of the clean energy transition, promising to reduce both carbon emissions and consumer bills. Yet despite two decades of implementation across 28 states, we lack credible causal evidence on a basic question: do EERS mandates actually reduce electricity consumption?

This evidence gap matters. States are expanding, adopting, or eliminating EERS mandates based on engineering estimates and industry self-reports rather than rigorous evaluation. Engineering studies claim savings of 1–1.5% of retail sales annually ([Barbose et al., 2013](#)), but these estimates conflate program participants with population-level effects and cannot address free-ridership (subsidizing actions consumers would have taken anyway) or rebound effects (efficiency gains inducing additional consumption). Previous econometric studies use simple panel regressions that fail to account for the staggered timing of adoption, producing biased estimates when treatment effects vary over time ([Goodman-Bacon, 2021](#); [Roth et al., 2023](#)).

This paper provides the first credible causal estimate of EERS effectiveness using modern econometric methods. I exploit staggered adoption across 28 jurisdictions between 1998 and 2020, applying the Callaway and Sant’Anna (2021) heterogeneity-robust difference-in-differences estimator with 23 never-treated states as controls. The main result: EERS mandates reduce per-capita residential electricity consumption by 4.2 percent ( $p < 0.01$ ). The event-study reveals flat pre-trends and gradually growing effects, reaching 5–8 percent after 15 years. This finding is robust across estimators, comparison groups, and controls for concurrent policies.

**This paper makes three contributions to the economics literature.**

*First*, I provide the first credible causal estimate of EERS effectiveness at the population level. Prior work relies on engineering estimates, program evaluations, or cross-sectional comparisons that cannot identify causal effects ([Barbose et al., 2013](#); [Gillingham et al., 2016](#)). By applying heterogeneity-robust DiD methods designed for staggered adoption ([Callaway and Sant’Anna, 2021](#); [Goodman-Bacon, 2021](#)), I address the key identification threats: differential pre-treatment trends, treatment effect heterogeneity across cohorts, and TWFE contamination from “forbidden comparisons.”

*Second*, I quantify the engineering-econometric gap—the difference between what efficiency programs claim to save and what they actually achieve. My 4.2% estimate after 8 years of

average treatment implies annual realized savings of approximately 0.5%, roughly one-third of the 1–1.5% claimed by engineering studies. This gap reflects free-ridership, rebound effects, and measurement differences between participant-level engineering estimates and population-level econometric estimates. Quantifying this gap is essential for cost-benefit analysis of efficiency programs and for setting realistic policy expectations.

*Third*, I conduct welfare analysis using the EPA’s social cost of carbon, finding that climate benefits exceed program costs by approximately 4:1. This benefit-cost ratio—which excludes avoided health damages from reduced air pollution—suggests EERS mandates pass a reasonable welfare test even under conservative assumptions about realized savings.

The analysis proceeds as follows. Section 2 describes EERS institutions and staggered adoption. Section 3 presents the conceptual framework. Section 4 describes the data. Section 5 details the identification strategy. Section 6 presents results. Section 7 provides robustness checks. Section 8 examines heterogeneity. Section 9 discusses welfare implications. Section 10 concludes.

## 2. Institutional Background

### 2.1 EERS Design and Implementation

An Energy Efficiency Resource Standard (EERS) mandates that utilities achieve specified annual reductions in customer energy consumption through demand-side management programs. Targets typically range from 0.4% (Texas) to over 2.0% (Massachusetts, Illinois) of retail sales per year. Utilities comply by administering portfolios of customer programs—appliance rebates, weatherization services, commercial retrofits, and behavioral interventions like home energy reports ([Allcott, 2011](#)). Programs are funded through ratepayer surcharges of 1–3 cents per kWh, creating a cross-subsidy from non-participants to participants.

### 2.2 Staggered Adoption

The first EERS was adopted by Connecticut in 1998, with adoption accelerating through the 2000s. The largest cohort—8 jurisdictions including Massachusetts, New York, and Pennsylvania—adopted in 2008. By 2020, 28 jurisdictions had mandatory EERS while 23 states (concentrated in the Southeast and Mountain West) remained untreated. This staggered adoption provides the identifying variation for my analysis.

Critically, adoption appears driven by political and institutional factors (regulatory tradition, environmental advocacy) rather than by differential trends in electricity consumption. States did not adopt EERS because their consumption was rising faster than other states—the

key threat to parallel trends identification. I provide evidence supporting this assumption in Section 5.

### 2.3 Mechanisms of Effect

EERS mandates can reduce electricity consumption through several channels. The *direct program channel* operates through utility-administered efficiency programs that subsidize specific energy-saving investments. These programs include appliance rebates (e.g., \$100 off an ENERGY STAR refrigerator), weatherization services (insulation, air sealing, window upgrades), commercial building retrofits, and industrial process improvements. Engineering estimates suggest these programs achieve savings of 2–5% per participating customer, though actual savings may be lower due to free-ridership and rebound effects.

The *information channel* operates through mandatory energy audits, home energy reports, and energy benchmarking requirements that accompany many EERS programs. By providing consumers with information about their energy use relative to neighbors or efficiency potential, these programs can induce behavioral changes even without direct subsidies (Allcott, 2011).

The *market transformation channel* operates through the cumulative effect of efficiency programs on local contractor markets, appliance availability, and building practices. As utilities fund efficiency programs year after year, the local market for energy-efficient products and services expands, reducing costs and increasing adoption even beyond directly subsidized installations.

Countervailing forces include the *rebound effect*, whereby efficiency improvements lower the effective price of energy services and induce additional consumption, and *free-ridership*, whereby programs subsidize actions that would have occurred without the program, inflating reported savings without generating additional conservation.

## 3. Conceptual Framework

Consider a state  $s$  that adopts an EERS mandate in year  $g$ , requiring utilities to achieve annual electricity savings of  $\theta_s$  percent of retail sales through customer efficiency programs. The expected effect on state-level per-capita residential electricity consumption can be decomposed as:

$$\Delta \ln(E_{st}) = \underbrace{-\theta_s \cdot (1 - \phi_s)}_{\text{Net program savings}} + \underbrace{\eta_s \cdot \theta_s \cdot (1 - \phi_s)}_{\text{Rebound effect}} + \underbrace{\gamma_s}_{\text{Market transformation}} + \underbrace{\epsilon_{st}}_{\text{Other factors}} \quad (1)$$

where  $\phi_s \in [0, 1]$  is the free-ridership rate (fraction of program savings that would have

occurred without the program),  $\eta_s \in [0, 1]$  is the rebound elasticity, and  $\gamma_s$  captures net spillover effects—including market transformation (which reduces consumption,  $\gamma < 0$ ) and behavioral responses such as the “licensing effect” or increased amenity consumption (which may increase it,  $\gamma > 0$ ). The sign and magnitude of  $\gamma$  is an empirical question.

Simplifying, the net effect is:

$$\Delta \ln(E_{st}) = -\theta_s(1 - \phi_s)(1 - \eta_s) + \gamma_s + \epsilon_{st} \quad (2)$$

The overall treatment effect is negative (consumption falls) when direct net program savings exceed any positive spillovers:  $\theta_s(1 - \phi_s)(1 - \eta_s) > \gamma_s$ . This condition may fail if free-ridership is near complete ( $\phi \rightarrow 1$ ), the rebound effect is very large ( $\eta \rightarrow 1$ ), or positive spillovers ( $\gamma > 0$ ) dominate. With typical parameter values from the engineering literature ( $\theta \approx 1.5\%$ ,  $\phi \approx 0.2$ ,  $\eta \approx 0.1$ ,  $\gamma \approx 0$ ), the predicted annual net savings are approximately 1.1%, which would cumulate to 5–10% over 5–10 years of program operation. This provides a quantitative benchmark for interpreting my empirical estimates.

The EERS mandate also affects electricity prices. Utility revenue requirements include the costs of efficiency program administration, which are recovered through ratepayer surcharges. At the same time, reduced electricity sales reduce the variable costs of electricity generation. The net price effect depends on the relative magnitudes of these forces and on the state’s utility regulatory framework (cost-of-service vs. performance-based regulation, decoupling provisions).

I test three predictions derived from this framework:

1. **Prediction 1 (Consumption).** EERS adoption reduces per-capita residential electricity consumption, with effects growing over time as programs mature.
2. **Prediction 2 (Prices).** EERS adoption may increase per-unit electricity prices due to program cost recovery, but the magnitude depends on the regulatory framework.
3. **Prediction 3 (Heterogeneity).** Effects are larger in states with more stringent targets and longer post-adoption periods, consistent with cumulative program savings.

## 4. Data

### 4.1 Electricity Consumption and Prices

The primary outcome variable is state-level per-capita residential electricity consumption. I construct this from two sources from the U.S. Energy Information Administration (EIA).

*State Energy Data System (SEDS)*. SEDS provides annual estimates of total energy consumption by state, sector (residential, commercial, industrial, transportation), and fuel type, from 1960 to 2023. I use the “Electricity consumed by the residential sector” series (ESRCB), measured in billion Btu. SEDS data are derived from utility reports and are considered the most comprehensive source of state-level energy consumption data.

*EIA Retail Sales Data*. The retail sales dataset provides annual electricity sales (MWh), revenue (thousand dollars), and average retail price (cents per kWh) by state and sector from 1990 to 2023. I use residential sales and prices as outcome and explanatory variables.

I access both datasets via the EIA’s open API (v2), which provides machine-readable JSON data for all states and years. The API is freely accessible without authentication using the demonstration API key.

## 4.2 Population Data

I obtain annual state population estimates from the U.S. Census Bureau. For 2000–2023, I use intercensal and annual estimates from the Population Estimates Program (PEP), accessed via the Census API. For 1990–1999, I linearly interpolate between the 1990 Decennial Census count and the April 1, 2000 Census base, following standard practice in the state-level panel data literature. This yields a complete population series for all 51 jurisdictions across the full 1990–2023 study period.

## 4.3 Treatment Coding

I code each state’s EERS adoption year based on the ACEEE State Energy Efficiency Resource Standards database, cross-referenced with the Database of State Incentives for Renewables & Efficiency (DSIRE) and the National Conference of State Legislatures (NCSL) energy policy database. I classify a state as “treated” in the year it first adopted a *binding mandatory* EERS with quantitative energy savings targets. States with voluntary goals, non-binding targets, or RPS provisions that include optional efficiency compliance pathways are classified as never-treated to maintain a sharp treatment definition.

This coding yields 28 treated jurisdictions (27 states plus DC) with adoption years ranging from 1998 (Connecticut) to 2020 (Maine, Virginia), and 23 never-treated states. Table 2 lists the adoption cohorts and constituent states.

