

California's cap-and-trade program and emission leakage in the Western Interconnection: comparing econometric and partial equilibrium model estimates

Chiara Lo Prete

Ashish Tyagi

Qingyu Xu*

December 18, 2022

*Lo Prete: John and Willie Leone Family Department of Energy and Mineral Engineering, The Pennsylvania State University, University Park, PA (chiaraloprete@psu.edu); Tyagi: Frankfurt School of Finance & Management, Frankfurt am Main (a.tyagi@fs.de); Xu: Tsinghua University (xuqingyu0610@gmail.com). We thank the editor and three anonymous referees for insightful comments and suggestions, which have greatly improved the quality of our paper. We are grateful for feedback from seminar participants at Penn State, Harvard University, Lehigh University, Sandia National Laboratories and Sant'Anna School of Advanced Studies, as well as conference participants at the 2016 USAEE/IAEE North American Conference, 2016 Annual Meetings of the Southern Economic Association, 2017 USAEE/IAEE North American Conference, 2017 International Workshop in Empirical Methods on Energy Economics, 2017 Rutgers Annual Eastern Conference, 2018 Atlantic Workshop on Energy and Environmental Economics, 2018 Mannheim Energy Conference, 2019 Annual Conference of the European Association of Environmental and Resource Economists, and 2019 USAEE/IAEE North American Conference. All errors are our own. Lo Prete gratefully acknowledges the support of The Pennsylvania State University, the Alfred P. Sloan Foundation (grant 2018-11301) and the National Science Foundation (grant 1832290). Tyagi thanks the Robert Bosch Foundation and the National Science Foundation (grant 1832290) for financial support. Xu gratefully acknowledges funding from The Johns Hopkins University and the Department of Energy (SETO EE0008215). The authors have no affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript.

Abstract

We compare the effects of California’s AB 32 cap-and-trade program on leakage in the electricity sector using two methods: a simulation-based partial equilibrium model that accounts for details of policy implementation and is parameterized using market data; and an econometric model applying a quasi-experimental design with matching and a robust inference method that does not require parallel trends to hold exactly. Based on the estimated shifts in electricity generation, we infer CO₂ emission leakage predictions in 2013 and 2016. The comparison allows us to identify critical assumptions driving the simulation results, and to benchmark the empirical results in a complex policy setting where threats to identification undermine attempts at statistical inference. Over the study period, we find significant leakage potential *ex ante* and empirical evidence that is consistent with some resource shuffling *ex post*. Limiting the ability of electricity importers to claim the default emission factor may reduce leakage risks.

1 Introduction

California’s Assembly Bill 32 (AB 32) established a statewide target of reducing GHG emissions to their 1990 levels by 2020. To achieve the expected emission reductions to meet the 2020 limit, the California Air Resources Board (CARB) outlined a mix of recommended actions combining direct regulations, market-based approaches and incentives in the initial Scoping Plan (California Air Resources Board, 2008) and First Update (California Air Resources Board, 2014). A key element of CARB’s emission reduction strategy was the development of a cap-and-trade program covering about 85% of the state’s emissions from industrial facilities, electricity generators and importers, and transportation fuel suppliers. Within the capped sectors, emission reductions would be accomplished through price incentives created by allowance prices, as well as direct regulations. CARB also recommended reduction measures for the uncapped sectors (e.g., agriculture, recycling and waste). The rationale for this combination of approaches was that complementary measures are needed to overcome market barriers that would persist, if the cap-and-trade system were the only policy employed to implement AB 32 (California Air Resources Board, 2008).¹

A central issue in the implementation of cap-and-trade programs is represented by the choice of the point of regulation. Given its reliance on imports to satisfy electricity consumption,² California opted for a first deliverer approach, whereby entities that own electricity at the first point of delivery in the state represent the point of regulation: in-state generators must monitor and report their emissions following a source-based paradigm, while electricity importers are responsible for emissions associated with in-state sales. The introduction of a border adjustment mechanism for the electricity sector was intended to mitigate concerns of leakage, defined as the shift in production and associated emissions from the region where climate regulations apply to surrounding unregulated jurisdictions (Stavins et al., 2010). Simulation-based studies quantifying the impacts of the prospective cap-and-trade scheme concluded that resource shuffling may enable substantial leakage (Bushnell et al., 2008b; Fowlie, 2009; Chen et al., 2011; Bushnell et al., 2014). For example, electricity contracts may be rearranged so that production from low emission sources serving out-of-state consumption (or load) is directed to California, while production from higher emission sources is assigned to serve out-of-state load (Burtraw et al., 2018). Contract shuffling would result in apparent emission reductions due to changes in the composition of imports to California,

¹Appendix A describes the policy context in greater detail.

²In 2016, California imported about a third of its total electricity consumption from out of state (California Energy Commission, 2017).

although emissions in exporting regions are unchanged or even increase. In recent years, the decrease in GHG emissions from the electric power sector in California has been attributed primarily to measured reductions in emissions from imports (California Air Resources Board, 2020b). This underscores the importance of assessing whether leakage has occurred and considering potential policy modifications to mitigate its impacts.

In this paper, we seek to identify CO₂ emission leakage in the electricity sector from California’s AB 32 cap-and-trade program in the first four years of policy implementation. We estimate shifts in electricity generation at baseload power plants in the Western Interconnection³ based on two models: a simulation-based partial equilibrium model of the electricity sector that includes salient features of the observed cap-and-trade program and is parameterized using market data in 2013-2016; and an econometric model applying a quasi-experimental design with matching and a robust inference method that does not require the parallel trends assumption to hold exactly. Based on the estimated shifts in electricity generation, we infer CO₂ emission leakage predictions in 2013 and 2016. We then compare the *ex ante* expected impacts of the policy to the *ex post* realized impacts. In 2013, the simulation model predicts a leakage rate of 96%, while the econometric model cannot rule out rates as low as 41% or as high as 218%, under credible restrictions on the sign of the bias of the post period event study coefficients and monotonicity of the difference in trends between the treated and the control groups. In 2016, the *ex ante* leakage rate is 83%, while the econometric estimates are in the 17%-177% range under the same restrictions. Earlier studies compared *ex ante* estimates of the effect of regional environmental policies with *ex post* empirical results (Carlson et al., 2000; Ellerman et al., 2000; Carbone et al., 2020). To our knowledge, we are the first to benchmark emission leakage predictions based on estimates from a quasi-experimental econometric model against the results of a partial equilibrium model designed to study the effects of a cap-and-trade program.

The potential for emission leakage in the electricity sector under regional climate policies has been analyzed using numerical models (Fowle, 2009; Chen et al., 2011; Bushnell et al., 2014; Caron et al., 2015; Carbone and Rivers, 2017), while empirical analyses of leakage are less common (Aichele and Felbermayr, 2013, 2015; Zhou and Huang, 2021). Appendix B reviews this literature in detail. Our paper is most closely related to Xu and Hobbs (2021) and

³California is part of the Western Interconnection, a synchronous electric grid that encompasses all or parts of 14 Western states in the U.S., the Canadian provinces of Alberta and British Columbia, and Northern Baja California in Mexico. Since reliability within the area is overseen by the Western Electric Coordinating Council, this synchronous grid is commonly referred to as WECC. Figure E1 in the Appendix presents the U.S. part of WECC.

Fell and Maniloff (2018), but differs from these earlier contributions in several important ways. Xu and Hobbs (2021) examine the potential cost and emission impacts of alternate border carbon adjustment (BCA) schemes under California’s AB 32 cap-and-trade system using the Johns Hopkins Stochastic Multistage Integrated Network Expansion (JHSMINE) model. They find that dynamically setting a facility-neutral deemed rate based on marginal units outside of California would provide efficiency gains, relative to facility-based schemes like the one that is currently implemented. The version of JHSMINE in this paper enables more direct comparisons with the econometric results than those allowed by previous simulation models making assumptions that do not align with actual policy implementation or using different data for parameterization. Specifically, we revise the model formulation in Xu and Hobbs (2021) to accommodate the observed cap-and-trade regime, and parameterize the model using actual market data in 2013-2016. Fell and Maniloff (2018) use a differences-in-differences model to estimate how the Regional Greenhouse Gas Initiative (RGGI) affected electricity generation in the regulated region and nearby states, and examine changes in electricity transmission flows into the RGGI region after policy implementation. They find that the cap-and-trade program led to a reduction in coal-fired generation in the regulated region and an increase in cleaner NGCC generation in the unregulated region, resulting in lower total emissions across regions and an implied leakage rate of approximately 50%. Relative to their paper, we strengthen the identification strategy using coarsened exact matching, adopt robust methods to conduct statistical inference under potential violations of parallel trends, and benchmark leakage predictions based on the empirical estimates against simulation-based results. Unlike Fell and Maniloff (2018), we are unable to estimate a model of inter-regional electricity transmission due to the lack of historical data on hourly power flows between balancing authorities in WECC over the period of our analysis.

The remainder of the paper is organized as follows. Section 2 presents the econometric model specification, data and results. Section 3 describes the simulation model and results. Section 4 compares the econometric and simulation results, and Section 5 provides concluding remarks.

2 Econometric model

We econometrically estimate shifts in electricity generation in the Western Interconnection after the introduction of California’s cap-and-trade program using a differences-in-differences framework with matching. Based on seasonal effects estimated for specific regions/technology

types/times of day and their confidence intervals, we infer CO₂ emission leakage predictions in 2013 and 2016. This section describes our empirical strategy and results.

2.1 Treated and control designation

The primary leakage mechanism would consist in replacing power generation in the regulated region (California) with generation in the unregulated regions (“leakers” in the Western Interconnection). Since contract shuffling and policy-induced changes in the dispatch order reallocate production among California plants and *all* out-of-state plants in WECC, every plant in the Western Interconnection may be a potential leaker. WECC’s footprint includes the Canadian provinces of Alberta and British Columbia, the northern portion of Baja California, Mexico, and all or portions of 14 Western U.S. states. Since data availability is limited for the two Canadian provinces and Northern Baja California, we exclude them from our analysis, and only consider leakers in the U.S. part of the Western Interconnection. Thus, the treated set consists of NGCC and coal-fired plants in the U.S. part of the Western Interconnection (California or leakers), while the control set consists of plants of the same technology type in five NERC regions in the Eastern and Texas Interconnections (FRCC, MRO-US, SERC, SPP and TRE) (Figure E1 in the Appendix).⁴

Power plants fall under the operational control of a balancing authority (BA), which is responsible for dispatching generation units and maintaining consumption-interchange-generation balance within a region of the electric grid (National Electric Reliability Council, 2022); WECC balancing authorities in the U.S. are presented in Figure E2 in the Appendix. Following the classification in WECC’s production cost model (WECC Staff, 2015), leaker balancing authorities are divided into three regions of contiguous connected electrical components. The Northwest (NW) region includes AVA, BPAT, NEVP, PACE, PACW, PGE and PSEI, as well as Utah plants in the LDWP footprint and Nevada plants in the CAISO footprint. Loosely speaking, this region corresponds to the Pacific Northwest, Nevada and Utah. The Southwest (SW) include AZPS, HGMA, SRP, TEPC, WALC, and plants within the CAISO footprint but located in Arizona. Loosely speaking, this region corresponds to Arizona. The Rest of WECC (RoW) includes all other balancing authorities in WECC, i.e. EPE, IPCO, NWMT, PNM, PSCO, WAUW, and WACM. Table F1 in the Appendix presents summary statistics for NGCC treated and control plants, Table F2 provides statistics for

⁴We do not include power plants in RFC and NPCC as controls because these NERC regions largely overlap with the footprint of RGGI, the cap-and-trade program for CO₂ emissions from power generation in the Northeastern and mid-Atlantic United States.

coal-fired treated and control plants, and Table F3 reports summary statistics for nuclear, hydro and renewable generation.

2.2 Matching

The simplest estimates of the treatment effects of interest can be obtained using an unconditional differences-in-differences (DID) estimator that measures the effect of California’s cap-and-trade program on average plant utilization. This approach has some drawbacks. First, constructing counterfactual outcomes using observations on plants from another interconnection poses a challenge, because these plants may have inherently different characteristics from the treated plants. Further, plants with similar average utilization over a period of time may be operated very differently.⁵ Constructing counterfactual estimates based on control plants that have similar average utilization over blocks of hours to the treated units allows us to identify pairs that, before policy implementation, held a similar position in the dispatch order of their respective balancing authority. The treatment effect of interest could then be obtained estimating a DID model in which the impact of other changes and shocks affecting plant utilization is captured by the covariates.

In order to mitigate potential bias in the unconditional DID estimates, we improve balance between treated and control groups by matching on pre treatment hourly variables. The basic idea of matching is to find untreated units that are similar to the treated ones in terms of variables that influence the outcome of interest (i.e., so called “matching variables”), except for treatment status. While earlier empirical work in energy and environmental economics relied on parametric and semi-parametric matching methods (Fowlie et al., 2012), we explore the use of coarsened exact matching (CEM) (Blackwell et al., 2009) to improve balance between treated and control observations before applying a differences-in-differences estimator. The first step of the CEM procedure is to identify observable variables to match members of the treatment and control populations. Each matching variable is coarsened to a discrete number of bins using a binning strategy, and each combination of bins across matching variables represents a stratum (or archetype). Based on their values for the matching variables, units in the sample are assigned to one stratum, which is used to exactly match members of the two

⁵To illustrate, consider two periods (1 and 2) and two plants (A in WECC and B in one of the control regions). Suppose that plant A does not produce electricity in period 1 and is operated at 80% of its capacity in period 2, while plant B is operated at 40% of its capacity in both periods. The two plants have the same average utilization, but serve a different role in their respective grid. As a result, A and B do not represent a suitable pair of treated-control observations.

populations. Only units with the same stratum are matched.⁶ To correct for the imbalance between the number of treated and control units in each stratum, matched control units receive a weight that normalizes the stratum to the distribution within the treatment group (Iacus et al., 2008). Unlike approximate matching methods (e.g., based on the propensity score), CEM bounds the maximum imbalance between treated and control groups by choosing the coarsening *ex ante*: as the bins for the matching variables become narrower, the bound on the maximum imbalance on the moments of the variables gets tighter. Recent applications are presented in Simcoe and Toffel (2014) and Ek and Miliute-Plepiene (2018).

The objective of our matching procedure is to achieve statistically indistinguishable means between treated and control plants across a set of exogenous covariates that are highly correlated with the outcomes of interest (i.e., daily, hourly or time-of-day measures of plant utilization). We use average capacity factors over four blocks of hours within the day, averaged over 2009-2010, as matching variables.⁷ Our matching strategy proceeds as follows. First, we choose 2009 and 2010 as pre treatment period, since 2011 was a wet hydrological year in which NGCC plants ran at much lower capacity factors than usual (Nyberg, 2018), and 2012 was the year before compliance obligations began.⁸ For each power plant, we average hourly capacity factors over four blocks of hours (morning, afternoon, evening, and night).⁹ Next, for each combination of technology type (coal-fired and NGCC) and region (treated and controls), we create a histogram of capacity factors by block of hours (averaged over 2009-2010) with 10 bins of equal width. The empirical distributions of capacity factors by technology type, treated/control region and block of hours are presented in Figure 1.¹⁰ We then define four matching bins corresponding to different levels of plant utilization in each block: matching bin

⁶To illustrate, consider two matching variables. The first variable is divided into 3 bins (A, B, and C), while the second variable is divided into 2 bins (D and E). The resulting strata are AD, BD, CD, AE, BE and CE. Control units in any stratum are matched to treated units in the same stratum.

