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*

August 1, 2021

Abstract

We seek to identify CO_2 emission leakage in the electricity sector from California's AB 32 cap-and-trade program based on two methods: a quasi-experimental econometric model and simulations conducted using a partial equilibrium model. Leakage predictions are based on the estimated effects on electricity generation in the Western Interconnection from the two models. Simulation results provide a benchmark against which to evaluate the results of the empirical analysis, in a policy setting where threats to identification undermine attempts at statistical inference. Both models suggest that leakage is occurring, with rates that are within the range of predictions from earlier studies analyzing the potential for leakage in this policy setting. There are, however, differences in the estimated effects on electricity generation and predicted source of leakage in WECC. These differences may be partly driven by modeling assumptions, but urge caution in lending a causal interpretation to the empirically estimated generation shifts.

*Lo Prete: John and Willie Leone Family Department of Energy and Mineral Engineering, The Pennsylvania State University, University Park, PA (email: chiaraloprete@psu.edu); Tyagi: Frankfurt School of Finance & Management, Frankfurt am Main (a.tyagi@fs.de); Xu: Andlinger Center For Energy And Environment at Princeton University (qingyux@princeton.edu). Cody Hohl provided excellent research assistance. We thank Karen Fisher-Vanden, Ben Hobbs and Oliver Schenker for helpful suggestions and encouragement. 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 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.

1 Introduction

California has been at the forefront of U.S. environmental policies for years. The Global Warming Solutions Act of 2006 (also known as Assembly Bill 32 or AB 32) set the state's target to reduce greenhouse gas (GHG) emissions to 1990 levels by 2020. In September 2016, California passed Senate Bill 32 (SB 32), which limited emissions to 40% below 1990 levels by 2030. Further, Executive Order S-3-05 set a GHG emission reduction target of 80% below 1990 levels by 2050. In order to achieve these ambitious goals, the state relies on a suite of policies, including a multi-sector cap-and-trade program that covers about 85% of the state's emissions from large industrial facilities, electricity generators and importers, and transportation fuel suppliers.

A central issue in the implementation of cap-and-trade programs is represented by the choice of the point of regulation. For example, the Regional Greenhouse Gas Initiative (RGGI), an emission trading system for CO_2 emissions from electricity generation in U.S. Northeastern and mid-Atlantic states, adopted a source-based approach where the point of regulation is at the generator level. Given its reliance on imports to satisfy electricity consumption,[1] California opted instead 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). However, simulation-based studies quantifying the impacts of the prospective cap-and-trade scheme concluded that contract reshuffling may enable substantial leakage (Bushnell, Peterman and Wolfram, 2008; Fowlie, 2009; Chen et al., 2011; Bushnell et al., 2014). Under resource shuffling, electricity contracts are 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). This would result in apparent emission reductions due to changes in the composition of imports to California, 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.

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

In this paper, we seek to identify CO_2 emission leakage in the electricity sector from California's AB 32 cap-and-trade program in the first four years of policy implementation. To this end, we estimate shifts in electricity generation at baseload power plants in the Western Interconnection[2] 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 CO_2 emission leakage predictions in 2013 and 2016. The comparison of results from simulation and the empirical analysis allows us to assess the plausibility of our *ex post* leakage predictions in a policy setting where threats to identification (i.e., the suite of changes that affected California's electricity market over the period of our study, and challenges associated with the construction of credible counterfactual outcomes) undermine attempts at statistical inference. 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, ours is the first attempt to benchmark 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 remainder of the paper is organized as follows. Section 2 reviews the literature on emission leakage. Section 3 provides an overview of California's cap-and-trade program and complementary emission reduction measures. Section 4 presents the econometric and simulation models, while Section 5 describes the sources of data and the steps taken to ensure comparability. Section 6 discusses the results, and Section 7 concludes.

2 Literature review

The potential for emission leakage in the electricity sector under regional climate policies has been analyzed using numerical models. A first strand of the literature includes simulation-based partial equilibrium models of the electricity sector. For example, Fowlie (2009), Chen et al. (2011), Bushnell and Chen (2012) and Bushnell et al. (2014) explore leakage in the context of California's prospective cap-and-trade program for GHG emissions (i.e., before regulations were finalized). Fowlie (2009) presents an analytical model to develop intuition about how industry features affect emission leakage and social welfare in an incompletely regulated

[2]California 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 1 presents the U.S. part of WECC.

and imperfectly competitive industry. Based on the theoretical framework, she develops a numerical model to simulate CO_2 emissions and other equilibrium outcomes in California's electricity sector under three carbon policy scenarios and alternative assumptions about firm conduct. Results suggest that a cap-and-trade program that only applies to in-state electricity producers would achieve about a third of emission reductions obtained under complete regulation at a higher cost per ton. Chen et al. (2011) formulate a market equilibrium model to compare source-based, load-based and first deliverer approaches for cap-and-trade regulation in California, and examine economic and emission implications on the electricity market. Under the first deliverer approach that was ultimately pursued in California, total emissions in the Western Interconnection decrease much less than regulated emissions, implying a leakage rate of 85%. Simulation results also indicate that significant reshuffling takes place under this scenario: while the emissions of electricity imports to California decrease due to changes in their composition, the emissions in the rest of the Western Interconnection actually increase under regulation. In a similar vein, Bushnell and Chen (2012) and Bushnell et al. (2014) simulate the impacts of CO_2 caps that only apply to California or cover all states in the Western U.S. under alternate assumptions, and conclude that a first deliverer approach in California is vulnerable to leakage due to laundering and reshuffling of import resources. Finally, 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 impacts of regional climate policies have also been studied with computable general equilibrium models (Carbone and Rivers, 2017). While the partial equilibrium models discussed above focus on short-run operations, general equilibrium models are useful to illustrate the impacts of carbon pricing policies on other sectors of the economy, and may consider the effects on long-run capital decisions (Shawhan et al., 2014). In the context of California's cap-and-trade program, Caron et al. (2015) find that the policy would result in only a small amount of emission leakage (9%), when imported electricity is included in the cap and provisions to prevent reshuffling are enforced. Without such measures, the authors estimate that 45% of emission reductions would be offset by leakage, in line with the range of predictions obtained by general equilibrium models for RGGI (Shawhan et al., 2014; Sue Wing and Kolodziej, 2009).

Empirical analyses of leakage are less common in the literature. For example, Aichele and Felbermayr (2013a), Aichele and Felbermayr (2013b) and Aichele and Felbermayr (2015) examine leakage in the context

of the Kyoto Protocol. With respect to RGGI, Kindle et al. (2011) analyze the relation between CO_2 permit prices and transmission power flows on seven high-voltage interties between New York and Pennsylvania between 2008 and 2010. Higher net flows from Pennsylvania to New York associated with a higher RGGI allowance price would indicate leakage. The authors do not find a significant impact of RGGI permit prices on PA-NY transmission flows, but prices may have been too low to affect leakage in the early years of the program. Fell and Maniloff (2018) use a differences-in-differences model to estimate how RGGI affected the operations of power plants 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. The implied leakage rate of approximately 50% is within the range of *ex post* leakage predictions from subsequent empirical analyses (Zhou and Huang, 2021). Chan and Morrow (2019) also investigate leakage from RGGI, but their analysis focuses on SO_2 emissions and associated damages, instead of CO_2 emissions.

Finally, a growing body of research in economics assesses the potential for leakage risk across sectors (Fowlie and Reguant, 2018), and explores how environmental regulation affects trade flows and the location choice of firms in the long run (Levinson and Taylor, 2008; Kahn and Mansur, 2013; Aldy and Pizer, 2015; Fowlie et al., 2016; Panhans et al., 2017; Saussay and Sato, 2018).

Our paper is most closely related to Xu and Hobbs (2021) and Fell and Maniloff (2018), but differs from these earlier contributions to the literature in several important ways. The version of JHSMINE in this paper differs from the one in Xu and Hobbs (2021): to enable more meaningful comparisons with the econometric results, we revise the model formulation to accommodate the observed cap-and-trade regime, and parameterize the model using actual market data in 2013-2016. Relative to Fell and Maniloff (2018), 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. However, 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.

3 Policy background

California adopted legislation limiting GHGs by passing AB 32, which established a statewide target of reducing GHG emissions to their 1990 levels (431 million metric tons of CO_2e (California Air Resources Board, 2014)) by 2020. To achieve the expected emission reductions to meet the 2020 limit (79 million metric tons of CO_2e), 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 to provide a firm cap on the sectors responsible for the majority of California's GHG emissions (i.e., transportation, electricity and industrial sectors). 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). The next sections provide an overview of the cap-and-trade-program and complementary measures aimed at reducing emissions from the energy sector.

3.1 Cap-and-trade program overview

California's cap-and-trade program regulates GHG emissions from large industrial facilities, electricity generators and importers, and transportation fuel suppliers. Covered entities emit at least 25,000 metric tons of CO_2e per year and are responsible for about 85% of the state's 2015 GHG emissions (California Air Resources Board, 2010, 2013). The first phase of compliance for the program began on January 1, 2013. The 2013 emission cap was set at approximately 98% of forecast 2012 emissions, with an annual decline of 2% in 2014 and 3% from 2015 through 2020. Based on CARB projections, the program was expected to drive about 23% of emission reductions needed to reach 1990 levels (California Air Resources Board, 2013). In July 2017, the scheme was extended through 2030 with bipartisan support.

CARB issues annual emission allowances equal to the cap, and each allowance represents a permit to emit one ton of carbon dioxide equivalent. Entities must monitor and annually report their emissions, and return an amount of allowances equivalent to their GHG emissions each year. Capped sources that keep emissions below the allowance amount can sell excess permits on the market, while sources that cannot cover total

emissions may take measures to reduce pollution and/or buy allowances on the market.[3] Emission allowances are distributed to covered entities through a mix of free allocation and quarterly auctions, with a declining share of free permits over time (California Air Resources Board, 2021b): in 2016, the last year in our period of study, 50% of allowances were given away for free, and 46% were auctioned (Legislative Analyst's Office, 2017).[4] California adopts an output-based benchmarking allocation approach for most industrial sectors, whereby the allocation of allowances is set at 90% of average emissions, based on benchmarks that reward efficient facilities, and updated annually according the production at each facility. Allowances are allocated to electric distribution utilities (EDUs) based on historical emissions, current generation mix, sales, and efforts at reducing emissions since the passage of AB 32 (Alcorn, 2013). Investor-owned utilities (IOUs) must consign their allowances to auction, while publicly owned utilities (POUs) may put their allowances up for auction or use them to meet compliance obligations. The revenue from these auctions must be used to provide rebates or bill relief to utility customers (Legislative Analyst's Office, 2017). CARB allows banking and borrowing of allowances, and the risk of unexpected price changes and excess volatility is mitigated through a price collar; secondary market allowance prices have generally hovered at or near the auction price floor from market launch to 2016 (Cullenward and Coghlan, 2016).

Under California's approach to regulate the electricity sector, the first entity that delivers load to the California grid has a compliance obligation. When electricity generation occurs within the state, generators must submit compliance instruments (allowances and offset credits) for the associated emissions. When electricity generation occurs out of state, the compliance obligation falls instead on the electricity importers, who must submit compliance instruments to cover the emissions generated for each MWh of imported electricity. Since energy entering the grid flows over the path of least resistance (rather than directly from an injection point to a withdrawal point), the CO_2 intensity of electricity imported in California from the rest of the Western Interconnection cannot generally be determined unambiguously.[5] To address the issue, CARB classifies imports as specified or unspecified source power.

Specified sources include generation resources owned by or under long-term contract to California's load serving entities, as well as generation resources owned by non-California entities that are approved and registered by CARB (California Air Resources Board, 2018b). First deliverers may claim facility-specific emission factors for power imports from out-of-state generation resources that are owned or under long-term

[3]Covered entities may also use carbon offsets (e.g., GHG emission reductions from projects outside the scope of the cap-and-trade regulation) to cover up to 8% of their emissions.

[4]The remaining 4% were made available at predetermined prices to reduce price volatility.

[5]The e-Tag functionality has been used to track proof of energy delivery within the Western Interconnection, but is complex to implement (WECC Staff, 2019).

contract. Further, CARB has developed the designation of Asset-Controlling Suppliers for out-of-state electric power entities that operate interconnected generating facilities. Once approved and registered by CARB, Asset-Controlling Suppliers are assigned a system emission factor for wholesale electricity procured from their systems and imported into California. For example, specified source power from Bonneville Power Administration (BPA) and Powerex (a subsidiary of BC Hydro) must be reported using CARB-approved emission factors reflecting the hydro-dominant resource portfolio of these systems (California Air Resources Board, 2018a). Specified sources mainly consist of coal, natural gas and nuclear power from the Southwest, and of hydro and wind power from the Northwest (California Energy Commission, 2017).[6]

In contrast, unspecified source power corresponds to wholesale market purchases from power plants that do not have a contract with a California utility and have not gone through the CARB process to become specified. Between 2013 and 2016, unspecified power represented about 26% of total imports, on average (California Air Resources Board, 2020a). Since in this case the generation source is unknown, unspecified sources are assigned a default emission factor of 0.428 metric ton CO_2/MWh, which was set by CARB based on the generation technology expected to be at the margin in WECC (Bushnell et al., 2014). Much of the Northwest spot market purchases are served by surplus hydro and gas-fired plants, while Southwest spot market purchases generally come from coal and natural gas combined cycles (California Energy Commission, 2017). The presence of a default emission factor creates 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 factor ("laundering"). This has been identified as one of the primary types of resource shuffling (Alcorn, 2013), defined by CARB as "any plan, scheme, or artifice undertaken by a First Deliverer of Electricity to substitute electricity deliveries from sources with relatively lower emissions for electricity deliveries from sources with relatively higher emissions to reduce its emissions compliance obligation" (Cal. Code Regs., Title 17, Article 5, § 95802(a)). As discussed in Section 1, contract shuffling would lead to apparent emission reductions due to changes in the composition of imports to California, although emissions in the exporting regions are unchanged or even increase. As a result, contract shuffling creates potentially severe leakage risks for the electricity sector in California. In response to these concerns, CARB released a guidance document listing a number of "safe harbor" exceptions to the regulatory ban on resource shuffling (i.e., transactions deemed not to be resource shuffling) (Cal. Code Regs., Title 17, Article 5, § 95852(b)(2)). This approach has been controversial because it is difficult to identify all potential violations *ex ante* (Bushnell et al., 2014).

[6]According to the Commission's definition, the Northwest region includes Alberta, British Columbia, Idaho, Montana, Oregon, South Dakota, Washington and Wyoming. The Southwest region includes Arizona, Baja California, Colorado, Mexico, Nevada, New Mexico, Texas and Utah (California Energy Commission, 2017).

Further, allowance prices hovering near the auction floor have been interpreted as evidence that contract shuffling is taking place, enabling regulated entities to avoid a significant part of their carbon liability and reducing demand for allowances (Borenstein et al., 2014; Cullenward and Coghlan, 2016).

3.2 Overview of AB 32 complementary measures in the energy sector

In addition to creating a cap-and-trade program, the initial Scoping Plan and its Update set forth a comprehensive list of recommended actions for improving energy efficiency, expanding the use of renewable energy resources and cleaner transportation, and reducing waste (California Air Resources Board, 2008, 2014). This section highlights significant measures to help achieve the state's 2020 target and support GHG emission reductions in the energy sector. However, measures that are designed to directly address GHG reductions in other sectors, which are not discussed here, may also have impacts on the energy sector (e.g., electrification in the transportation sector increases electricity demand).

A key measure to support the high-level objectives for the energy sector is California's renewable portfolio standard (RPS). Originally enacted in 2002, the RPS set a goal of achieving a statewide renewable energy mix of 20% by 2010. In 2011, Senate Bill 2 created one of the most aggressive renewable energy goals in the United States, requiring California's electric utilities and retail sellers to serve 33% of customer needs with clean renewable energy by 2020, with intermediate requirements of 20% by 2013 and 25% by 2016. The RPS, combined with programs like the California Solar Initiative and federal tax credits, spurred significant growth in utility-scale projects and customer installations over the period of our study (Bushnell and Novan, 2018). CARB estimated that achieving a 33% renewable mix by 2020 would avoid about 11 million metric tons of CO_2e (California Air Resources Board, 2013), and counted the avoided emissions towards the reductions needed to achieve the 2020 target established by AB 32.