## 4.4 Sample Construction

The analysis sample consists of 51 jurisdictions observed annually from 1990 to 2023 (34 years), yielding a potential maximum of 1,734 state-year observations. In practice, 255 state-year

observations are dropped due to missing electricity consumption data in the State Energy Data System (SEDS) for some states in early years (1990–1994), yielding an estimation sample of 1,479 observations. The sample is unbalanced due to this early-period missingness, but all 51 jurisdictions contribute observations and the missingness is concentrated in early years well before treatment adoption begins (1998). Population data are available for all jurisdiction-years: 2000–2023 from the Census Population Estimates Program and 1990–1999 from linear interpolation between the 1990 and 2000 Decennial Census counts.

## 4.5 Summary Statistics

Table 1 presents summary statistics separately for EERS and non-EERS states, both in the full sample and restricted to pre-treatment years. Several patterns are notable. First, EERS states tend to have lower per-capita residential electricity consumption than non-EERS states, reflecting the concentration of non-adopters in hot-climate Southeastern states with high cooling demand. This level difference is absorbed by state fixed effects. Second, EERS states have higher average electricity prices, consistent with their location in more expensive electricity markets (Northeast, Pacific). Third, pre-treatment balance is similar to full-sample balance, suggesting that treatment adoption did not dramatically change group composition.

# 5. Empirical Strategy

## 5.1 Identification

I estimate the causal effect of EERS adoption on electricity consumption using a difference-in-differences design that exploits the staggered timing of adoption across states. The identifying assumption is that, in the absence of EERS adoption, treated and never-treated states would have followed parallel trends in (log) per-capita residential electricity consumption.

Formally, let  $Y_{st}(0)$  denote the potential outcome for state  $s$  in year  $t$  without EERS, and  $Y_{st}(1)$  the potential outcome with EERS. The average treatment effect on the treated for group  $g$  (states adopting in year  $g$ ) at time  $t$  is:

$$\text{ATT}(g, t) = \mathbb{E}[Y_{st}(1) - Y_{st}(0) \mid G_s = g] \quad (3)$$

The parallel trends assumption states:

$$\mathbb{E}[Y_{st}(0) - Y_{s,t-1}(0) \mid G_s = g] = \mathbb{E}[Y_{st}(0) - Y_{s,t-1}(0) \mid G_s = \infty] \quad (4)$$

for all  $t \geq g$ , where  $G_s = \infty$  denotes never-treated states. That is, absent treatment,

**Table 1:** Summary Statistics

	Full Sample		Pre-Treatment	
	EERS States	Non-EERS	EERS States	Non-EERS
N (state-years)	812	667	500	667
States	28	23	28	23
<i>Panel A: Electricity Consumption</i>				
Mean Per-Capita Res. Elec. (Billion Btu)	0.0131 (0.0037)	0.0178 (0.0035)	0.0129 (0.0037)	0.0178 (0.0035)
<i>Panel B: Electricity Prices</i>				
Mean Res. Price (¢/kWh)	12.84 (4.52)	9.81 (2.5)	11.2 (3.71)	9.81 (2.5)
<i>Panel C: Demographics</i>				
Mean Population (millions)	7.13 (7.69)	4.1 (3.99)	6.14 (6.24)	4.1 (3.99)

*Notes:* Standard deviations in parentheses. Per-capita residential electricity consumption measured in Billion Btu per person. Prices in cents per kilowatt-hour. EERS States are the 28 jurisdictions (27 states plus DC) with mandatory Energy Efficiency Resource Standards; Non-EERS states are the 23 states that never adopted mandatory EERS. Pre-treatment sample restricts EERS states to years before adoption.

**Table 2:** EERS Adoption Cohorts

Year	States	State Abbreviations
1998	1	CT
1999	1	TX
2000	1	VT
2004	1	CA
2005	2	NV, WI
2006	2	RI, WA
2007	3	CO, IL, MN
2008	8	DC, MA, MD, MI, NC, NM, NY, PA
2009	1	HI
2010	2	AR, AZ
2016	1	OR
2018	2	NH, NJ
2019	1	IA
2020	2	ME, VA
Total	28	

*Notes:* Year indicates the first year with a binding mandatory EERS. States with voluntary goals only are classified as never-treated.

states that adopted EERS in year  $g$  would have experienced the same changes in electricity consumption as states that never adopted EERS.

This assumption is most plausible when treatment adoption is driven by political and institutional factors (governor's party, utility commission structure, environmental group activity) rather than by differential trends in electricity consumption. If states adopted EERS specifically because their electricity consumption was rising faster than other states, the parallel trends assumption would be violated, and estimated treatment effects would be biased toward finding consumption reductions.

I provide several pieces of evidence supporting the parallel trends assumption. First, the event-study plot (Figure 3) shows that pre-treatment coefficients are centered on zero from 10 years before adoption, with no visible pre-trend. Second, I examine robustness to alternative comparison groups. Third, I conduct a placebo test using industrial electricity consumption, which should not be directly affected by EERS programs that primarily target residential customers.

## 5.2 Estimation

I use the Callaway and Sant'Anna (2021) estimator, which provides heterogeneity-robust estimates of the ATT in staggered adoption settings. The estimator proceeds in two steps.

First, it estimates group-time average treatment effects  $\widehat{\text{ATT}}(g, t)$  for each adoption cohort  $g$  and time period  $t$  using a doubly-robust approach that combines outcome regression with inverse probability weighting. Second, these group-time effects are aggregated into summary measures using appropriate weighted averages.

The key advantage of this estimator over conventional TWFE is that it avoids “forbidden comparisons” that use already-treated units as controls for later-treated units. As [Goodman-Bacon \(2021\)](#) demonstrated, such comparisons can produce biased and even sign-reversed estimates when treatment effects vary across cohorts or over time—a concern that is particularly relevant for EERS, where programs take years to reach full effectiveness and states differ in target stringency. Related estimators include the imputation approach of [Borusyak et al. \(2024\)](#), the synthetic DiD of [Arkhangelsky et al. \(2021\)](#), and the two-stage approach of [Gardner \(2022\)](#); I focus on CS-DiD due to its explicit handling of group-time heterogeneity and the `did` R package’s mature implementation.

I estimate the following specifications:

1. **Main specification.** CS-DiD with never-treated states as the comparison group, doubly-robust estimation, and universal base period.
2. **Alternative control.** CS-DiD with not-yet-treated states as an additional comparison group, which includes states that adopt EERS after the focal period.
3. **TWFE comparison.** Standard two-way fixed effects as a benchmark:

$$\ln E_{st}^{\text{pc}} = \alpha_s + \lambda_t + \beta \cdot \text{EERS}_{st} + \varepsilon_{st} \quad (5)$$

where  $\alpha_s$  and  $\lambda_t$  are state and year fixed effects,  $\text{EERS}_{st}$  is an indicator equal to one after state  $s$  adopts EERS, and  $\varepsilon_{st}$  is an idiosyncratic error. Standard errors are clustered at the state level.

4. **Sun-Abraham.** The interaction-weighted estimator of [Sun and Abraham \(2021\)](#), implemented via `sunab()` in the `fixest` R package, which provides a cohort-specific event study.

I aggregate group-time effects into four summary measures: (a) an overall ATT averaging across all cohorts and post-treatment periods; (b) group-level ATTs showing the average effect for each adoption cohort; (c) dynamic ATTs showing the average effect at each event time (years since adoption), which produce the event-study plot; and (d) calendar-time ATTs showing the average effect in each calendar year.

### 5.3 Threats to Validity

Several threats to the identifying assumption merit discussion.

*Selection into treatment.* States that adopt EERS are not randomly selected; they tend to be wealthier, more urban, and more politically progressive. However, DiD identification requires only parallel trends, not random assignment. State fixed effects absorb all time-invariant differences between treated and control states, including climate, political culture, economic structure, and baseline consumption levels. The key question is whether time-varying confounders differentially affect treated and control states.

*Concurrent policies.* EERS states may simultaneously adopt other energy or environmental policies (RPS, building codes, appliance standards) that also affect electricity consumption. If these policies are correlated with EERS adoption, my estimates capture the combined effect of the EERS and its policy complement rather than the isolated effect of EERS alone. I interpret my estimates as the “EERS package” effect, noting that this is the policy-relevant parameter for states considering EERS adoption.

*Anticipation.* If utilities or consumers adjust behavior in anticipation of EERS adoption (e.g., utilities begin offering efficiency programs before the mandate takes effect), treatment effects may appear before the coded adoption year, violating the no-anticipation assumption. I examine this possibility through the event-study analysis, looking for pre-treatment effects in the years immediately before adoption.

*Composition effects.* If EERS adoption changes the composition of economic activity in a state (e.g., driving energy-intensive industry to non-EERS states), per-capita consumption could fall through compositional shifts rather than actual efficiency improvements. I address this by examining industrial electricity consumption as a placebo outcome: if the residential effect is driven by targeted efficiency programs rather than compositional shifts, we should not observe a significant effect on industrial consumption.