⁷Capacity factors are a percentage measure of plant utilization over a period of time and represent the dependent variable of our DID model, as discussed in Section 2.3. Historic capacity factors over blocks of hours in 2009-2010 are expected to be a good predictor of capacity utilization in the following years: hence, we choose them as matching variables. We also experiment with matching on other observable factors that could be correlated with plant utilization, such as heat rate (a measure of efficiency) and age. Since including these factors as additional matching variables reduces the control sample size without substantially improving the quality of our matches, we do not take this approach.

⁸We also remove from the matching dataset outliers (i.e., plants for which generation from CEMS is greater than generating capacity from EIA) and plants that were operating for less than 3 years over the period of our study.

⁹Following NREL’s classification (National Renewable Energy Laboratory, 2019), the morning block is from 6am to 1pm, the afternoon block is from 1pm to 5pm, the evening block is from 5pm to 10pm, and the night block is from 10pm to 6am.

¹⁰Figure E3 in the Appendix presents additional detail by WECC region.

1 (low utilization) includes the first three histogram bins in the lower tail of the distribution; matching bin 2 (medium-low utilization) includes histogram bins 4 and 5; matching bin 3 (medium-high utilization) includes histogram bins 6 and 7; matching bin 4 (high utilization) includes the last three histogram bins in the upper tail of the distribution. In order to improve the quality of our matches, we create smaller matching bins in the middle of the distribution. Further, we coarsen each matching variable according to cut points given by the upper and lower limits of the matching bins for the treated plants. The final step is to perform exact matching on these bins and discard observations from bins that do not contain both treated and control observations.¹¹

2.3 Differences-in-differences

After pruning observations that have no close matches on pre treatment variables in both treated and control groups, we econometrically estimate changes in power plant utilization in the Western Interconnection using the following DID model specification:

$$Y_{jt} = \alpha_C TREAT_{jt}^C + \sum_L \alpha_L TREAT_{jt}^L + \mathbf{X}_{jt}' \underline{\beta} + \gamma_j + \gamma_y + \gamma_{dw} + \gamma_{sm} + \epsilon_{jt} \quad (1)$$

where j indexes a plant-technology, t indicates a period, L denotes a leaker region, and y , dw and sm stand for year, day-of-week and state by month-of-year respectively. We focus on two baseload technology types that are most likely affected by the policy (natural gas combined cycle or NGCC plants and coal-fired plants), and run separate regressions by technology.¹² The dependent variable Y_{jt} is the capacity factor of plant-technology j in period t (hour or day), defined as the ratio of net generation over operating capacity multiplied by total number of hours in the period.¹³

¹¹We experiment with an alternate two-step strategy that matches by annual average capacity factors on blocks of hours first, and then by seasonal average capacity factors on those blocks. We also test robustness to an alternate binning strategy with five matching bins, where matching bin 1 includes the first two histogram bins in the lower tail of the distribution; matching bin 2 includes histogram bins 3 and 4; and so on. Both strategies yield results that are similar to those in the baseline, and thus we do not present them in the paper.

¹²Natural gas steam turbines represent a small fraction of generating capacity in the WECC region. Other technology types like natural gas combustion turbines and oil turbines that are used as peaker plants during high load periods are unlikely to have responded to California’s carbon policy, given the modest level of permit prices over the period of our study.

¹³Capacity factors provide a measure of capacity utilization that is independent of plant size. Based on the estimated shifts in capacity utilization, we infer CO₂ emission leakage effects. An alternate (and more direct) approach consists in estimating a model with net generation or emissions at the plant as dependent variable, and capacity as an additional covariate to control for plant size. During the period of our study, there is no significant variation in installed capacity, which is thus correlated with plant fixed effects, raising concerns of

The treatment of interest is the introduction of California’s cap-and-trade program on January 1, 2013. Subject to the identification assumptions in Section 2.5.2, the estimated treatment effects α_C and α_L measure the average effect of the cap-and-trade program on capacity factors of matched facilities in California and the leaker regions, conditional on the covariates. $TREAT_{jt}^C$ is a treatment dummy equal to 1 if plant j is in California and t is January 2013 or later; $TREAT_{jt}^L$ is similarly defined for plants in leaker region L . The construction of a credible counterfactual against which to measure the effects of the cap-and-trade program is difficult because a suite of coincident changes affected California’s electricity market over the study period. For example, the increase in solar generation brought about by the aggressive renewable portfolio standard significantly impacted California’s wholesale electricity market outcomes (Bushnell and Novan, 2018). Other complementary measures in the Scoping Plan under AB 32 also affected utilization of baseload power plants in the Western Interconnection. We control for the impacts of these coincident changes through a broad set of determinants of capacity factors in $\mathbf{X}'_{jt}$, as discussed below.

Hydro, nuclear and renewable generation affect capacity utilization at NGCC and coal-fired plants, and are available at the monthly level over the period of our study. Since BAs are responsible for dispatching generation units and maintaining consumption-interchange-generation balance within a region of the electric grid, we might conceivably control for non-fossil generation within the plant’s BA. Instead, we consider in-state (not in-BA) non-fossil generation for two reasons. First, renewable portfolio standards are defined at the state level, and state policies may impact nuclear power use and capacity. Second, our dataset includes many small BAs (e.g., GRMA and HGMA), as well as large BAs spanning several states (e.g., MISO and SPP): when BAs are too small, hydro, nuclear and renewable generation is sparse, and does not allow for the use of cubic splines; when BAs are too large, in-BA nuclear and renewable generation conflate the effects of various policies at the state level. We model in-state hydro, nuclear and renewable generation using cubic splines, with coefficients that may vary by NERC interconnection and time of day.

We also control for non-fossil substitutes of a plant at two additional geographic scales. First, we calculate monthly shares for hydro-nuclear generation and renewable generation outside the plant’s BA, but in the same region.¹⁴ These shares take a non-zero value for all

collinearity. For this reason, we estimate DID models with capacity factor (rather than net generation or emissions) as dependent variable.

¹⁴Since nuclear generation at the BA level is sparse, we merge nuclear and hydro generation into one series. With regard to regional classification, for treated units in the Western Interconnection, BAs fall into one of four regions (California and three leaker regions), as discussed in Section 2.1. For control units in the Eastern Interconnection, we follow the regional classification by the U.S. Energy Information Administration (2022d).

plants in the Eastern and Western Interconnections, and a value of zero for plants in the Texas Interconnection, which largely overlaps with the ERCOT BA. We model the shares using linear functions, with coefficients that do not vary across markets.¹⁵

Second, we focus on the indirect effects of non-fossil generation in the North and South regions on capacity utilization at NGCC plants in California, and calculate monthly shares of hydro-nuclear and renewable generation in each of the two regions. These shares take a non-zero value for all plants in California, and a value of zero for plants in the leaker or control regions. We model the shares using linear functions. Other indirect effects in WECC are not considered for two reasons: California imports a substantial share of its electricity consumption from out of state (but does not often export electricity), and the net interchange between other regions in WECC is not significant in the period of our analysis. In line with the strategy adopted for WECC, we do not consider cross-regional effects in the Eastern Interconnection. Lastly, cross-regional effects are not present in the Texas Interconnection, which largely overlaps with ERCOT. Table F4 in the Appendix describes our control strategy for hydro, nuclear and renewable generation in each model specification.

Electricity consumption in the plant’s planning area is modeled with a logarithmic functional form implying low responsiveness of capacity factors when electricity consumption is high (Bushnell et al., 2008a; Davis and Hausman, 2016). We control for power imports into CAISO and imports from Canada into MRO-US and WECC. Specifically, imports into CAISO have a positive value for plants in California, and are equal to zero otherwise. Imports from Alberta and British Columbia are assigned to plants in the NW and RoW regions of WECC, while imports from Manitoba and Saskatchewan are assigned to plants in MRO-US; imports from Canada take a value of zero for all other plants in the dataset. We also account for factors that may affect plant productivity, like temperature (measured by heating and cooling degree days in the plant’s climate division) and precipitation (measured by the Standardized Precipitation Index in the plant’s climate division). In order to assess plant competitiveness, we calculate monthly fuel cost ratios. For coal plants, the coal-to-gas cost ratio divides plant-specific variable cost of generation by state average variable cost of natural gas for power generation. Similarly, for natural gas plants the gas-to-coal ratio divides plant-specific variable cost of generation by state average variable cost of coal for

¹⁵Cubic splines with invariant coefficients by region yield similar results. Creating splines varying by interconnection presents some challenges in the coal regression: specifically, splines cannot be created for shares of solar-wind-other renewable generation in the Eastern Interconnection because more than half of the observations have a share equal to 0. A hybrid approach based on a combination of cubic splines and a linear function for solar-wind-other renewable generation in the Eastern Interconnection yields estimates that are in line with the baseline, but parallel trends for this specification are not as robust.

power generation. We include the fuel cost ratios with both linear and quadratic terms to account for potential nonlinear responses to input prices (Cullen and Mansur, 2017). Further, we consider local economic activity and different recovery rates from the Great Recession through percent changes in state monthly seasonally-adjusted employment levels in energy intensive sectors (mining and logging, manufacturing, and construction). Finally, we include individual, regional and time fixed effects in the regressions. Plant specific effects, γ_j , may be associated with time-invariant differences in plant characteristics, like ownership (private utilities or political subdivision) and vintage. γ_y and γ_{dw} capture differential changes in average utilization that are common to all plants in a given year or day of the week. State by month-of-year fixed effects γ_{sm} account for seasonality (which is important when plants are part of a vast interconnection like WECC) and control for differential changes that are common to all plants within a state in a given month. Finally, the error term ϵ_{jt} is assumed independent of the covariates and treatment indicators.

2.4 Data

The econometric model uses a novel panel dataset built from publicly available sources including the U.S. Department of Energy’s Energy Information Administration (EIA), the U.S. Environmental Protection Agency (EPA), the Federal Energy Regulatory Commission (FERC) and the California Independent System Operator (CAISO). The period of our study spans January 2009 through December 2016, including four years before and four years after the treatment date (January 1, 2013).

2.4.1 EIA data

U.S. electric generating facilities with more than one MW of capacity are required to complete an annual survey to report plant characteristics. Form EIA-860 collects information on the status of existing plants in the U.S., while EIA-923 gathers information on plant operations. Relying on these surveys, we assemble a dataset for power plants within the U.S. portion of six NERC regions (FRCC, MRO-US, SERC, SPP, TRE and WECC) from 2009 to 2016 (Figure E1 in the Appendix). A plant consists of at least one, but typically several, generating units, which may be added to or retired from service over its lifetime. Although energy output, operating capacity and fuel input are available at the unit level, we aggregate units of the same technology to plants to provide an accurate representation of capacity factors

and heat rates for combined cycle plants.¹⁶ The advantage of EIA data is that its coverage is comprehensive, including not only large thermal plants, but also nuclear, hydro and renewable facilities. Plant-level characteristics reported at the annual level include primary fuel type, operating capacity, month and year when each unit was in service, NERC region and subregion, balancing authority and planning area. In addition, the EIA provides monthly plant operating statistics like energy output (measured by megawatt-hours or MWh of net electricity generation),¹⁷ consumption and heat content by fuel type, and cost of fuel delivered to the plant. We rely on Form EIA-860 for primary fuel type and operating capacity (U.S. Energy Information Administration, 2022b), and Form EIA-923 for other plant characteristics (U.S. Energy Information Administration, 2022c). We exclude plants with operating capacity below 25 MW.¹⁸

Plant fuel costs are used to calculate monthly ratios to assess competitiveness (Section 2.3). Fuel costs are not publicly available for non-regulated plants and plants with nameplate capacity below 50 MW. In these instances, we use state average costs of fossil fuels for electricity generation provided by the U.S. Energy Information Administration (2022a). If state average costs are also not available, we impute the fuel costs assuming the same growth rate of Rocky Mountain Colorado Rail coal prices (with a heat rate of 11,700 Btu/lb and a sulfur content of 0.8 lb/MMBtu) and NW Opal WY natural gas prices from SNL Energy.

2.4.2 CEMS data

We assemble a database of hourly gross electricity generation, heat input and CO₂ emissions for NGCC and coal-fired plants from the EPA’s Continuous Emissions Monitoring System (U.S. Environmental Protection Agency, 2022). CEMS represents the only publicly available information on high-frequency operating data for thermal power plants in the U.S., and has

¹⁶In combined cycle plants, gas is burned in a combustion turbine that generates electricity, and the waste heat from the turbine is captured and used to create steam that runs a second generator (the steam turbine) to produce additional electricity. The EIA reports energy output, operating capacity and fuel input for the combustion turbine part (denoted as CT) and the steam part (denoted as CA) separately but, in general, the CT of a NGCC plant cannot operate independently from its CA. Calculating capacity factors and heat rates for individual units that report separate output does not provide an accurate representation of plant utilization and efficiency, since the CT and CA parts of a NGCC plant cannot operate independently. Therefore, we aggregate energy output, operating capacity and fuel input for CT and CA units within the same combined cycle plant, and calculate plant-level capacity factors and heat rates. For consistency, we use plant-level data for the other technology types.

¹⁷Net generation excludes power consumption for plant operations.

¹⁸25 MW corresponds to the minimum size of generators subject to requirements for monitoring and reporting emissions under the EPA’s Continuous Emissions Monitoring System (U.S. Environmental Protection Agency, 2022). Plants with capacity below 25 MW generally use renewable energy sources and represent less than 5% of generating capacity in our sample.

been widely used in empirical studies (Davis and Hausman, 2016; Cullen and Mansur, 2017). We match units in CEMS to EIA generators using a 2015 crosswalk provided by the EPA (personal communication), and aggregate unit-level information from CEMS at the plant level by EIA site code and technology type. This step allows us to assign operating capacity to each power plant for which EPA data is available. We convert CEMS gross generation to net generation using technology-specific parasitic loss factors from the U.S. Environmental Protection Agency (2020). Finally, as noted above only thermal plants with capacity above 25 MW are required to report to CEMS; cogeneration, industrial and commercial facilities are also generally not in CEMS. These exceptions do not result in a substantial loss of coverage for our analysis: net generation of NGCC (coal-fired) plants from CEMS represents about 86% (97%) of EIA generation in WECC over the period of our study.

2.4.3 Other data

We complement information on the operations and status of electric power plants with data from other sources. We collect hourly scheduled net power imports into the California ISO on twelve transmission interfaces connecting the state’s electrical grid to the rest of WECC, which are identified based on the analysis of annual reports on the frequency of import congestion on each intertie (U.S. Department of Energy, 2014; California Independent System Operator, 2022a). Data are available from April 2009 to October 2015 from the Open Access Same-time Information System (California Independent System Operator, 2022c). In addition, hourly total power imports into the California grid are available from April 2010 to December 2016 from the California Independent System Operator (2022b). We merge these sources to create a time series of daily power imports into the California grid from April 1, 2009 to December 31, 2016. We also collect monthly net imports of power from Alberta, British Columbia, Manitoba and Saskatchewan by U.S. destination (National Energy Board of Canada, 2022), and aggregate them to the interconnection level (MRO-US and WECC). Hourly power flows between balancing authorities in WECC and the other interconnections are not publicly available over the period of our analysis.