Energy efficiency recommendations set new targets for statewide annual energy demand reductions of 32,000 GWh and 800 million therms. Strategies include, among others, more stringent building and appliance efficiency standards, and utility energy efficiency programs in the residential and non-residential sectors. Based on CARB projections, energy efficiency measures would drive approximately 12 million metric tons of CO_2e of emission reductions by 2020 (California Air Resources Board, 2013). The Scoping Plan also requires increased use of combined heat and power units by setting a goal for 4,000 MW of new installed capacity by 2020, enough to displace about 30,000 GWh of demand from other power generation sources. Finally, the Plan establishes incentive programs to promote the installation of 200,000 solar water heaters and 3,000 MW of new rooftop solar capacity (Million Solar Roofs initiative) by 2017.

4 Models

4.1 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 these estimates and robust confidence intervals, we infer CO_2 emission leakage predictions in 2013 and 2016. This section describes the designation of treated and control units, the use of coarsened exact matching to improve balance between treated and control groups, and the differences-in-differences model.

4.1.1 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 between. 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 1).[7]

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, 2021); WECC balancing authorities in the U.S. are presented in Figure 2. 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,

[7] We do not include power plants in RFC and NPCC as controls because some states in these NERC regions participate in the Regional Greenhouse Gas Initiative, the cap-and-trade program for CO_2 emissions from power generation in the Northeastern and mid-Atlantic United States.

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 1 presents summary statistics for NGCC treated and control plants, while Table 2 presents statistics for coal-fired treated and control plants. Since the period of our study was characterized by significant fuel mix changes in the California market, we also report summary statistics for nuclear, hydro and renewable generation in Table A1 of the Appendix.

4.1.2 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 capacity factors at the plant level. 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 capacity factors may be operated very differently. To 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 has zero net generation in period 1 and is operated at 80% capacity factor in period 2, while plant B is operated at 40% capacity factors in both periods. The two plants have the same average capacity factor, 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. 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 capacity factors 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. When a matching estimator (like a nearest neighbor estimator or a propensity score-based estimator) is applied, counterfactual outcomes for treated plant j are then inferred using a weighted average of the outcomes of units that are comparable to j, but receive a different treatment. Control units whose observable characteristics are closer to those of plant j are weighed more heavily in the construction of the counterfactual estimate. While earlier empirical work

in energy and environmental economics relied on parametric and semi-parametric matching methods,[8] we explore the use of coarsened exact matching (CEM) (Blackwell et al., 2009; Iacus et al., 2011, 2012) 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 on. 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 populations. Only units with the same stratum are matched.[9] 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 is a monotonic imbalance bounding method that 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. Further, unlike model dependent methods, CEM does not extrapolate counterfactual outcomes when there is limited overlap in the distributions of covariates across treatment and control groups, because matched data are restricted to areas of common empirical support. Recent applications of this matching method are presented in Simcoe and Toffel (2014), Guignet et al. (2018) 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., monthly, daily, hourly or time-of-day capacity factors). We use average capacity factors over four blocks of hours within the day, averaged over 2009-2010, as matching variables. 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.[10] For each power plant, we average hourly capacity factors over

[8] In the context of Southern California's RECLAIM program, Fowlie et al. (2012) use a semi-parametric DID matching estimator of the ATT that compares differences between post and pre treatment NO_x emissions across treated and control plants, and a regression-based adjustment to mitigate bias introduced by poor match quality (Heckman et al., 1997, 1998). As a robustness check, the authors implement a propensity score matching estimator, which relies on a parametric regression model to estimate the propensity score. One disadvantage of this approach is that a misspecified matching model may produce greater imbalance in variables that are omitted from the matching procedure.

[9] To 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.

[10] We 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.

four blocks of hours (morning, afternoon, evening, and night).[11] 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, region and block of hours are presented in Figure 3. We then define four matching bins corresponding to different levels of plant utilization in each block: matching bin 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.

4.1.3 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:

$$Y_{jt} = \alpha_C TREAT^C_{jt} + \sum_L \alpha_L TREAT^L_{jt} + \mathbf{X}^{'}_{jt}\underline{\beta} + \gamma_j + \gamma_y + \gamma_{dw} + \gamma_{sm} + \epsilon_{jt} \tag{1}$$

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 policies (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 (month, day, hour or time of day), defined as the ratio of net generation over operating capacity multiplied by total number of hours in the period. Capacity factors provide a measure of utilization that is independent of plant size, thus allowing for more accurate comparisons than if we used net generation at the plant as dependent variable in our model.[13] In most model specifications, capacity factors

[11] Following 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.

[12] Natural 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.

[13] Using net generation as dependent variable would also require the addition of plant capacity as a covariate in the model. However, since we examine an 8-year period, there is not significant variation in capacity over time, making this variable

have daily frequency (i.e., t is one day). Some of our robustness checks examine the use of monthly, hourly or time-of-day capacity factors (i.e., t corresponds instead to one month, hour or day/night, respectively).

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 6.1.2, 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 made extremely difficult by a suite of changes that 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. While these coincident changes pose a threat to identification, we seek to control for their impacts through a broad set of determinants of capacity factors in $\mathbf{X}_{jt}^{'}$.

To model the relationship between plant capacity factors and in-state hydro, nuclear and renewable generation, we use a cubic spline with three knot points for each variable in the baseline. In addition, we create splines whose coefficients vary by NERC interconnection (EAST, TRE and WECC) and by NERC region (FRCC, MRO-US, SERC, SPP, TRE and WECC), and use them to test robustness of the results. Since non-fossil substitutes of a plant may not be in the same balancing authority as the plant, we consider two additional metrics for hydro-nuclear and renewable generation.[14] The first metric controls for the growth of hydro-nuclear and renewable generation in the same region, but outside the plant's balancing authority. EIA monthly generation for hydro-nuclear and renewable plants is aggregated by BA, and BAs are assigned to regions in the interconnection. For treated units in the Western Interconnection, BAs fall into one of four regions (California and three leaker regions), as discussed in Section 4.1.1. For control units in the Eastern Interconnection, balancing authorities are divided into regions, following the Energy Information Administration (2021d). Next, for each BA we calculate monthly shares of hydro-nuclear and solar-wind-other renewable generation in the region, excluding generation in the BA. Finally, we assign shares to plants, based on the balancing authority they fall in. This metric takes a value of zero for plants in ERCOT, as this BA largely overlaps with the Texas Interconnection.

The second metric controls for the growth of hydro-nuclear and renewable generation in neighboring

correlated with plant fixed effects and raising concerns of collinearity in the estimation.

[14]Since nuclear generation at the balancing authority level is sparse, we merge nuclear and hydro generation into one series.

regions that may affect utilization of California's NGCC plants. We construct monthly shares of hydro-nuclear and solar-wind-other renewable generation in the North and South regions in WECC, and assign them to NGCC plants in California.[15] This metric takes a value of zero for plants in the leaker or the control regions.

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, Mansur and Saravia, 2008; 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 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.

[15] To illustrate, consider a plant in CAISO: the first metric captures the growth of nuclear and renewable generation in the rest of California, while the second metric captures the growth of out-of-state nuclear and renewable generation in the North and South regions.

4.2 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 CO_2 emission leakage predictions in 2013 and 2016. This section discusses distinctive features of the version of JHSMINE presented here, the model formulation and the two scenarios.

4.2.1 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 to the transmission planning process (Hobbs et al., 2016; Xu and Hobbs, 2019) 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 carbon price is assumed, rather than determined endogenously.[16]

The version of JHSMINE in this paper differs from the one in Xu and Hobbs (2021) in several important ways. While the latter is 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 (Gabriel et al., 2013). The resulting DC OPF uses a linearized approximation of the alternating current (AC) power flow equations (Schweppe et al., 1988), 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),

[16]Since California's cap-and-trade system is a multi-sector cap-and-trade system, but JHSMINE is a partial equilibrium model, representing the carbon price as an exogenous tax on in-state electricity generators and power importers from out of state is the preferred modeling approach. An emission cap formulation would require estimating one or more demand functions for emission allowances from the other sectors of the economy.

or a scenario in which all imports are considered unspecified and assigned the default emission factor of 0.428 metric ton CO_2/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 where 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 CO_2/MWh.

4.2.2 Formulation

A market model consists of submodels for individual market participants and a set of market clearing conditions linking participant decisions (e.g., supply = demand for electricity and other commodities). Equilibrium models search for a set of solutions that satisfy a) each submodel's first order conditions, subject to participant expectations about how the rest of the market will react if it changes its decisions, and b) market clearing conditions. Under some conditions, it is possible to define a single-objective problem yielding a solution equivalent to the equilibrium for competitive and oligopolistic models. For example, under the assumption of perfectly inelastic demand a competitive equilibrium among power producers is equivalent to minimization of total generation costs. In this section we present the optimization problems of the market players in JHSMINE (the system operator, generation companies and load serving entities) and the market clearing conditions. Under the assumptions of perfectly inelastic demand and perfect competition, the market equilibrium is obtained by solving an equivalent single optimization problem, whose objective is to minimize the sum of individual objectives. The constraints of this single optimization are formed by the union of constraints of each individual player and the market clearing conditions. The model has hourly resolution and is solved for eight representative days in 2013 and 2016. The nomenclature of JHSMINE is presented in

the Appendix.

A. System Operator

The system operator (SO) arbitrages any differences in nodal prices on the network by buying power at one location and selling it to the other. This can be viewed as spatial arbitrage. The SO's objective is to maximize the annual profit from spatial arbitrage across the nodes of the network:[17]

$$\sum_{h} HW_h \cdot \sum_{l} \left(\lambda^{LMP}_{h,i^{To}_l} - \lambda^{LMP}_{h,i^{From}_l} \right) \cdot pf_{h,l} \tag{2}$$

where HW_h is the number of hours represented by hour h, $\lambda^{LMP}_{h,i}$ is the locational marginal price at hour h and node i, and $\left(\lambda^{LMP}_{h,i^{To}_l} - \lambda^{LMP}_{h,i^{From}_l} \right)$ is the price difference between the receiving node and the sending node of transmission line l. Constraint (3) is the DC power flow at hour h through transmission line l:[18]

$$pf_{h,l} = B_l \cdot BP \cdot \left(\theta_{h,i^{From}_l} - \theta_{h,i^{To}_l} \right) \quad \forall h, l \notin L^{DC} \tag{3}$$

Based on equations (4) and (5), power flows cannot exceed transmission capacity limits.

$$pf_{h,l} \leq LTM_l \quad \forall h, l \tag{4}$$

$$-pf_{h,l} \leq LTM_l \quad \forall h, l \tag{5}$$

B. Generation Companies

Each generation company (GenCO) k in JHSMINE owns and operates one power plant.[19] The GenCO sells energy, as well as the non-electrical attributes associated with its power generation (emissions and renewable energy credits), whose demand is created through regulation. Energy (denoted by *gopt* in the model) is sold to the system operator at the nodal electricity price, λ^{LMP}. The non-electrical attributes of power are traded instead through bilateral contracts, and remunerated separately from the energy output. In particular, the GenCO may enter bilateral contracts with load serving entities in its own state, neighbor

[17] System operators are non-profit entities that operate but do not own network or generation assets. Although maximization of profits from spatial arbitrage is not the objective of real-world system operators, the SO problem formulation in JHSMINE is equivalent to one in which the SO adjusts demand and arbitrage variables to maximize consumer benefit from power consumption, subject to fixed values of generator sales and output (Hobbs and Helman, 2004).

[18] Note that the DC approximation applies to AC transmission lines. Power flows radially on HVDC lines in the model (e.g., the Intermountain HVDC between Utah and California and the Pacific DC Intertie).

[19] The formulation may be generalized to allow for GenCOs that own and operate multiple generators.

states or states adjacent to their neighbors: the variable $cpfs^S$ refers to the emission attribute associated with a contract for specified source power (in MW) sold by generator k at a price λ^{SEC}. Further, if GenCO k is outside of California, it may also enter contracts to sell unspecified source power to a pool. Thus, the variable $cpfs^U$ refers to the emissions associated with a contract for unspecified source power (in MW) sold by out-of-state generator k at a price λ^{UEC}; the emission rate of unspecified power is 0.428 ton/MWh. In addition, any GenCO in WECC may enter bilateral contracts with load serving entities in its own state, neighbor states or states adjacent to their neighbors to sell its renewable energy credits (denoted by $recs$ in the model).

We present the problem of in-state electricity producers and producers in the rest of WECC separately, because the former can only sell specified power and are subject to the cap-and-trade program.

The objective of GenCOs in California ($k \in K_{CA}$) is to maximize the annual profit equal to the revenues from the electricity market, specified power contracts and renewable energy contracts, minus the variable costs of generation and the cost of emission allowances:

$$\begin{aligned} &\sum_h HW_h \cdot \Bigg(\lambda^{LMP}_{h,i_k} \cdot gopt_{h,k} + \sum_w \lambda^{SEC}_{w,h,k} \cdot cpfs^S_{w,h,k} + \sum_w \lambda^{REC}_{w,h,k} \cdot recs_{w,h,k} \\ &-GVC_{h,k} \cdot gopt_{h,k} - CTAX \cdot GER_k \cdot gopt_{h,k} \Bigg) \end{aligned} \tag{6}$$

where $\lambda^{SEC}_{w,h,k}$ is the price of the bilateral power contract between GenCO k and load serving entity w at hour h, $\lambda^{REC}_{w,h,k}$ denotes the hourly price of the bilateral REC contract between k and w, and $CTAX$ is the carbon price. The GenCO's objective is subject to several constraints. Based on equation (7), GenCOs in California must sell their power output to the state load serving entity through specified energy contracts:

$$gopt_{h,k} = cpfs^S_{CA,h,k} \quad \forall h, k \in K_{CA} \tag{7}$$

Equation (8) is the generation capacity limit, accounting for forced outage rates and (in the case of intermittent sources) hydro, wind, or solar availability:

$$gopt_{h,k} \leq GNPL_k \cdot GHAV_{h,k} \quad \forall h, k \in K_{CA} \tag{8}$$

Equation (9) limits the total amount of RECs sold to be lower than power generation:

$$\sum_{w} recs_{w,h,k} \leq gopt_{h,k} \quad \forall h, k \in K_{CA} \tag{9}$$

Finally, the GenCO's problem is subject to non-negativity constraints on the specified energy contracts sold $cpfs^{S}_{w,h,k}$, power output $gopt_{h,k}$, REC contracts sold $recs_{w,h,k}$, unit commitment and minimum up/down time constraints (not shown).

The objective of GenCOs outside of California ($k \notin K_{CA}$) is given by:

$$\sum_{h} HW_h \cdot \Bigg(\lambda^{LMP}_{h,i_k} \cdot gopt_{h,k} + \sum_{w} \lambda^{SEC}_{w,h,k} \cdot cpfs^{S}_{w,h,k} + \lambda^{UEC}_{h} \cdot cpfs^{U}_{h,k} + \sum_{w} \lambda^{REC}_{w,h,k} \cdot recs_{w,h,k} \\ -GVC_{h,k} \cdot gopt_{h,k} \Bigg) \tag{10}$$

where λ^{UEC}_{h} is the hourly price at which k sells its unspecified power to the pool. Unlike the GenCOs in California, electricity producers in the rest of WECC can also sell unspecified power and are not subject to the carbon policy. As a result, equation (7) is modified as follows:

$$gopt_{h,k} = \sum_{w} cpfs^{S}_{w,h,k} + cpfs^{U}_{h,k} \quad \forall h, k \notin K_{CA} \tag{11}$$

The objective is also subject to a generation capacity constraint (like equation (8), but applied to $k \notin K_{CA}$) and a limit on the total amount of RECs sold (like equation (9), but applied to $k \notin K_{CA}$). Finally, the GenCO's problem is subject to non-negativity constraints on the energy contracts sold $cpfs^{S}_{w,h,k}$ and $cpfs^{U}_{h,k}$, power output $gopt_{h,k}$, REC contracts sold $recs_{w,h,k}$, unit commitment and minimum up/down time constraints (not shown).

C. Load Serving Entities

Each load serving entity (LSE) w in JHSMINE corresponds to one state in the U.S. (or province in Canada). LSEs serve load by purchasing energy from the SO at the nodal price, and all load served must be bought through specified contracts or from the unspecified pool. Thus, for each MW bought a LSE pays the price of the contract (λ^{SEC} or λ^{UEC}) and the energy price at its node (λ^{LMP}). In addition, the LSE buys RECs to meet its RPS obligation, or pays a penalty for noncompliance. As for the GenCOs, we present the problem of the California LSE ($w = CA$) and LSEs in the rest of WECC separately, because only the former is subject to the cap-and-trade program and pays for the emissions associated with power imports.