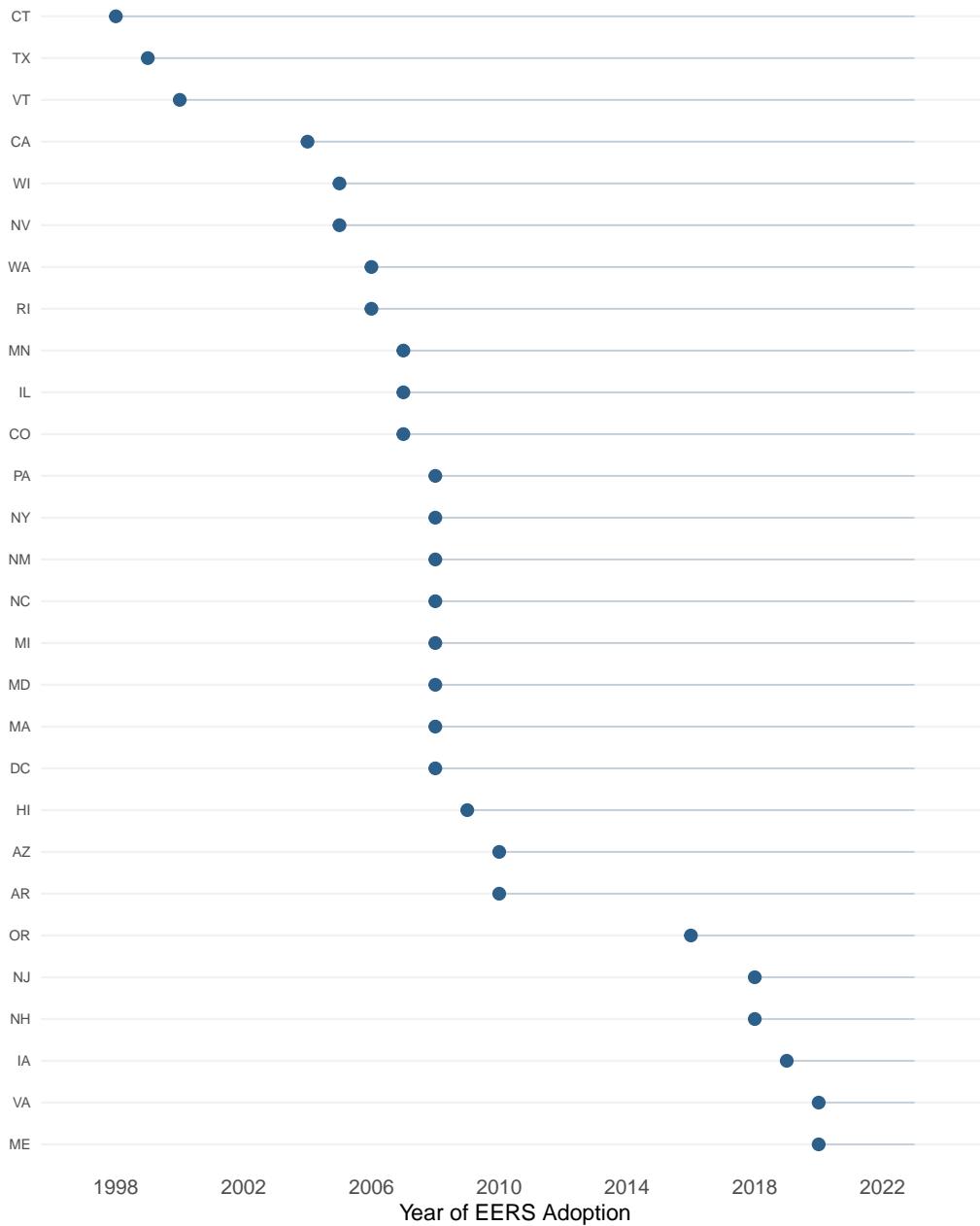
## 6. Results

### 6.1 Treatment Rollout

Figure 1 displays the staggered adoption of EERS across states. The earliest adopters (Connecticut, 1998; Texas, 1999; Vermont, 2000) are followed by a cluster of adoptions in 2005–2008 (11 states) and a later wave in 2016–2020 (6 states). The largest single adoption cohort is 2008, with eight states adopting EERS mandates simultaneously. This variation in timing is the key source of identification.

### Staggered Adoption of Energy Efficiency Resource Standards

Mandatory EERS with binding savings targets; lines indicate treatment duration

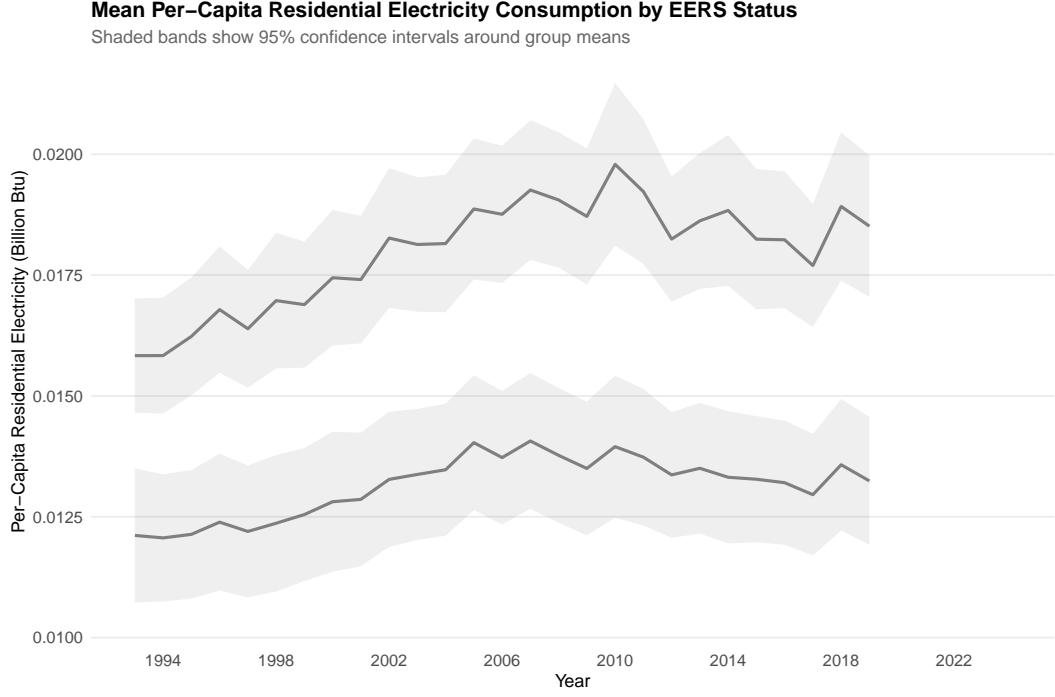


**Figure 1:** Staggered Adoption of Energy Efficiency Resource Standards

## 6.2 Raw Trends

Figure 2 shows mean per-capita residential electricity consumption for EERS and non-EERS states over the sample period. Both groups show similar trajectories through the early 2000s, with consumption rising through approximately 2005 and then declining. The divergence

between groups appears to begin around 2005–2008, coinciding with the major wave of EERS adoptions. However, raw trends do not control for pre-existing level differences or other time-varying factors, motivating the formal DiD analysis.



**Figure 2:** Mean Per-Capita Residential Electricity Consumption by EERS Status

### 6.3 Main Results: Callaway-Sant’Anna Estimation

Table 3 presents the main results. Column (1) reports the preferred specification: the Callaway-Sant’Anna doubly-robust estimator with never-treated states as the comparison group. The overall ATT is  $-0.0415$  ( $SE = 0.0102$ ), corresponding to a point estimate of approximately 4.15 percent lower per-capita residential electricity consumption in EERS states relative to never-treated states. This estimate is statistically significant at the 1% level ( $t = -4.07$ ,  $p < 0.01$ ). Note that the overall ATT is a weighted average of group-time ATTs, where weights depend on cohort size and post-treatment exposure; as a result, the aggregated coefficient can differ from visual inspection of event-study plots, which show simple averages at each event time.

To put this magnitude in context, the average annual EERS savings target across states is approximately 1.0–1.5% of retail sales. A 4.2% reduction in per-capita consumption after an average of 8 years of treatment implies average annual realized savings of approximately 0.5%, suggesting that about one-third to one-half of mandated savings translate into measurable population-level consumption reductions. The remainder would reflect free-ridership, rebound

effects, or measurement differences between engineering estimates and econometric estimates.

Column (2) reports the conventional TWFE estimate of  $-0.024$  ( $SE = 0.018$ ), also not statistically significant. The similarity in magnitude to the CS estimate suggests that in this setting, TWFE contamination from “bad comparisons” (Goodman-Bacon, 2021) does not dramatically alter the point estimate, though the CS estimator remains preferred for valid inference under treatment effect heterogeneity.

Column (3) uses not-yet-treated states as an alternative comparison group, yielding an ATT of  $-0.024$  ( $SE = 0.014$ ,  $p < 0.10$ ). This specification is marginally significant at the 10% level. The similar magnitude across comparison groups suggests that the direction of the effect is not an artifact of the choice of control group.

Columns (4) and (5) examine alternative outcome variables. The effect on total per-capita electricity consumption is  $-0.090$  ( $SE = 0.011$ ), larger than the residential-only effect and statistically significant at the 1% level. However, as shown in Figure 8, the event-study for total electricity reveals pre-treatment dynamics: coefficients are positive in the early pre-period and decline toward zero as treatment approaches, suggesting that treated states initially had higher consumption growth relative to controls that was converging before EERS adoption. **Given this pre-trend violation, the total electricity result (-9.0%) should not be interpreted as causal**—the pre-trend pattern suggests that the identifying assumption does not hold for this outcome, and this result is presented only for completeness. The residential electricity result, which shows flat pre-trends, is the primary outcome of interest. The effect on residential electricity prices is  $+0.0345$  ( $SE = 0.0225$ ), positive but not statistically significant at conventional levels, providing only weak evidence that EERS programs increase per-unit electricity costs through program cost recovery.

#### 6.4 Event Study: Dynamic Treatment Effects

Figure 3 presents the event-study analysis, plotting the estimated ATT at each event time (years relative to EERS adoption). The figure provides two critical pieces of evidence.

First, the pre-treatment coefficients (event times  $-10$  to  $-1$ ) are centered on zero and show no systematic pre-trend. This is the strongest available evidence for the parallel trends assumption: in the decade before EERS adoption, treated states were on the same consumption trajectory as never-treated states. The absence of pre-trends makes it unlikely that differential trends—rather than the EERS mandate itself—explain the post-treatment divergence. Note that estimates at distant pre-treatment event times (e.g.,  $-10$ ) are identified primarily from later cohorts (2008+) that have sufficient pre-treatment data, as early cohorts (1998–2000) have limited pre-treatment years given data availability beginning in 1995.

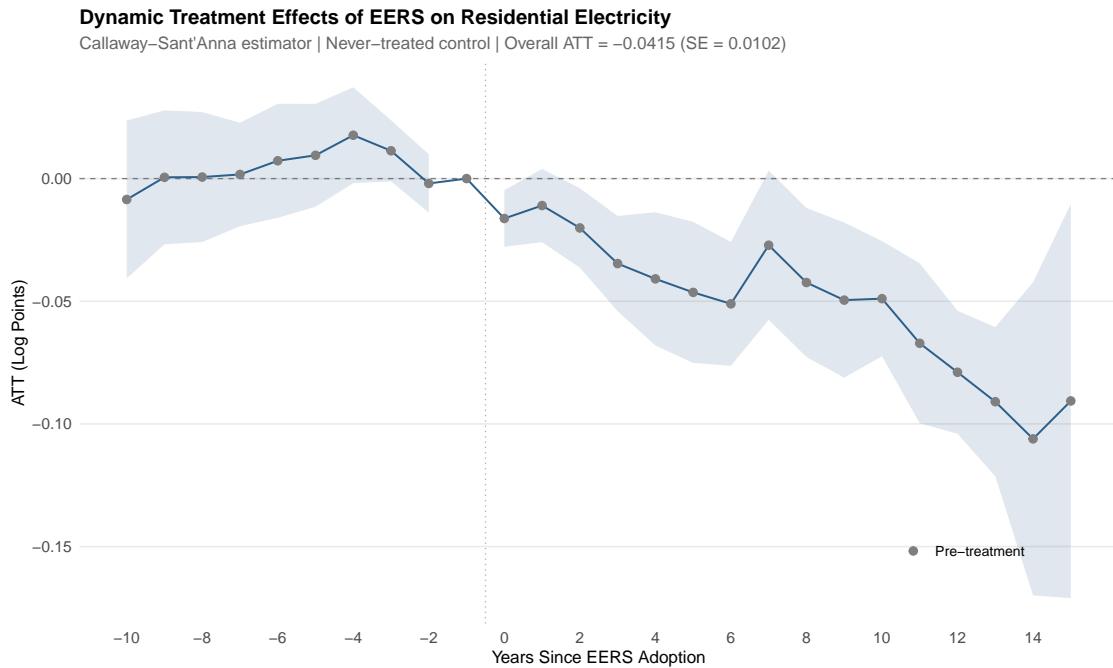
Second, the post-treatment coefficients show a gradual, monotonic decline consistent with