Electricity consumption comes from the Federal Energy Regulatory Commission (FERC). FERC Form 714 provides hourly load information by planning area (Federal Energy Regulatory Commission, 2022). We aggregate load to the monthly and daily level, and assign it to power plants based on their planning area. Monthly population-weighted heating and cooling degree days, as well as measures of water scarcity by state climate division are from the National Oceanic and Atmospheric Administration (2022). The monthly seasonally-adjusted

employment level in the mining and logging, construction and manufacturing sectors by state is from the Bureau of Labor Statistics (2022). Finally, we obtain daily carbon futures prices for year vintage allowances expiring in December of the same year, in \$/ton, from the California Carbon Dashboard (Climate Policy Initiative, 2022).

2.5 Results

2.5.1 Shifts in electricity generation

Table 1 shows the estimates based on equation (1), using hourly capacity factors from CEMS as dependent variable (i.e., t in equation (1) corresponds to one hour). Covariates include the natural log of electricity consumption in the plant’s planning area; temperature and precipitation variables; cubic splines with three knots for each of the state-level hydro, nuclear and renewable generation variables; linear and quadratic terms for the fuel cost ratio; change in state employment levels in energy intensive sectors; power imports from Canada; and power imports into CAISO. Covariates are at the monthly level, except electric load by planning area and power imports into CAISO, which are available at the hourly and daily level, respectively. Robust standard errors are clustered at the plant level.

The treatment effects in equation (1) are interacted with a time-of-day indicator (equal to 1 between 7am and 7pm) to yield separate estimates for day and night. Time-of-day fixed effects are also included in the regression, in addition to those described in Section 2.3. Leakage would result in lower natural gas generation in California and higher coal and/or natural gas generation in the rest of WECC. Therefore, in the presence of leakage we would expect a negative and statistically significant α_C , and positive and statistically significant α_L for the leaker regions. Provided that the DID identification assumptions hold, our empirical results suggest a statistically significant reduction of NGCC capacity factors in California by 8.9% during daylight hours, relative to matched control facilities. In contrast, relative increases in capacity utilization at Western coal plants (outside California) occur throughout the day (i.e., during both daylight and evening hours), and are higher in the evening hours.¹⁹ We also estimate separate effects for every hour of the day by interacting each treatment indicator with 24 hourly indicators. Results are presented in Figure 2, and suggest a statistically significant reduction in NGCC capacity factors in California between 8am

¹⁹To address concerns that imports may be “bad controls” in this setting because they are themselves outcomes of the cap-and-trade policy, we removed imports as a control variable to test robustness. Results are in line with those in Table 1, suggesting that controlling for CAISO imports is not biasing the results. This is consistent with imports not increasing significantly as a result of the cap-and-trade policy, though their composition may have changed, and in line with the findings of Davis and Hausman (2016).

and 8pm. Increases in capacity utilization at Western coal plants are statistically significant during all hours of the day in the NW. Further, relative increases in capacity utilization in the NW and RoW regions happen in the evening hours. Note that changes in utilization rates do not tend to occur around the same hour, but are concentrated during daylight hours for California and in the evening hours for the leakers. These patterns urge caution in lending a causal interpretation to the empirical estimates, which may be biased due to confounding factors.

Specifications (2)-(4) are robustness checks that control for nuclear and renewable generation more flexibly, as described in Table F4 in the Appendix.²⁰ The dependent and independent variables are measured at the same frequency as in (1). The comparison of results from (1) and (2)-(4) yields two insights. First, the reduction of NGCC capacity factors in California during the day remains statistically significant but decreases in absolute value (i.e., the treatment effect gets closer to zero) in the more flexible specifications. Thus, (1) tends to overestimate the reduction in NGCC plant utilization from the cap-and-trade program during daylight hours, due to the symmetric structure imposed on the relationship between NGCC capacity factors and non-fossil generation. Second, statistically significant coal treatment effects in the leaker regions generally increase in magnitude, as we control for non-fossil generation more flexibly. Robustness checks also suggest a statistically significant increase in utilization of NGCC plants in the NW and coal-fired plants in the SW, albeit only at a 10% level. Thus, (1) tends to slightly underestimate the policy-induced increase in plant utilization outside California. Overall, the bias in the estimates from the coal regressions appears to be smaller than that associated with the estimates from the NGCC regressions.

Figure E4 in the Appendix shows the changes in plant utilization rates across hours of the day, relative to matched counterparts, based on specification (2). Here, bias from solar is expected to be less of a concern, because we allow for more flexible estimation of in-state non-fossil generation and add controls for non-fossil generation in the neighbouring regions. However, the estimated patterns look similar to those in Figure 2. Lastly, Appendix C describes a companion set of results using daily capacity factors from CEMS as dependent variable (i.e., t in equation (1) corresponds to one day), and Table F5 presents these results.

²⁰Specification (4) can only be run for NGCC because the number of observations in the coal database is not sufficient to create nuclear splines that vary by interconnection.

2.5.2 Evaluating the identification assumptions

Several assumptions must hold for our empirical estimates to provide an unbiased measure of the effect of California’s cap-and-trade program on baseload power plant operations in the Western Interconnection. We examine the plausibility of each assumption in turn.

Unconfoundedness. Our empirical strategy assumes that, conditional on observable plant characteristics, the distribution of the outcome is the same among treated and control plants. If this holds, biases in the unconditional differences-in-differences estimates are removed. As noted above, we match on capacity factors over four blocks of hours within the day, averaged over 2009-2010. Table F6 in the Appendix presents the t-statistics of tests of identical means of capacity factors (by hour and block of hours) in the treated and control groups, based on the matching procedure described in Section 2.2. Tables F7 and F8 in the Appendix show balance results for two additional plant characteristics: heat rate (a measure of efficiency) and age. The balancing tests confirm that matching achieves statistically indistinguishable means between treated plants in WECC and control plants. Before matching, there are significant differences between plant characteristics, particularly with respect to plant efficiency; after matching, the null of identical means in both groups is no longer rejected for any of the variables. This suggests that our matching procedure removes much of the potential bias.

Parallel trends. A second key assumption is that utilization of matched treated and control plants would follow parallel trajectories over time, in the absence of the treatment (Angrist and Pischke, 2009). Constructing counterfactual outcomes using observations on plants from another interconnection poses a challenge, because these plants do not “share the same economic environment” (Heckman et al., 1997) as the WECC plants; in particular, California’s electricity market was transformed at a rapid pace over the period of our study. The parallel trends assumption cannot be directly tested, but we assess its plausibility in several ways. Figure E5 in the Appendix shows the capacity factor trajectories of matched treated and control plants by technology type between 2009 and 2016.²¹ We conduct two tests to examine whether treated and control plants follow systematically different trends in the outcome variable before treatment. A common approach in the literature is to test the equivalence of time trends between treated and control groups prior to the intervention

²¹Note that these trajectories are unconditional, and the inclusion of covariates in the DID model serves to adjust for observable differences between treated and control groups in these plots.

(Autor, 2003; Fell and Maniloff, 2018). We use the following regression:

$$Y_{jt} = \sum_{qd} \alpha_C^{qd} B^q D_j^C T_t^d + \sum_L \sum_{qd} \alpha_L^{qd} B^q D_j^L T_t^d + \mathbf{X}_{jt}' \underline{\beta} + \gamma_j + \gamma_y + \gamma_{dw} + \gamma_t + \epsilon_{jt} \quad (2)$$

where B^q is a seasonal dummy equal to 1 if t lies in season q and 0 otherwise, $D_j^C = 1$ if plant j is in California, $D_j^L = 1$ if plant j is in one of the leaker regions in WECC, T_t^d is a time-of-day indicator to yield separate estimates for day and night (where index d is equal to 1 if t is between 7am and 7pm, and 0 otherwise), and α_C^{qd} and α_L^{qd} are the seasonal effects (or event study coefficients) estimated for specific times of the day.²² Other variables are defined as in specification (1) in Table 1. Pre treatment seasonal effects that are statistically significantly different from zero would support the assumption of parallel trends between treated and control groups prior to the intervention. Figures 3 and 4 present the event study coefficients by treated region, technology type and time of day. While in most cases the pre treatment effects are not individually statistically different from zero, these coefficients are imprecisely estimated. Therefore, the test results do not provide conclusive evidence to rule out the possibility of statistically significant pre treatment effects.

Next, we conduct a parallel trends test that compares the treatment effects in the baseline to the treatment effects in a specification that includes group-specific trends (Kearney and Levine, 2016; Kahn-Lang and Lang, 2020). If adding a trend changes the interpretation of the coefficients of interest, trend differences between treated and control groups prior to the intervention cannot be ruled out. We introduce linear and quadratic trends for each of the treatment group in hourly specification (1). If a treatment effect is statistically significant, we examine how the introduction of a trend affects its sign and significance, and present the results in Table 2. Adding a trend does not change the sign and significance of the estimated treatment effects for coal-fired plants in NW. The change in utilization rates at RoW coal-fired plants in the evening hours also remains positive and statistically significant when group-specific trends are included. However, the estimated effects for NGCC in California and Coal in RoW during daylight hours are not robust to the inclusion of trends, raising concerns about the causal interpretation of our results. Based on the evidence from the two parallel trends tests, we cannot rule out that treated and control groups were trending differentially before 2013.

²²We use seasonal dummies to account for cyclical factors that may affect plant utilization. In line with the National Renewable Energy Laboratory (2011), seasons are defined as follows: Summer = June, July, and August; Fall = September and October; Winter = November, December, January, and February; Spring = March, April, and May.

To address this challenge, we adopt the robust inference method proposed by Rambachan and Roth (2022a) to test sensitivity of statistically significant average treatment effects to violations of parallel trends.²³ Their approach builds on the intuition that, even if pre trends are not parallel, the difference in trends observed before treatment is informative about post treatment differences that would have occurred absent treatment. The researcher chooses the extent to which the counterfactual difference in trends post treatment deviates from the extrapolation of the pre-existing difference in trends by specifying a parameter M , which may be informed by context-specific knowledge: the bigger M is, the larger the deviation from the pre-existing difference in trends. Given a value of M , we can construct a robust confidence interval for the treatment effect. Further, we can examine robustness of the estimated treatment effect under varying assumptions on potential violations of parallel trends. For example, we can examine what deviation from the pre-existing difference in trends is needed to render a treatment effect statistically insignificant. Tighter bounds on the confidence intervals may be obtained by imposing sign and monotonicity restrictions that draw on context-specific knowledge.

Using equation (2), we estimate seasonal treatment effects for all regions and baseload technology types in WECC. Next, we construct robust confidence intervals for the seasonal treatment effects that do not pass the test in Table 2 (i.e., CA NGCC Day and RoW Coal Day), using the R code HonestDiD (Rambachan and Roth, 2022b). To illustrate, Figure 5 presents sensitivity analyses for these event study coefficients in the first period after treatment (Jan and Feb 2013). We compare the OLS confidence intervals (in blue) to the 95% confidence intervals from Rambachan and Roth’s method (in red), under varying restrictions and for different values of M . Each panel represents a specific set of restrictions on the sign of the bias of the post period event study coefficients (which are appropriate in cases with simultaneous policy changes) and monotonicity of the underlying difference in trends. For example, our treatment effects for California may overestimate the reduction in NGCC plant utilization from the cap-and-trade program, due to potential confounders that would have a coincident negative effect on capacity factors (e.g., complementary measures under AB 32). Therefore, in panel (b) we impose that the bias of California’s event study coefficients after treatment is negative. On the other hand, the net effect of the confounders on NGCC capacity factors may also be positive (e.g., if the SONGS replacement strategy empirically identified by Davis and Hausman (2016) continued after the introduction of the cap-and-trade

²³In a similar vein, Ang (2021) uses this method to conduct robust inference on statistically significant average treatment effects.

policy, and gas utilization in California to meet the lost generation from SONGS increased more than it decreased due to the effect of complementary measures under AB 32). Although this seems less likely, in light of the discussion in Section 2.5.1, we test the robustness of our results by considering positive bias in panel (c). Lastly, it is reasonable to assume that the downward sloping pre trend in California’s NGCC utilization in Figure 3 would have continued in the absence of the cap-and-trade program, due to the effect of other policies promoting renewable investment and generation in 2013-2016. This motivates our restriction of a monotone decreasing trend for California. Following Rambachan and Roth (2022a)’s recommendation, we use fixed length confidence intervals (FLCIs) when no restrictions are imposed, and conditional FLCIs under sign and monotonicity restrictions.

Turning to the results presented in each panel, a value of M equal to zero corresponds to a linear extrapolation of the pre-existing trend to the post treatment period; higher values of M reflect group-specific deviations from the pre existing trends that are calibrated on empirical estimates, and are driven by the evolution of factors that affect NGCC plant utilization in California (solar generation) and coal-fired plant utilization in the leaker regions (natural gas prices and thus coal-to-gas ratios) beyond the climate legislation. We benchmark M following Rambachan and Roth (2022a). First, we run a regression of capacity factor on standardized nuclear and renewable covariates and other determinants in $\mathbf{X}'_{jt}$ for each region and technology type. We find that a 1 standard deviation increase in solar generation corresponds to a 0.010 decrease in NGCC generation in California over the period of our study. Further, a 1 standard deviation increase in the coal-to-gas ratio corresponds to a 0.0763 (0.02361) decrease in NW (RoW) coal capacity factors. Next, we use these estimates to benchmark the value of M in each region. For California, a value of M equal to 0.0003 (0.001) {0.004} corresponds to changes in the differential slope of solar generation of about one fortieth (one tenth) {one third} of a standard deviation. For the RoW region, a value of M equal to 0.0006 (0.002) would correspond with allowing for changes in the differential slope of the coal-to-gas ratio of about one fortieth (one tenth) of a standard deviation. We also construct robust confidence intervals for an intermediate value of 0.0013.

The estimated treatment effect for California in January-February 2013 is negative, and the OLS confidence intervals rule out zero. When we assume a linear extrapolation of the pre-existing trend to the post treatment period ($M=0$), our conclusions are similar, but confidence intervals are tighter. As M grows larger, confidence intervals become less informative, as expected. However, the estimated confidence intervals exclude zero for all values of M , indicating that, given plausible non-linear deviations from the pre-existing

differences in trends, we cannot rule out a statistically significant treatment effect of the policy in the first period after treatment. In the RoW, the OLS estimate is slightly negative but the confidence interval includes zero. When we allow for linear violations of parallel trends, we cannot rule out a statistically significant increase in coal-fired capacity factors in the first period after treatment.

Robust confidence intervals for all treatment effects under varying restrictions and for different values of M are presented in the Appendix (Figures E6-E7). For $M = 0$, we find that 9 seasonal daily effects in California (out of 17 post treatment effects) are statistically significant under no restrictions; 16 seasonal effects are statistically significant under monotonicity and negative bias; and 10 seasonal effects are statistically significant under monotonicity and positive bias. Further, 8 seasonal daily effects are statistically significant in RoW, and 7 of these are in the first part of the post treatment period (Jan-Feb 2013 till Summer 2014).

Stability of unit treatment values. The empirical framework assumes that plant-level capacity utilization depends on the treatment status of the corresponding plant, but is independent of the treatment status of other plants. This is the stable unit treatment value assumption. By designating control plants outside of WECC, we assume that the policy does not affect facilities in other NERC regions. This is plausible, because the Western, Eastern and Texas Interconnections operate largely independently from each other and power transfers between them are limited. As a result, spillovers and market equilibrium effects on the designated control plants in the Eastern and Texas Interconnection are unlikely. Although not testable in principle, we believe that the SUTVA holds in our study.