The objective of the California LSE ($w = CA$) is to minimize the cost of serving inelastic electricity demand, which includes the cost of energy bought from the electricity market, specified power contracts, unspecified power, renewable energy contracts, unserved energy, noncompliance with the RPS policy, and imported emissions:

$$\sum_h HW_h \cdot \Bigg[\sum_{i \in I_w} \lambda_{h,i}^{LMP} \left(LOAD_{h,i} - n_{h,i}^{Load} \right) + \sum_k \lambda_{w,h,k}^{SEC} \cdot cpfb_{w,h,k}^{S} + \lambda_h^{UEC} \cdot cpfb_{w,h}^{U} + \sum_k \lambda_{w,h,k}^{REC} \cdot recb_{w,h,k}$$
$$+ \sum_{i \in I_w} VOLL \cdot n_{h,i}^{Load} + ACP_w \cdot n_{w,h}^{RPS} + CTAX \cdot \left(\sum_{k \notin K_w} DR_{w,k}^{S} \cdot cpfb_{w,h,k}^{S} + DR_w^{U} \cdot cpfb_{w,h}^{U} \right) \Bigg] \quad (12)$$

Constraints (13)-(18) apply for $w = CA$. Based on equation (13), the LSE must buy power directly from GenCOs k or from the unspecified power pool:

$$\sum_k cpfb_{w,h,k}^{S} + cpfb_{w,h}^{U} = \sum_{i \in I_w} \left(LOAD_{h,i} - n_{h,i}^{Load} \right) \quad \forall h \quad (13)$$

The California LSE also faces a RPS constraint:

$$\sum_h HW_h \cdot \left(n_{w,h}^{RPS} + \sum_k RE_{w,k} \cdot recb_{w,h,k} \right) \geq RPS_w \cdot \left(\sum_h HW_h \sum_{i \in I_w} \left(LOAD_{h,i} - n_{h,i}^{Load} \right) \right) \quad (14)$$

Equation (15) sets a lower bound on the renewable energy coming from resources within California and/or in states directly adjacent to California (Official California Legislative Information, 2015):

$$\sum_h HW_h \cdot \left(n_{w,h}^{RPS} + \sum_{\substack{k \in K_{CA} \\ \cup K_{OR} \\ \cup K_{NV} \\ \cup K_{AZ}}} RE_{w,k} \cdot recb_{w,h,k} \right) \geq RES_{CA} \cdot RPS_w \cdot \left(\sum_h HW_h \sum_{i \in I_w} \left(LOAD_{h,i} - n_{h,i}^{Load} \right) \right) \quad (15)$$

where RES_{CA} is the minimum annual share of renewable energy supplied to the California grid, located within or proximate to the state. To make sure that the level and composition of power imports to California in JHSMINE are comparable to historical values, we take two steps. First, we introduce a constraint on the annual share of specified imports over total power imports to California:

$$\sum_h HW_h \cdot \left(\sum_{k \notin K_w} cpfb_{w,h,k}^{S} \right) = SSI_y \cdot \sum_h HW_h \left(\sum_{k \notin K_w} cpfb_{w,h,k}^{S} + cpfb_{w,h}^{U} \right) \quad (16)$$

Second, we introduce constraints on the share of California imports by fossil fuel generation type. For specified generation from natural gas ($f = NG$), the constraint is:

$$\sum_h HW_h \left(\sum_{\substack{k \in K_{NG} \\ k \notin K_w}} cpfb^S_{w,h,k} \right) = SSI_{f,y} \cdot \sum_h HW_h \left(\sum_{k \notin K_w} cpfb^S_{w,h,k} + cpfb^U_{w,h} \right) \tag{17}$$

where K_{NG} refers to natural gas-fired generators, and $SSI_{f,y}$ is the fuel-specific annual share of imports. For specified generation from coal and oil products ($f = CO$), the constraint is:

$$\sum_h HW_h \left(\sum_{\substack{k \in K_{CO} \\ k \notin K_w}} cpfb^S_{w,h,k} \right) = SSI_{f,y} \cdot \sum_h HW_h \left(\sum_{k \notin K_w} cpfb^S_{w,h,k} + cpfb^U_{w,h} \right) \tag{18}$$

where K_{CO} refers to coal- and oil-fired generators. Finally, the LSE's problem is subject to non-negativity constraints on the contracts bought $cpfb^S_{w,h,k}$ and $cpfb^U_{w,h}$, the load shedding amount $n^{Load}_{h,i}$, REC contracts bought $recb_{w,h,k}$, and non-compliance amount with RPS policy $n^{RPS}_{w,h}$.

The objective of LSEs in the rest of WECC ($w \neq CA$) is given by:

$$\sum_h HW_h \cdot \Bigg[\sum_{i \in I_w} \lambda^{LMP}_{h,i} \left(LOAD_{h,i} - n^{Load}_{h,i} \right) + \sum_k \lambda^{SEC}_{w,h,k} \cdot cpfb^S_{w,h,k} + \lambda^{UEC}_h \cdot cpfb^U_{w,h} + \sum_k \lambda^{REC}_{w,h,k} \cdot recb_{w,h,k} + \sum_{i \in I_w} VOLL \cdot n^{Load}_{h,i} + ACP_w \cdot n^{RPS}_{w,h} \Bigg] \tag{19}$$

Unlike the California LSE, load serving entities in the rest of WECC do not incur a cost of imported emissions (i.e., last term in (12)). The LSEs must buy specified or unspecified power (equation (13)) and are subject to RPS constraints (equation (14)), plus non-negativity constraints on the contracts bought $cpfb^S_{w,h,k}$ and $cpfb^U_{w,h}$, the load shedding amount $n^{Load}_{h,i}$, REC contracts bought $recb_{w,h,k}$, and non-compliance amount with RPS policy $n^{RPS}_{w,h}$.

D. Market Clearing Conditions

Market clearing conditions for electricity ensure that hourly demand equals supply at each location in the network, and the associated dual variable presents the nodal electricity price $\lambda^{LMP}_{h,i}$:

$$\sum_{k \in K_i} gopt_{h,k} + \sum_{l \in L^{In}_i} pf_{l,h} - \sum_{l \in L^{Out}_i} pf_{l,h} = LOAD_{h,i} - n^{Load}_{h,i} \quad \forall h, i \tag{20}$$

There are two clearing conditions for the contract market. First, the unspecified power sold from GenCOs equals the amount by the LSEs at every hour. The associated dual variable is the price of unspecified power at hour h, λ_h^{UEC}:

$$\sum_k cpfs_{h,k}^U = \sum_w cpfb_{w,h}^U \quad \forall h \tag{21}$$

Second, for every bilateral contract between GenCO k and LSE w at hour h, demand must equal supply equation (22). The associated dual variable is the price of the specified energy contract $\lambda_{w,h,k}^{SEC}$:

$$cpfs_{w,h,k}^S = cpfb_{w,h,k}^S \quad \forall w, h, k \tag{22}$$

Finally, since emission allowances and RECs are unbundled in the current formulation of JHSMINE, a REC market clearing is given by equation (23). The corresponding price variable is $\lambda_{w,h,k}^{REC}$.

$$recs_{w,h,k} = recb_{w,h,k} \quad \forall w, h, k \tag{23}$$

4.2.3 Scenarios

We consider two scenarios: (a) no regulation of GHG emissions (benchmark); (b) a scenario where California generators and the California LSE are subject to a first deliverer cap-and-trade program. 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 CO_2/MWh. Further, both scenarios include RPSs and assume the same share and composition of specified imports into California.

5 Data

This section describes the sources of data used for each model and the steps taken to ensure comparability.

5.1 Econometric model

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).

5.1.1 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 1). 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.[20] 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[21]), 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 (Energy Information Administration, 2021b), and EIA-923 for other plant characteristics (Energy Information Administration, 2021c). We exclude plants with operating capacity below 25 MW.[22]

Plant fuel costs are used to calculate monthly ratios to assess competitiveness (Section 4.1.3). 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 Energy

[20] In 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, if the CT and CA parts of a NGCC plant cannot operate independently. As a result, 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.

[21] Net generation excludes power consumption for plant operations.

[22] 25 MW corresponds to the minimum size of generators subject to requirements for monitoring and reporting emissions under the EPA's Continuous Emissions Monitoring System (Environmental Protection Agency, 2021). Plants with capacity below 25 MW generally use renewable energy sources and represent less than 5% of generating capacity in our sample.

Information Administration (2021a). 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.

5.1.2 CEMS data

We assemble a database of hourly gross electricity generation, heat input and CO_2 emissions for NGCC and coal-fired plants from the EPA's Continuous Emissions Monitoring System (Environmental Protection Agency, 2021). CEMS represents the only publicly available information on high frequency operating data for thermal power plants in the U.S., and has been widely used in empirical studies (Joskow and Kahn, 2002; Mansur, 2007; Puller, 2007; Graff Zivin et al., 2014; Kotchen and Mansur, 2014; 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 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.

5.1.3 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 (Department of Energy, 2014; California Independent System Operator, 2021a). Data are available from April 2009 to October 2015 from the Open Access Same-time Information System (California Independent System Operator, 2021c). In addition, hourly total power imports into the California grid are available from April 2010 to December 2016 from the California Independent System Operator (2021b). 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, 2021), and aggregate them to the interconnection level (MRO-US and WECC).

Electricity consumption comes from the Federal Energy Regulatory Commission (FERC). FERC Form 714 provides hourly load information by planning area (Federal Energy Regulatory Commission, 2021). 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 (2021). The monthly seasonally-adjusted employment level in the mining and logging, construction and manufacturing sectors by state is from the Bureau of Labor Statistics (2021). 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, 2021).

5.2 Simulation model

To ensure comparability of the data used for both models, 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 import shares in Section 4.2.2 based on historical data from CARB to make sure that the level and composition of power imports into California in JHSMINE are comparable to historical values.

As noted in Section 4.2.1, JHSMINE's test system is a network reduction of the Western Interconnection that consists of 361 buses, 712 transmission lines and 1,504 existing aggregate generators. There is no one-to-one correspondence between aggregate generators in JHSMINE and power plants in the empirical analysis. However, unit-level data is the basis for modeling the aggregate generators. Thus, we replace the existing generation capacity in JHSMINE (based on WECC's 2026 Common Case database) with unit-level operating capacity from EIA Form 860 in 2013 and 2016 (Energy Information Administration, 2021b). Average fuel costs by state, technology type and month-year are from the econometric model database, while CO_2 emission rates by fuel type are from the Energy Information Administration (2021e).[23] Minimum up/down times for coal plants, NGCCs and combustion turbines are drawn from Herrero et al. (2018). JHSMINE assumes that all transmission reinforcements in the WECC 2026 Common Case have been brought online. We are unable to replicate the network topology in 2013-2016 because the reduction algorithm (Shi et al., 2012)

[23]The CO_2 emission rates from the EIA are 117 lb/MMBtu for natural gas and 205.70 lb/MMBtu for bituminous coal. These values are closely aligned to the average emission rates from the econometric model dataset (118.05 lb/MMBtu for natural gas in California, 208.37 lb/MMBtu for coal in NW WECC, and 208.74 lb/MMBtu for coal in RoW WECC).

does not provide a one-to-one correspondence between the aggregated branches and the original transmission lines. Consequently, removing the lines added between 2013 to 2026 would not be possible without recreating the transmission network database. However, the use of the 2026 network topology does not significantly affect our results, because reinforcements on high-voltage transmission lines only add 88.5 GW, expanding capacity to 4,040.9 GW.[24]

Electricity consumption in JHSMINE is from WECC's long term planning tool (LTPT) (Xu and Hobbs, 2021). To ensure comparability with the demand levels in our econometric analysis, we replace load from the LTPT with hourly consumption by planning area from FERC Form 714 (Federal Energy Regulatory Commission, 2021). Next, we identify eight representative days to run the model in 2013 and 2016. Days are first grouped into eight categories representing different combinations of seasons (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec) and day type (weekend and weekday). For each category, we calculate the daily average net load. The representative days are chosen so that the net load of each day is closest to the net load in that category.

Table A2 in the Appendix summarizes the assumed state-level RPS requirements for 2013 and 2016, which are drawn from the DSIRE database (DSIRE, 2021).[25] Load serving entities in states with an RPS are subject to an alternative compliance penalty of 100 \$/MWh if they fall short of renewable energy credits. The shares of specified imports over total power imports in California (SSI_y), as well as the shares of imports by fossil fuel generation type ($SSI_{f,y}$), are calculated based on historical data from the California Air Resources Board (2020a); these shares are presented in Table A3 in the Appendix. Finally, the carbon price is set equal to average historical values in California (\$ 13.53/metric ton CO_2e in 2013, and \$12.84/metric ton CO_2e in 2016).

6 Results

6.1 Econometric model

6.1.1 Shifts in electricity generation

Tables 3-5 show the results corresponding to the estimation of equation (1). Our main specification uses daily capacity factors as dependent variable. 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

[24] The list of line additions is available on the Release Notes for WECC 2026 Common Case, Version 1.5, p. 23.

[25] When a state sets RPS targets for multiple types of utilities, we apply the target for investor-owned utilities to all utilities in that state. This is unlikely to significantly affect our results since the target mostly impacts long-term investment.

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 daily level. Robust standard errors are clustered at the plant level. Baseline results are presented in the first column of Tables 3 and 4.

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_D and α_I for the leaker regions. Provided that the DID identification assumptions hold, our empirical results suggest that daily capacity factors for matched NGCC plants in California decreased by 5.8% in response to the introduction of the cap-and-trade program, relative to similar control facilities. In contrast, capacity factors for matched coal-fired plants increased by 4.7% in the Pacific Northwest, Nevada and Utah, and by 4.4% in Central and Eastern WECC. Other estimates are not statistically significant.

In columns (2a)-(2c) of Tables 3 and 4, we examine robustness to alternate specifications for hydro-nuclear and renewable generation. The dependent and independent variables are measured at the same frequency as in the baseline. In (2a), we use splines whose coefficients vary by NERC region. Relative to the base specification, daily capacity factors for matched NGCC plants in California decrease less and at a lower level of statistical significance; in addition, we find evidence of increased NGCC generation in the Pacific Northwest, Nevada and Utah. Results from the specification where spline coefficients vary by NERC interconnection yield similar insights. In (2b) and (2c), we include the metrics presented in Section 4.1.3 as additional covariates to account for the indirect effects of out-of-state generation on plant utilization. In (2b), we consider the effect of hydro-nuclear and renewable generation in the same region (but outside the plant's balancing authority). Results are broadly consistent with the main specification. In (2c), we examine the effect of hydro-nuclear and renewable generation in neighboring regions on California's NGCC plants. We note that the reduction in NGCC plant utilization in California is not statistically significant when we control for both in-state and out-of-state hydro-nuclear and renewable effects in California, a point to which we return below when evaluating treatment effect heterogeneity between day and night.[26]

In columns (3a)-(3b) of Tables 3 and 4, we examine whether lower or higher frequency measures of plant utilization yield different insights. The specification in (3a) uses monthly capacity factors as dependent variable, while the one in (3b) uses hourly capacity factors. Electric load by planning area is measured at the same frequency as capacity factors, while all other covariates retain the same frequency as in the baseline. In

[26]Table 4 does not present results for specification (2c) because California has no coal-fired generation capacity.

columns (4a)-(4b), we evaluate robustness of the baseline to more conservative clustering. In (4a), standard errors are clustered by NERC region because treatment is assigned at the interconnection level. In (4b), we consider clustering by balancing authority as a compromise between widening the scope of clustering to be more in line with the true level of treatment and having a sufficient number of clusters for inference. In (5), we present results based on an alternate matching set. As in the baseline, we use pre treatment average capacity factors over four hour blocks during the day as matching variables. However, we coarsen the matching variables based on an alternate set of cut points (0.3, 0.5 and 0.7 for NGCC plants, 0.6 and 0.8 for coal-fired plants), which are based on visual inspection of the empirical distribution of the 2009-2010 average capacity factors by hour.[27] In specifications (3)-(5), we find results to be broadly consistent with the baseline, implying a policy-induced decrease in capacity factors of NGCC generators in California between 5.9% and 6.9%, and an increase in coal-fired plant utilization in the Pacific Northwest, Nevada and Utah, as well as in Central/Eastern WECC, subject to the maintained assumptions of the DID estimator.