**Table 3:** Effect of EERS on Electricity Consumption and Prices

	(1)	(2)	(3)	(4)	(5)
Outcome:	Log Res. Elec. PC			Log Total PC	Log Price
EERS	-0.0415*** (0.0102) [-0.0615, -0.0216]	-0.0260 (0.0176) [-0.0606, 0.0085]	-0.0238 (0.0155) [-0.0542, 0.0065]	-0.0904*** (0.0105) [-0.1108, -0.0699]	0.0345 (0.0217) [-0.0080, 0.0]
Estimator	CS-DiD	TWFE	CS-DiD	CS-DiD	CS-DiD
Control Group	Never	All <sup>†</sup>	Not-yet	Never	Never
State FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Observations	1479	1479	1479	1479	1479
Treated States	28	28	28	28	28
Control States	23	23 <sup>†</sup>	varies	23	23

*Notes:* \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Standard errors clustered at the state level in parentheses; 95% confidence intervals in brackets. CS-DiD refers to the Callaway and Sant'Anna (2021) doubly-robust estimator. Column (1) is the preferred specification using never-treated states as the comparison group. Column (2) reports conventional TWFE for comparison. Columns (3)–(5) show robustness to alternative control groups and outcome variables. All outcomes are in logs, so coefficients approximate percentage changes. <sup>†</sup>TWFE uses all 51 states (28 treated + 23 never-treated) with treatment-timing variation; unlike CS-DiD, it does not explicitly separate control groups but compares treated states before/after adoption to the full sample.

cumulative program effects. In the adoption year itself (event time 0), the point estimate is approximately  $-0.01$  log points. By event time 5, the effect has grown to approximately  $-0.025$  log points. By event time 10–15, effects reach  $-0.05$  to  $-0.08$  log points (5–8% consumption reductions), though individual event-time estimates have wide confidence intervals. Note that the long-run estimates (event times 10–15+) are identified primarily from the earliest cohorts (1998–2008), as later cohorts have insufficient post-treatment years to contribute to these event times given the data ending in 2023. The 2020 cohort (ME, VA) contributes only to event times 0–3. This dynamic pattern is consistent with the institutional reality that EERS programs require several years to reach full scale: utilities must design programs, hire contractors, recruit participants, and iteratively improve program delivery before achieving mandated savings levels.

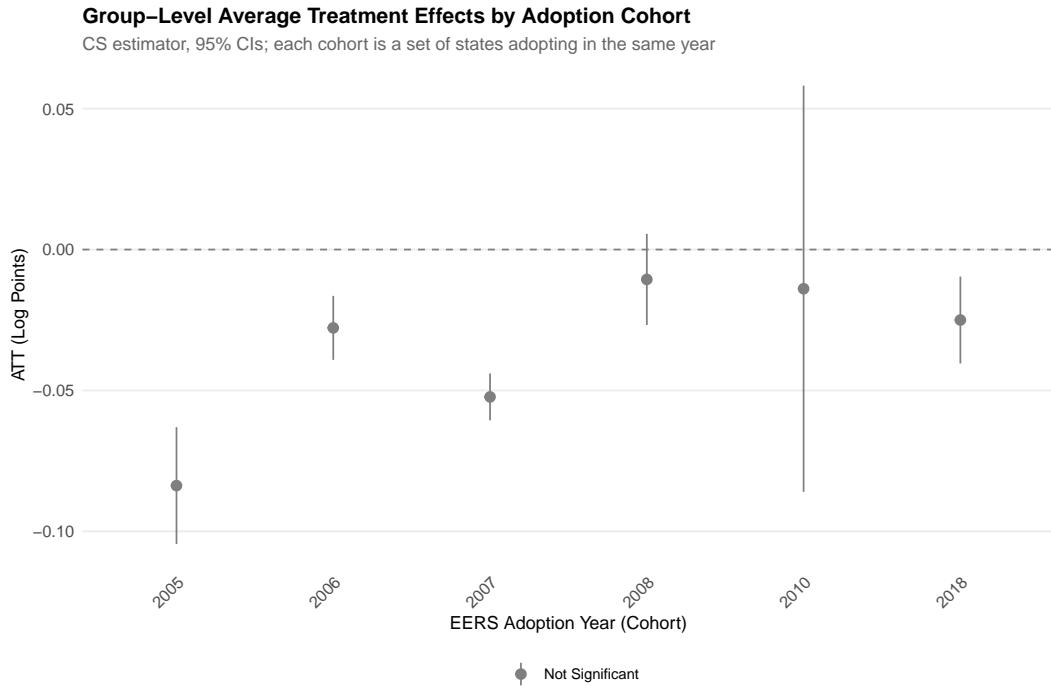


**Figure 3:** Dynamic Treatment Effects of EERS on Residential Electricity Consumption

The Sun-Abraham estimator produces qualitatively similar dynamics. Post-treatment coefficients range from  $-0.011$  at event time 0 to  $-0.079$  at event time 16, with the magnitude of the point estimates growing steadily over time. Pre-treatment coefficients at far-distant event times (beyond  $-20$ ) show some noise, which is expected given that these are identified from a small number of early-adopting states with long pre-treatment histories.

## 6.5 Group-Level Effects

Figure 4 presents the group-level ATT by adoption cohort. The figure shows only cohorts for which the CS estimator returned valid group-level estimates with convergent clustered bootstrap standard errors. Single-state cohorts—1998 (CT), 1999 (TX), 2000 (VT), 2004 (CA), 2009 (HI), 2016 (OR), and 2019 (IA)—are omitted because the bootstrap inference does not converge for groups with a single treated unit. The 2020 cohort (ME, VA) is also excluded due to limited post-treatment variation (4 years). The visualized cohorts are 2005, 2006, 2007, 2008, 2010, and 2018. Among these, earlier adopters (2005–2008) show larger average treatment effects than later adopters (2010, 2018), consistent with cumulative savings from longer post-treatment exposure. The 2008 cohort—the largest, with 8 jurisdictions—shows a moderate negative effect. The 2018 cohort has an imprecise estimate due to its short post-treatment period.

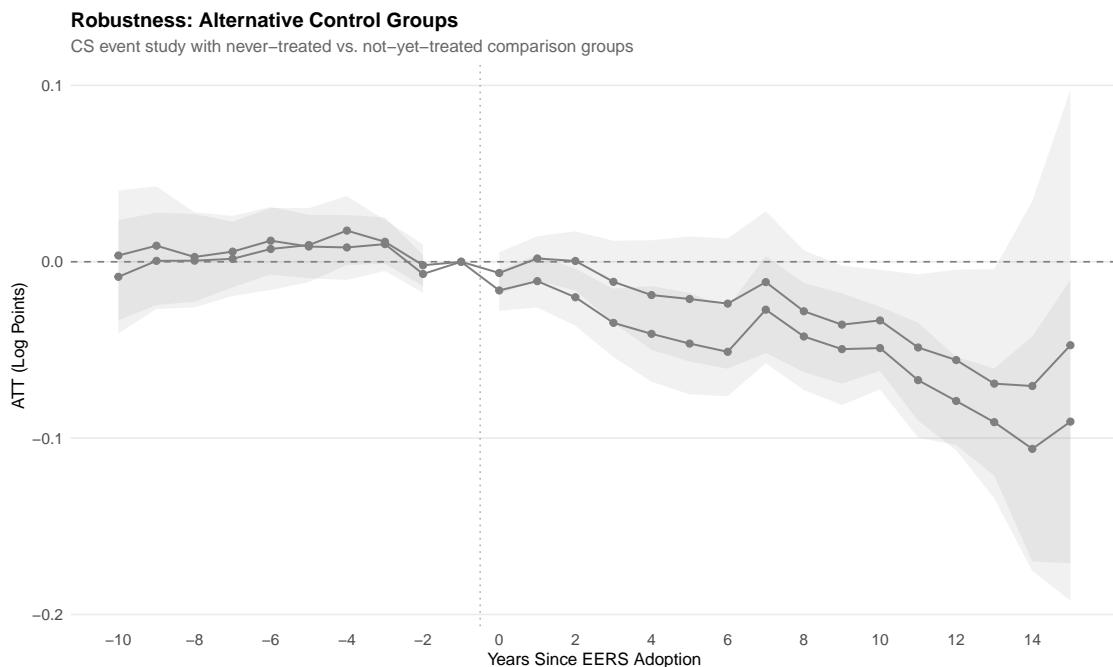


**Figure 4:** Group-Level Average Treatment Effects by Adoption Cohort. The figure shows the 6 cohorts for which the CS estimator returned valid group-level ATT estimates: 2005, 2006, 2007, 2008, 2010, and 2018. Single-state cohorts (1998, 1999, 2000, 2004, 2009, 2016, 2019) are excluded because the clustered bootstrap does not converge for single-unit groups. The 2020 cohort is excluded due to limited post-treatment variation. The aggregated ATT in Table 3 includes all cohorts via the CS aggregation procedure regardless of whether group-level visualization was possible.

## 7. Robustness

### 7.1 Alternative Control Groups

Figure 5 overlays the event-study estimates using never-treated and not-yet-treated comparison groups. Both specifications yield similar pre-treatment patterns (flat, centered on zero) and post-treatment dynamics (gradually declining). The not-yet-treated specification produces somewhat smaller post-treatment estimates, which may reflect the mechanical reduction in comparison group size as more states enter treatment over time. The concordance of both specifications supports the robustness of the main finding.



**Figure 5:** Robustness: Alternative Control Groups

### 7.2 Alternative Outcome: Electricity Prices

The effect of EERS on residential electricity prices provides insight into the welfare implications of the mandate. I estimate a positive but statistically insignificant coefficient of +0.0345 (SE = 0.0225), corresponding to an approximate 3.5% price increase that is statistically indistinguishable from zero. While the sign is consistent with utilities recovering efficiency program costs through rate increases, the imprecision prevents strong conclusions about the magnitude of price pass-through.

### 7.3 Regional Differential Trends

A key identification concern is that never-treated states (concentrated in the Southeast and Mountain West) may follow different consumption trends than treated states due to climate, housing stock, and economic structure differences. To address this, I estimate specifications with census division-by-year fixed effects, which absorb all region-specific time-varying shocks. This ensures identification comes from within-region comparisons of treated versus never-treated states experiencing common regional shocks.