Treatment exogeneity and overlap. Two additional assumptions that are required for identification are treatment exogeneity and overlap. In our setting, treatment is exogenous because participation in the cap-and-trade program does not depend on the outcomes. The overlap assumption requires the support of the distribution of covariates in the treated group to overlap the support of the distribution of these covariates in the control group. Coarsened exact matching automatically restricts the matched data to areas of common support, as discussed in Section 2.2: this helps avoid making inferences based on extrapolation, which are known to be highly model dependent. Thus, we believe that the overlap condition is satisfied in our study.

2.5.3 Leakage estimates

Based on seasonal effects estimated for specific regions/technology types/times of day and their confidence intervals, we infer CO₂ emission leakage predictions in 2013 and 2016. Our

focus on these years is intended to enable more direct comparisons with the simulation results. The starting point is given by the estimates for specification (1) in Table 1: regions and technology types that are statistically significant for at least one time of the day (day or night) are considered in our leakage calculations. This criterion yields CA NGCC, NW Coal and RoW Coal.²⁴ Next, we estimate seasonal daytime/nighttime effects and relative confidence intervals for the selected regions and technology types. For the reasons discussed in Section 2.5.2, we construct robust confidence intervals that allow for linear violations of parallel trends ($M=0$) for CA NGCC Day and RoW Coal Day, assuming negative bias and monotonicity restrictions for CA NGCC Day, and no restrictions for RoW Coal Day. For all other seasonal treatment effects (CA NGCC Night, NW Coal Day and Night, and RoW Coal Night), we rely instead on OLS confidence intervals.

Not all seasonal estimates by time of day for the selected regions are included in the leakage calculations implied by the econometric model. Instead, we only include seasons and times of day with contemporaneous changes in capacity factors that are consistent with leakage (negative for California, and positive for at least one of the leaker regions), regardless of statistical significance of the seasonal effects. Thus, if on a given season/time of day the estimated change in capacity factors is negative for California and positive for at least one of the leaker regions, we include that season/time of day into our leakage calculations. In contrast, if on a given season/time of day the estimated changes are not in the expected direction (i.e., positive changes in California’s capacity factors, or negative changes in California’s capacity factors coupled with zero or negative changes in capacity factors for all leaker regions), we do not consider that season/time-of-day in our leakage calculations. We use the lower (upper) bound of the 95% confidence interval for the estimated seasonal effects to calculate a lower (upper) bound for the generation, emissions and leakage rates associated with the econometric estimates.

Next, we find the estimated generation leakage by multiplying the seasonal treatment effects by the total generation capacity of matched plants by region, year and technology, and the number of hours in that season. Based on these generation leakage estimates, we calculate the change in local CO₂ emissions in California (E_1) and WECC-NonCA emissions (E_4), based on region-, year- and technology-specific heat rates and CO₂ emission rates. The resulting change in WECC emissions ($E_5 = E_1 + E_4$) is between -10.49 and 15.26 million metric tons in 2013, and between -15.60 and 11.35 million metric tons in 2016 (Table 3).

²⁴Leakage bounds based on seasonal impacts for all regions and baseload technology types (NGCC and Coal) are too wide to be informative, and thus we do not present them in the paper.

As noted above, emissions subject to the cap-and-trade regulation include not only in-state emissions, but also emissions associated with power imports into California. As a result, the change in regulated emissions includes the change in local emissions in California (E_1), as well as the change in emissions associated with power imports into California, relative to a counterfactual (E_2). Since the change in import emissions E_2 cannot be obtained from the econometric estimates, we proceed as follows. First, denote as I_1 the “Total Covered Emissions” for electricity importers reported by CARB in its annual GHG emission inventories (California Air Resources Board, 2022). These emissions (36.20 million metric tons CO₂e in 2013 and 21.02 million metric tons CO₂e in 2016) form the basis to determine compliance obligations in the cap-and-trade program. Next, we construct the year-specific counterfactual import emissions I_2 assuming the same percentage change between counterfactual emissions and emissions under the carbon cap predicted by JHSMINE. The difference between I_1 and I_2 yields the estimated emission reduction E_2 on Table 3.²⁵

Finally, we calculate the implied leakage rates. A common metric used in the literature (e.g., Bushnell et al. (2014); Caron et al. (2015); Fell and Maniloff (2018)) is the physical leakage rate, which reflects the share of local emission reductions that is offset by emission increases in the rest of the system. In our setting, this would be defined as $100\% \times (-E_4/E_1)$. Given California’s first deliverer approach, we adopt an alternate leakage metric that considers the difference between the decrease in regulated emissions and the decrease in system-wide emissions in WECC. In line with Chen et al. (2011) and Xu and Hobbs (2021), we define leakage as $100\% - E_5/E_3 = (1 - E_5/E_3) \times 100\%$, where $E_3 = E_1 + E_2$. A positive leakage rate indicates a mismatch between WECC emissions and California’s regulated emissions. In particular, if the leakage rate is positive but below 100%, regulated emissions decrease, but total emissions in WECC fall by a lower amount; if the leakage rate exceeds 100%, regulated emissions fall, but WECC emissions actually increase. Given robust confidence intervals that allow for linear violations of parallel trends, the leakage rates implied by our econometric estimates are between 40.8% and 217.9% in 2013, and between 17.4% and 177.4% in 2016. The lower bounds of these intervals are within the range of earlier empirical estimates: for example, in the context of RGGI, Fell and Maniloff (2018) find an electricity-sector specific leakage of about 50%, while the leakage interval predicted by Zhou and Huang (2021) is

²⁵To illustrate, in 2013 JHSMINE predicts import emissions of 51.25 million metric tons for the no cap scenario and 39.41 million metric tons for the carbon cap scenario. This implies that no cap emissions would be 30% higher than emissions under the cap. Hence, the counterfactual import emissions for the empirical analysis are 47.07 ($= 36.20 \times 1.30$) million metric tons CO₂e, and the estimated import emission reduction is 10.87 ($= 36.20 - 47.07$) million metric tons CO₂e in 2013.

43%-85%. However, direct comparisons are difficult due to the use of different metrics that are less relevant in our setting: for instance, both studies cited above calculate a physical leakage rate.

3 Simulation model

We use a partial equilibrium model of the electricity sector (JHSMINE) to simulate shifts in electricity generation in the Western Interconnection in response to the introduction of California’s cap-and-trade program. Based on these estimates, we infer CO₂ emission leakage predictions in 2013 and 2016. This section presents an overview of the model, describes the scenarios and steps taken to ensure that the data used for both models are comparable, and discusses the simulation results.

3.1 Overview

The Johns Hopkins Stochastic Multistage Integrated Network Expansion model is a long-term transmission-generation-storage expansion planning model of the electricity sector based on scenario-based stochastic programming. The model was applied to the Western Electricity Coordinating Council using a reduced network based on the WECC 2026 Common Case (WECC Staff, 2016) to provide insights into the transmission planning process and efficiency of border carbon adjustment schemes in the Western U.S. (Xu and Hobbs, 2021). The reduced network consists of 361 buses, 712 transmission lines, and 1,504 existing aggregate generators of various technology types, including coal steam plants and combined cycles. Key modeling assumptions include perfectly inelastic demand, perfect competition, and perfect foresight of market participants.

The version of JHSMINE in this paper builds on the one in Xu and Hobbs (2021), but differs from it in several important ways. While Xu and Hobbs (2021) consider a capacity expansion planning model, we run a production cost model that simulates hourly commitment and dispatch decisions under alternate carbon pricing scenarios, taking generation capacity as given. Further, in order to generate plausible leakage predictions, we introduce features that enhance realism in the model formulation. First, we approximate power flows on the transmission network by a direct current (DC) load flow. The resulting DC OPF uses a linearized approximation of the alternating current (AC) power flow equations, and allows for a more accurate representation of power flows than the transshipment model in earlier formulations, which ignores Kirchhoff’s Voltage Law. Second, we include relaxed (non-integer)

unit commitment variables in the model. Third, with respect to power imports into California, the original model can only simulate a scenario in which all imports are considered specified power and assigned facility-specific emission factors (100% specified), or a scenario in which all imports are considered unspecified and assigned the default emission factor of 0.428 metric ton CO₂/MWh (0% specified). The observed regime in California is a hybrid of the two, where source specification was not possible for about 26% of electricity imports, on average between 2013 and 2016 (California Air Resources Board, 2020a). To make simulation results more directly comparable with the empirical estimates, we revise the formulation of JHSMINE by unbundling the non-electrical attributes of power generation (emissions and renewable energy credits, or RECs). Xu and Hobbs (2021) model these attributes with one variable, *cpf*, representing the emissions and RECs associated with a contract (in MW) sold by a generator to a load serving entity. In contrast, we allow for emissions and RECs to be traded through separate contracts. This change allows us to model a regime where (a) electricity producers can enter bilateral contracts in which power is specified, or sell unspecified power to a pool, and (b) load serving entities can buy specified power through bilateral contracts, or unspecified power from the pool. To obtain the emissions of power imports to California, energy contracts between the California LSE and out-of-state generation companies are assigned an emission rate. When imports are considered specified power, the emission rate is plant-specific. When imports are considered unspecified power, the emission rate is set equal to the default emission factor of 0.428 metric ton CO₂/MWh. The model formulation is presented in Appendix D1.

3.2 Scenarios

We consider two scenarios: (a) a benchmark scenario with no regulation of GHG emissions (“No cap”); (b) a scenario where California generators and the California LSE are subject to a first deliverer cap-and-trade program (“Carbon cap”). The carbon price is assumed and set equal to average historical values over the period of our study, rather than determined endogenously in the model. In both scenarios, specified electricity imports to California are assigned facility-specific emission factors, while unspecified imports are assigned the default emission factor of 0.428 metric ton CO₂/MWh. Further, both scenarios include RPSs and assume the same share and composition of specified imports into California.

3.3 Data

To ensure comparability of the data used for the econometric model and the simulation model, we modify the JHSMINE dataset in Xu and Hobbs (2021) making use of the installed generation capacity, average fuel costs and load from the econometric model dataset. In addition, we parameterize the shares of California imports by fossil fuel generation type based on historical data from CARB, in order to ensure that the level and composition of power imports into California in JHSMINE are comparable to historical values. Appendix D2 provides additional details about the sources of data used for JHSMINE.

3.4 Results

3.4.1 Shifts in electricity generation

Table 4 presents the predicted impact of California’s cap-and-trade program on capacity factors in WECC. We run JHSMINE for forty-eight representative days in 2013 and 2016, under a no cap and a carbon cap scenario. As noted above, WECC Regions and seasons are defined as in the econometric model. The introduction of the carbon policy mainly affects utilization at NGCC and coal-fired plants in WECC, supporting our choice to focus on these technology types in the empirical analysis. Comparing counterfactual (no cap) scenarios across years, generation shifts are due to lower natural gas prices and higher RPS requirements in 2016. For example, lower natural gas prices in 2016 determine an increase in NGCC generation in all WECC regions, and solar capacity factors are higher in regions with a more stringent RPS requirement in 2016 (e.g., California, Arizona and New Mexico). Comparing counterfactual and policy scenarios for the same year, the introduction of the carbon policy yields minor generation shifts in 2013: capacity utilization decreases by about 0.7% at California NGCC plants, and the highest increase outside California is in the NW region (+1%). Note that natural gas prices are relatively high in 2013: as a result, coal-fired plants are heavily utilized, and power imports into California in the carbon cap scenario are sourced at NGCC plants elsewhere in WECC. Generation shifts are more evident in 2016, when coal and carbon prices are at comparable levels but natural gas prices are lower, relative to 2013. The counterfactual scenario in 2016 is thus characterized by lower generation at coal-fired plants, leaving more room for leakage from this technology type after the introduction of a carbon price. This yields a large policy-induced decrease in NGCC capacity factors in California (9.6%), a combined increase in capacity utilization by 3.8% at Western coal plants outside California (mainly in the NW region), and by 12.8% at Western NGCC plants outside

California (mainly in the SW region).

JHSMINE is well suited to identifying average changes in hourly capacity utilization that are solely due to the effects of the cap-and-trade policy, *ceteris paribus*. Thus, we calculate average changes in capacity factors between the two simulated scenarios (carbon cap - no cap) at hours 0-23 for each WECC region and technology type (Figures 6 and 7). In 2013, California’s NGCC capacity factors decrease between 8am and 7pm due to the cap-and-trade policy. Note that bias from solar is not a concern in these results, because solar generation is the same under both scenarios. Lower capacity factors in California are mainly offset by coincident changes in utilization rates at NGCC units in the NW and RoW leaker regions. Moreover, in the evening hours NGCC capacity factors at California’s gas plants decrease slightly while coal capacity utilization in the NW and RoW increases, albeit by a smaller magnitude than predicted by the econometric model. In 2016, the average reduction of in-state capacity utilization predicted by JHSMINE is more substantial, particularly between 2pm and 7pm. This is offset by coincident positive changes in utilization rates at NGCC plants in the rest of WECC, mainly in the SW and RoW. Coal capacity factors also increase throughout the day in NW, and in the early morning hours in SW and RoW.

3.4.2 Leakage estimates

Table 5 presents the distribution of emissions among WECC regions, as well as the implied leakage rates (as defined in Section 2.5.3) in 2013 and 2016. We emphasize two important differences, relative to the results in Table 3. First, when calculating leakage based on empirical estimates, we use seasonal average impacts estimated for specific times of the day. As noted above, we consider seasons and times of day that exhibit contemporaneous changes in capacity factors in the expected direction, regardless of statistical significance. Further, we consider seasonal effects for regions and technology types that are statistically significant for at least one time of the day in specification (5) (i.e., CA NGCC, NW Coal and RoW Coal). In contrast, the results in Table 5 are based on daily averages for all technology types and WECC regions. Second, for a given technology type and region, the econometric model only accounts for changes in emissions associated with the subset of treated units for which good matches exist among available controls, while JHSMINE considers all aggregate generators.

We find that local emissions in California (E_1) decrease by only 1.10 million metric tons in 2013, but by 6.58 million metric tons in 2016. Given the estimated changes in import emissions relative to the no cap scenario (11.83 million metric tons CO₂e in 2013 and 13.88 million metric tons CO₂e in 2016), regulated emissions in California decrease in both 2013

and 2016 (by 12.94 and 20.46 million metric tons, respectively), but total emissions in WECC fall by much less (in 2013) or slightly increase (in 2016) due to higher unregulated emissions out-of-state. In particular, JHSMINE suggests that most policy-induced change in out-of-state generation and emissions takes place in the NW and SW regions of WECC. Plants in the RoW only adjust their output slightly in response to California’s cap-and-trade program, leading to small emission increases. This contrasts with our empirical estimates, which suggest that coal-fired generation increased in the Northwest and RoW regions, while other treatment effects are not statistically significant. The leakage rates implied by JHSMINE are 94.3% in 2013 and 110% in 2016, in line with the predictions from earlier simulation-based partial equilibrium models that use the same leakage metric (Chen et al., 2011; Xu and Hobbs, 2021).

4 Comparison of results

JHSMINE is well suited to isolating the effects of California’s cap-and-trade program on power plant operations in WECC, but yields *ex ante* predictions based on assumed firm behavior under perfectly inelastic demand, perfect competition, and perfect foresight of market participants. Although we parameterize the model using historical data, it is not surprising that its predictions may differ greatly from seasonal changes in plant utilization that are observed in the data. For example, a comparison of estimated and simulated coal-fired capacity factors in NW in Spring 2016 shows that JHSMINE does not predict a drop in coal utilization due to record low natural gas prices in the U.S. (U.S. Energy Information Administration, 2020), which is captured by a negative and statistically significant seasonal effect. In contrast, the econometric model measures the *ex post* realized effects of the policy, but the empirical estimates may be imprecise due to threats to identification in this policy setting and inaccurate representation of transmission network constraints.