Finally, we explore potential treatment heterogeneity between day and night using hourly measures of plant utilization based on the CEMS data. Electric load by planning area also has an hourly frequency, and all other covariates are measured at the same frequency as in the baseline. 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. Table 5 presents the results. The specification in (6) is similar to the baseline, but uses hourly (rather than daily) capacity factors as dependent variable. In (7), we allow spline coefficients to vary by NERC region, as in (2a). Specification (8) adds the effect of renewables in the same subregion (but outside the plant's balancing authority), as in (2b), while (9) adds the effect of renewables in the Northern and Southern WECC subregions on California plants only, as in (2c). Results suggest a statistically significant reduction of NGCC capacity factors in California between 6% and 9.5% during the day, relative to similar control facilities. In contrast, capacity factors increased across both periods (day and night) for matched coal-fired plants in the leaker regions. This is consistent with the heavy utilization of Western U.S. coal plants, which tend to ramp more slowly than NGCC plants (Herrero et al., 2018). We also estimate separate effects for every hour of the day by interacting each treatment indicator with 24 hourly indicators in the hourly regressions. Results for the baseline are presented in Figure 4, and suggest a statistically significant reduction in NGCC capacity factors in California between 8am and 8pm. Coal-fired plant utilization in the Pacific Northwest, Nevada and Utah increases during all hours of the day, while coal plants in Central and Eastern WECC are more heavily utilized only during the night hours.

[27] Coal-fired plants (particularly in the Southwest region of WECC) tend to be more heavily utilized than NGCC plants, motivating our choice of higher capacity factors as cut points for this technology type.

6.1.2 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 6 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 4.1.2. Tables A4 and A5 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 A1 in the Appendix shows the capacity factor trajectories of matched treated and control plants by technology type between 2009 and 2016.[28] 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 (Autor, 2003; Kearney and Levine, 2015; Fell and Maniloff, 2018; Jaeger et al., 2020). We use the following regression:

$$Y_{jt} = \alpha_{Ct} B_t D_j^C + \sum_L \alpha_{Lt} B_t D_j^L + \mathbf{X}_{jt}^{'} \underline{\beta} + \gamma_j + \gamma_y + \gamma_{dw} + \gamma_{sm} + \epsilon_{jt} \tag{24}$$

[28]Note 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.

where B_t is a seasonal dummy equal to 1 after January 2013 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, and α_{Ct} and α_{Lt} are the estimated coefficients associated with group specific time trends.[29] Other variables are defined in Section 4.1.3. Estimated pre treatment trend coefficients α_{Ct} and α_{Lt} that are statistically significantly different from zero would support the assumption of parallel trends between treated and control groups prior to the intervention. In our regression, not all the α_{Ct} can be identified since the D_j^C dummies are perfectly collinear in the presence of state effects. Hence, we omit one period for which we have complete data for all treated groups (the last season of our pre treatment matching period, i.e. Winter 2011). Figure 5 presents our estimated coefficients by treated region and technology type. While in most cases the pre treatment coefficients 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 significant pre treatment trends.

Next, we conduct a parallel trends test that compares the treatment effects in the base specification 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, 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 the regression model (equation (1)). If a treatment effect is statistically significant in the baseline, we examine how the introduction of a trend affects its sign and significance, and present the results in Table 7. Adding a trend does not change the sign and significance of the estimated treatment effect for coal-fired plants in the Northwest region of WECC, and changes the significance of the estimated coefficient for coal-fired plants in Central/Eastern WECC when we add a linear trend. However, the treatment effect in California is no longer statistically significant when group-specific trends are included, 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 already trending differentially before 2013.

To address this challenge, we adopt the robust inference method proposed by Rambachan and Roth (2020) to test sensitivity of statistically significant average treatment effects to violations of parallel trends.[30] 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.

[29] We use seasonal dummies to account for cyclical factors that may affect plant utilization. In line with 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.

[30] In a similar vein, Ang (2021) and Rose (2021) use this method to conduct robust inference on statistically significance average treatment effects.

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 also draw on context-specific knowledge.

Using equation (24), we estimate seasonal treatment effects for the regions and technology types with at least one statistically significant ATE in Table 7 (CA NGCC, NW Coal and RoW Coal). Next, we construct robust confidence intervals for the seasonal treatment effects using the R code HonestDiD written by the authors (Rambachan and Roth, 2021). To illustrate, Figure 6 presents sensitivity analyses for the event study coefficients in the first period after treatment (Jan and Feb 2013). For each treated region, we compare the OLS confidence intervals (in blue) to the 95% confidence intervals from the Rambachan and Roth 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 likely 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, we impose that the bias of California's event study coefficients after treatment is positive. In addition, it seems reasonable to assume that the downward sloping pre trend in California's NGCC utilization in Figure 5 would have continued even 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. For the leaker regions, since decreasing natural gas prices would have a negative effect on utilization of coal-fired plants over the period of our study, we assume that the sign of the bias is negative. We also introduce a monotone increasing trend for the NW region region, for the reason discussed below. Following the authors' 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 (2020). 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 NW region, a value of M equal to 0.002 (0.008) 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 0.005. Finally, 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. 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. For the coal leakers, the OLS estimates are positive but the confidence interval includes zero. When we allow for linear violations of parallel trends (M=0) in the NW region, the estimated coefficient becomes negative, and the robust confidence interval only includes negative values. This result mirrors the sensitivity analysis for the treatment effect on female employment 15 years after the passage of a duty-to-bargain law in Rambachan and Roth (2020), and suggests a pre-existing upward trend: if we linearly extrapolated this pre period trend in the post period, the estimated coefficient would lie below the linear extrapolation. We again include zero in the robust confidence interval for M around 0.005. This result motivates our decision to include a monotone increasing trend for the NW region. Finally, when we allow for linear or non-linear violations of parallel

trends in the RoW region, 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 Figures A2-A4 in the Appendix. To show how often robust inference changes the conclusions based on OLS confidence intervals for each region and technology type, we calculate the number of post treatment periods in which OLS statistically significant treatment effects become insignificant for at least one value of M, under a given set of restrictions. OLS statistical significance no longer holds between 29% and 47% of the times under no restrictions; and up to 6% of the times under bias and bias/monotonicity restrictions. As a result, when we allow for linear violations of parallel trends (or non-linear violations calibrated on the data), leakage is implied by our econometric estimates when we impose bias (or monotonicity and bias) restrictions, but does not necessarily follow under no restrictions. We return to this point below.

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 4.1.3: 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.

6.1.3 Leakage estimates

Based on the shifts in electricity generation implied by the seasonal treatment effects and their robust confidence intervals, we infer CO_2 emission leakage predictions in 2013 and 2016. Table 8 focuses on the first and last year after treatment over the period of our study to allow for more direct comparisons with

the JHSMINE results (Section 6.3). Further, the results in Table 8 are based on robust confidence intervals that assume sign restrictions and allow for linear violations of parallel trends. To calculate the leakage rates implied by the econometric model, we proceed as follows.

First, we identify the statistically significant post period event study coefficients. If the robust 95% confidence interval for a post period event study coefficient from equation (24) excludes zero, we include the confidence interval in our leakage calculations. In contrast, if the robust confidence interval includes zero, the estimated generation shift for that period is not statistically different from zero and we exclude it from the leakage calculations. We use the lower (upper) bound of the robust 95% confidence interval for statistically significant coefficients to calculate a lower (upper) bound for the generation, emissions and leakage rates associated with the econometric estimates.[31]

Next, we find the estimated generation leakage by multiplying the statistically significant seasonal treatment effects by the total generation capacity of matched plants by region, year and technology, and the number of hours in that season. Per these estimates, the change in NGCC generation of matched units in California was between -9.04 and 0.80 TWh in 2013, and between -11.18 and -2.78 TWh in 2016. Coal generation of matched units in the leaker regions increased between 4 and 9.99 TWh in 2013, and between 2.28 and 17.82 TWh in 2016. Based on these generation leakage estimates, we calculate the change in local CO_2 emissions in California (E_1) and WECC-NonCA emissions (E_4), based on region-, year- and technology-specific heat rates and CO_2 emission rates. The resulting change in WECC emissions ($E_5 = E_1 + E_4$) is between 0.38 and 12.12 million metric tons in 2013, and between -2.26 and 20.02 million metric tons in 2016.

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 counterfactual import emissions, and I_2 the actual import emissions subject to the cap-and-trade regulation. Under California's first deliverer approach, electricity importers are required to purchase compliance instruments to offset any emissions outside of California that are associated with in-state sales. Thus, the total instruments (allowances and offset credits) surrendered by the importers correspond to the import emissions subject to the cap-and-trade regulation (18.88 million metric tons CO_2e in 2013 and 19.04 million metric tons CO_2e in 2016). To derive I_2 in 2013 (or 2016), we identify the electricity importers in each

[31] Note that the OLS point estimate may be outside of the robust confidence interval. Hence, we do not rely on the OLS estimates, but only consider the robust confidence intervals to derive leakage predictions from the econometric model.

year (California Air Resources Board, 2021c), and obtain the total instruments surrendered for compliance by these entities (California Air Resources Board, 2021a). Next, we construct the year-specific counterfactual import emissions I_1 assuming the same percentage change between counterfactual emissions and emissions under the carbon cap predicted by JHSMINE.[32] Finally, the difference between I_1 and I_2 yields the estimated emission reduction E_2 that is presented on Table 8.

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 $100\% - E_5/E_3 = (1 - E_5/E_3) \times 100\%$. 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 104.0% and 344.4% in 2013, and between 79.9% and 376.6% in 2016. While the upper bound estimates are too high to reflect realistic leakage rates, the lower bound estimates are within the range of predictions from simulation-based partial equilibrium models in the literature that use the same leakage metric. For example, Chen et al. (2011) predict a leakage rate of about 85% under California's cap-and-trade program. More recently, Xu and Hobbs (2021) find that the facility-based scheme currently implemented in California leads to leakage rates above 100% due to the increase in WECC emissions, as implied by our 2013 estimates. Direct comparisons with earlier econometric estimates of leakage are difficult due to the use of different metrics that are less relevant in our setting. For example, Fell and Maniloff (2018) and Zhou and Huang (2021) estimate a leakage rate of about 50% in the context of the Regional Greenhouse Gas Initiative, but these studies calculate a physical leakage rate.

[32] To illustrate, in 2013 JHSMINE predicts import emissions of 54.50 million metric tons for the no cap scenario and 42.48 million metric tons for the cap scenario. This implies a percentage change of 28%. Hence, the counterfactual import emissions for the empirical analysis are 24.22 ($= 18.88 \cdot 1.28$) million metric tons CO_2e, and the estimated import emission reduction is 5.34 million metric tons CO_2e in 2013.

6.2 Simulation model

6.2.1 Shifts in electricity generation

Table 9 presents the predicted impact of California's cap-and-trade program on capacity factors in WECC from JHSMINE. We run the model for eight representative days in 2013 and 2016 under a no cap and a carbon cap scenario. WECC Regions are defined as in the econometric model (Section 4.1.1). 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 a cap yields minor generation shifts in 2013, relative to the baseline. Capacity factors of NGCC plants in California decrease by about 0.3%; in the leaker regions, coal-fired plant capacity factors decrease by 0.04% while NGCC capacity factors increase by 0.3%, with the highest increase in plant utilization occurring in the SW. 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 (11.2%) and an increase of similar magnitude for NGCC utilization in the SW (12.3%). The model also predicts a moderate increase in NGCC and coal-fired generation in the NW (3.4% and 2.4%, respectively), and a small increase in generation from the RoW.

6.2.2 Leakage estimates

Table 10 presents the distribution of emissions among WECC regions, as well as the implied leakage rate (as defined in Section 6.1.3) in 2013 and 2016. These results differ from the ones in Table 8 in three ways. First, when we calculate the leakage predictions of the econometric model, we only consider generation shifts and associated emission changes for technology types and regions for which robust inference is possible (i.e., CA NGCC, NW coal and RoW coal). In contrast, the JHSMINE results in Table 10 include 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. Third, leakage estimates from the econometric model are based on a subset of seasons, i.e. the post treatment seasons whose robust confidence intervals exclude zero. In contrast, the JHSMINE results in Table 10 are based on representative days covering 8,760 hours in a year.[33]

We find that local emissions in California (E_1) decrease by only 0.55 million metric tons in 2013, but by 7.69 million metric tons in 2016. Given the estimated changes in import emissions relative to the no cap scenario (12.02 million metric tons in 2013 and 7.88 million metric tons in 2016), regulated emissions in California decrease in both 2013 and 2016 (by 12.56 and 15.57 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 balancing authorities in the Northwest and Southwest regions of WECC. Plants in the RoW only adjust their output by small amounts in response to California's cap-and-trade program, leading to small emission increases. This is contrast 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 96.8% in 2013 and 110.4% 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).

6.3 Comparison of results

JHSMINE is well suited to isolating the effects of California's cap-and-trade program on power plant operations in WECC. In contrast, isolating the same effects in our empirical analysis is difficult, due to the suite of coincident changes affecting California's electricity market and challenges associated with the construction of credible counterfactual outcomes. This raises questions about the validity of a quasi-experimental design to isolate the effect of the cap-and-trade program, and casts doubts on our *ex post* leakage predictions, which are based on the empirically estimated generation shifts. To assess the plausibility of the *ex post* results, we

[33]Note that the definition of seasons is slightly different in the two models and, as discussed in Section 5.2, we select one representative day type (weekend or weekday) per season in JHSMINE. As a result, there is no one-to-one correspondence between seasons in the econometric model and representative days in JHSMINE, and there may be statistically significant periods in the econometric model for which we do not have a corresponding representative day in JHSMINE. For example, the empirical analysis yields statistically significant treatment effects for January-February 2013. However, there is no representative day that specifically refers to this two-month period in JHSMINE, as the selected weekday (weekend) for the first season of 2013 is March 6 (March 10).

benchmark the leakage predictions based on the empirical estimates against the JHSMINE predictions. Since leakage rates in Tables 8 and 10 differ for the reasons discussed in Section 6.2.2, we take two steps to enable more direct comparisons in Table 11.

First, we calculate the leakage rates implied by JHSMINE considering only changes in emissions associated with NGCC plants in California and coal-fired plants in the NW and RoW leaker regions (i.e., regions and technology types for which robust inference is possible). Relative to Table 10, this does not significantly affect results for 2013, but leads to lower leakage in 2016, as the new estimate does not account for higher emissions in the Southwest region of WECC. Because there is no one-to-one correspondence between seasons in the econometric model and representative days in JHSMINE, we do not make adjustments along the timing dimension: thus, as noted above, the simulation results in Table 11 are based on representative days covering 8,760 hours in a year, while the econometric model results are based on the post period seasons whose robust confidence intervals exclude zero. Second, we document how robust confidence intervals from the econometric model change under various assumptions, and how that affects the *ex post* leakage predictions. Specifically, we report confidence intervals that restrict the sign of the bias (positive for California, negative for the leakers, as discussed in Section 6.1.2), and allow for linear ($M = 0$) or non-linear ($M > 0$) violations of parallel trends. In the latter case, we construct confidence intervals based on the first positive, region- and technology-specific value of M in Section 6.1.2 (and Figures A2-A4 in the Appendix) to obtain confidence intervals that are not too wide, and thus uninformative. Under no restrictions, all confidence intervals for the 2016 California estimates include zero, implying no empirical evidence of leakage for 2016; hence, we do not report confidence intervals under no restrictions in Table 11.

The CO_2 emission leakage rates predicted by JHSMINE are close to the lower bound estimates from the econometric model. Both models suggest higher potential for leakage from the Pacific Northwest and Central/Eastern WECC in 2013 than in 2016, with rates ranging between 98.7% and 118.9% in 2013, and between 67% and 79.9% in 2016. Despite a similar magnitude for the predicted leakage impacts from the Pacific Northwest and Central/Eastern WECC, the two models yield different predictions about electricity generation shifts, which in turn imply different changes in emissions. In 2013, JHSMINE shows a reduction in California's local emissions from NGCC plants (0.22 million metric tons of CO_2) and a small emission increase from coal-fired plants in the NW and RoW regions (0.06 million metric tons of CO_2), in line with the generation shifts in Table 9. The confidence intervals from the econometric model reflect instead a substantial decline in NGCC utilization in California in the first half of 2013 (followed by statistically insignificant seasonal treatment effects in the rest of the year), and higher coal-fired plant utilization in both leaker regions.