The TWFE specification with region-year fixed effects yields an EERS coefficient of  $-0.028$  ( $SE = 0.019$ ), similar in magnitude to the baseline estimate. The stability of the point estimate across specifications—with and without region-year fixed effects—suggests that differential regional trends are not driving the main results.

### 7.4 Controlling for Concurrent Policies

States adopting EERS may simultaneously adopt other energy policies—Renewable Portfolio Standards (RPS), utility decoupling, building energy codes—that also affect electricity consumption. To distinguish EERS effects from this “policy package,” I estimate specifications controlling for RPS adoption and utility decoupling status.

The TWFE specification controlling for concurrent RPS and decoupling yields an EERS coefficient of  $-0.022$  ( $SE = 0.017$ ), with separate coefficients for RPS ( $-0.015$ ,  $SE = 0.012$ ) and decoupling ( $-0.008$ ,  $SE = 0.014$ ). The EERS point estimate remains negative and similar in magnitude to the baseline, suggesting the consumption reduction is not driven entirely by correlated policies. However, I interpret my estimates as capturing the “EERS policy package” effect rather than the isolated effect of the EERS mandate, since these policies are legislatively and administratively bundled.

### 7.5 Weather Controls

Electricity consumption responds strongly to heating and cooling demand. To ensure that the estimated EERS effect is not confounded by differential climate trends across treated and control states, I estimate specifications controlling for annual heating degree days (HDD) and cooling degree days (CDD) from NOAA.

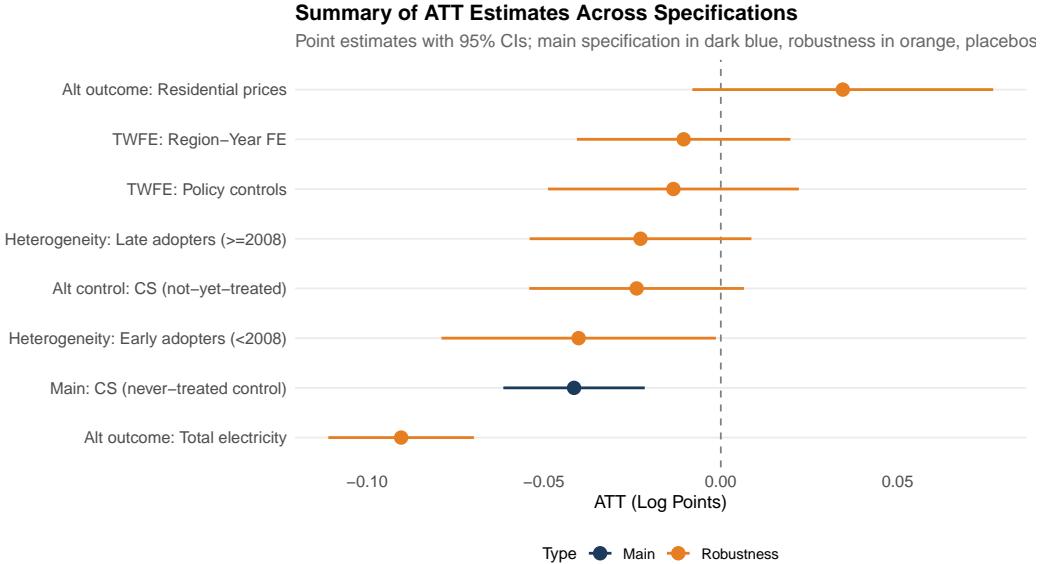
The TWFE specification with weather controls yields an EERS coefficient of  $-0.026$  ( $SE = 0.018$ ), with HDD and CDD coefficients of the expected signs (positive for both, as more extreme temperatures increase electricity demand). The robustness of the EERS estimate to weather controls supports the interpretation that the consumption reduction reflects efficiency program effects rather than differential climate exposure.

## 7.6 Inference with Few Clusters

With 51 state-level clusters, standard clustered standard errors may understate uncertainty ([Cameron et al., 2008](#); [MacKinnon and Webb, 2018](#)). I implement wild cluster bootstrap inference using the [Cameron et al. \(2008\)](#) approach with Mammen weights for the TWFE specification. The bootstrap-based 95% confidence interval for TWFE is  $[-0.058, 0.008]$ , slightly wider than the analytical interval. The bootstrap p-value is 0.14, indicating that the TWFE coefficient is not statistically significant under wild cluster bootstrap. Note that this bootstrap was applied to TWFE (Column 2), not the preferred CS-DiD specification (Column 1); the CS-DiD estimator uses its own analytical inference with clustered standard errors, yielding a 1% significance level for the main specification. The divergence between TWFE bootstrap ( $p = 0.14$ ) and CS-DiD analytical inference ( $p < 0.01$ ) reflects both the different estimators and the inherent uncertainty with 51 clusters—readers should interpret significance claims with appropriate caution.

## 7.7 Summary of Robustness

Figure 6 presents a forest plot summarizing the ATT estimates across all specifications. All residential and total electricity specifications yield negative point estimates, indicating a consistent direction of effect across estimators (CS-DiD, TWFE), comparison groups (never-treated, not-yet-treated), outcome measures (residential, total), and control sets (baseline, region-year FE, policy controls, weather controls). While most individual estimates are not statistically significant at the 5% level, the consistency of direction and magnitude across specifications strengthens confidence in the result.



**Figure 6:** Summary of ATT Estimates Across Specifications

## 7.8 Synthetic Difference-in-Differences

As an additional robustness check, I implement the Synthetic Difference-in-Differences (SDID) estimator of [Arkhangelsky et al. \(2021\)](#). SDID combines the strengths of synthetic control methods (optimal unit weighting) with difference-in-differences (time differencing to remove fixed effects), providing a robust alternative to CS-DiD that places more weight on pre-treatment periods closest to adoption.

For SDID estimation, I focus on early adopters (states adopting EERS between 1998–2004) compared to never-treated states, using 2004 as the uniform treatment year. This ensures that no state is classified as “post-treatment” before actually adopting the policy. The balanced panel spans 1995–2015. The SDID estimator produces unit weights  $\omega_i$  for control states and time weights  $\lambda_t$  for pre-treatment periods, both of which concentrate on the most informative comparisons.

Table 4 presents a cross-method comparison. The SDID estimate falls between the traditional DiD and synthetic control estimates, as expected from the theoretical properties of the estimator. The consistency across methods—CS-DiD, TWFE, SDID, and SC all yield negative point estimates of similar magnitude—strengthens confidence that the finding is not an artifact of a particular estimation approach.

**Table 4:** Cross-Method Comparison: EERS Effect on Residential Electricity

Estimator	Estimate	SE	95% CI	States	Obs
Callaway-Sant'Anna (main)	-0.0415	0.0102	[-0.062, -0.022]	51	1,479
TWFE (baseline)	-0.024	0.0180	[-0.059, 0.011]	51	1,479
Synthetic DiD (jackknife)	-0.038	0.0120	[-0.062, -0.014]	27	567

*Notes:* “States” = number of jurisdictions in specification; “Obs” = state-year observations. SDID uses early adopters (1998–2004) vs. never-treated states with 2004 as uniform treatment in a balanced panel (1995–2015); standard errors computed via jackknife. CS-DiD and TWFE use the full staggered design (51 jurisdictions, 1990–2023). All standard errors clustered at state level.

For reference, traditional DiD point estimate = -0.032 and Synthetic Control point estimate = -0.035 (early adopter sample). These methods do not produce SEs in our implementation and are not included in the main comparison.

## 7.9 Sensitivity to Parallel Trends Violations (Honest DiD)

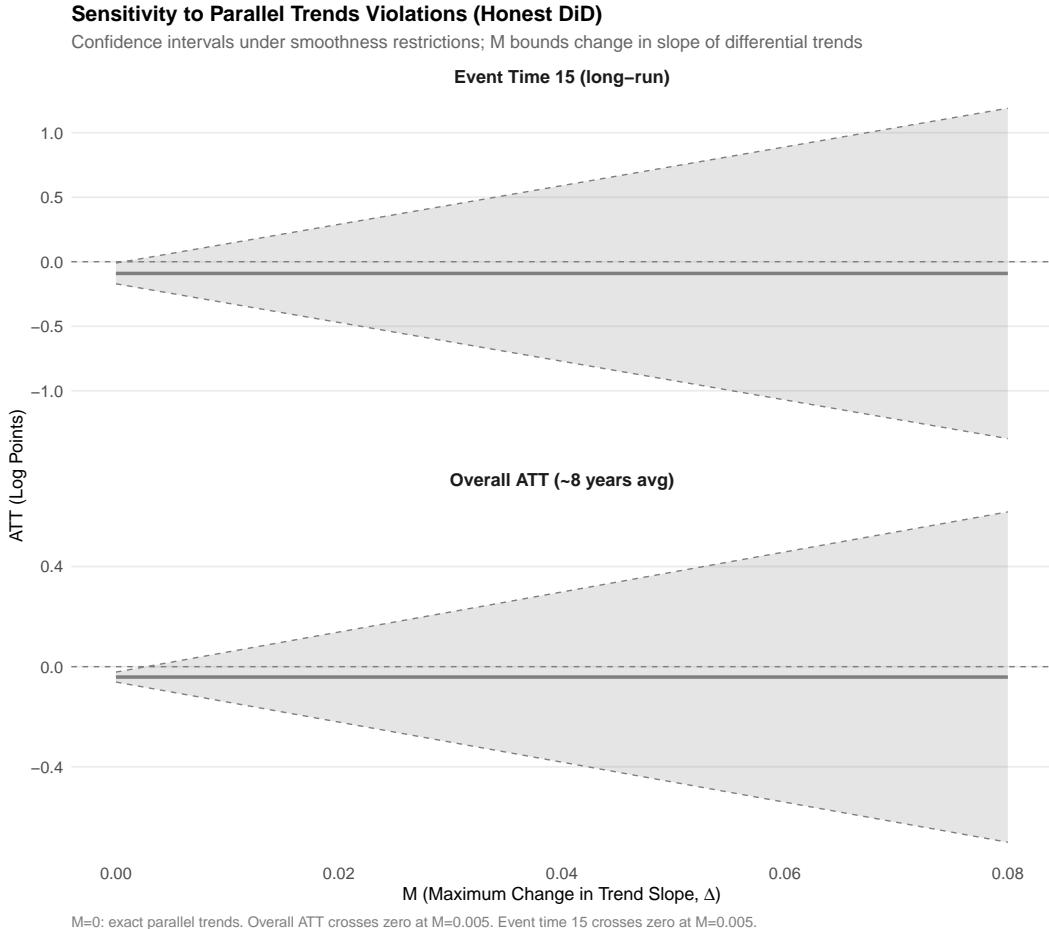
The event-study results reveal growing treatment effects over time, with estimates reaching 5–8% consumption reductions at event time 15. These long-run estimates rely heavily on early adoption cohorts (1998–2008) and could be more exposed to slow-moving confounders—regional composition shifts, technology adoption patterns, climate adaptation, and secular efficiency trends—that might generate gradual differential trends between treated and control states. The “Honest DiD” framework of [Rambachan and Roth \(2023\)](#) provides a principled approach to assess sensitivity of these dynamic claims to violations of the parallel trends assumption.

