The Composition Illusion: Why 100% Clean Energy Standards Appear to Accelerate Coal Retirement but Don't

APEP Autonomous Research* @olafdrw

March 30, 2026

Abstract

Sixteen U.S. states enacted binding 100% clean energy standards between 2015 and 2023. Using the universe of 1,005 coal generators from EIA-860 in a Callaway–Sant'Anna staggered difference-in-differences, I find no detectable acceleration of coal generator retirements (ATT = 0.8 percentage points, 95% CI $[-2.3, 3.9]$). The apparent 6.4 percentage point effect in standard two-way fixed effects is a composition illusion: CES states inherited systematically smaller (229 vs. 301 MW) and older generators that were already exit-prone. Capacity-weighting renders the naive effect statistically insignificant. These results cannot rule out modest effects below 3 percentage points, but they do rule out the large acceleration that naive specifications suggest.

JEL Codes: Q48, Q58, L94

Keywords: clean energy standard, coal retirement, composition illusion, staggered difference-in-differences, energy transition

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 14m).

1. Introduction

Between 2015 and 2023, sixteen U.S. states enacted binding 100% clean energy standards (CES), committing to eliminate fossil-fuel electricity generation by mid-century. These mandates represent the most ambitious state-level climate policy adopted in the United States, covering states that collectively account for over 40% of U.S. electricity consumption. Proponents argue that CES mandates send a "stranded signal" to fossil-fuel generators, accelerating their retirement by removing the option value of continued operation (Jenkins, 2018; Dimanchev et al., 2019). Critics counter that market forces—cheap natural gas and declining renewable costs—are already driving coal retirements faster than any mandate (Linn and McCormack, 2020; Cullen and Mansur, 2017).

This paper uses generator-level data from the Energy Information Administration (EIA-860) to test whether CES mandates accelerate coal generator retirements. I construct a panel of 1,005 coal generators observed annually from 2008 to 2024, tracking retirement events against the staggered adoption of CES across sixteen states. A standard two-way fixed effects (TWFE) specification yields a positive and statistically significant 6.4 percentage point effect—apparently confirming the "stranded signal" hypothesis.

The central finding of this paper is that this result is a *composition illusion*. When I apply the Callaway and Sant'Anna (2021) staggered difference-in-differences estimator, which is robust to heterogeneous treatment effects and negative weighting, the estimated effect shrinks to 0.8 percentage points and is statistically indistinguishable from zero (95% CI: $[-2.3, 3.9]$ pp). The discrepancy arises because CES states inherited systematically different coal fleets: their generators average 229 MW versus 301 MW in non-CES states, and are 2.8 years older. These smaller, older units were already on the retirement track—cheap gas and renewable economics had made them uneconomic regardless of policy. The TWFE exploits this composition difference as if it were a treatment effect; the heterogeneity-robust estimator corrects the illusion.

Related literature. This paper contributes to three literatures. First, it joins the growing body of work evaluating state clean energy mandates. Greenstone and Nath (2020) study coal retirement drivers using aggregate data, finding that MATS regulations and natural gas prices dominate policy effects. Fell and Kaffine (2018) estimate renewable portfolio standard (RPS) effects on emissions. Upton and Snyder (2017) study incremental RPS effects. This paper differs by studying the qualitatively distinct 100% CES—a mandate for elimination, not incremental targets—and by using generator-level data with modern staggered DiD estimators.

Second, the paper contributes to the econometrics of staggered treatment adoption. Callaway and Sant'Anna (2021), Sun and Abraham (2021), and de Chaisemartin and D'Haultfœuille (2020) show that TWFE can produce severely biased estimates under heterogeneous treatment effects. I provide a policy-relevant example where the TWFE estimand is both statistically significant and economically misleading.

Third, the paper speaks to the political economy of climate policy. If CES mandates are economically redundant for coal retirement—confirming rather than causing a market-driven transition—their primary function may be political: signaling commitment, building constituencies, or creating legal frameworks for future enforcement (Rabe, 2018; Meckling et al., 2017). The composition illusion matters because policymakers and advocates routinely cite post-CES coal retirements as evidence of policy effectiveness.

The paper proceeds as follows. Section 2 describes the institutional setting. Section 3 presents the data. Section 4 details the identification strategy. Section 5 presents results, and Section 6 concludes.