In this section, we compare the *ex ante* expected impacts of the policy with its *ex post* realized impacts, starting with the effects on electricity generation shifts. With regard to the predicted source of leakage in WECC, JHSMINE suggests that the policy-induced reduction in capacity factors at California NGCC plants was offset by higher NGCC capacity utilization in the NW and SW regions of WECC, as well as coal capacity utilization in the NW; in contrast, the econometric model suggests increases in capacity utilization at coal plants in the NW and RoW. What factors may explain these differences? In the simulation model, the shares of specified imports into California by fossil fuel generation type are parameterized based on

historical values, but the composition of unspecified imports is unknown. Thus, the findings of the econometric model may be consistent with higher levels of imports of out-of-state coal generation as unspecified power (relative to the *ex ante* predictions), and an incentive for electricity importers to not report the emission content of out-of-state higher-emitting generation resources in order to attain the lower default emission rate for GHG compliance obligations (“laundering”).

With regard to heterogeneity between day and night, the changes in plant utilization rates across hours of the day based on the simulation model (Figures 6 and 7) are generally consistent with the expected substitution patterns at regulated and leaker units (i.e., decreases/increases in utilization rates that tend to occur at the same hour, given limited energy storage capacity over the study period). In contrast, the changes based on the econometric model (Figure 2) suggest that capacity factor reductions at California’s gas plants mainly occur in the middle of the day, while coal capacity factor increases outside California are concentrated in the evening hours. Direct comparisons between these sets of results are difficult because the changes based on the econometric model are relative to matched controls, while the changes based on the simulation model are relative to the benchmark scenario with no regulation of GHG emissions. Further, the empirical model yields average effects throughout the post treatment period (2013-2016), while JHSMINE yields results for specific years (2013 and 2016). Bearing these caveats in mind, the changes based on the econometric model are likely to be confounded by the effects of coincident policies and market developments, such as increased solar generation brought about by California’s renewable portfolio standard. However, the diurnal patterns observed in Figure 2 are not solely driven by NGCC being crowded out by solar generation, because the changes based on the simulation model (where bias from solar and other confounding factors is not a concern) also suggest a policy-induced reduction in utilization at California’s gas plants during daylight hours.

Lastly, we compare the impacts of the policy on emissions and leakage. The results in Tables 3 and 5 differ for the reasons discussed in Section 3.4.2. To enable more direct comparisons, we calculate the leakage rates implied by JHSMINE when considering only emission changes associated with NGCC plants in California and coal-fired plants in the NW and RoW leaker regions. Table 6 presents the results, and shows robustness of confidence intervals from the econometric model to removing bias and monotonicity restrictions, under linear ($M = 0$) violations of parallel trends. The CO₂ emission leakage rates predicted by JHSMINE are 95.5% in 2013 and 83.3% in 2016. Note that the adjustment does not significantly affect the 2013 rate, relative to Table 5, but yields lower leakage in 2016, because

it does not consider higher emissions in the Southwest region of WECC. With respect to the confidence intervals implied by our empirical strategy, the lower bound estimates are in the 40.8%-42% range in 2013 and 17.4%-23.1% range in 2016. *Ex post* rates below *ex ante* rates are consistent with contracts not being shuffled as easily as predicted by the simulation model, which does not account for transaction costs that would discourage these rearrangements. This result is also in line with previous findings in the literature: for example, in the context of RGGI, Chen (2009) predicts relative leakage rates of 90%-100%, while empirical estimates from Zhou and Huang (2021) are in the 43%-85% range. On the other hand, the upper bound estimates from the econometric model are around 200%. This level is outside the range of predictions from simulation-based partial equilibrium models in the literature that use the same leakage metric, but is not entirely unrealistic: similar rates would be observed if the reduction in NGCC generation in California was offset by an increase in coal-fired generation outside California of the same magnitude, and there was no change in emissions associated with power imports.²⁶

5 Concluding remarks

In this paper, we seek to identify CO₂ emission leakage in the electricity sector from California’s AB 32 cap-and-trade program in the first four years of policy implementation. We estimate shifts in electricity generation at baseload power plants in the Western Interconnection based on two models: a simulation-based partial equilibrium model of the electricity sector (JHSMINE) that includes salient features of the observed cap-and-trade program and is parameterized using market data in 2013-2016; and an econometric model applying a quasi-experimental design with matching and a robust inference method that does not require the parallel trends assumption to hold exactly. Based on the estimated shifts in electricity generation, we infer CO₂ emission leakage predictions in 2013 and 2016. We then compare the *ex ante* expected impacts of the policy to the *ex post* realized impacts. This allows us to identify critical assumptions driving the simulation results, and benchmark the empirical results in a setting where threats to identification undermine attempts at statistical inference.

Both models predict reduced utilization at California’s gas plants, but insights differ with respect to the predicted source of leakage at Western plants outside California. JHSMINE suggests relative increases in the utilization of gas plants in the NW and SW regions of WECC

²⁶Coal combustion emits almost twice as much carbon dioxide per unit of energy as does the combustion of natural gas.

and, to a more limited extent, of coal plants in the NW; in contrast, the econometric model finds that capacity utilization mostly increases at coal plants in the NW and RoW regions over this study period. As discussed in the paper, the effects of coincident policy changes and market developments likely confound our empirical estimates. However, the composition of unspecified imports is not parameterized based on historical data in the simulation model. Thus, the *ex post* findings may be consistent with higher levels of imports of out-of-state coal generation as unspecified power (relative to the *ex ante* predictions), and an incentive for electricity importers to not report the emission content of out-of-state higher-emitting generation resources in order to attain the lower default emission rate for GHG compliance obligations (“laundering”). This suggests that limiting the ability of electricity importers to claim the default emission factor may reduce leakage risks (Fowle et al., 2021; Fowle and Reguant, 2022).

With regard to policy impacts on emissions, JHSMINE finds a significant potential for leakage in WECC, with predicted rates of 95.5% in 2013 and 83.3% in 2016. Predictions from the econometric model have wide confidence intervals, but suggest some empirical evidence of leakage. We cannot rule out leakage rates as low as 41% (or as high as 218%) in 2013, and as low as 17% (or as high as 177%) in 2016. These results are based on credible assumptions on the sign of the bias of the post period event study coefficients and monotonicity of the difference in trends between the treated and control groups, and are robust to the exclusion of bias and monotonicity restrictions. In particular, *ex post* rates below *ex ante* rates are consistent with contracts not being shuffled as easily as predicted by the simulation model, which does not account for transaction costs discouraging these rearrangements.

Our study shows that simulation models and econometric models can play complementary roles in the evaluation of carbon policy impacts. To support comparisons between simulation results and empirical estimates, future research in this area could enhance representation of network effects in empirical analyses and simulate power market outcomes under relaxed assumptions on the degree of market competition, demand elasticity and foresight of market participants.

References

- Aichele, Rahel, and Gabriel Felbermayr. 2013. The Effect of the Kyoto Protocol on Carbon Emissions. *Journal of Policy Analysis and Management* 32(4): 731-757.
- , 2015. Kyoto and Carbon Leakage: an Empirical Analysis of the Carbon Content of Bilateral Trade. *Review of Economics and Statistics* 97 (1): 104-115.
- Ang, Desmond. 2021. The Effects of Police Violence on Inner-City Students. *The Quarterly Journal of Economics* 136 (1): 115-168.
- Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.
- Autor, David. 2003. Outsourcing at Will: the Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics* 21 (1): 1-42.
- Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. CEM: Coarsened Exact Matching in Stata. *The Stata Journal* 9 (4): 524-546.
- Bureau of Labor Statistics. 2022. *Employment, Hours, and Earnings - State and Metro Area*. <http://www.bls.gov/data/#employment>.
- Burtraw, Dallas, Ann Carlson, Danny Cullenward, Quentin Foster, and Meredith Fowle. 2018. *2018 Annual Report of the Independent Emissions Market Advisory Committee*. http://calepa.ca.gov/wp-content/uploads/sites/6/2018/10/Final_2018_IEMAC_Annual_Report_10-22-2018.pdf.
- Bushnell, James, Erin Mansur, and Celeste Saravia. 2008a. Vertical Arrangements, Market Structure and Competition: an Analysis of Restructured U.S. Electricity Markets. *American Economic Review* 98 (1): 237-266.
- Bushnell, James, Carla Peterman, and Catherine Wolfram. 2008b. Local Solutions to Global Problems: Climate Change Policies and Regulatory Jurisdiction. *Review of Environmental Economics and Policy* 2 (2): 175-193.

- Bushnell, James, Yihsu Chen, and Matthew Zaragoza-Watkins. 2014. Downstream Regulation of CO₂ Emissions in California's Electricity Sector. *Energy Policy* 64: 313–323.
- Bushnell, James, and Kevin Novan. 2018. *Setting with the Sun: the Impacts of Renewable Energy on Wholesale Power Markets*. NBER Working Paper 24980, National Bureau of Economic Research, Cambridge, MA.
- California Air Resources Board. 2008. *Climate Change Scoping Plan*. https://ww2.arb.ca.gov/sites/default/files/classic/cc/scopingplan/document/adopted_scoping_plan.pdf.
- . 2014. *First Update to the Climate Change Scoping Plan*. https://ww2.arb.ca.gov/sites/default/files/classic/cc/scopingplan/2013_update/first_update_climate_change_scoping_plan.pdf.
- . 2020a. *2000-2018 GHG Inventory (2020 Edition)*. <https://ww2.arb.ca.gov/ghg-inventory-archive>.
- . 2020b. *California Greenhouse Gas Emissions for 2000 to 2018 - Trends of Emissions and Other Indicators*. https://ww3.arb.ca.gov/cc/inventory/pubs/reports/2000_2018/ghg_inventory_trends_00-18.pdf.
- . 2022. *Mandatory GHG Reporting - Reported Emissions. 2013 GHG Facility and Entity Emissions (Updated 11/4/2019), and 2016 GHG Facility and Entity Emissions (Updated 11/4/2022)*. <https://ww2.arb.ca.gov/mrr-data>.
- California Energy Commission. 2017. *2016 Total System Electric Generation*. <http://www.energy.ca.gov/data-reports/energy-almanac/california-electricity-data/2018-total-system-electric-generation/2016>.
- California Independent System Operator. 2022a. *Annual and Quarterly Reports*. <http://www.caiso.com/market/Pages/MarketMonitoring/AnnualQuarterlyReports/Default.aspx>.
- . 2022b. *Daily Renewables Watch*. <http://www.caiso.com/market/Pages/ReportsBu>

[lletins/RenewablesReporting.aspx](#).

-----, 2022c. *Open Access Same-time Information System (OASIS)*. <http://oasis.caiso.com/mrioasis/logon.do>.

Carbone, Jared, and Nicholas Rivers. 2017. The Impacts of Unilateral Climate Policy on Competitiveness: Evidence from Computable General Equilibrium Models. *Review of Environmental Economics and Policy* 11 (1): 24–42.

Carbone, Jared, Nicholas Rivers, Akio Yamazaki, and Hidemichi Yonezawa. 2020. Comparing Applied General Equilibrium and Econometric Estimates of the Effect of an Environmental Policy Shock. *Journal of the Association of Environmental and Resource Economists* 7 (4): 687–719.

Carlson, Curtis, Dallas Burtraw, Maureen Cropper, and Karen Palmer. 2000. Sulfur Dioxide Control by Electric Utilities: What Are the Gains from Trade? *Journal of Political Economy* 108 (6): 1292–1326.

Caron, Justin, Sebastian Rausch, and Niven Winchester. 2015. Leakage from Sub-National Climate Policy: the Case of California’s Cap-and-Trade Program. *The Energy Journal* 36 (2): 167–190.

Chen, Yihsu. 2009. Does a regional greenhouse gas policy make sense? A case study of carbon leakage and emissions spillover. *Energy Economics* 31 (5): 667–675.

Chen, Yihsu, Andrew Liu, and Benjamin Hobbs. 2011. Economic and Emissions Implications of Load-Based, Source-Based, and First-Seller Emissions Trading Programs under California AB 32. *Operations Research* 59 (3): 696–712.

Climate Policy Initiative. 2022. *California Carbon Dashboard*. <http://calcarbondash.org/>.

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

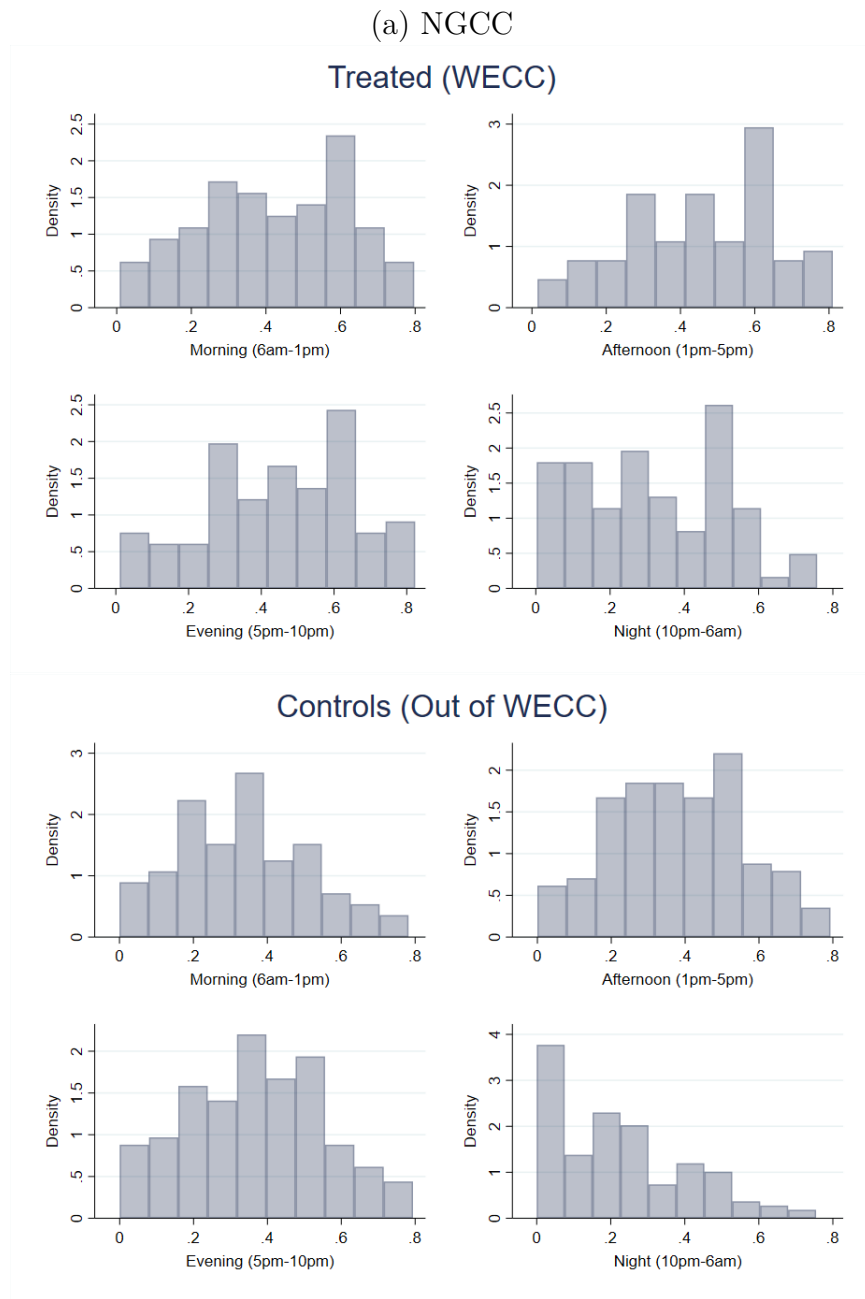
- Davis, Lucas, and Catherine Hausman. 2016. Market Impacts of a Nuclear Power Plant Closure. *American Economic Journal: Applied Economics* 8 (2): 92–122.
- Ek, Claes, and Jurate Miliute-Plepiene. 2018. Behavioral Spillovers from Food-Waste Collection in Swedish Municipalities. *Journal of Environmental Economics and Management* 89: 168–186.
- Ellerman, Denny, Paul Joskow, Richard Schmalensee, Elizabeth Bailey, and Juan-Pablo Montero. 2000. *Markets for Clean Air: The U.S. Acide Rain Program*. Cambridge University Press.
- Federal Energy Regulatory Commission. 2022. *Form No. 714 - Annual Electric Balancing Authority Area and Planning Area Report*. <https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-no-714-annual-electric/data>.
- Fell, Harrison, and Peter Maniloff. 2018. Leakage in Regional Environmental Policy: the Case of the Regional Greenhouse Gas Initiative. *Journal of Environmental Economics and Management* 87: 1–23.
- Fowlie, Meredith. 2009. Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage. *American Economic Journal: Economic Policy* 1 (2): 72–112.
- Fowlie, Meredith, Stephen Holland, and Erin Mansur. 2012. What Do Emissions Markets Deliver and To Whom? Evidence from Southern California’s NO_x Trading Program. *American Economic Review* 102 (2): 965–993.
- Fowlie, Meredith, Claire Petersen, and Mar Reguant. 2021. Border Carbon Adjustments When Carbon Intensity Varies Across Producers: Evidence from California. *American Economic Review Papers and Proceedings* 111: 401–405.
- Fowlie, Meredith, and Mar Reguant. 2022. Mitigating Emissions Leakage in Incomplete Carbon Markets. *Journal of the Association of Environmental and Resource Economists* 9 (2): 307–343.

- Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies* 64 (4): 605–654.
- Iacus, Stefano, Gary King, and Giuseppe Porro. 2008. *Matching for Causal Inference Without Balance Checking*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1152391.
- Kahn-Lang, Ariella, and Kevin Lang. 2020. The Promise and Pitfalls of Differences-in-Differences: Reflections on *16 and Pregnant* and Other Applications. *Journal of Business & Economics Statistics* 38 (3): 613–620.
- Kearney, Melissa, and Phillip Levine. 2016. *Does Reality TV Induce Real Effects? A Response to Jaeger, Joyce and Kaestner*. IZA Discussion Paper 10318, Bonn, Germany.
- National Electric Reliability Council. 2022. *Glossary of Terms Used in NERC Reliability Standards*. https://www.nerc.com/pa/Stand/Glossary%20of%20Terms/Glossary_of_Terms.pdf.
- National Energy Board of Canada. 2022. *Commodity Statistics*. <http://apps.neb-one.gc.ca/CommodityStatistics/Statistics.aspx?language=english>.
- National Renewable Energy Laboratory. 2011. *Regional Energy Deployment System (ReEDS)*. <https://www.nrel.gov/docs/fy12osti/46534.pdf>.
- , 2019. *Regional Energy Deployment System (ReEDS) Model Documentation: Version 2019*. <https://www.nrel.gov/docs/fy20osti/74111.pdf>.
- National Oceanic and Atmospheric Administration. 2022. *National Centers for Environmental Information*. <http://www.ncdc.noaa.gov/>.
- Nyberg, Michael. 2018. *Thermal Efficiency of Natural Gas-Fired Generation in California: 2017 Update*. <https://www.energy.ca.gov/publications/2018/thermal-efficiency-natural-gas-fired-generation-california-2017-update>.
- Rambachan, Ashesh and Jonathan Roth. 2022a. *A More Credible Approach to Parallel*

- Trends*. https://jonathandroth.github.io/assets/files/HonestParallelTrends_Main.pdf.
- , 2022b. *HonestDiD R package*. <https://github.com/asheshrambachan/HonestDiD>.
- Simcoe, Timothy, and Michael Toffel. 2014. Government Green Procurement Spillovers: Evidence from Municipal Building Policies in California. *Journal of Environmental Economics and Management* 68: 411–434.
- Stavins, Robert, Jonathan Borck, and Todd Schatzki. 2010. *Options for Addressing Leakage in California’s Climate Policy*. <https://www.analysisgroup.com/Insights/publishing/options-for-addressing-leakage-in-california-s-climate-policy/>.
- U.S. Department of Energy. 2014. *Transmission Constraints and Congestion in the Western and Eastern Interconnections*. <https://www.energy.gov/oe/downloads/transmission-constraints-and-congestion-western-and-eastern-interconnections-2009-2012>.
- U.S. Energy Information Administration. 2020. *As U.S. Coal-Fired Capacity and Utilization Decline, Operators Consider Seasonal Operation*. <https://www.eia.gov/todayinenergy/detail.php?id=44976>.
- , 2022a. *Average Cost of Fossil Fuels for Electricity Generation*. <http://www.eia.gov/electricity/data.cfm>.
- , 2022b. *Form EIA-860*. <http://www.eia.gov/electricity/data/eia860/>.
- , 2022c. *Form EIA-923*. <http://www.eia.gov/electricity/data/eia923/>.
- , 2022d. *Form EIA-930*. <https://www.eia.gov/beta/electricity/gridmonitor/about>.
- U.S. Environmental Protection Agency. 2020. *AVoided Emissions and geneRation Tool (AVERT)*. https://www.epa.gov/sites/production/files/2020-09/documents/avert_user_manual_09-12-20_508.pdf.
- , 2022. *Air Market Program Data*. <http://ampd.epa.gov/ampd/>.

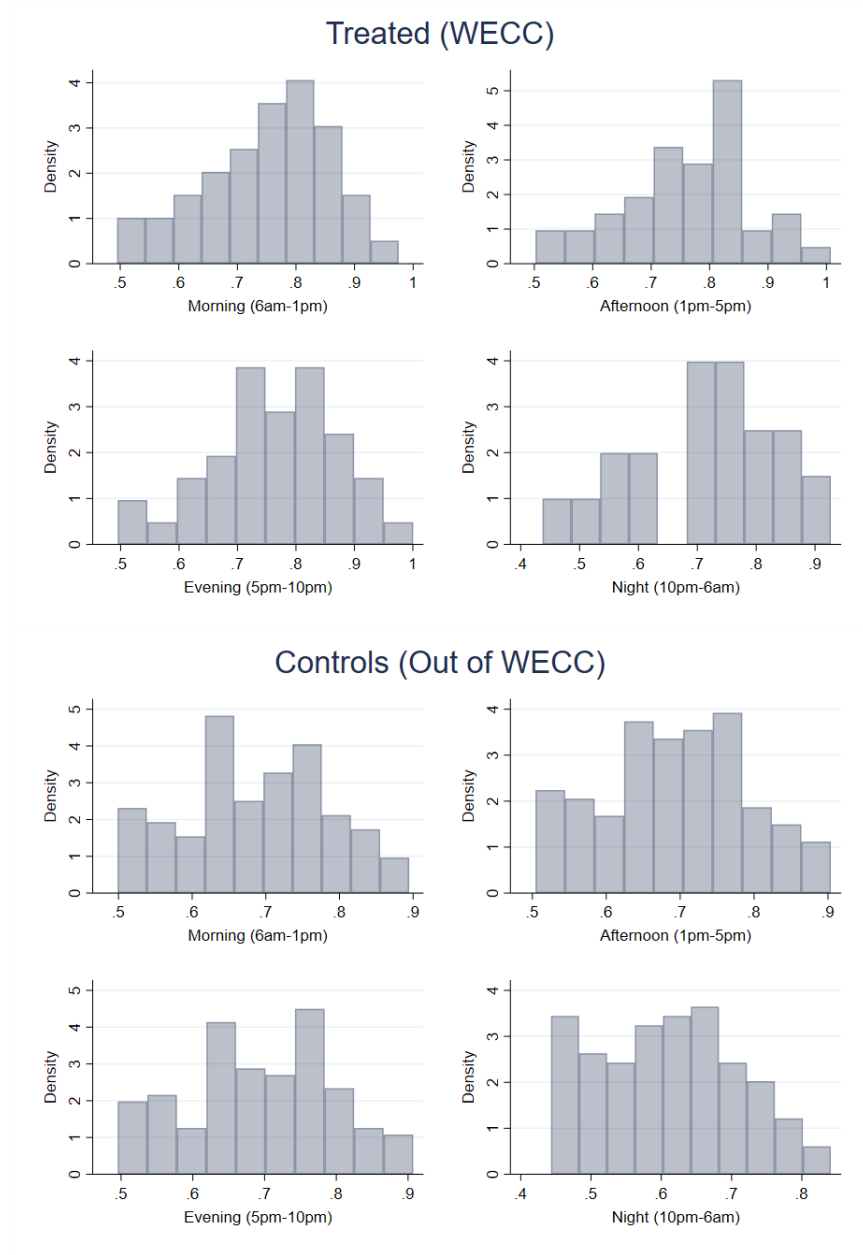
- WECC Staff. 2015. *Loads and Resources Methods and Assumptions*. http://www.wecc.biz/Reliability/2015LAR_MethodsAssumptions.pdf.
- , 2016. *WECC 2026 Common Case Version 1.5*. <http://www.wecc.org/SystemAdequacyPlanning/Pages/Datasets.aspx>.
- Xu, Qingyu, and Benjamin Hobbs. 2021. Economic Efficiency of Alternative Border Carbon Adjustment Schemes: A Case Study of California Carbon Pricing and the Western North American Power Market. *Energy Policy* 156, article 112463.
- Zhou, Yihsu, and Ling Huang. 2021. How Regional Policies Reduce Carbon Emissions in Electricity Markets: Fuel Switching or Emission Leakage. *Energy Economics* 97, article 105209.

Figure 1: Empirical distribution of 2009-10 average capacity factors by technology, region and block of hour



Note: Histograms for the control plants are constrained to include capacity factors below 0.797, 0.810, 0.823, and 0.758 corresponding to the highest capacity factor of WECC plants (i.e., the upper limit of the last matching bin for the treated plants) in the morning, afternoon, evening and night period, respectively.

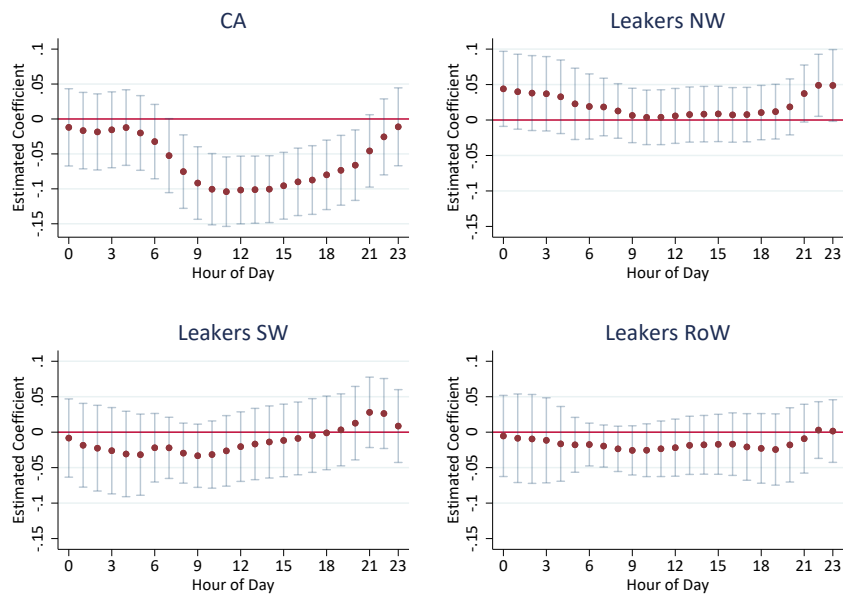
(b) Coal



Note: Histograms for the control plants are constrained to include capacity factors above 0.495, 0.502, 0.495, and 0.437 corresponding to the lowest capacity factor of WECC plants (i.e., the lower limit of the first matching bin for the treated plants) in the morning, afternoon, evening and night period, respectively.

Figure 2: Treatment heterogeneity by hour of day based on specification (1)

(a) NGCC



(b) Coal

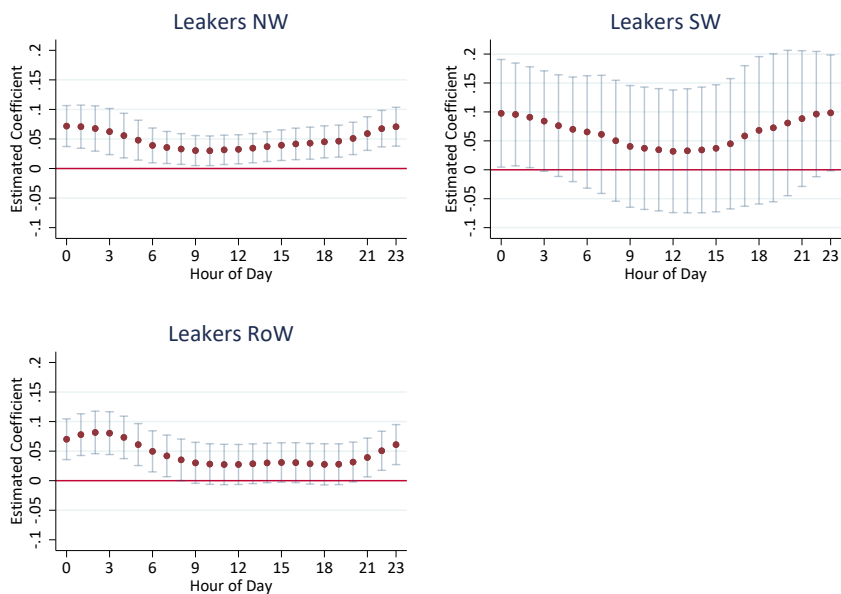
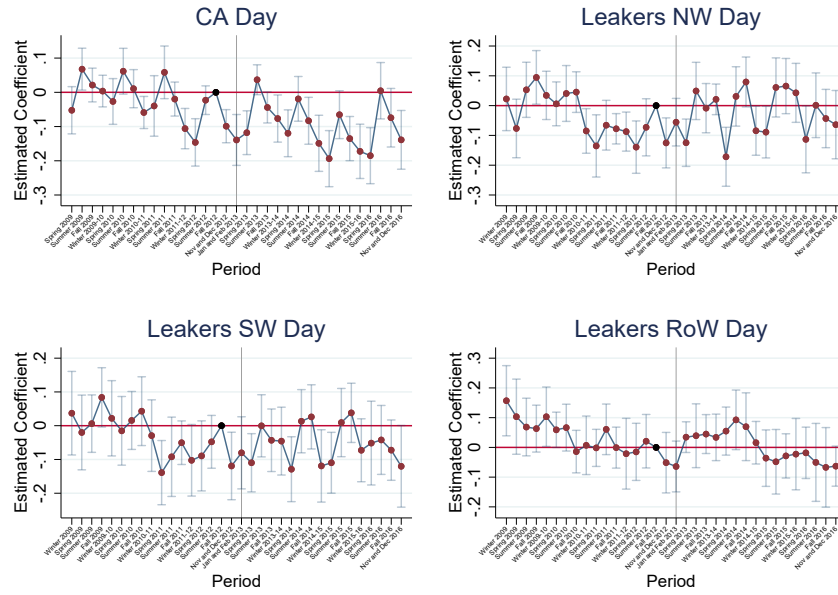
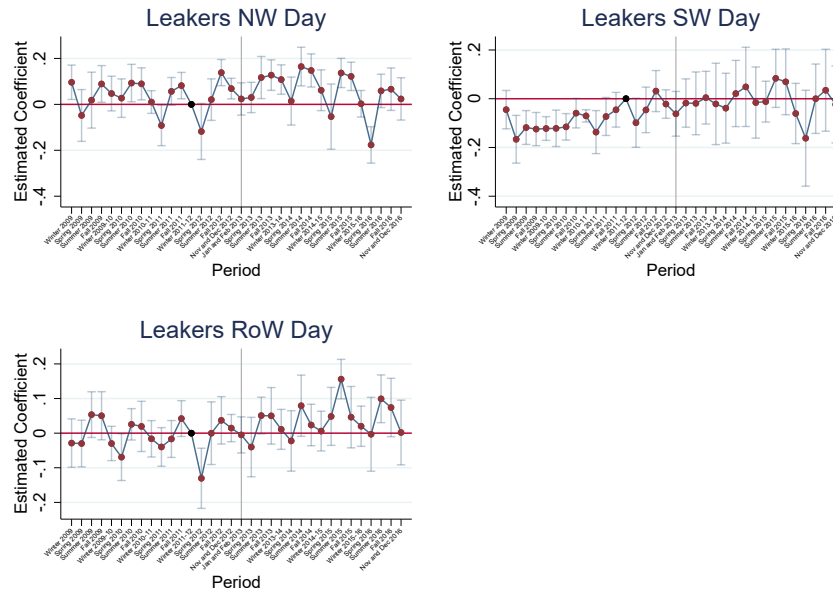