In 2016, both models predict a significant reduction in California's NGCC plant utilization, on average (-11.2% from JHSMINE and -8.2% from the econometric model). However, while JHSMINE's emissions are based on 8,760 hours and generation capacity for all NGCC plants, the seasonal treatment effects associated with summer and fall 2016 are not statistically significant, and only matched plants are considered in the econometric model. As a result, JHSMINE predicts an emission reduction in California that falls outside the robust confidence interval from the econometric model. For the leaker regions, the simulation model predicts a 2.5% increase in coal-fired capacity factors, mostly in the NW, leading to an emission increase by 2.42 million metric tons of CO_2. On the other hand, the econometric model estimates a more substantial increase in coal-fired utilization (and associated emissions), particularly from the RoW.

Overall, differences between JHSMINE results and empirical estimates may be explained in two ways. First, threats to identification may not allow for rigorous causal inference in this policy setting. As a result, we urge caution in lending a causal interpretation to our empirical estimates. Second, some features of the econometric and simulation models cannot be easily reconciled, driving differences in the results. We have discussed the lack of one-to-one correspondence between seasons in the econometric model and representative days in JHSMINE. Further, both models rely on assumptions that may bias the simulated or empirically estimated output adjustments. On the one hand, JHSMINE simulates the behavior of aggregate generators on eight representative days under the assumptions of perfectly inelastic demand, perfect competition, and perfect foresight of market participants. Although we parameterize the simulation model using historical data, it is not surprising to find that its prediction are loosely connected to 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. (Energy Information Administration, 2020), which is reflected instead in a negative and statistically significant treatment effect from the econometric model. On the other hand, assuming away transmission constraints in most of WECC may bias estimated output adjustments in the econometric model, relative to JHSMINE. While we control for imports into CAISO from neighboring regions, we are unable to account for transmission constraints between balancing authorities in WECC due to lack of available data on the hourly interchange between directly interconnected balancing authorities over the period of our study. This may bias the empirically estimated generation shifts, particularly at WECC plants outside California. In this regard, the comparison between the results in Tables 8 and 10 is instructive, as insights differ with respect to the predicted source of leakage in WECC: JHSMINE suggests that most policy-induced change in out-of-state generation takes place at NGCC plants in the Southwest, as well as NGCC and coal-fired plants

in the Northwest, while the empirical estimates suggest that plant utilization mostly increase at coal-fired plants in the Northwest and RoW. It seems more plausible that most policy-induced change in out-of-state generation takes place in balancing authorities in the Northwest and Southwest regions of WECC, while plants in other WECC regions adjust their generation by small amounts in response to the policy: for example, plants in Colorado or Wyoming are less likely to trade day-ahead or have long-term contracts with California, since obtaining a transmission path to California may lead to pancaking of transmission charges.

7 Concluding remarks

In this paper, we seek to identify CO_2 emission leakage in the electricity sector from California's AB 32 cap-and-trade program in the first four years of policy implementation. To this end, 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 coarsened exact 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 CO_2 emission leakage predictions in 2013 and 2016. The leakage metric presented in our paper accounts for the border adjustment mechanism that is currently implemented in California. 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.

JHSMINE predicts that most policy-induced change in out-of-state generation takes place at NGCC plants in the Southwest, as well as NGCC and coal-fired plants in the Pacific Northwest, yielding a leakage rate of 96.8% in 2013 and 110.4% in 2016. The econometric model predicts instead that higher emissions outside of California are due to increased utilization at coal-fired plants in the Pacific Northwest and Central/Eastern WECC. The leakage rates implied by the lower bound of the robust confidence intervals for the estimated treatment effects are between 104.0% and 118.9% in 2013, and between 67.0% and 79.9% in 2016. Leakage is implied by the empirical estimates when we allow for violations of parallel trends and under restrictions on the sign of the bias of the post period event study coefficients. While these restrictions seem plausible, in light of known simultaneous policy changes that affected generation outcomes over the period of our study,

leakage implications critically hinge upon them: for example, it is not possible to draw definitive conclusions about leakage in 2016, when no restrictions are imposed.

Simulation-based models are well suited to generating counterfactual scenarios that may be used to quantify the effects of environmental policies on outcomes of interest. In addition, JHSMINE relies on a reduced network and dataset that have been vetted by the Western Electricity Coordinating Council. We have also taken steps to enhance realism in the formulation to accommodate the observed cap-and-trade regime, and parameterized the model to match the observed market conditions. Thus, JHSMINE can not only isolate the effects of California's carbon policy on power plant operations in WECC, but also generate plausible leakage predictions that are aligned with comparable estimates in the literature. In contrast, isolating the effects of the policy is difficult in our empirical analysis, due to the suite of coincident changes that affected California's electricity market and challenges associated with the construction of a credible counterfactual. To confront these threats to identification, we control for a broad set of determinants of capacity factors, enhance the construction of counterfactuals through matching, and conduct statistical inference using a robust method that does not require the assumption of parallel trends to hold exactly. There remain, however, reasons to question the validity of a quasi-experimental design in this policy setting.

To assess the plausibility of our *ex post* results, we benchmark the emission leakage predictions based on the empirical estimates against the JHSMINE predictions. To enable more direct comparisons, we calculate the leakage rates implied by the simulation model considering only emission changes associated with technology types and leaker regions for which robust inference is possible. The CO_2 emission leakage rates predicted by JHSMINE are close to the rates implied by the lower bound of the robust confidence intervals from the econometric model. Both models suggest higher potential for leakage in 2013 than in 2016. In 2013, JHSMINE predicts a leakage rate of 98.7%, while the econometric model estimates are between 104.0% and 118.9%; in 2016, JHSMINE predicts a leakage rate of 67.8%, while the econometric model estimates are between 67.0% and 79.9%. Despite a similar magnitude for the predicted leakage impacts from the Pacific Northwest and Central/Eastern WECC, the two models yield different predictions about electricity generation shifts and resulting changes in emissions.

Overall, differences between JHSMINE results and empirical estimates may be explained in two ways. First, threats to identification may not allow for rigorous causal inference in this policy setting, urging caution in lending a causal interpretation to our estimated treatment effects. Second, modeling assumptions may drive different predictions about generation shifts. For example, inaccurate representation of transmission network constraints in the econometric model may bias estimated output adjustments and resulting emissions relative

to JHSMINE, particularly at WECC plants outside California. Similarly, assumptions of the simulation model such as representative days, perfect competition, and inelastic demand, may also bias its output adjustments relative to what is observed in the data.

Our study shows that, in quasi-experimental research designs where threats to identification undermine attempts at statistical inference, results from a simulation-based model that is parameterized based on actual market data may be used to interpret the empirical estimates and assess the plausibility of policy implications. To support comparisons between simulation results and empirical estimates, an important role for future research in this area would be to 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. 2013a. The Effect of the Kyoto Protocol on Carbon Emissions. *Journal of Policy Analysis and Management* 32(4): 731-7-57.

------. 2013b. Estimating the Effects of Kyoto on Bilateral Trade Flows Using Matching Econometrics. *The World Economy*: 36(3): 303–330.

------. 2015. Kyoto and Carbon Leakage: an Empirical Analysis of the Carbon Content of Bilateral Trade. *Review of Economics and Statistics* 97 (1): 104–115.

Alcorn, Thomas. 2013. The Constitutionality of California's Cap-and-Trade Program and Recommendations for Design of Future State Programs. *Michigan Journal of Environmental and Administrative Law* 3 (1): 87–177.

Aldy, Joseph, and William Pizer. 2015. The Competitiveness Impacts of Climate Change Mitigation Policies. *Journal of the Association of Environmental Resource Economists* 2 (4): 565–595.

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.

Borenstein, Severin, James Bushnell, Frank Wolak, and Matthew Zaragoza-Watkins. 2014. *Report of the Market Simulation Group on Competitive Supply/Demand Balance in the California Allowance Market and the Potential for Market Manipulation*. http://www.arb.ca.gov/cc/capandtrade/simulationgroup/msg_final_v25.pdf.

Bureau of Labor Statistics. 2021. *Employment, Hours, and Earnings - State and Metro Area*. http://www.bls.gov/data/#employment.

Burtraw, Dallas, Ann Carlson, Danny Cullenward, Quentin Foster, and Meredith Fowlie. 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, and Yihsu Chen. 2012. Allocation and Leakage in Regional Cap-and-Trade Markets for CO_2. *Resource and Energy Economics* 34: 647–668.

Bushnell, James, Yihsu Chen, and Matthew Zaragoza-Watkins. 2014. Downstream Regulation of CO_2 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.

Bushnell, James, Carla Peterman, and Catherine Wolfram. 2008. Local Solutions to Global Problems: Climate Change Policies and Regulatory Jurisdiction. *Review of Environmental Economics and Policy* 2 (2): 175–193.

Bushnell, James, Erin Mansur, and Celeste Saravia. 2008. Vertical Arrangements, Market Structure and Competition: an Analysis of Restructured U.S. Electricity Markets. *American Economic Review* 98 (1): 237–266.

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.

------. 2010. *California GHG Emissions - Forecast (2008-2020)*. https://ww3.arb.ca.gov/cc/inventory/data/tables/2020_ghg_emissions_forecast_2010-10-28.pdf

------. 2013. *Semi-Annual Report to the Joint Legislative Budget Committee on Assembly Bill 32*. https://ww2.arb.ca.gov/sites/default/files/classic//research/apr/reports/california-global-warming-solutions-act-2006.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.

------. 2018a. *Mandatory GHG Reporting - Asset Controlling Supplier*. http://ww2.arb.ca.gov/mrr-acs.

------. 2018b. *Mandatory GHG Reporting - Electric Power Entities*. http://ww2.arb.ca.gov/mrr-epe.

------. 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.

------. 2021a. *Cap-and-Trade Program Data*. https://ww2.arb.ca.gov/our-work/programs/cap-and-trade-program/cap-and-trade-program-data.

------. 2021b. *Electrical Distribution Utility and Natural Gas Supplier Allowance Allocation*. https://ww2.arb.ca.gov/our-work/programs/cap-and-trade-program/allowance-allocation/edu-ngs.

------. 2021c. *Mandatory GHG Reporting - Reported Emissions*. 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. 2021a. *Annual and Quarterly Reports*. http://www.caiso.com/market/Pages/MarketMonitoring/AnnualQuarterlyReports/Default.aspx.

------. 2021b. *Daily Renewables Watch*. http://www.caiso.com/market/Pages/ReportsBulletins/RenewablesReporting.aspx.

------. 2021c. *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.

Chan, Nathan, and John Morrow. 2019. Unintended Consequences of Cap-and-Trade? Evidence from the Regional Greenhouse Gas Initiative. *Energy Economics* 80: 411–422.

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. 2021. *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.

Cullenward, Danny, and Andy Coghlan. 2016. Structural Oversupply and Credibility in California's Carbon Market. *The Electricity Journal* 29: 7–14.

Davis, Lucas, and Catherine Hausman. 2016. Market Impacts of a Nuclear Power Plant Closure. *American*

Economic Journal: Applied Economics 8 (2): 92–122.

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.

DSIRE. 2021. *Database of State Incentives for Renewables and Efficiency*. https://www.dsireusa.org/.

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.

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.

------. 2021a. *Average Cost of Fossil Fuels for Electricity Generation*. http://www.eia.gov/electricity/data.cfm.

------. 2021b. *Form EIA-860*. http://www.eia.gov/electricity/data/eia860/.

------. 2021c. *Form EIA-923*. http://www.eia.gov/electricity/data/eia923/.

------. 2021d. *Form EIA-930*. https://www.eia.gov/beta/electricity/gridmonitor/about.

------. 2021e. *How Much Carbon Dioxide Is Produced When Different Fuels Are Burned?*. https://www.eia.gov/tools/faqs/faq.php?id=73&t=11.

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.

------. 2021. *Air Market Program Data*. http://ampd.epa.gov/ampd/.

Federal Energy Regulatory Commission. 2021. *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, and Mar Reguant. 2018. Challenges in the Measurement of Leakage Risk. *AEA Papers and Proceedings* 108: 124–129.

Fowlie, Meredith, Mar Reguant, and Stephen Ryan. 2016. Market-Based Environmental Regulation and Industry Dynamics. *Journal of Political Economy* 124 (1): 249–302.

Gabriel, Steven, Antonio Conejo, J. David Fuller, Benjamin Hobbs, and Carlos Ruiz. 2013. *Complementarity Modeling in Energy Markets*. Springer.

Graff Zivin, Joshua, Matthew Kotchen, and Erin Mansur. 2014. Spatial and Temporal Heterogeneity of Marginal Emissions: Implications for Electric Cars and Other Electricity-Shifting Policies. *Journal of Economic Behavior and Organization* 107 (Part A): 248–268.

Guignet, Dennis, Robin Jenkins, Matthew Ranson, and Patrick Walsh. 2018. Contamination and Incomplete Information: Bounding Implicit Prices using High-Profile Leaks. *Journal of Environmental Economics and Management* 88: 259–282.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica* 66 (5): 1017–1098.

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.

Herrero, Ignacio, Pablo Rodilla, and Carlos Batlle. 2018. Enhancing Intraday Price Signals in U.S. ISO Markets for a Better Integration of Variable Energy Resources. *The Energy Journal* 39 (3): 141–165.

Hobbs, Benjamin, and Udi Helman. 2004. Complementarity-Based Equilibrium Modeling for Electric Power Markets, in *Modeling Prices in Competitive Electricity Markets* (D. W. Bunn, ed.), 69–97. 2004. John Wiley and Sons.

Hobbs, Benjamin, Qingyu Xu, Jonathan Ho, Pearl Donohoo, Saamrat Kasina, Jasmine Ouyang, Sang Woo Park, Joseph Eto, and Vijay Satyal. 2016. Adaptive Transmission Planning: Implementing a New Paradigm for Managing Economic Risks in Grid Expansion. *IEEE Power & Energy Magazine* July-August: 30–40.

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.

------. 2011. Multivariate Matching Methods that are Monotonic Imbalance Bounding. *Journal of the American Statistical Association* 106 (493): 345–361.

------. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20:

1–24.

Jaeger, David, Theodore Joyce and Robert Kaestner. 2020. A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing? *Journal of Business & Economics Statistics* 38 (2): 317–326.

Joskow, Paul, and Edward Kahn. 2002. A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000. *The Energy Journal* 23 (4): 1–35.

Kahn, Matthew, and Erin Mansur. 2013. Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment? *Journal of Public Economics* 101: 105–114.

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. 2015. Media Influences on Social Outcomes: The Impact of MTV's *16 and Pregnant* on Teen Childbearing. *American Economic Review* 105: 3597–3632.

------. 2016. *Does Reality TV Induce Real Effects? A Response to Jaeger, Joyce and Kaestner*. IZA Discussion Paper 10318, Bonn, Germany.

Kindle, Andrew, Daniel Shawhan, and Michael Swider. 2011. *An Empirical Test for Inter-State Carbon-Dioxide Emissions Leakage Resulting from the Regional Greenhouse Gas Initiative*. Unpublished manuscript.

Kotchen, Matthew, and Erin Mansur. 2014. How Stringent Are the U.S. EPA's Proposed Carbon Pollution Standards for New Power Plants? *Review of Environmental Economics and Policy* 8 (2): 290–306.

Legislative Analyst's Office. 2017. *The 2017-18 Budget: Cap-and-Trade*. http://www.lao.ca.gov/reports/2017/3553/cap-and-trade-021317.pdf.

Levinson, Arik, and M. Scott Taylor. 2008. Unmasking the Pollution Haven Effect. *International Economic Review* 49 (1): 223–254.

Mansur, Erin. 2007. Do Oligopolists Pollute Less? Evidence from a Restructured Electricity Market. *Journal of Industrial Economics* 55 (4): 661–689.

National Electric Reliability Council. 2021. *Glossary of Terms Used in NERC Reliability Standards*. http://www.nerc.com/files/glossary_of_terms.pdf.

National Energy Board of Canada. 2021. *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. 2021. *National Centers for Environmental Information*. http://www.ncdc.noaa.gov/.

Nyberg, Michael. 2018. *Thermal Efficiency of Natural Gas-Fired Generation in California: 2017 Update*. http://www.energy.ca.gov/2017publications/CEC-200-2017-003/CEC-200-2017-003.pdf.