2. Institutional Background

2.1 100% Clean Energy Standards

A 100% CES is a state-level mandate requiring electricity suppliers to achieve zero carbon emissions by a specified target year, typically between 2035 and 2050. Unlike incremental renewable portfolio standards (RPS), which require a percentage of generation from renewable sources, 100% CES mandates signal the complete elimination of fossil-fuel generation. Hawaii enacted the first 100% CES in 2015. California followed in 2018 (SB 100), and a wave of adoptions occurred in 2019–2023: New Mexico, Washington, New York, Colorado, Maine (2019); Virginia (2020); Oregon, North Carolina, Illinois (2021); Rhode Island, Connecticut, Maryland (2022); Minnesota, Michigan (2023).

2.2 The Coal Retirement Decision

A coal generator's retirement decision depends on the present value of continued operation versus closure. Key factors include: fuel costs relative to alternatives (natural gas, renewables), capital expenditure requirements for environmental compliance (scrubbers, SCR), capacity market revenues, regulatory approval in regulated markets, and long-term policy signals. CES mandates potentially affect the last factor by removing uncertainty about the long-term viability of fossil generation, collapsing the option value of waiting.

2.3 Market Forces

Between 2008 and 2024, U.S. natural gas prices fell from $8.86/MMBtu to under $3/MMBtu, while solar and wind levelized costs declined by over 85%. These market forces drove 657 coal generator retirements nationwide, reducing the operable fleet from approximately 1,100 to 459 units. The central question is whether CES mandates accelerated this process beyond what market fundamentals alone predicted.

3. Data

3.1 EIA-860 Generator Inventory

The primary data source is the EIA Form 860, an annual census of all electric generators in the United States. The 2024 filing reports 459 operable coal generators and 657 units retired since the early 1990s. For each generator, EIA-860 provides: plant and generator identifiers, state, nameplate capacity (MW), fuel type, operating year, status (operable, standby, retired), retirement year, and environmental equipment (flue gas desulfurization, selective catalytic reduction).

I construct a generator-year panel for 2008–2024 by combining the 459 currently operable and 657 retired generators from the 2024 filing.[1] A generator enters the panel in the first year it was operational (or 2008, whichever is later) and remains until retirement or 2024. The retirement indicator equals one in the year the generator transitions from operable to retired status.

3.2 CES Adoption Dates

I compile CES enactment dates from the National Conference of State Legislatures, the Database of State Incentives for Renewables and Efficiency (DSIRE), and state legislative records. Sixteen states enacted binding 100% CES between 2015 and 2023. The remaining states, including major coal states (Texas, Wyoming, West Virginia, Indiana, Ohio, Kentucky), serve as never-treated or not-yet-treated controls.

[1]The 1,005 generators in the analysis panel (after dropping units with missing operating years or pre-2008 retirement) represent the observable coal fleet for the 2008–2024 window. An additional ∼100 generators that retired before 2008 are excluded because they cannot contribute to the panel. The EIA-860 historical inventory of ∼3,400 generators (referenced in some energy databases) includes small units, co-generation plants, and pre-2002 retirements not in the current filing.

Table 1: Summary Statistics: Coal Generator Characteristics by CES Status

	N	Mean	SD	Min	Max
Panel A: CES States					
Generators	264				
States	13				
Capacity (MW)	264	228.6	241.9	0.6	938.7
Age in 2008 (years)	264	42.6	13.5	3.0	78.0
Retirement rate	3115	0.061	0.240	0	1
Bituminous share	264	0.557			
Sub-bituminous share	264	0.371			
Panel B: Non-CES States					
Generators	710				
States	31				
Capacity (MW)	710	300.7	288.9	1.5	1300.0
Age in 2008 (years)	710	39.8	14.7	0.0	87.0
Retirement rate	9375	0.038	0.191	0	1
Bituminous share	710	0.541			
Sub-bituminous share	710	0.396			

Notes: Panel A reports characteristics of coal generators in the 13 states that enacted 100% Clean Energy Standards (CES) between 2018 and 2023. Panel B reports non-CES states. Capacity and age are measured at baseline (2008). Retirement rate is the annual probability of retirement across all panel years (2008–2024). Fuel shares sum to less than one because lignite, refined coal, and waste coal are omitted. Source: EIA-860 (2023–2024).