Figure 3: Parallel trend tests between treated and control regions - Day

(a) NGCC



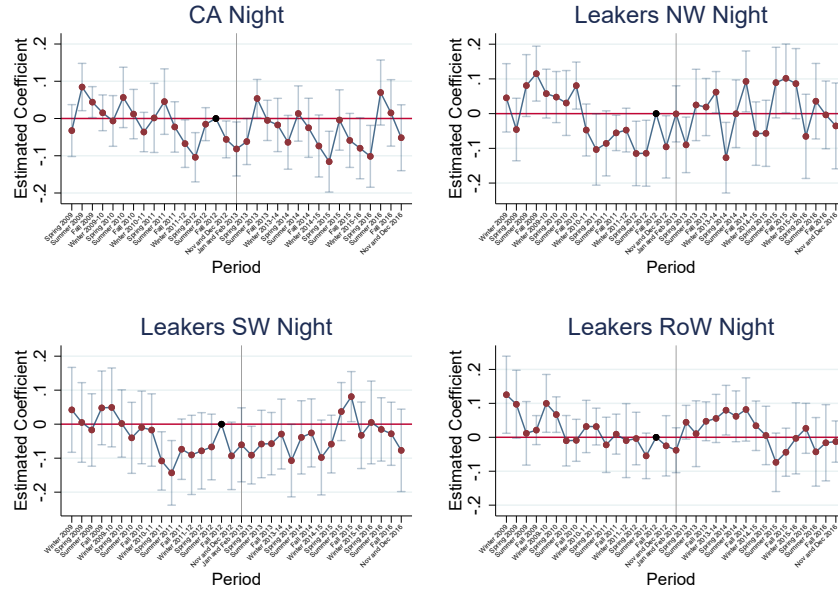
(b) Coal



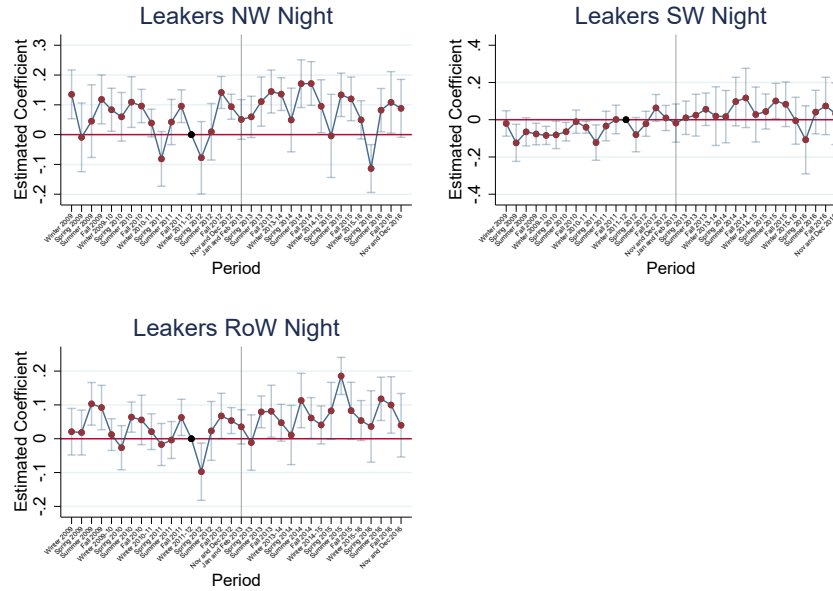
Note: The vertical line indicates the start date of the cap-and-trade program. In the CA Day plot, Winter 2009 is dropped due to lack of import data until April 1, 2009. The reference period (Fall 2012 for NGCC and Winter 2011-12 for Coal) is represented by a black dot at zero and no confidence interval, and corresponds to the reference period of choice for robust inference using Rambachan and Roth (2022a)'s approach.

Figure 4: Parallel trend tests between treated and control regions - Night

(a) NGCC

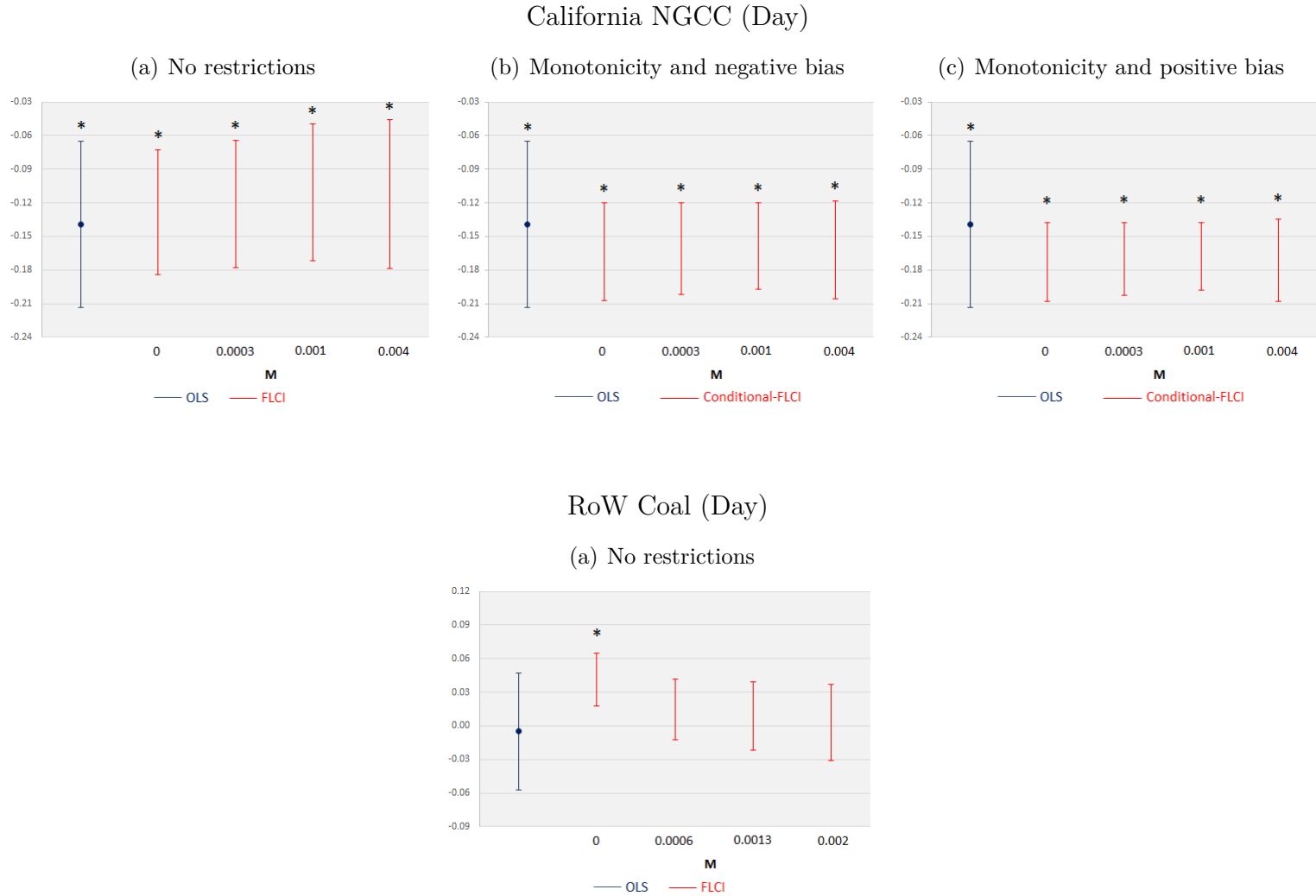


(b) Coal



Note: The vertical line indicates the start date of the cap-and-trade program. In the CA Night plot, Winter 2009 is dropped due to lack of import data until April 1, 2009. The reference period (Fall 2012 for NGCC and Winter 2011-12 for Coal) is represented by a black dot at zero and no confidence interval, and corresponds to the reference period of choice for robust inference using Rambachan and Roth (2022a)'s approach.

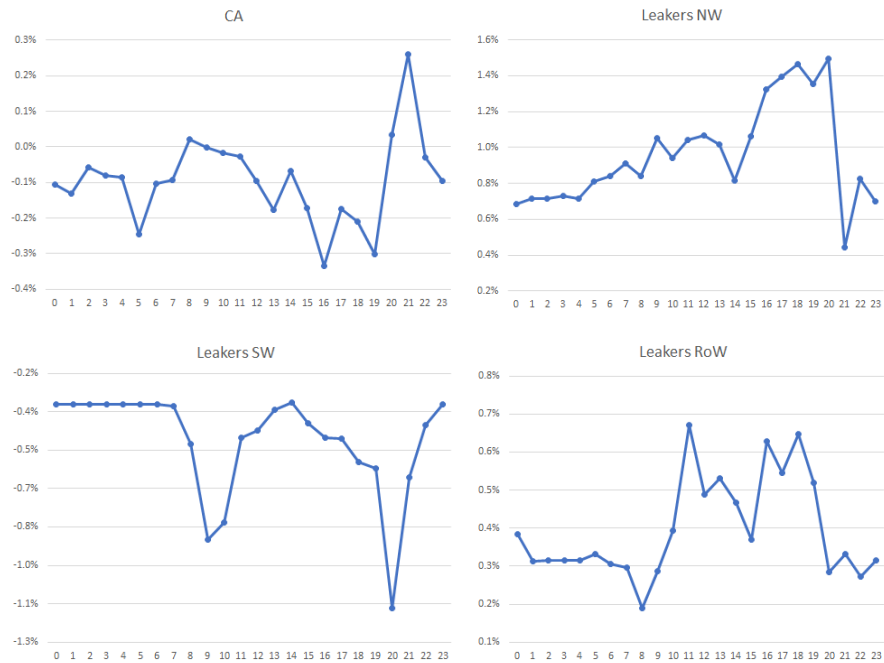
Figure 5: Sensitivity analysis for the event study coefficients in Jan-Feb 2013



Note: In each panel, “OLS” refers to the 95% confidence intervals for Jan-Feb 2013 treated effect estimated using OLS. “FLCI” (“Conditional FLCI”) indicates the 95% fixed length confidence interval (conditional fixed length confidence interval) using the Rambachan and Roth (2022a) robust inference method. Stars indicate intervals that do not cross zero.

Figure 6: Change in average capacity factors by region, tech type and hour of day based on 2013 simulation results

(a) NGCC



(b) Coal

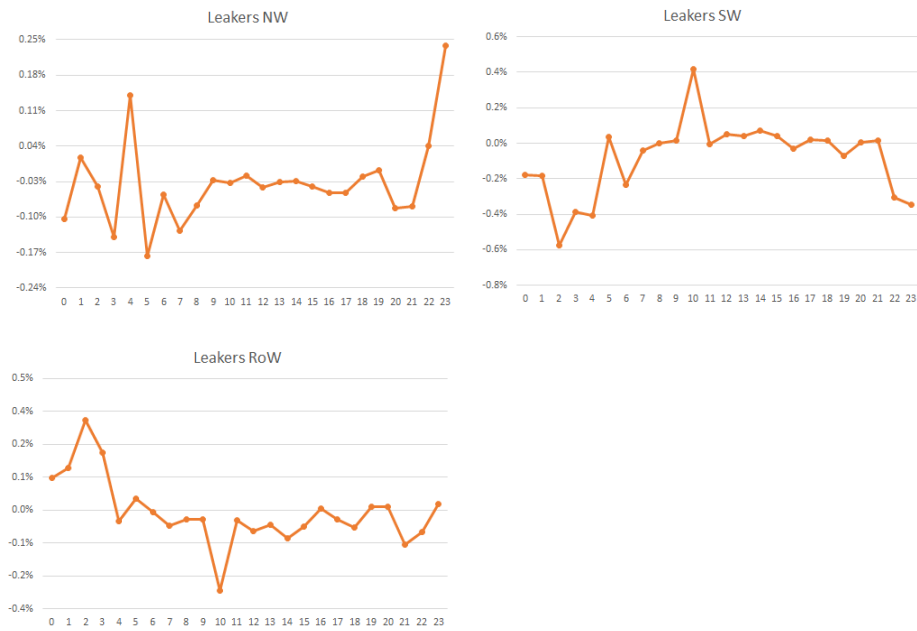
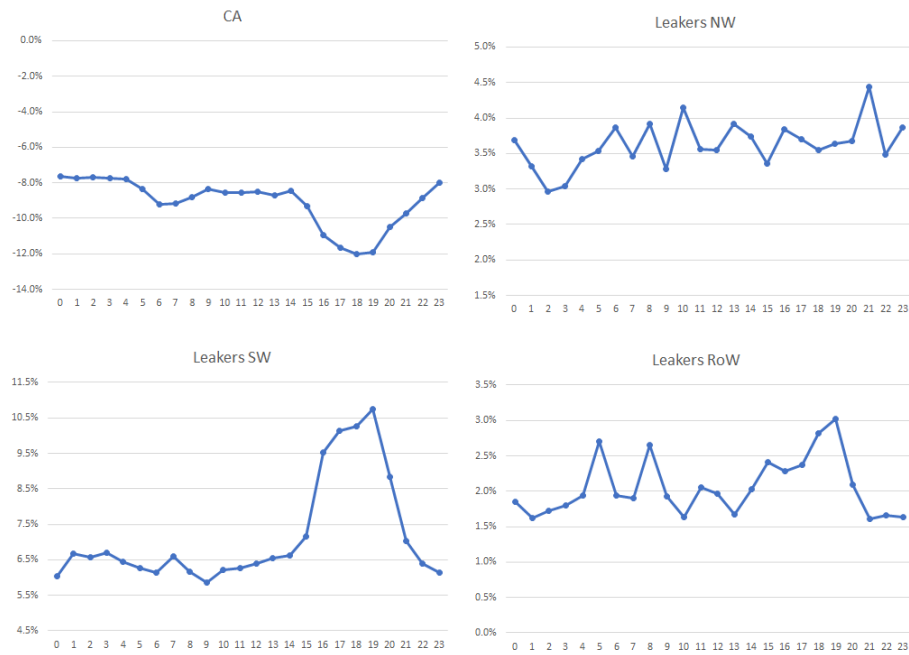