Official California Legislative Information. 2015. *Senate Bill No. 350*. http://www.leginfo.ca.gov/pub/15-16/bill/sen/sb_0301-0350/sb_350_cfa_20150910_200142_asm_comm.html.

Panhans, Matthew, Lucia Lavric, and Nick Hanley. 2017. The Effects of Electricity Costs on Firm Re-location Decisions: Insights for the Pollution Haven Hypothesis. *Environmental and Resource Economics* 68 (4): 893–914.

Puller, Steven. 2007. Pricing and Firm Conduct in California's Deregulated Electricity Market. *Review of Economics and Statistics* 89 (1): 75–87.

Rambachan, Ashesh and Jonathan Roth. 2020. *An Honest Approach to Parallel Trends*. https://jonathandroth.github.io/assets/files/HonestParallelTrends_Main.pdf.

------. 2021. *HonestDiD R package*. https://github.com/asheshrambachan/HonestDiD.

Rose, Evan. 2021. Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders. *The Quarterly Journal of Economics* 136 (2): 1199–1253.

Saussay, Aurélien and Misato Sato. 2018. *The Impacts of Energy Prices on Industrial Foreign Investment Location: Evidence from Global Firm Level Data*. Centre for Climate Change Economics and Policy Working Paper 344, London School of Economics, London, U.K.

Schweppe, Fred, Michael Caramanis, Richard Tabors, and Roger Bohn. 1988. *Spot Pricing of Electricity*. Kluwer.

Shawhan, Daniel, John Taber, Di Shi, Ray Zimmerman, Jubo Yan, Charles Marquet, Yingying Qi, Biao Mao, Richard Schuler, William Schulze, and Daniel Tylavsky. 2014. Does a Detailed Model of the Electricity Grid Matter? Estimating the Impacts of the Regional Greenhouse Gas Initiative. *Resource and Energy Economics* 36 (1): 191–207.

Shi, Di, Daniel Shawhan, Nan Li, Daniel Tylavsky, John Taber, Ray Zimmerman, and William Schulze. 2012. Optimal Generation Investment Planning: Pt. 1: Network Equivalents. In *2012 North American Power Symposium (NAPS)*, 1–6.

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/.

Sue Wing, Ian and Marek Kolodziej. 2009. *The Regional Greenhouse Gas Initiative: Emission Leakage and the Effectiveness of Interstate Border Adjustments*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1448748.

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.

------. 2019. *WREGIS System e-Tag User Training*. http://www.wecc.org/Administrative/WREGIS%20System%20e-Tag%20User%20Training%20-%20Updated%20022019.pdf.

Xu, Qingyu, and Benjamin Hobbs. 2019. Value of Model Enhancements: Quantifying the Benefit of Improved Transmission Planning Models. *IET Generation, Transmission & Distribution* 13 (13): 2836–2845.

------. 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.

Figures

Figure 1: NERC regions in the United States, 2016

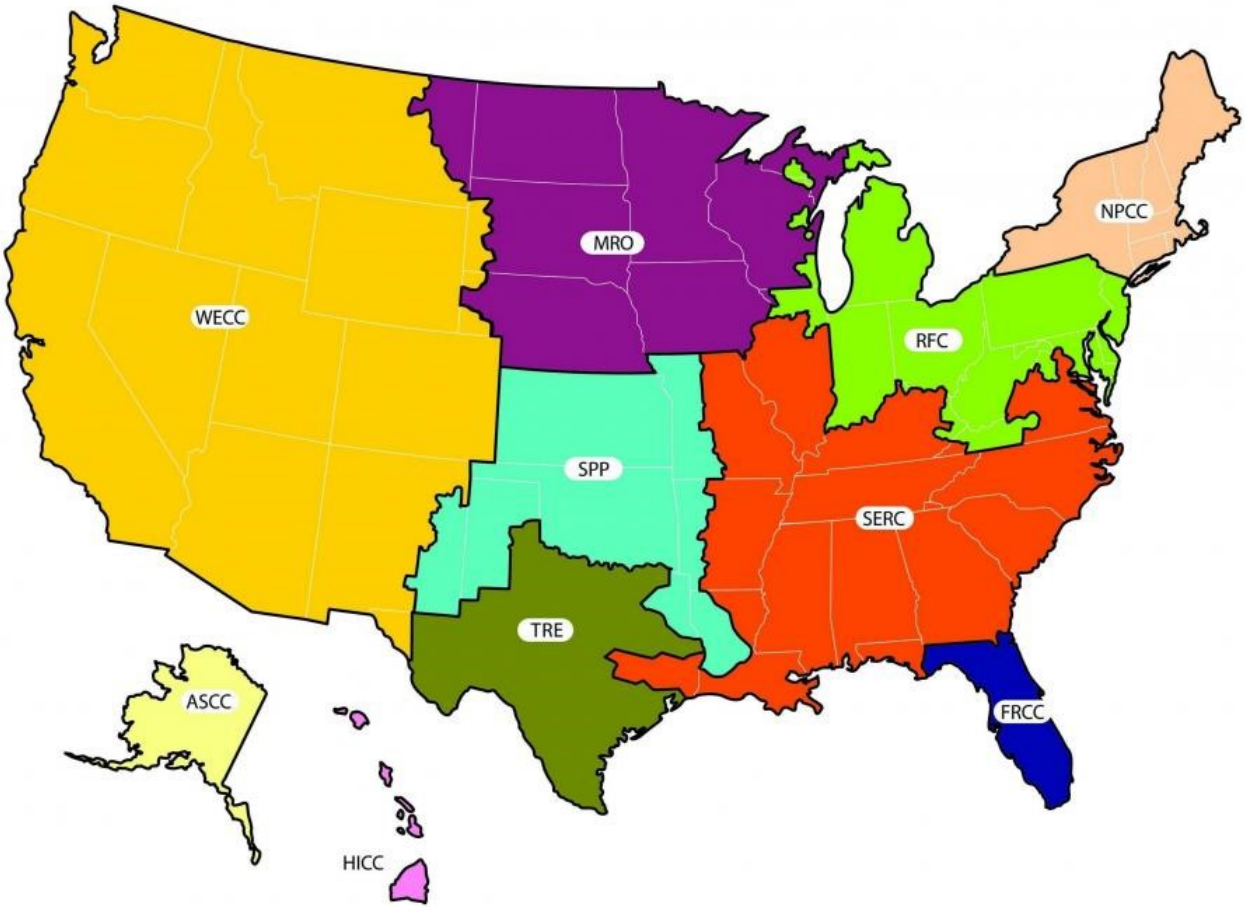
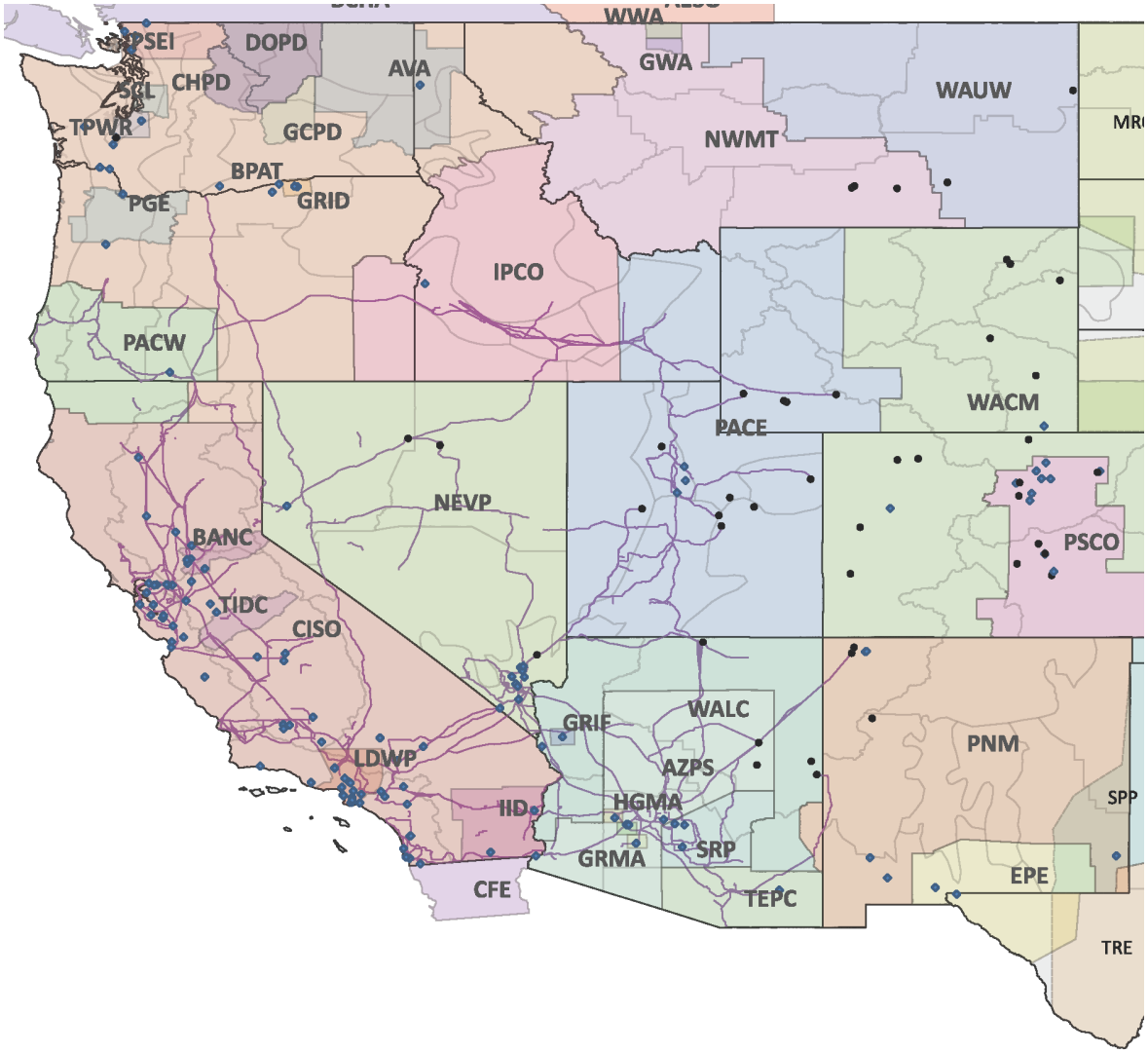


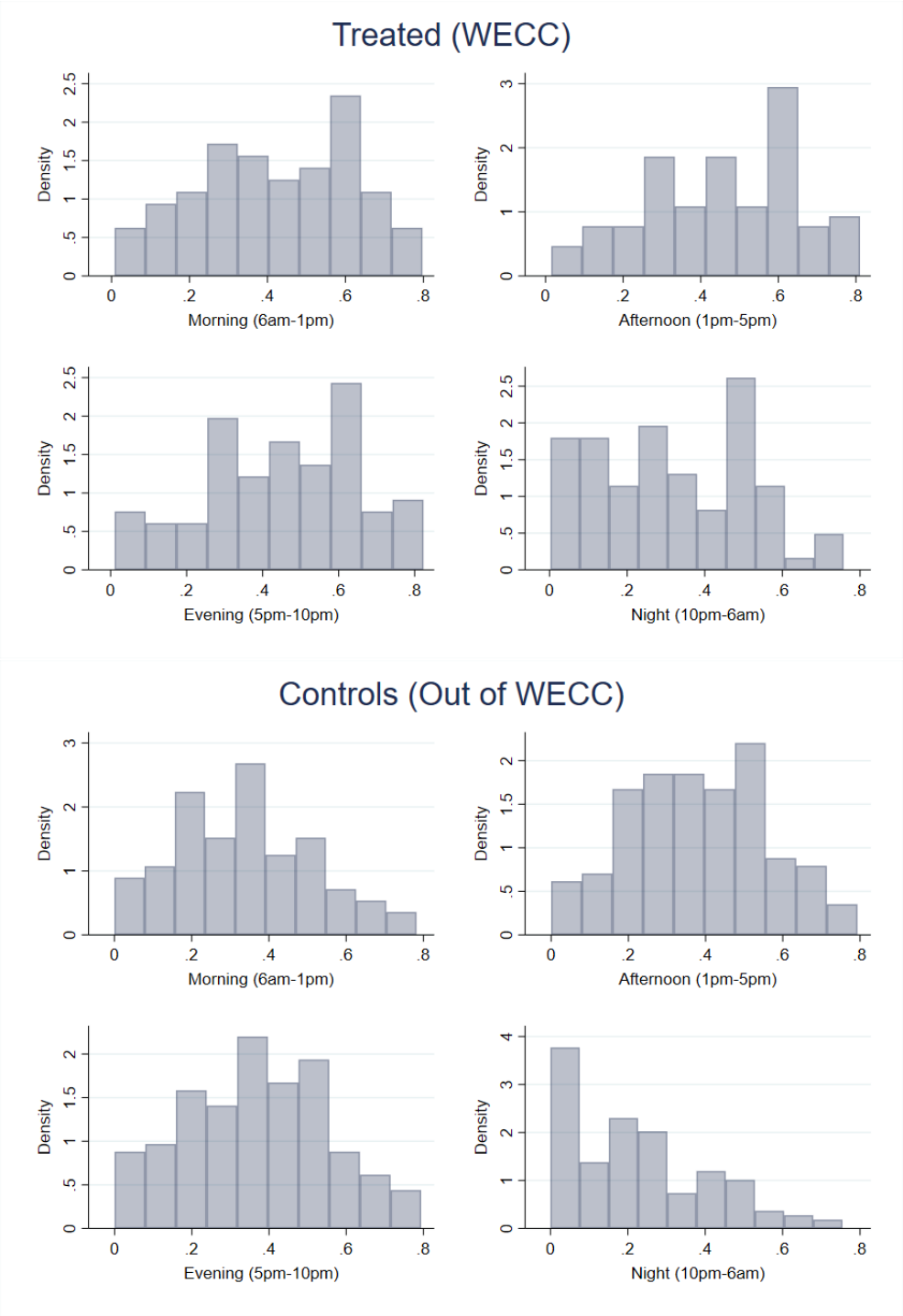
Figure 2: WECC balancing authorities in the United States, 2016



Note: Black dots represent coal-fired power plants, blue diamonds represent NGCC plants.

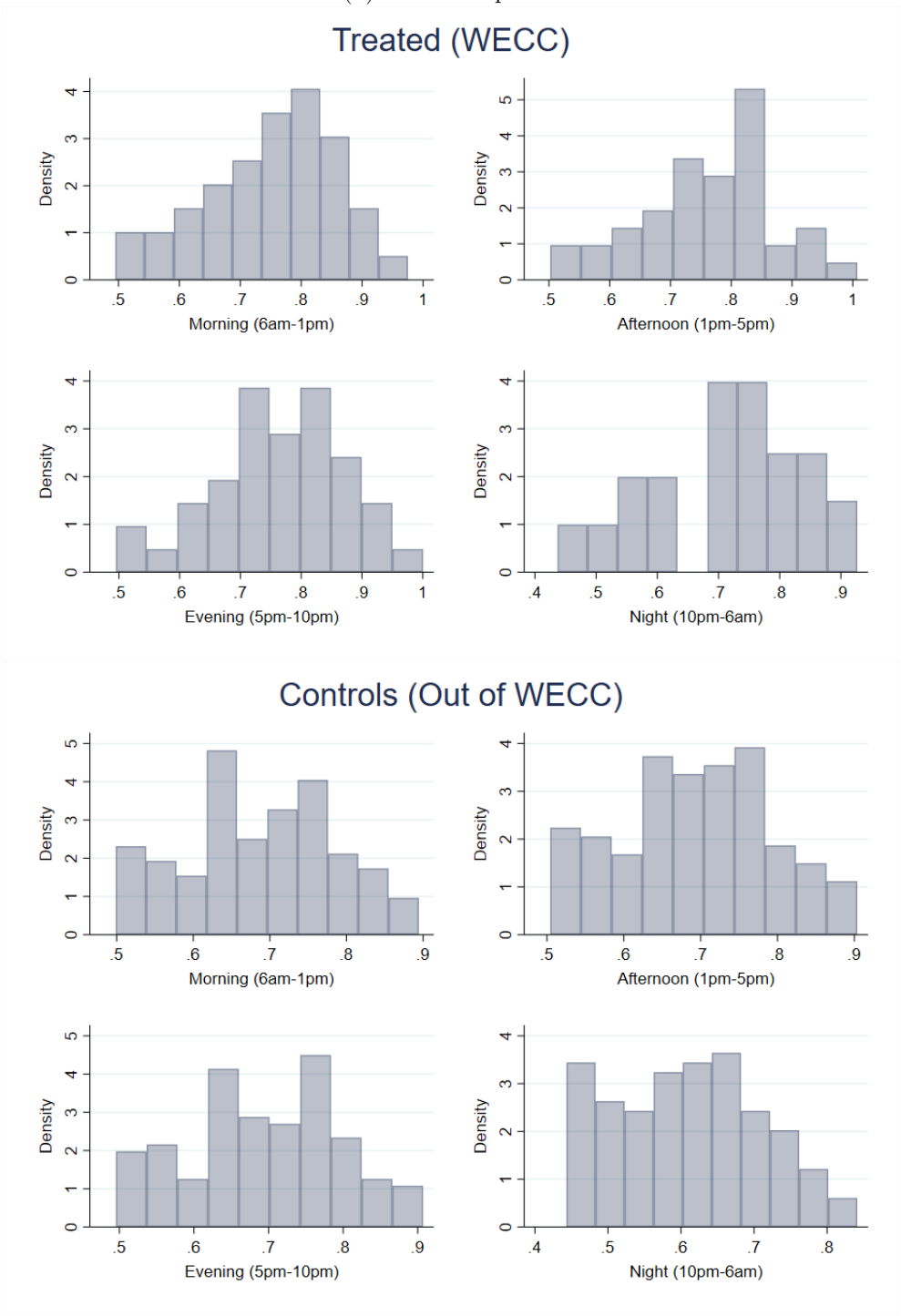
Figure 3: Empirical distribution of 2009-10 average capacity factors by technology, region and block of hour

(a) NGCC plants



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.