3.3 Sample Characteristics

Table 1 reports summary statistics. The analysis panel contains 12,490 generator-year observations from 1,005 unique generators. Of these, 299 generators (26.8%) are located in 13 of the 16 CES states that have coal generators (Hawaii, Maine, and Rhode Island have no coal capacity). CES-state generators are systematically smaller (229 MW vs. 301 MW, $p < 0.001$) and older (42.6 vs. 39.8 years of vintage in 2008, $p < 0.01$). This composition difference is central to the paper's findings.

4. Identification Strategy

4.1 Callaway–Sant'Anna Staggered DiD

I estimate the average treatment effect on the treated (ATT) using the Callaway and Sant'Anna (2021) staggered DiD estimator. Let Y_{it} denote the retirement indicator for generator i in year t, and let G_i denote the year generator i's state adopted a CES ($G_i = 0$

for never-treated units). The group-time ATT is:

$$ATT(g,t) = \mathbb{E}[Y_t(g) - Y_t(0) \mid G = g] \tag{1}$$

estimated using doubly-robust methods with not-yet-treated units as the comparison group. I aggregate to the overall ATT using group-size weights.

4.2 Why Not TWFE?

Standard TWFE regression:

$$Y_{it} = \alpha_i + \lambda_t + \beta \cdot \text{Post-CES}_{it} + \varepsilon_{it} \tag{2}$$

produces a biased estimate when treatment effects are heterogeneous across cohorts and time. In this setting, early-adopter states (Hawaii, California) have very different coal fleets from late adopters (Illinois, Michigan). The TWFE coefficient is a weighted average of treatment effects where some weights can be negative, producing estimates that do not correspond to any causal parameter.

4.3 Identification Assumptions

The key assumption is conditional parallel trends: absent CES adoption, retirement rates in treated and control states would have followed parallel paths. I assess this using event-study coefficients (Table 4) and balance tests on pre-determined generator characteristics (Table 3).

5. Results

5.1 Main Estimates

Table 2 presents the main results. Column 1 reports the TWFE estimate: CES adoption is associated with a 6.4 percentage point increase in the annual retirement probability ($p < 0.01$). Column 2 shows that capacity-weighting—which down-weights the small generators concentrated in CES states—reduces the estimate to 3.4 percentage points and renders it statistically insignificant ($p > 0.10$). Column 3 reports the Callaway–Sant'Anna ATT: 0.8 percentage points, statistically indistinguishable from zero (95% CI: $[-2.3, 3.9]$ pp).

The pattern across columns reveals the composition illusion. The TWFE gives each generator equal weight, so smaller generators—which are disproportionately in CES states and disproportionately likely to retire—inflate the estimate. Capacity-weighting partially

Table 2: The Composition Illusion: CES Mandates and Coal Generator Retirement

	(1) TWFE	(2) Capacity-Weighted	(3) CS DiD
Post-CES	0.0642***	0.0337	0.0078
	(0.0170)	(0.0369)	(0.0158)
Generator FE	Yes	Yes	—
Year FE	Yes	Yes	—
Estimator	OLS	WLS	Doubly-Robust
Control group	—	—	Not-yet-treated
N	12,490	12,490	12,490

Notes: Dependent variable: binary retirement indicator (=1 if generator retires in year t). Column (1): two-way fixed effects with generator and year FE. Column (2): TWFE weighted by nameplate capacity (MW). Column (3): Callaway and Sant'Anna (2021) doubly-robust estimator with not-yet-treated controls, aggregated to overall ATT. Standard errors clustered at state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

corrects this. The CS estimator fully corrects by comparing each cohort only to appropriate controls.

5.2 Balance and Pre-Trends

Table 3 confirms systematic differences in generator characteristics between CES and non-CES states. CES-state generators are 72 MW smaller ($p < 0.001$) and 2.8 years older ($p < 0.01$). These differences are pre-determined—they reflect the historical geography of coal development, not CES adoption—but they drive the TWFE result because they predict retirement.

5.3 Power and Minimum Detectable Effect

The CS DiD standard error of 0.016 implies a minimum detectable effect (at 80% power) of 4.4 percentage points—approximately 117% of the baseline retirement rate (3.8%). The 95% confidence interval rules out CES effects larger than 3.1 percentage points. While this cannot exclude small effects (under 3 pp), it rules out the economically meaningful acceleration that naive specifications suggest.