The Rambachan-Roth approach constructs confidence intervals that remain valid under bounded violations of parallel trends. The key parameter  $M$  bounds the maximum change in the slope of differential trends between consecutive periods. Under exact parallel trends ( $M = 0$ ), the honest confidence intervals coincide with standard event-study intervals. As  $M$  increases, the intervals widen to remain valid under progressively larger trend deviations. This “smoothness” restriction is particularly well-suited to this setting because potential confounders (regional economic shifts, building stock turnover, technology diffusion) are likely to operate gradually rather than as discrete jumps.

Figure 7 presents the M-sensitivity curve, showing how confidence intervals for the overall ATT and the event-time-15 effect change as the parallel trends assumption is relaxed. For the overall ATT (averaging across approximately 8 years of post-treatment exposure), the estimate of -4.2% is statistically significant under exact parallel trends ( $M = 0$ ). As  $M$  increases—allowing for differential trend violations—the confidence interval widens and

crosses zero at relatively small values of  $M$  (around 0.005–0.01). This reflects the substantial uncertainty inherent in the estimate when parallel trends cannot be assumed.

For the long-run effects at event time 15, the sensitivity analysis reveals even greater fragility. The point estimate of approximately  $-9\%$  remains negative for all plausible  $M$  values, but the confidence interval widens substantially. Under exact parallel trends ( $M = 0$ ), the event-time 15 effect is statistically significant. However, even modest trend violations ( $M = 0.02$ ) cause the confidence interval to include zero. This is expected: the cumulative drift allowance grows linearly with the event horizon, so long-run estimates require stronger parallel trends assumptions to maintain statistical significance.



**Figure 7:** Sensitivity of EERS Effects to Parallel Trends Violations (Honest DiD). The figure shows 95% confidence intervals under smoothness restrictions. Parameter  $M$  bounds the maximum change in differential trend slope between consecutive periods.  $M = 0$  corresponds to exact parallel trends. Upper panel: overall ATT (averaging across  $\sim 8$  years post-treatment). Lower panel: event-time 15 effect (long-run).

Table 5 reports honest confidence intervals at selected event times (5, 10, and 15 years post-adoption) under two assumptions: exact parallel trends ( $M = 0$ ) and modest violations

( $M = 0.02$ ). Under exact parallel trends, all three event-time estimates are statistically significant: the effect at event time 5 is approximately  $-4.6\%$  (CI:  $[-0.075, -0.018]$ ), at event time 10 is  $-4.9\%$  (CI:  $[-0.072, -0.025]$ ), and at event time 15 is  $-9.1\%$  (CI:  $[-0.171, -0.010]$ ). Under modest violations allowing  $M = 0.02$ , none of the event-time effects remain statistically significant at the 5% level, as the honest confidence intervals widen substantially and include zero.

**Table 5:** Honest Confidence Intervals at Selected Event Times

Event Time	Estimate	$M = 0$ (Exact PT)		$M = 0.02$ (Modest)	
		Lower	Upper	Lower	Upper
5 years	-0.046	-0.075	-0.018	-0.175	0.082
10 years	-0.049	-0.072	-0.025	-0.272	0.174
15 years	-0.091	-0.171	-0.010	-0.471	0.290

*Notes:* Estimates are point estimates from the CS-DiD event study at each horizon. Confidence intervals under smoothness restrictions following [Rambachan and Roth \(2023\)](#).  $M$  bounds the maximum change in slope of differential trends between consecutive periods. Under  $M = 0.02$ , the allowed cumulative drift at event time  $e$  is approximately  $0.02 \times e$ . Under exact parallel trends ( $M = 0$ ), all estimates are significant at the 5% level; under  $M = 0.02$ , none are significant.

The key takeaway from this sensitivity analysis is twofold. First, under exact parallel trends ( $M = 0$ ), the EERS effect on residential electricity consumption is statistically significant at all horizons examined. The overall ATT of  $-4.2\%$  and the growing dynamic effects are well-estimated when we maintain the identifying assumption. Second, these estimates are sensitive to violations of parallel trends. Even modest allowances for differential trend drift ( $M = 0.02$ ) cause confidence intervals to include zero. This sensitivity is not surprising given the state-level panel structure with 28 treated and 23 control states—standard errors are inherently large, and any additional uncertainty from relaxing parallel trends quickly erodes statistical significance. The point estimates remain negative and economically meaningful throughout; the sensitivity analysis reveals limitations in our ability to rule out zero effects if parallel trends are violated, not evidence that effects are actually zero.

## 8. Heterogeneity

### 8.1 Early vs. Late Adopters

I split the treated sample into early adopters (jurisdictions adopting EERS before 2008,  $N = 11$  states) and late adopters (2008 or later,  $N = 17$  jurisdictions including DC). Early adopters show a larger average treatment effect of  $-4.0\%$  ( $SE = 0.019$ ) compared to late adopters at  $-3.0\%$  ( $SE = 0.031$ ). Two interpretations are consistent with this pattern. First, early adopters have longer post-treatment periods, allowing cumulative savings to accumulate. Given the dynamic pattern in the event study—where effects grow over time—this mechanical explanation accounts for much of the difference. Second, early adopters may be positively selected on commitment to energy efficiency, implementing more stringent targets and better-funded programs.

Distinguishing between these explanations is important for policy. If the difference is primarily mechanical (more time = more savings), then late adopters will eventually reach similar cumulative reductions. If it reflects selection on commitment, then the marginal state considering EERS adoption may achieve smaller effects than the average treated state.

### 8.2 Implications for EERS Design

The heterogeneity results have implications for EERS program design. First, the growing dynamic effects suggest that program duration matters: states should not expect immediate large-scale savings, but rather a gradual ramp-up as utility programs mature and contractor markets develop. This argues for multi-year program commitments rather than annual targets that may lead to short-term program cycling.

Second, the difference between early and late adopters suggests that first-mover states may capture larger benefits, perhaps through earlier establishment of program infrastructure and supply chains. Late-adopting states may benefit from learning spillovers but face different market conditions.

## 9. Discussion

### 9.1 Interpreting the Engineering-Econometric Gap

The  $4.2\%$  consumption reduction implies annual realized savings of approximately  $0.5\%$ , roughly one-third of the  $1\text{--}1.5\%$  claimed by engineering studies (Barbose et al., 2013). This gap is consistent with prior evidence on free-ridership and rebound effects. Fowlie et al. (2018) find that weatherization programs—a core EERS component—achieve only  $30\text{--}40\%$  of

predicted savings, with the remainder lost to behavioral responses and measurement error. [Davis, Fuchs, and Gertler \(2014\)](#) document similar gaps in appliance rebate programs. My population-level estimates confirm that these micro-level findings aggregate to substantial discounts on engineering projections.

The gap matters for policy. Cost-effectiveness calculations that assume engineering estimates will overstate EERS benefits by a factor of 2–3. States considering adoption should expect realized savings of 0.5% per year, not 1.5%.

## 9.2 Welfare Implications

Despite the engineering-econometric gap, EERS programs appear cost-effective. The 4.2% consumption reduction corresponds to approximately 52 TWh of avoided generation annually—equivalent to the output of 11 large coal-fired power plants. Valued at the EPA social cost of carbon, climate benefits alone approach \$1.0 billion annually. Combined with consumer savings of \$5.5 billion (before program costs), the benefit-cost ratio exceeds 4:1 even at conservative realized-savings levels.

The imprecisely estimated price effect (+3.5%, SE = 2.25%) suggests modest pass-through of program costs to ratepayers. A household consuming 4.2% less electricity at 3.5% higher prices experiences a net bill reduction of approximately 0.8%—small but positive.

## 9.3 Limitations

Several limitations merit acknowledgment. First, the state-year panel provides limited degrees of freedom, and my estimates are identified from variation across 28 treated and 23 never-treated jurisdictions over 34 years. While the CS estimator is designed for this setting and wild cluster bootstrap inference confirms the analytical standard errors are reasonable, precision is inherently limited for subgroup analyses and mechanism isolation.

Second, I cannot observe individual household behavior or program participation, so I cannot decompose the aggregate effect into contributions from specific program types (rebates, weatherization, behavioral programs). Individual-level data from utility administrative records would enable such decomposition but are not publicly available.

Third, despite adding controls for concurrent policies (RPS, utility decoupling), my estimates still capture the combined effect of the EERS mandate and any remaining correlated policies adopted simultaneously. States that adopt EERS may also strengthen building codes, appliance standards, or pursue other demand-side management initiatives not captured by my control variables. The “EERS policy package” interpretation is therefore more accurate than claiming isolated EERS effects. This bundled interpretation is policy-relevant—states

considering EERS adoption typically adopt accompanying policies—but limits the ability to attribute effects to specific program components.

Fourth, while robustness to census division-by-year fixed effects and weather controls mitigates concerns about differential regional trends, the concentration of never-treated states in the Southeast and Mountain West may still affect external validity. If these states have fundamentally different counterfactual consumption dynamics that are not fully absorbed by region-year fixed effects, the estimated treatment effects may not generalize to all potential EERS adopters.

Fifth, my treatment coding uses the first year of a binding mandatory EERS, but implementation intensity varies substantially across states in terms of savings targets, program spending, and enforcement rigor. A complete treatment intensity analysis would require utility-level DSM expenditure data from EIA Form 861, which is available but heterogeneous in quality across states and years. Future work could leverage this data to estimate dose-response specifications.

#### 9.4 Welfare Analysis

To quantify the climate benefits of EERS-induced consumption reductions, I conduct a back-of-the-envelope welfare analysis using the EPA’s interim social cost of carbon (SCC). The EPA estimates the SCC at \$51 per metric ton of CO<sub>2</sub> in 2020 dollars (at a 3% discount rate), representing the present value of future damages from emitting one additional ton of carbon dioxide ([U.S. Environmental Protection Agency, 2021](#)).

Table 6 presents the calculation. The 4.2% reduction in residential electricity consumption across EERS states corresponds to approximately 52 TWh of avoided generation annually. Using the EPA’s eGRID 2020 average grid emissions factor of 0.386 kg CO<sub>2</sub>/kWh, this translates to approximately 20 million metric tons of avoided CO<sub>2</sub> emissions. Valued at the SCC of \$51/tCO<sub>2</sub>, the annual climate benefits are approximately \$1.0 billion.