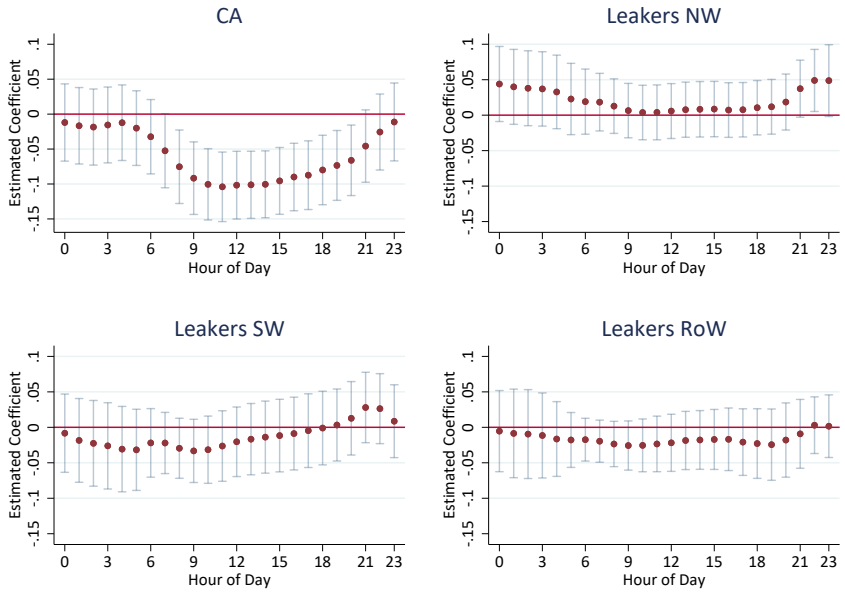
(b) Coal-fired plants



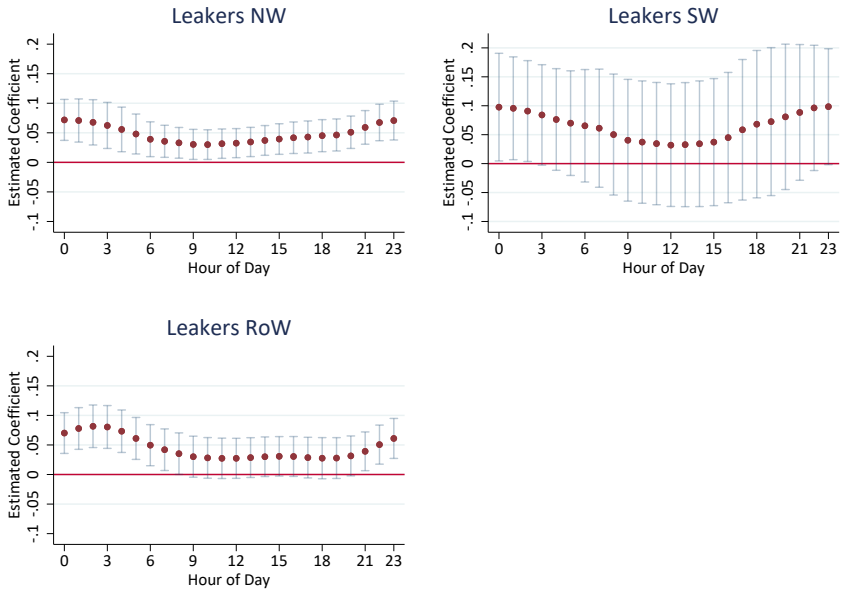
Note: Histograms for the control plants are constrained to include capacity factors above 0.495, 0.502, 0.495, and 0.437 corresponding to the lowest capacity factor of WECC plants (i.e., the lower limit of the first matching bin for the treated plants) in the morning, afternoon, evening and night period,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

Figure 4: Treatment heterogeneity by hour of day

(a) NGCC plants



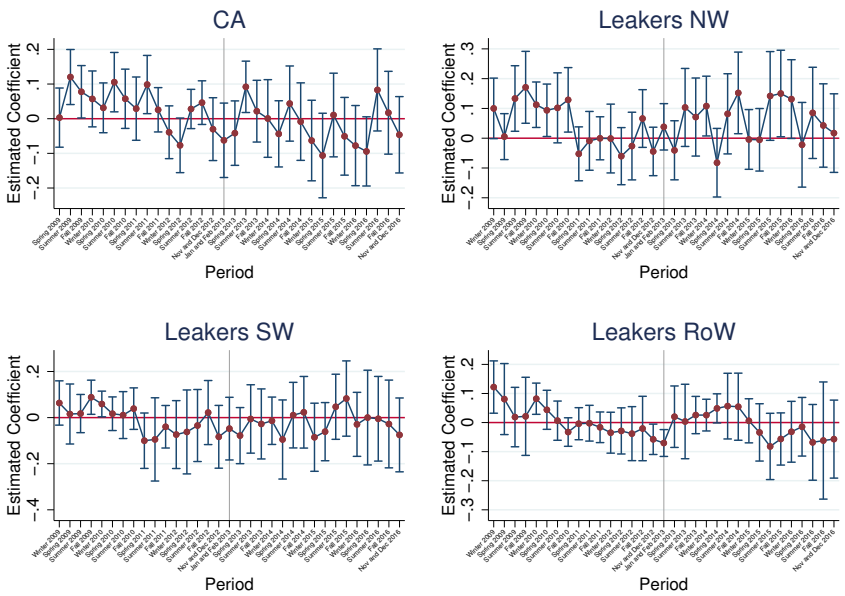
(b) Coal-fired plants



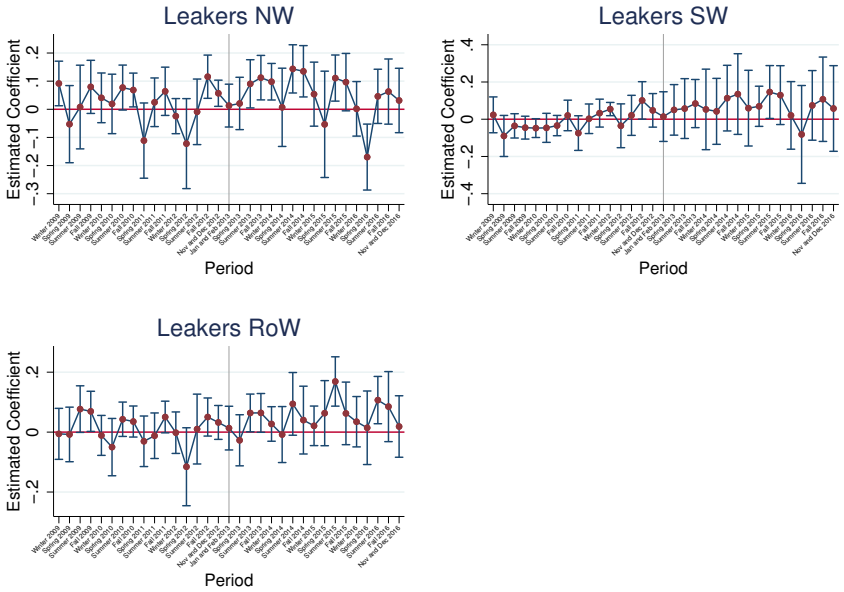
Note: Each graph presents the estimated parameters associated with treatment group specific hourly time trends in equation (1), along with 95% confidence intervals.

Figure 5: Parallel trend tests between treated and control regions

(a) NGCC plants



(b) Coal-fired plants

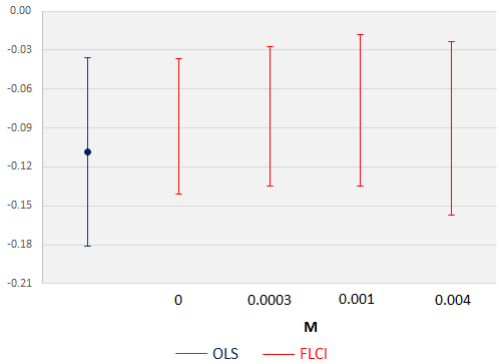


Note: Each graph presents the estimated parameters associated with treatment group specific time trends in equation (24), along with 99% confidence intervals. The vertical line indicates the start date of the cap-and-trade program. In the CA plot, Winter 2009 is dropped due to lack of import data until April 1, 2009.

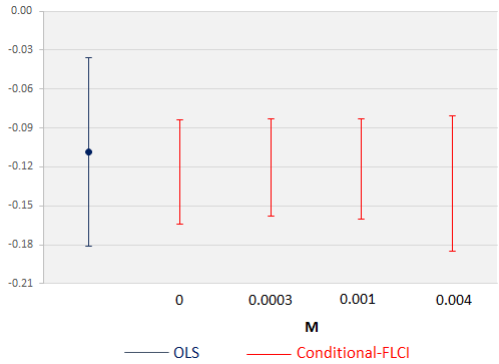
Figure 6: Sensitivity analysis for the event study coefficients in Jan-Feb 2013

California NGCC

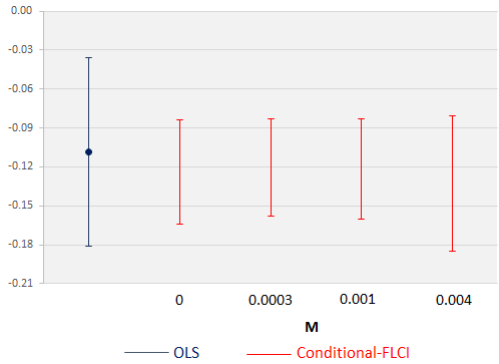
(a) No restrictions



(b) Positive bias

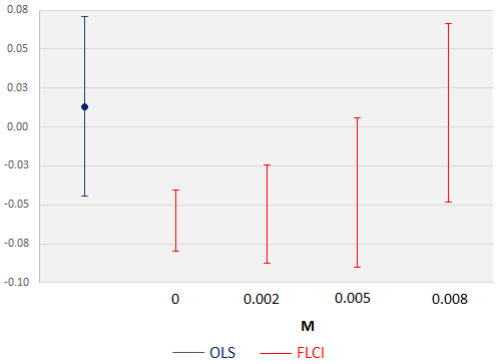


(c) Monotonicity and positive bias

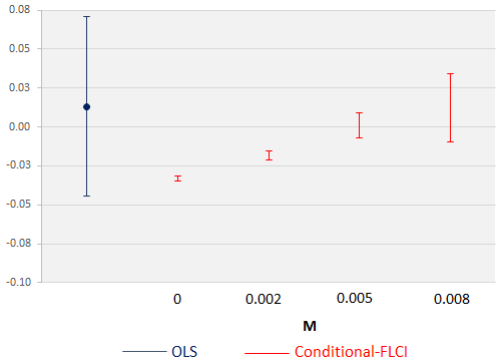


NW coal

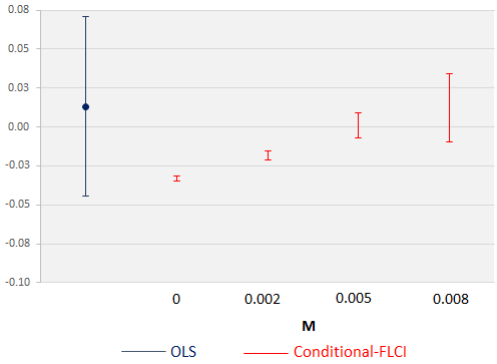
(a) No restrictions



(b) Negative bias

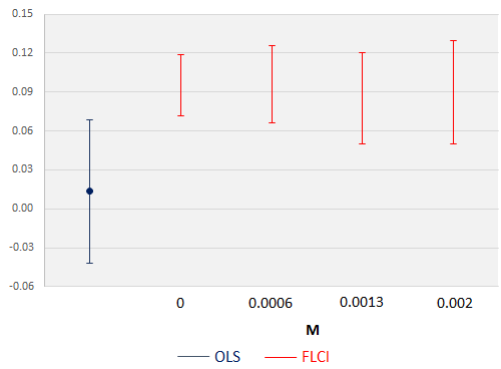


(c) Monotonicity and negative bias

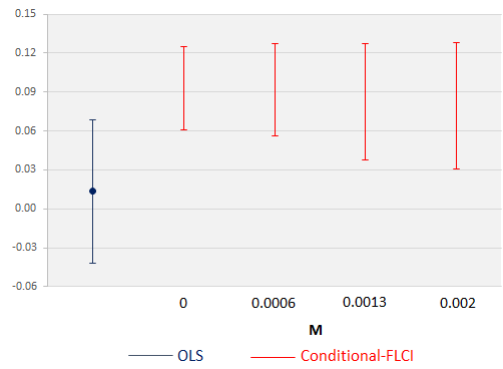


RoW coal

(a) No restrictions



(b) Negative bias



Note: In each panel, "OLS" refers to the 95% confidence intervals for Jan-Feb 2013 treatmed effect estimated using OLS. "FLCI" ("Conditional FLCI") indicates the 95% fixed length confidence interval (conditional fixed length confidence interval) using the Rambachan and Roth (2020) robust inference method.

Tables

Table 1: Summary statistics for NGCC plants

Region	Pre ETS						Post ETS					
	Number of plants	Nameplate capacity (MW)	Capacity factor (%)	Heat rate (Btu/kWh)	CO_2 emission rate (lb/MMBtu)	Age (years)	Number of plants	Nameplate Capacity (MW)	Capacity factor (%)	Heat rate (Btu/kWh)	CO_2 emission rate (lb/MMBtu)	Age (years)
WECC												
CA	67	309	0.54	9,028	118.14	10.67	69	332	0.49	8,913	117.98	11.88
		(334)	(0.31)	(2,246)	(1.45)	(11.30)		(329)	(0.31)	(2,356)	(2.37)	(7.87)
NW	34	456	0.46	8,259	118.41	8.34	34	479	0.50	7,967	118.51	11.98
		(280)	(0.32)	(2,667)	(1.41)	(4.46)		(308)	(0.30)	(1,246)	(0.73)	(4.77)
RoW	15	361	0.23	9,924	118.23	11.30	15	371	0.28	9,373	118.38	11.76
		(260)	(0.20)	(4,178)	(1.06)	(8.33)		(253)	(0.25)	(3,732)	(0.81)	(5.99)
SW	13	875	0.36	8,176	118.44	7.65	13	905	0.34	8,138	118.50	11.69
		(673)	(0.26)	(1,276)	(0.75)	(1.77)		(665)	(0.25)	(1,777)	(0.78)	(1.78)
Controls												
FRCC	31	835	0.49	8,139	118.78	16.27	31	911	0.52	7,915	118.84	18.96
		(877)	(0.21)	(1,637)	(1.07)	(12.55)		(964)	(0.24)	(1,621)	(1.02)	(12.70)
MRO-US	17	381	0.13	9,054	118.55	12.70	17	361	0.21	9,186	118.54	16.11
		(217)	(0.15)	(3,426)	(1.15)	(13.98)		(216)	(0.20)	(3,717)	(1.03)	(14.58)
SERC	58	670	0.46	8,777	118.91	9.53	58	671	0.51	8,723	118.88	13.22
		(476)	(0.28)	(3,198)	(11.23)	(5.15)		(473)	(0.27)	(3,082)	(9.65)	(5.94)
SPP	21	598	0.38	8,256	118.54	19.15	21	597	0.37	8,306	118.56	23.14
		(388)	(0.26)	(1,424)	(0.71)	(16.96)		(385)	(0.26)	(1,596)	(0.68)	(16.95)
TRE	58	652	0.44	9,147	118.39	11.60	58	656	0.45	9,090	118.04	15.09
		(381)	(0.24)	(2,721)	(1.29)	(8.88)		(387)	(0.24)	(2,772)	(4.14)	(9.21)

Note: Pre ETS refers to January 2009-December 2012 and post ETS to January 2013-December 2016. Standard deviations are reported in parentheses.