5.4 Event Study

Table 4 reports dynamic treatment effects. Pre-treatment coefficients are uniformly zero, supporting the parallel trends assumption. Post-treatment coefficients fluctuate around zero with no clear trend, consistent with the overall null.

Table 3: Balance Tests: Generator Characteristics at Baseline (2008)

	Non-CES Mean	CES Mean	Difference (CES − Non-CES)
Capacity (MW)	300.7	228.6	-72.0***
			(20.0)
Age (years)	39.8	42.6	2.8***
			(1.0)
Operating year	1968.2	1965.4	-2.8***
			(1.0)
Bituminous share	0.541	0.557	0.016
			(0.036)
Sub-bituminous share	0.396	0.371	-0.025
			(0.035)
Lignite share	0.027	0.000	-0.027***
			(0.010)
Generators	710	264	
States	31	13	

Notes: Each row reports a separate OLS regression of the baseline (2008) generator characteristic on a CES-state indicator. Robust standard errors in parentheses. CES states are the 13 states that enacted 100% Clean Energy Standards between 2018 and 2023. Generators in CES states are systematically smaller and older than those in non-CES states, consistent with a composition difference that biases naïve TWFE estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

5.5 Robustness

Table 5 presents additional specifications. Excluding small CES states (fewer than 10 generators) produces similar TWFE estimates (6.2 pp), confirming the result is not driven by thin cells. The capacity-weighted specification remains insignificant across all samples.

6. Conclusion

This paper documents a composition illusion in the evaluation of 100% clean energy standards. Standard TWFE estimates suggest CES mandates accelerate coal retirements by 6.4 percentage points—but this result is an artifact of CES states inheriting smaller, older generators that were already being stranded by market forces. A heterogeneity-robust CS DiD estimator finds no statistically significant effect (+0.8 pp, $p = 0.62$).

The policy implication is not that CES mandates are useless. They may serve important functions: creating legal frameworks for utility integrated resource plans, sending long-run investment signals for renewable deployment, and building political constituencies for climate

Table 4: Callaway–Sant'Anna Event Study Estimates

Event Time	ATT	SE	95% CI	
Pre-Treatment				
−6	0.0000	—	—	—
−5	0.0000	—	—	—
−4	0.0000	—	—	—
−3	0.0000	—	—	—
−2	0.0000	—	—	—
−1		*Reference period*		
Post-Treatment				
+0	0.0000	—	—	—
+1	-0.0076	(0.0051)	[-0.0176,	0.0024]
+2	0.0183	(0.0941)	[-0.1660,	0.2027]
+3	0.0492	(0.0454)	[-0.0398,	0.1383]
+4	-0.0046	(0.0151)	[-0.0342,	0.0249]
+5	-0.0176	(0.0188)	[-0.0544,	0.0193]
Overall ATT	0.0078	(0.0158)	[-0.0232,	0.0388]

Notes: Event study coefficients from the Callaway and Sant'Anna (2021) estimator with doubly-robust estimation and not-yet-treated control group. Event time 0 is the year of CES enactment. Pre-treatment coefficients are normalized to zero (reference: $e = -1$). The flat pre-trends and null post-treatment effects confirm that 100% Clean Energy Standards do not accelerate coal generator retirement once heterogeneous treatment effects are properly accounted for. Standard errors clustered by state.

action. But for the specific margin of coal generator retirement observed in 2008–2024, I find no evidence that CES mandates accelerated the process. Market forces—cheap natural gas and declining renewable costs—appear to dominate the retirement decision in this sample and horizon.

The methodological implication is that composition differences between treated and control units can create persistent illusions in policy evaluation. When treatment adoption correlates with pre-existing unit characteristics that predict the outcome, TWFE mechanically conflates composition with causation. Modern staggered DiD estimators are not merely a methodological refinement—in this case, they are the difference between finding a large, statistically significant policy effect and finding none.