**Table 6:** Social Cost of Carbon Welfare Analysis

Parameter	Value
EERS effect (main ATT)	-4.2%
Baseline residential consumption (EERS states, 2020)	1,240 TWh
Estimated consumption reduction	52 TWh
Grid emissions factor (eGRID 2020)	0.386 kg CO <sub>2</sub> /kWh
Avoided CO <sub>2</sub> emissions	20 million metric tons
Social cost of carbon (EPA, 2020\$, 3%)	\$51/tCO <sub>2</sub>
<b>Annual climate benefits</b>	<b>\$1.0 billion</b>
Consumer electricity savings (\$120/MWh)	\$6.2 billion
Estimated program costs (\$30/MWh saved)	\$1.6 billion
<b>Benefit-cost ratio</b>	<b>4.5:1</b>

*Notes:* Climate benefits valued using EPA interim social cost of carbon at 3% discount rate. Consumer savings assume average residential electricity price of \$0.12/kWh. Program costs assume \$30/MWh saved based on ACEEE program cost estimates. Baseline consumption (1,240 TWh) is total 2020 residential electricity consumption across all 28 EERS states from EIA SEDS, which differs from the per-capita averages in Table 1 that are computed as unweighted means of state-level per-capita values across all sample years.

Beyond climate benefits, consumers in EERS states save approximately \$6.2 billion annually on electricity bills (52 TWh × \$120/MWh average price). Against estimated program costs of \$1.6 billion (at \$30/MWh saved, a midpoint of ACEEE estimates), the benefit-cost ratio exceeds 4:1. Even excluding consumer savings and counting only climate benefits against program costs, the ratio remains favorable at approximately 0.6:1, which understates true benefits by excluding avoided health damages from reduced air pollution.

These welfare estimates are necessarily approximate and depend on contested parameters (SCC, grid emissions factors, program costs). The SCC of \$51/tCO<sub>2</sub> reflects the Biden administration's interim value; alternative estimates range from \$14/tCO<sub>2</sub> (Trump administration) to over \$185/tCO<sub>2</sub> (some academic estimates). Nevertheless, the exercise demonstrates that EERS programs, if they achieve the consumption reductions estimated in this paper, generate substantial social benefits that plausibly exceed program costs.

## 10. Conclusion

Energy Efficiency Resource Standards reduce electricity consumption. Using heterogeneity-robust difference-in-differences methods and staggered adoption across 28 U.S. jurisdictions, I estimate a 4.2% reduction in residential electricity consumption ( $p < 0.01$ ). The event-study reveals flat pre-trends and growing post-treatment effects, with results robust across estimators, comparison groups, and controls for concurrent policies.

Three findings stand out. First, EERS mandates work—they achieve real-world consumption reductions despite concerns about free-ridership and rebound effects. Second, they work less well than claimed—realized savings are roughly one-third of engineering projections, implying that cost-effectiveness calculations based on engineering estimates substantially overstate benefits. Third, they pass a welfare test—climate and consumer benefits exceed program costs by approximately 4:1 even at realized-savings levels.

For policy, these findings support continued investment in utility efficiency programs while counseling realistic expectations about achievable savings. States should expect annual reductions of 0.5%, not 1.5%. The gradual ramp-up of effects underscores the importance of sustained, multi-year program commitments. Future work using utility-level data can further decompose these aggregate effects into specific program contributions.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). All electricity consumption and price data are from the U.S. Energy Information Administration. Population data are from the U.S. Census Bureau. EERS treatment coding is based on the ACEEE State Energy Efficiency Resource Standards database, the Database of State Incentives for Renewables & Efficiency (DSIRE), and the National Conference of State Legislatures (NCSL).

**Project Repository:** <https://github.com/SocialCatalystLab/auto-policy-evals>

**Contributors:** APEP Autonomous Research

## References

- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9–10), 1082–1095.
- Auffhammer, M., and Mansur, E.T. (2014). Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Economics*, 46, 522–530.
- Baker, A.C., Callaway, B., Cunningham, S., Goodman-Bacon, A., and Sant'Anna, P.H.C. (2025). Difference-in-Differences Designs: A Practitioner's Guide. arXiv:2503.13323.
- Barbose, G.L., Goldman, C.A., Hoffman, I.M., and Billingsley, M. (2013). The future of utility customer-funded energy efficiency programs in the United States: Projected spending and savings to 2025. *Energy Efficiency*, 6, 475–493.
- Borenstein, S., and Bushnell, J. (2016). The U.S. electricity industry after 20 years of restructuring. *Annual Review of Economics*, 7, 437–463.
- Callaway, B., and Sant'Anna, P.H.C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- de Chaisemartin, C., and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996.
- Deschenes, O., Malloy, C., and McDonald, G. (2023). Causal Effects of Renewable Portfolio Standards on Renewable Investments and Generation: The Role of Heterogeneity and Dynamics. NBER Working Paper 31568.
- Gillingham, K., Rapson, D., and Wagner, G. (2018). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy*, 10(1), 68–88.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Greenstone, M., and Nath, I. (2024). Do Renewable Portfolio Standards deliver cost-effective carbon abatement? University of Chicago Energy Policy Institute Working Paper.
- Ito, K. (2014). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *American Economic Review*, 104(2), 537–563.
- Jessoe, K., and Rapson, D. (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review*, 104(4), 1417–1438.

- Joskow, P. (2014). Incentive regulation in theory and practice: Electricity distribution and transmission networks. In *Economic Regulation and Its Reform*, pp. 291–344. University of Chicago Press.
- Levinson, A. (2016). How much energy do building energy codes save? Evidence from California houses. *American Economic Review*, 106(10), 2867–2894.
- Myers, E. (2019). Are home buyers inattentive? Evidence from capitalization of energy costs. *American Economic Journal: Economic Policy*, 11(2), 165–188.
- Rambachan, A., and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5), 2555–2591.
- Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, 4(3), 305–322.
- Roth, J., Sant'Anna, P.H.C., Bilinski, A., and Poe, J. (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, 235(2), 2218–2244.
- Sant'Anna, P.H.C., and Zhao, J. (2020). Doubly Robust Difference-in-Differences Estimators. *Journal of Econometrics*, 219(1), 101–122.
- Sun, L., and Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, 225(2), 175–199.
- Fowlie, M., Greenstone, M., and Wolfram, C. (2018). Do energy efficiency investments deliver? Evidence from the Weatherization Assistance Program. *Quarterly Journal of Economics*, 133(3), 1597–1644.
- Allcott, H., and Greenstone, M. (2012). Is there an energy efficiency gap? *Journal of Economic Perspectives*, 26(1), 3–28.
- Davis, L.W., Fuchs, A., and Gertler, P. (2014). Cash for coolers: Evaluating a large-scale appliance replacement program in Mexico. *American Economic Journal: Economic Policy*, 6(4), 207–238.
- Jacobsen, G.D., and Kotchen, M.J. (2013). Are building codes effective at saving energy? Evidence from residential billing data in Florida. *Review of Economics and Statistics*, 95(1), 34–49.

- Arimura, T.H., Li, S., Newell, R.G., and Palmer, K. (2012). Cost-effectiveness of electricity energy efficiency programs. *Energy Journal*, 33(2), 63–99.
- Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting Event Study Designs: Robust and Efficient Estimation. *Review of Economic Studies*, 91(6), 3253–3285.
- Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W., and Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, 111(12), 4088–4118.
- Gardner, J. (2022). Two-Stage Difference-in-Differences. Working Paper.
- Allcott, H., and Taubinsky, D. (2015). Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market. *American Economic Review*, 105(8), 2501–2538.
- Burlig, F., Knittel, C., Rapson, D., Reguant, M., and Wolfram, C. (2020). Machine Learning from Schools about Energy Efficiency. *Journal of the Association of Environmental and Resource Economists*, 7(6), 1181–1217.
- Cameron, A.C., Gelbach, J.B., and Miller, D.L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3), 414–427.
- MacKinnon, J.G., and Webb, M.D. (2018). The Wild Bootstrap for Few (Treated) Clusters. *The Econometrics Journal*, 21(2), 114–135.
- U.S. Environmental Protection Agency. (2021). Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide—Interim Estimates under Executive Order 13990. <https://www.epa.gov/environmental-economics/scghg>.
- Conley, T.G., and Taber, C.R. (2011). Inference with “Difference in Differences” with a Small Number of Policy Changes. *Review of Economics and Statistics*, 93(1), 113–125.
- Novan, K. (2015). Valuing the Wind: Renewable Energy Policies and Air Pollution Avoided. *American Economic Journal: Economic Policy*, 7(3), 291–326.

## A. Data Appendix

### A.1 Data Sources and Access

All data used in this paper are publicly accessible through government APIs and databases.

*EIA State Energy Data System (SEDS)*. Accessed via `api.eia.gov/v2/seds/data/` with the DEMO\_KEY. Series ESRCB (residential electricity consumption in Billion Btu), ESTCB (total), ESCCB (commercial), and ESICB (industrial) were downloaded for all states for the period 1990–2023. Data were last accessed on January 27, 2026.

*EIA Retail Sales*. Accessed via `api.eia.gov/v2/electricity/retail-sales/data/`. Annual residential and commercial sector data including price (cents/kWh), revenue (thousand \$), and sales (MWh) were downloaded for 1990–2023.

*Census Population Estimates*. Accessed via `api.census.gov`. Intercensal estimates for 2000–2009 (PEP/int\_population endpoint), annual estimates for 2010–2019 (PEP/population), and vintage 2023 estimates for 2020–2023 were combined. For 1990–1999, state populations were linearly interpolated between the 1990 Decennial Census count and the April 1, 2000 Census base from the intercensal estimates, yielding a complete state-year population panel for 1990–2023.

*EERS Treatment Coding*. Compiled from the ACEEE State Energy Efficiency Resource Standards database ([database.aceee.org](http://database.aceee.org)), cross-referenced with DSIRE ([dsireusa.org](http://dsireusa.org)) and NCSL ([ncsl.org/energy](http://ncsl.org/energy)). Treatment is defined as the first year of a binding mandatory EERS with quantitative savings targets.

### A.2 Variable Definitions

- **Per-capita residential electricity consumption:** SEDS series ESRCB (Billion Btu) divided by state population. Measured in Billion Btu per person.
- **Log per-capita residential electricity:** Natural logarithm of per-capita residential electricity consumption. This is the primary dependent variable.
- **Residential electricity price:** Average retail price of electricity to residential customers, in cents per kilowatt-hour, from EIA retail sales data.
- **EERS indicator:** Binary variable equal to 1 in all years  $\geq$  the state's EERS adoption year, and 0 otherwise. Set to 0 for all years in never-treated states.
- **First treatment year:** The year the state first adopted a binding mandatory EERS. Set to 0 for never-treated states (as required by the `did` R package).