Figure 7: Change in average capacity factors by region, tech type and hour of day based on 2016 simulation results

(a) NGCC



(b) Coal

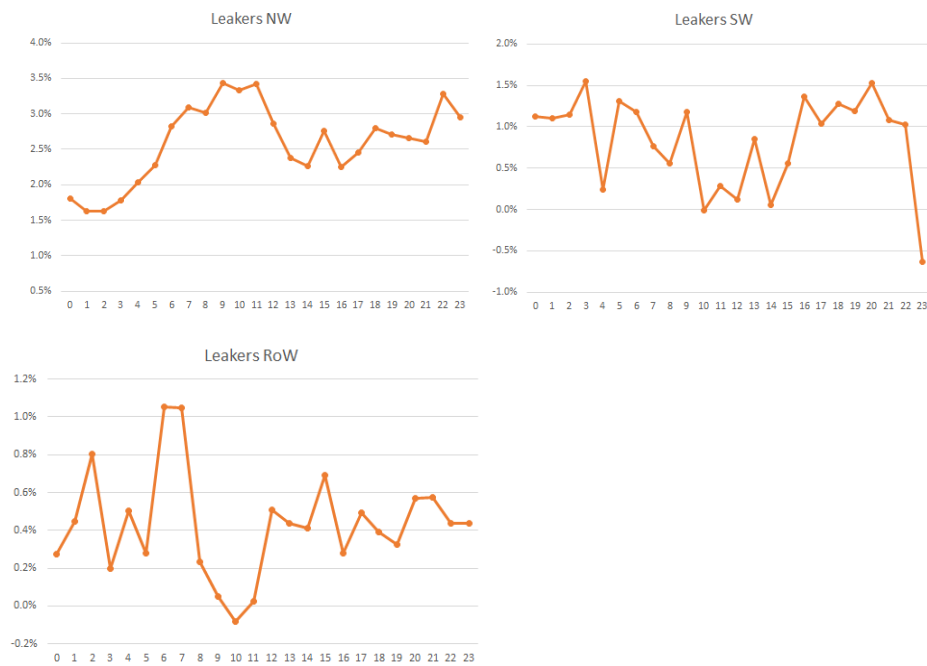


Table 1: Econometric model results: Treatment effects based on hourly capacity factors

	(1)		(2)		(3)		(4)
	NGCC	Coal	NGCC	Coal	NGCC	Coal	NGCC
	<i>Spline coeff. invariant across markets</i>		<i>Spline coeff. vary by interconnection, controls for hydro, nuclear and renewable generation at two additional geographic scales (outside-BA but in-region, and outside-CA but in-WECC)</i>		<i>Spline coeff. vary by interconnection, renewable spline coeff. also vary by time of day</i>		<i>Hydro spline coeff. invariant across markets, nuclear spline coeff. vary by interconnection, renewable spline coeff. vary by interconnection and time of day</i>
<i>Day</i>							
CA	−0.089*** (0.024)	- -	−0.057** (0.025)	- -	−0.063** (0.026)	- -	−0.049** (0.024)
NW	0.008 (0.019)	0.035*** (0.012)	0.019 (0.020)	0.037** (0.014)	0.028 (0.020)	0.045*** (0.014)	0.019 (0.020)
RoW	−0.020 (0.019)	0.031* (0.017)	−0.014 (0.021)	0.038** (0.016)	0.001 (0.020)	0.043*** (0.016)	−0.003 (0.020)
SW	−0.018 (0.023)	0.050 (0.061)	−0.004 (0.026)	0.060 (0.056)	−0.001 (0.026)	0.075 (0.056)	−0.008 (0.025)
<i>Night</i>							
CA	−0.028 (0.026)	- -	0.002 (0.026)	- -	−0.026 (0.026)	- -	−0.012 (0.023)
NW	0.033 (0.022)	0.058*** (0.015)	0.043* (0.026)	0.060*** (0.015)	0.047* (0.024)	0.048*** (0.014)	0.038 (0.024)
RoW	−0.011 (0.020)	0.059*** (0.017)	−0.004 (0.024)	0.066*** (0.016)	0.002 (0.023)	0.054*** (0.016)	−0.002 (0.023)
SW	−0.006 (0.025)	0.091 (0.057)	0.006 (0.027)	0.102* (0.051)	0.011 (0.027)	0.091* (0.052)	0.003 (0.026)
Observations	13,619,137	9,096,375	13,619,137	9,096,375	13,619,137	9,096,375	13,619,137
Clusters	201	134	201	134	201	134	201

Note: The unit of observation is plant-hour for all specifications. All regressions include plant, time-of-day, day-of-week, year and state by month-year fixed effects, where time-of-day is defined as in Section 2.5.1. Standard errors are clustered by plant and reported in parentheses.

Table 2: Parallel trends test: Treatment effects with and without trends

	No trend		Linear trend		Quadratic trend	
	NGCC	Coal	NGCC	Coal	NGCC	Coal
<i>Day</i>						
CA	−0.089*** (0.024)	- -	−0.018 (0.027)	- -	−0.013 (0.027)	- -
NW	0.008 (0.019)	0.035*** (0.012)	- -	0.098*** (0.030)	- -	0.099*** (0.030)
RoW	−0.020 (0.019)	0.031* (0.017)	- -	0.026 (0.021)	- -	0.025 (0.021)
SW	−0.018 (0.023)	0.050 (0.061)	- -	- -	- -	- -
<i>Night</i>						
CA	−0.028 (0.026)	- -	- -	- -	- -	- -
NW	0.033 (0.022)	0.058*** (0.015)	- -	0.121*** (0.026)	- -	0.122*** (0.026)
RoW	−0.011 (0.020)	0.059*** (0.017)	- -	0.054*** (0.020)	- -	0.053*** (0.020)
SW	−0.006 (0.025)	0.091 (0.057)	- -	- -	- -	- -

Note: If a treatment effect is statistically significant in the baseline regression (“No trend”), we augment the model with a group-specific linear (quadratic) trend for all treated regions. The table presents the estimated coefficients in the augmented models. Standard errors are clustered by plant and reported in parentheses.

Table 3: Econometric model results: Emissions and leakage based on the econometric estimates, 2013 and 2016

<i>2013</i>	Lower bound of the robust 95% CI	Upper bound of the robust 95% CI
Change in CA local emissions (E_1)	-6.84	-2.07
Change in CA import emissions (E_2)	-10.87	-10.87
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	-17.71	-12.94
Change in WECC-NonCA emissions (E_4)	-3.65	17.33
- NW	1.26	12.82
- RoW	-4.91	4.51
Change in WECC emissions ($E_5 = E_1 + E_4$)	-10.49	15.26
<i>Leakage</i> $[(1 - E_5/E_3) \times 100\%]$	<i>40.8%</i>	<i>217.9%</i>
<i>2016</i>	Lower bound of the robust 95% CI	Upper bound of the robust 95% CI
Change in CA local emissions (E_1)	-7.09	-2.86
Change in CA import emissions (E_2)	-11.81	-11.81
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	-18.90	-14.67
Change in WECC-NonCA emissions (E_4)	-8.51	14.21
- NW	-4.54	7.27
- RoW	-3.97	6.94
Change in WECC emissions ($E_5 = E_1 + E_4$)	-15.60	11.35
<i>Leakage</i> $[(1 - E_5/E_3) \times 100\%]$	<i>17.4%</i>	<i>177.4%</i>

Note: Emissions are in million metric tons of CO₂ per year. Results are based on robust confidence intervals for CA NGCC Day and RoW Coal Day that allow for linear violations of parallel trends (M=0), and OLS confidence intervals for CA NGCC Night, NW Coal Day and Night, and RoW Coal Night. We assume negative bias and monotonicity restrictions for CA NGCC Day, and no restrictions for RoW Coal Day. Only contemporaneous (i.e., same season and time of day) changes in capacity factors in the regulated and unregulated regions are included for the leakage calculation.

Table 4: Simulation model results: Effect of California’s cap-and-trade program on capacity factors in WECC, 2013 and 2016

		2013			2016		
		No cap	Carbon cap	Δ	No cap	Carbon cap	Δ
CA	Hydro	24.9%	24.9%	-	30.1%	30.1%	-
	NGCC	50.2%	49.5%	-0.7%	43.3%	33.8%	-9.6%
	NGCT	6.1%	5.5%	-0.6%	4.4%	4.4%	-
	Nuclear	95.0%	95.0%	-	95.0%	95.0%	-
	Oil	5.3%	5.3%	-	5.3%	5.3%	-
	Solar	9.8%	9.8%	-	19.2%	19.2%	-
	Wind	19.9%	19.9%	-	19.7%	19.7%	-
NW	Coal	87.5%	87.5%	-	76.6%	79.2%	+2.6%
	Hydro	29.9%	29.9%	-	31.0%	31.0%	-
	NGCC	56.9%	57.9%	+1.0%	59.0%	62.5%	+3.5%
	NGCT	1.6%	2.0%	+0.4%	1.6%	2.8%	+1.2%
	Nuclear	95.0%	95.0%	-	95.0%	95.0%	-
	Oil	94.0%	94.0%	-	94.0%	94.0%	-
	Solar	21.4%	21.4%	-	20.5%	20.5%	-
RoW	Wind	16.8%	16.8%	-	17.6%	17.6%	-
	Coal	87.1%	87.1%	-	79.2%	79.6%	+0.4%
	Hydro	32.0%	32.0%	-	38.5%	38.5%	-
	NGCC	19.5%	19.9%	+0.4%	33.1%	35.2%	+2.1%
	NGCT	0.9%	1.0%	+0.1%	4.1%	4.7%	+0.6%
	Nuclear	-	-	-	-	-	-
	Oil	18.2%	18.2%	-	18.2%	18.2%	-
SW	Solar	16.7%	16.7%	-	15.0%	15.0%	-
	Wind	29.9%	29.9%	-	31.9%	31.9%	-
	Coal	83.3%	83.2%	-0.1%	59.4%	60.2%	+0.8%
	Hydro	32.1%	32.1%	-	37.0%	37.0%	-
	NGCC	19.5%	19.0%	-0.5%	34.1%	41.3%	+7.2%
	NGCT	1.0%	1.1%	+0.1%	1.0%	1.3%	+0.3%
	Nuclear	95.0%	95.0%	-	95.0%	95.0%	-
	Oil	-	-	-	-	-	-
	Solar	17.5%	17.5%	-	26.0%	26.0%	-
	Wind	9.3%	9.3%	-	9.9%	9.9%	-

Note: The oil-fired capacity in the NW region is only 13.80 MW, resulting in high capacity factors for this peak technology.

Table 5: Simulation model results: Effect of California’s cap-and-trade program on emissions and leakage, 2013 and 2016

<i>2013</i>	No cap	Carbon cap	
Local emissions in CA	39.79	38.68	
Emissions of CA imports	51.25	39.41	
Regulated emissions in CA	91.03	78.10	
Emissions in WECC-NonCA	265.42	265.78	
- NW	113.28	113.78	
- RoW	90.99	91.06	
- SW	61.15	60.94	
Total emissions in WECC	305.21	304.47	
Change in CA local emissions (E_1)			-1.10
Change in CA import emissions (E_2)			-11.83
Change in CA regulated emissions ($E_3 = E_1 + E_2$)			-12.94
Change in WECC-NonCA emissions (E_4)			+0.36
Change in WECC emissions (E_5)			-0.74
<i>Leakage</i> $[(1 - E_5/E_3) \times 100\%]$			<i>94.3%</i>
<i>2016</i>	No cap	Carbon cap	
Local emissions in CA	35.85	29.27	
Emissions of CA imports	38.59	24.71	
Regulated emissions in CA	74.45	53.99	
Emissions in WECC-NonCA	232.81	241.44	
- NW	102.58	107.15	
- RoW	83.39	84.31	
- SW	46.84	49.98	
Total emissions in WECC	268.67	270.71	
Change in CA local emissions (E_1)			-6.58
Change in CA import emissions (E_2)			-13.88
Change in CA regulated emissions ($E_3 = E_1 + E_2$)			-20.46
Change in WECC-NonCA emissions (E_4)			+8.62
Change in WECC emissions (E_5)			+2.04
<i>Leakage</i> $[(1 - E_5/E_3) \times 100\%]$			<i>110.0%</i>

Note: Emissions are in million metric tons of CO₂ per year.

Table 6: Comparison of emissions and leakage results, 2013 and 2016

<i>2013</i>	Simulation model	Econometric model 95% CI, No restrictions		Econometric model 95% CI, With restrictions	
		LB	UB	LB	UB
Change in CA local emissions (E_1)	−0.51	−6.33	−0.35	−6.84	−2.07
Change in CA import emissions (E_2)	−11.83	−10.87	−10.87	−10.87	−10.87
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	−12.35	−17.20	−11.22	−17.71	−12.94
Change in WECC-NonCA emissions (E_4)	−0.05	−3.65	17.33	−3.65	17.33
Change in WECC emissions ($E_5 = E_1 + E_4$)	−0.56	−9.98	16.98	−10.49	15.26
<i>Leakage</i> $[(1 - E_5/E_3) \times 100\%]$	<i>95.5%</i>	<i>42.0%</i>	<i>251.4%</i>	<i>40.8%</i>	<i>217.9%</i>
<i>2016</i>	Simulation model	Econometric model 95% CI, No restrictions		Econometric model 95% CI, With restrictions	
		LB	UB	LB	UB
Change in CA local emissions (E_1)	−6.35	−5.73	0.31	−7.09	−2.86
Change in CA import emissions (E_2)	−13.88	−11.81	−11.81	−11.81	−11.81
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	−20.24	−17.54	−11.50	−18.90	−14.67
Change in WECC-NonCA emissions (E_4)	2.97	−7.75	11.36	−8.51	14.21
Change in WECC emissions ($E_5 = E_1 + E_4$)	−3.38	−13.48	11.67	−15.60	11.35
<i>Leakage</i> $[(1 - E_5/E_3) \times 100\%]$	<i>83.3%</i>	<i>23.1%</i>	<i>201.5%</i>	<i>17.4%</i>	<i>177.4%</i>

Note: Changes in emissions are in million metric tons of CO₂ per year. When calculating leakage from the empirical estimates, we use robust confidence intervals for CA NGCC Day and RoW Coal Day (allowing for linear violations of parallel trends, or $M = 0$), and OLS confidence intervals for CA NGCC Night, NW Coal Day and Night, and RoW Coal Night. LB indicates the lower bound of the 95% confidence interval, while UB refers to its upper bound. Only contemporaneous (i.e., same season and time of day) changes in capacity factors in the regulated and unregulated regions are included for the leakage calculation. In the model with restrictions, we assume negative bias and monotonicity restrictions for CA NGCC Day, and no restrictions for RoW Coal Day.