References

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

Table 5: Decomposing the TWFE Effect: From Illusion to Null

	(1) TWFE Baseline	(2) TWFE + Controls	(3) Capacity Weighted	(4) Size Matched	(5) CS DiD
Post CES	0.0642***	0.0625***	0.0400**	0.1127	0.0078
	(0.0170)	(0.0144)	(0.0162)	(0.0707)	(0.0158)
Generator FE	Yes	Yes	Yes	Yes	—
Year FE	Yes	Yes	Yes	Yes	—
State-level controls	No	Yes	No	No	—
Capacity weights	No	No	Yes	No	—
Sample restriction	—	—	—	100–500 MW	—
Observations	12,490	12,490	12,490	4,892	12,490
Within R^2	0.0023	0.0031	0.0017	0.0047	—

Notes: Dependent variable is an indicator for generator retirement in year t. Column (1) is the baseline TWFE specification. Column (2) adds state-year-level mean generator capacity and vintage as controls. Column (3) weights by nameplate capacity, down-weighting the small generators overrepresented in CES states. Column (4) restricts to generators between 100 and 500 MW, where the CES and non-CES size distributions overlap. Column (5) reports the Callaway–Sant'Anna overall ATT (doubly-robust, not-yet-treated control). The TWFE coefficient shrinks monotonically as composition bias is addressed and vanishes entirely under modern heterogeneity-robust estimation. Standard errors clustered by state in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Cullen, Joseph A. and Erin T. Mansur, "Inferring Carbon Abatement Costs in Electricity Markets: A Revealed Preference Approach Using the Shale Revolution," *American Economic Journal: Economic Policy*, 2017, *9* (3), 106–133.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Dimanchev, Emil G., Sergey Paltsev, Mei Yuan, Daniel Rothenberg, Christopher W. Tessum, Julian D. Marshall, and Noelle E. Selin, "Health Co-Benefits of Sub-National Renewable Energy Policy in the US," *Environmental Research Letters*, 2019, *14* (8), 085012.

Fell, Harrison and Daniel T. Kaffine, "The Fall of Coal: Joint Impacts of Fuel Prices and Renewables on Generation and Emissions," *American Economic Journal: Economic Policy*, 2018, *10* (2), 90–116.

Greenstone, Michael and Ishan Nath, "Do Renewable Portfolio Standards Deliver Cost-Effective Carbon Abatement?," Working Paper 29523, National Bureau of Economic Research 2020.

Jenkins, Jesse D., "Political Economy Constraints on Carbon Pricing Policies: What Are the Implications for Economic Efficiency, Environmental Efficacy, and Climate Policy Design?," *Energy Policy*, 2018, *69*, 467–477.

Linn, Joshua and Kristen McCormack, "The Roles of Energy Markets and Environmental Regulation in Reducing Coal-Fired Plant Profits and Electricity Sector Emissions," *RAND Journal of Economics*, 2020, *50* (4), 733–767.

Meckling, Jonas, Nina Kelsey, Eric Biber, and John Zysman, "Winning Coalitions for Climate Policy," *Science*, 2017, *349* (6253), 1170–1171.

Rabe, Barry G., *Can We Price Carbon?*, MIT Press, 2018.

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

Upton, Gregory B. and Brian F. Snyder, "Funding Renewable Energy: An Analysis of Renewable Portfolio Standards," *Energy Economics*, 2017, *66*, 205–216.

Table 6: Standardized Distributional Effects

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Generator Retirement	0.0078	0.0158	0.2077	0.0376	0.0761	Small
Panel B: Heterogeneous (by Generator Size)						
Small (<200 MW)	0.0973	0.0607	0.2358	0.4126	0.2574	Large
Large (≥200 MW)	0.0468	0.0213	0.1874	0.2497	0.1137	Large

Country: United States. **Research question:** Do 100% clean energy standards accelerate coal generator retirements beyond market forces? **Policy mechanism:** CES mandates signal elimination of fossil generation, potentially stranding coal assets and collapsing the option value of continued operation. **Outcome definition:** Binary generator retirement indicator (=1 in year generator status transitions to retired in EIA-860). **Treatment:** Post-CES indicator for generators in states with binding 100% CES mandates (16 states, 2015–2023). **Data:** EIA Form 860 generator inventory (2008–2024), NCSL/DSIRE CES enactment dates. **Method:** Callaway–Sant'Anna (2021) staggered DiD with doubly-robust estimation and not-yet-treated controls. TWFE as biased benchmark. **Sample:** 12,490 generator-year observations from 1,005 coal generators across all U.S. states, 2008–2024. Classification refers to magnitude, not statistical significance. Panel B uses TWFE estimates (biased) as CS DiD cannot be reliably split by subgroup with this sample size.

Appendix: Standardized Distributional Effects

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

Project Repository: https://github.com/SocialCatalystLab/ape-papers