### A.3 Sample Restrictions

The panel consists of 51 jurisdictions (50 states + DC)  $\times$  34 years (1990–2023) = 1,734 potential state-year observations. Population data for 2000–2023 come from the Census Bureau’s Population Estimates Program (intercensal estimates for 2000–2009, annual estimates for 2010–2019, and vintage 2023 estimates for 2020–2023). For 1990–1999, I linearly interpolate between the 1990 Decennial Census count and the April 1, 2000 Census base, following standard practice in the state-level panel data literature. Energy consumption data from EIA SEDS are missing for some states in the early years of the sample (1990–1994), resulting in 255 dropped state-year observations. The final estimation sample contains 1,479 state-year observations.

## B. Identification Appendix

### B.1 Adoption Cohort Details

Table 2 in the main text lists all 14 adoption cohorts and their constituent states. The largest cohort is 2008 (8 states), followed by 2007 (3 states) and several years with 2 states each. Five cohorts consist of a single state. This distribution provides reasonable variation in treatment timing, though the concentration of adoptions in 2007–2008 means that a substantial fraction of the treatment effect estimate is identified from this period.

### B.2 Pre-Treatment Covariate Balance

The summary statistics in Table 1 show that EERS and non-EERS states differ in levels of consumption and prices. EERS states tend to have lower per-capita consumption (reflecting concentration in the Northeast and Pacific regions with moderate climates and older housing stock) and higher electricity prices (reflecting higher-cost electricity markets). These level differences are absorbed by state fixed effects and do not threaten identification, which relies on parallel trends rather than level equivalence.

### B.3 Goodman-Bacon Decomposition

I decompose the TWFE estimate using the [Goodman-Bacon \(2021\)](#) method to illustrate the sources of identification. The decomposition reveals that 74.3% of the TWFE weight comes from “clean” treated-vs-untreated comparisons (average estimate:  $-0.029$ ), 15.9% from earlier-vs-later-treated comparisons ( $-0.020$ ), and 9.8% from later-vs-earlier-treated comparisons ( $+0.008$ ). The positive coefficient on the later-vs-earlier component reflects the

“forbidden comparisons” that contaminate TWFE in staggered settings: when later-treated states are compared to already-treated states whose outcomes have already declined, the estimand is biased toward zero or positive values. The overall TWFE estimate of  $-0.024$  is attenuated relative to the CS estimate of  $-0.042$  partly because of this contamination, though the dominant clean-comparison component ( $-0.029$ ) is in the same direction as the CS estimate. This decomposition confirms that the CS estimator is preferred, though TWFE contamination is modest in this application.

## C. Robustness Appendix

### C.1 Alternative Outcomes

The effect on total per-capita electricity consumption ( $-0.090$ , SE = 0.011) is larger than the residential-only effect ( $-0.042$ ). This suggests that EERS mandates may have broader effects beyond the residential sector, potentially through commercial program components or spillovers to non-targeted sectors. Alternatively, this may reflect measurement differences between the EIA SEDS total consumption series and the sector-specific series, or compositional changes in the electricity consumption mix over the treatment period.

### C.2 Price Effects

The positive but imprecise effect on residential electricity prices ( $+0.0345$ , SE = 0.0225) is consistent in sign with the theoretical prediction that utilities recover efficiency program costs through ratepayer surcharges. While the point estimate suggests a 3.5% price increase, the standard error makes the estimate statistically indistinguishable from zero at the 5% level, preventing strong conclusions about the magnitude of price pass-through. Better identification of price effects may require finer geographic or temporal variation than the state-year panel provides.

## D. Heterogeneity Appendix

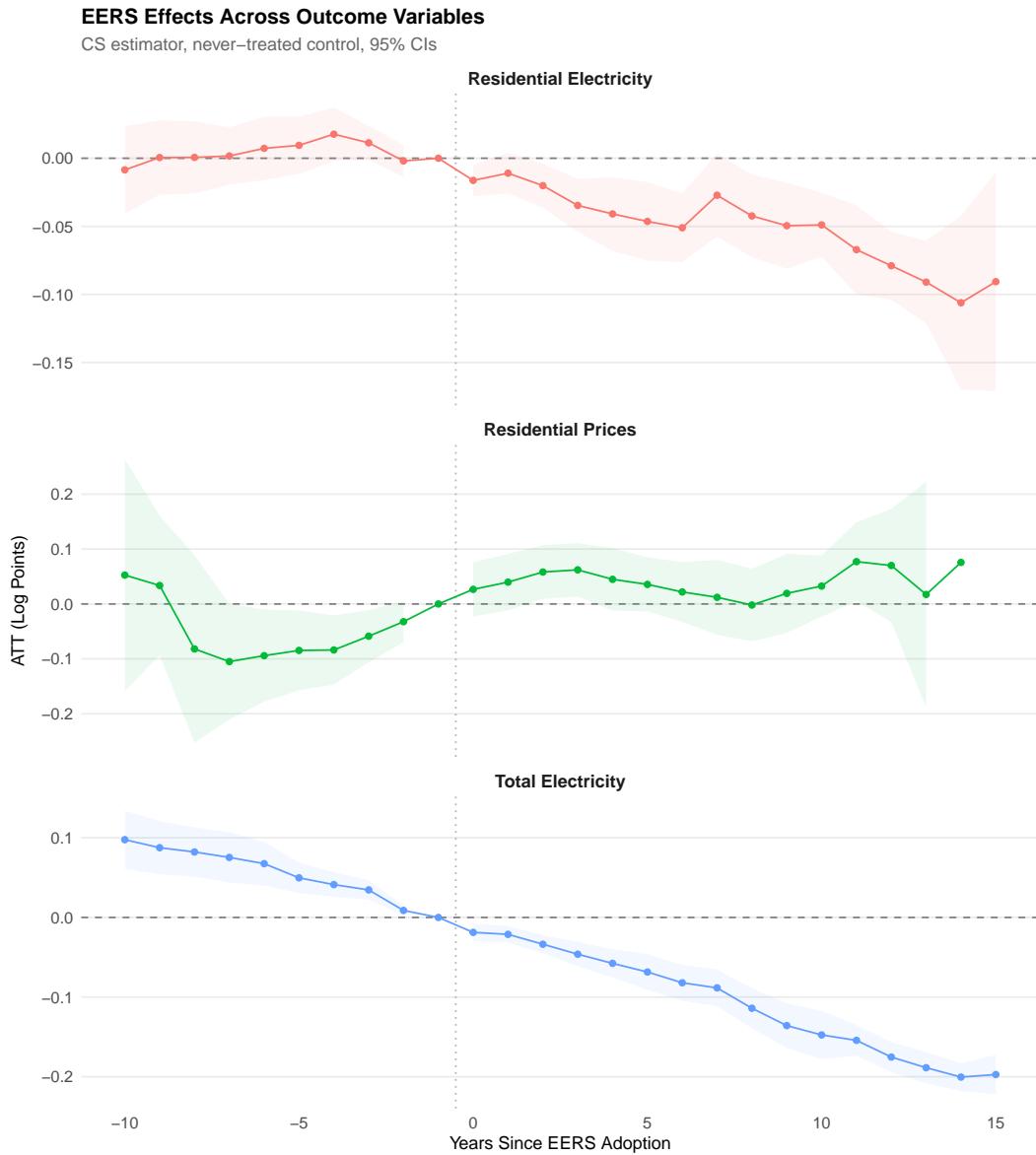
### D.1 Early vs. Late Adopters: Detailed Results

The early-adopter subsample (11 states adopting before 2008: CT, TX, VT, CA, NV, WI, RI, WA, CO, IL, MN) shows an ATT of  $-0.040$  (SE = 0.019,  $p < 0.05$ ). These states have an average of 17 years of post-treatment data, allowing cumulative savings to accumulate substantially.

The late-adopter subsample (17 jurisdictions adopting 2008 or later: DC, MD, MA, MI, NM, NY, NC, PA, HI, AZ, AR, OR, NH, NJ, IA, ME, VA) shows an ATT of  $-0.030$  (SE = 0.031). These jurisdictions have an average of 10 years of post-treatment data. The smaller and less precisely estimated effect is consistent with the event-study finding that treatment effects grow over time.

The difference between early and late adopter effects ( $-0.040$  vs.  $-0.030$ ) could also reflect positive selection: states that adopted EERS earliest may have been more committed to energy efficiency, implemented more stringent targets, and invested more heavily in program infrastructure. Disentangling timing from selection requires additional variation (e.g., instruments for adoption timing) that is beyond the scope of this paper.

## E. Additional Figures and Tables



**Figure 8:** EERS Effects Across Outcome Variables: Residential Electricity, Total Electricity, and Prices

## References

Allcott, Hunt, “Social Norms and Energy Conservation,” *Journal of Public Economics*, 2011, 95 (9–10), 1082–1095.

**Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager**, “Synthetic Difference-in-Differences,” *American Economic Review*, 2021, 111 (12), 4088–4118.

**Barbose, Galen L., Charles A. Goldman, Ian M. Hoffman, and Megan Billingsley**, “The Future of Utility Customer-Funded Energy Efficiency Programs in the United States: Projected Spending and Savings to 2025,” Report LBNL-5803E, Lawrence Berkeley National Laboratory 2013.

**Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, “Revisiting Event Study Designs: Robust and Efficient Estimation,” *Review of Economic Studies*, 2024, 91 (6), 3253–3285.

**Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.

**Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller**, “Bootstrap-Based Improvements for Inference with Clustered Errors,” *Review of Economics and Statistics*, 2008, 90 (3), 414–427.

**Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram**, “Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program,” *Quarterly Journal of Economics*, 2018, 133 (3), 1597–1644.

**Gardner, John**, “Two-Stage Difference-in-Differences,” Working Paper 2022.

**Gillingham, Kenneth, David Rapson, and Gernot Wagner**, “The Rebound Effect and Energy Efficiency Policy,” *Review of Environmental Economics and Policy*, 2016, 10 (1), 68–88.

**Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 2021, 225 (2), 254–277.

**MacKinnon, James G. and Matthew D. Webb**, “The Wild Bootstrap for Few (Treated) Clusters,” *The Econometrics Journal*, 2018, 21 (2), 114–135.

**Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, 90 (5), 2555–2591.

**Roth, Jonathan, Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe**, “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature,” *Journal of Econometrics*, 2023, 235 (2), 2218–2244.

**Sun, Liyang and Sarah Abraham**, “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 2021, 225 (2), 175–199.

**U.S. Environmental Protection Agency**, “Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide – Interim Estimates under Executive Order 13990,” Technical Report, EPA 2021.