Table 2: Summary statistics for coal-fired plants

Region	Pre ETS						Post ETS					
	Number of plants	Nameplate capacity (MW)	Capacity factor (%)	Heat rate (Btu/kWh)	CO_2 emission rate (lb/MMBtu)	Age (years)	Number of plants	Nameplate Capacity (MW)	Capacity factor (%)	Heat rate (Btu/kWh)	CO_2 emission rate (lb/MMBtu)	Age (years)
WECC												
NW	18	779	0.73	11,472	207.58	32.48	18	765	0.69	12,527	208.13	35.56
		(648)	(0.23)	(4,801)	(2.33)	(11.53)		(660)	(0.24)	(7,566)	(2.14)	(11.05)
RoW	25	537	0.74	12,484	208.57	29.53	25	558	0.72	11,878	208.55	29.84
		(659)	(0.21)	(4,911)	(3.01)	(15.44)		(669)	(0.22)	(3,404)	(4.28)	(15.74)
SW	7	1,289	0.72	12,229	206.22	31.09	6	1,376	0.70	10,497	206.45	34.75
		(837)	(0.17)	(4,652)	(3.29)	(8.92)		(688)	(0.17)	(418)	(2.45)	(8.91)
Controls												
FRCC	10	1,081	0.53	11,197	201.53	26.85	10	1,078	0.43	11,287	198.50	30.92
		(678)	(0.22)	(2,415)	(26.34)	(7.67)		(678)	(0.22)	(2,092)	(17.29)	(7.64)
MRO-US	49	459	0.54	13,814	209.84	38.18	56	491	0.52	13,591	210.55	40.55
		(515)	(0.23)	(6,538)	(3.17)	(11.06)		(512)	(0.23)	(6,409)	(3.98)	(11.43)
SERC	71	1,232	0.59	11,675	205.92	38.68	71	1,247	0.51	11,801	205.78	41.26
		(836)	(0.23)	(4,764)	(8.47)	(10.80)		(835)	(0.24)	(4,722)	(9.64)	(11.96)
SPP	35	689	0.63	11,730	208.44	33.70	35	706	0.55	11,680	208.87	34.47
		(546)	(0.24)	(2,980)	(5.73)	(9.86)		(529)	(0.26)	(3,395)	(7.78)	(13.06)
TRE	18	1,134	0.72	10,870	213.14	26.14	20	1,167	0.64	11,170	213.06	26.97
		(762)	(0.26)	(1,051)	(5.22)	(8.91)		(739)	(0.26)	(2,203)	(5.28)	(12.21)

Note: Pre ETS refers to January 2009-December 2012 and post ETS to January 2013-December 2016. Standard deviations are reported in parentheses.

Table 3: Econometric model results: Estimated treatment effects for NGCC plants

	(1) Baseline	(2) Robustness to hydro-nuclear and renewable generation			(3) Robustness to outcome frequency		(4) Robustness to clustering		(5) Robustness to matching set
		(a)	(b)	(c)	(a)	(b)	(a)	(b)	
CA	-0.058^{**}	-0.042^{*}	-0.064^{**}	-0.030	-0.069^{**}	-0.059^{**}	-0.058^{***}	-0.058^{**}	-0.060^{**}
	(0.025)	(0.025)	(0.027)	(0.022)	(0.030)	(0.025)	(0.002)	(0.025)	(0.027)
NW	0.021	0.039^{*}	0.019	0.022	-0.001	0.020	0.021	0.021	0.015
	(0.020)	(0.022)	(0.019)	(0.020)	(0.023)	(0.019)	(0.011)	(0.017)	(0.019)
RoW	-0.015	0.002	-0.018	-0.014	-0.007	-0.016	-0.015	-0.015	-0.034
	(0.018)	(0.021)	(0.018)	(0.018)	(0.020)	(0.018)	(0.009)	(0.016)	(0.025)
SW	-0.011	0.006	-0.013	-0.010	-0.008	-0.012	-0.011	-0.011	-0.014
	(0.023)	(0.026)	(0.023)	(0.023)	(0.025)	(0.023)	(0.011)	(0.029)	(0.023)
Before matching									
CA plants	40	40	40	40	70	40	40	40	40
Leaker plants	48	48	48	48	62	48	48	48	48
Control plants	153	153	153	153	185	153	153	153	153
After matching									
CA plants	33	33	33	33	33	33	33	33	34
Leaker plants	40	40	40	40	40	40	40	40	44
Control plants	128	128	128	128	128	128	128	128	140
Number of obs	567,484	567,484	567,484	567,484	18,199	13,619,137	567,484	567,484	616,666
Number of clusters	201	201	201	201	201	201	6	38	218

Note: The unit of observation is plant-day for specifications (1), (2), (4) and (5), plant-month for specification (3a), and plant-hour for specification (3b). All regressions include plant, year, day-of-week and state by month-year fixed effects. Specification (3b) also includes hour-of-day fixed effects. Standard errors are reported in parentheses, and clustered by plant in specifications (1), (2), (3) and (5), by NERC region in (4a), and by balancing authority in (4b). *, **, and *** indicate statistical significance at 10%, 5% and 1% level, respectively. The number of plants before matching includes existing, new and retired facilities between 2009 and 2016, and thus differs from Table 1. In specifications (1), (2), (3b), (4) and (5), this number refers to plants reporting to CEMS; in specification (3a), it refers to plants completing the EIA-923 survey.

Table 4: Econometric model results: Estimated treatment effects for coal-fired plants

	(1) Baseline	(2) Robustness to hydro-nuclear and renewable generation		(3) Robustness to outcome frequency		(4) Robustness to clustering		(5) Robustness to matching set
		(a)	(b)	(a)	(b)	(a)	(b)	
NW	0.047***	0.050***	0.051***	0.051***	0.047***	0.047**	0.047***	0.041***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)	(0.017)	(0.011)
RoW	0.044***	0.051***	0.052***	0.047**	0.044***	0.044**	0.044**	0.043***
	(0.016)	(0.016)	(0.015)	(0.019)	(0.016)	(0.012)	(0.020)	(0.015)
SW	0.059	0.075	0.063	0.030	0.064	0.059**	0.059	0.065
	(0.047)	(0.046)	(0.047)	(0.033)	(0.052)	(0.018)	(0.054)	(0.055)
Before matching								
Leaker plants	42	42	42	50	42	42	42	42
Control plants	170	170	170	192	170	170	170	170
After matching								
Leaker plants	40	40	40	40	40	40	40	41
Control plants	94	94	94	94	94	94	94	150
Number of obs	379,028	379,028	379,028	12,455	9,096,375	379,028	379,028	533,642
Number of clusters	134	134	134	134	134	6	28	191

Note: The unit of observation is plant-day for specifications (1), (2), (4) and (5), plant-month for specification (3a), and plant-hour for specification (3b). All regressions include plant, year, day-of-week and state by month-year fixed effects. Specification (3b) also includes hour-of-day fixed effects. Standard errors are reported in parentheses, and clustered by plant in specifications (1), (2), (3) and (5), by NERC region in (4a), and by balancing authority in (4b). *, **, and *** indicate statistical significance at 10%, 5% and 1% level, respectively. The number of plants before matching includes existing, new and retired facilities between 2009 and 2016, and thus differs from Table 2. In specifications (1), (2), (3b), (4) and (5), this number refers to plants reporting to CEMS; in specification (3a), it refers to plants completing the EIA-923 survey.

Table 5: Econometric model results: Treatment heterogeneity between day and night

		(6)		(7)		(8)		(9)
		NGCC	Coal	NGCC	Coal	NGCC	Coal	NGCC
Day								
	CA	-0.089^{***}	-	-0.075^{***}	-	-0.095^{***}	-	-0.060^{***}
		(0.024)	-	(0.025)	-	(0.026)	-	(0.022)
	NW	0.008	0.035^{***}	0.025	0.035^{**}	0.005	0.039^{***}	0.009
		(0.019)	(0.012)	(0.020)	(0.014)	(0.018)	(0.012)	(0.019)
	RoW	-0.020	0.031^{*}	-0.003	0.034^{**}	-0.024	0.039^{**}	-0.020
		(0.019)	(0.017)	(0.020)	(0.016)	(0.019)	(0.015)	(0.019)
	SW	-0.018	0.050	-0.001	0.062	-0.020	0.054	-0.016
		(0.023)	(0.061)	(0.026)	(0.056)	(0.023)	(0.061)	(0.023)
Night								
	CA	-0.028	-	-0.014	-	-0.035	-	0.001
		(0.026)	-	(0.026)	-	(0.028)	-	(0.023)
	NW	0.033	0.058^{***}	0.050	0.058^{***}	0.030	0.062^{***}	0.034
		(0.022)	(0.015)	(0.025)	(0.015)	(0.022)	(0.015)	(0.022)
	RoW	-0.011	0.059^{***}	0.006	0.062^{***}	-0.014	0.067^{***}	-0.010
		(0.020)	(0.017)	(0.024)	(0.016)	(0.020)	(0.015)	(0.020)
	SW	-0.006	0.091	0.010	0.104^{**}	-0.008	0.095^{*}	-0.005
		(0.025)	(0.057)	(0.027)	(0.051)	(0.025)	(0.057)	(0.025)
Number of obs		13,619,137	9,096,375	13,619,137	9,096,375	13,619,137	9,096,375	13,619,137
Number of 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. Standard errors are clustered by plant and reported in parentheses.

Table 6: Capacity factor balance in full and matched samples

	NGCC		Coal	
	Before matching	After matching	Before matching	After matching
Hour 0	2.075**	0.192	6.364***	1.004
Hour 1	1.932*	0.064	6.670***	1.207
Hour 2	1.856*	−0.001	6.779***	1.263
Hour 3	1.835*	−0.041	6.707***	1.171
Hour 4	1.869*	−0.051	6.357***	0.892
Hour 5	1.772*	−0.160	5.832***	0.564
Hour 6	1.859*	−0.120	5.064***	0.126
Hour 7	2.001**	−0.043	4.636***	−0.028
Hour 8	1.917*	−0.185	4.381***	−0.057
Hour 9	1.621	−0.509	4.244***	−0.021
Hour 10	1.377	−0.662	4.215***	0.055
Hour 11	1.266	−0.692	4.217***	0.058
Hour 12	1.298	−0.601	4.249***	0.047
Hour 13	1.380	−0.529	4.350***	0.135
Hour 14	1.399	−0.502	4.468***	0.195
Hour 15	1.392	−0.502	4.542***	0.272
Hour 16	1.339	−0.542	4.528***	0.229
Hour 17	1.383	−0.503	4.380***	0.151
Hour 18	1.426	−0.488	4.218***	0.048
Hour 19	1.500	−0.405	4.172***	0.005
Hour 20	1.659*	−0.203	4.427***	0.133
Hour 21	2.100**	0.391	4.852***	0.290
Hour 22	2.320**	0.567	5.442***	0.590
Hour 23	2.283**	0.441	5.929***	0.771
Morning	1.646	−0.402	4.447***	0.026
Afternoon	1.378	−0.519	4.473***	0.208
Evening	1.627	−0.240	4.419***	0.128
Night	2.009**	0.129	6.308***	0.945

Note: The table reports t statistics of a two-sided test of mean comparisons between treated and control groups before and after matching. *, **, and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

Table 7: Parallel trends test: Treatment effects with and without trends

	No trend		Linear trend		Quadratic trend	
	NGCC	Coal	NGCC	Coal	NGCC	Coal
CA	-0.058^{**}	-	0.012	-	0.017	-
	(0.025)	-	(0.027)	-	(0.028)	-
NW	0.021	0.047^{***}	-	0.110^{***}	-	0.112^{***}
	(0.020)	(0.013)	-	(0.028)	-	(0.028)
RoW	-0.015	0.044^{**}	-	0.040^{*}	-	0.040^{*}
	(0.018)	(0.016)	-	(0.021)	-	(0.021)
SW	-0.011	0.059	-	-	-	-
	(0.023)	(0.047)	-	-	-	-

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. The table presents the estimated coefficients in the augmented models.

Table 8: Econometric model results: Emissions and leakage based on the econometric estimates, 2013 and 2016

2013	Lower bound of the robust 95% CI	Upper bound of the robust 95% CI
Change in CA local emissions (E_1)	-4.31	0.38
Change in CA import emissions (E_2)	-5.34	-5.34
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	-9.65	-4.96
Change in WECC-NonCA emissions (E_4)	4.69	11.74
- NW	2.29	4.91
- RoW	2.40	6.83
Change in WECC emissions ($E_5 = E_1 + E_4$)	0.38	12.12
Leakage $[(1 - E_5/E_3) \times 100\%]$	*104.0%*	*344.4%*

2016	Lower bound of the robust 95% CI	Upper bound of the robust 95% CI
Change in CA local emissions (E_1)	-5.33	-1.33
Change in CA import emissions (E_2)	-5.91	-5.91
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	-11.23	-7.23
Change in WECC-NonCA emissions (E_4)	3.06	21.34
- NW	-6.75	-1.46
- RoW	9.81	22.80
Change in WECC emissions ($E_5 = E_1 + E_4$)	-2.26	20.02
Leakage $[(1 - E_5/E_3) \times 100\%]$	*79.9%*	*376.7%*

Note: Emissions are in million metric tons of CO_2 per year.

Table 9: 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	26.8%	26.8%	0	30.0%	30.0%	0
	NGCC	39.0%	38.6%	-0.003	42.4%	31.2%	-0.112
	NGCT	21.2%	21.0%	-0.001	20.5%	20.6%	0.001
	Nuclear	95.0%	95.0%	0	95.0%	95.0%	0
	Oil	5.3%	5.3%	0	5.3%	5.3%	0
	Solar	9.4%	9.4%	0	20.0%	20.0%	0
	Wind	26.0%	26.0%	0	22.0%	22.0%	0
NW	Coal	84.9%	84.8%	-0.0003	77.5%	79.9%	0.024
	Hydro	29.8%	29.8%	0	30.9%	30.9%	0
	NGCC	55.2%	55.2%	0	59.1%	62.4%	0.034
	NGCT	3.1%	3.2%	0.0005	3.5%	3.8%	0.003
	Nuclear	95.0%	95.0%	0	95.0%	95.0%	0
	Oil	94.0%	94.0%	0	94.0%	94.0%	0
	Solar	22.0%	22.0%	0	24.6%	24.6%	0
	Wind	21.1%	21.1%	0	15.0%	15.0%	0
RoW	Coal	86.0%	86.0%	0.001	78.5%	78.6%	0.001
	Hydro	31.5%	31.5%	0	39.9%	39.9%	0
	NGCC	18.3%	18.3%	0	31.9%	32.8%	0.010
	NGCT	3.7%	3.7%	0	5.7%	6.4%	0.007
	Nuclear	-	-	-	-	-	-
	Oil	18.2%	18.2%	0	18.2%	18.2%	0
	Solar	16.8%	16.8%	0	15.9%	15.9%	0
	Wind	30.8%	30.8%	0	29.9%	30.4%	0.005
SW	Coal	81.0%	80.9%	-0.001	62.9%	63.6%	0.007
	Hydro	31.3%	31.3%	0	36.3%	36.3%	0
	NGCC	24.6%	24.9%	0.003	31.4%	43.6%	0.123
	NGCT	2.8%	2.8%	0	2.5%	2.6%	0.001
	Nuclear	95.0%	95.0%	0	95.0%	95.0%	0
	Oil	0%	0%	0	0%	0%	0
	Solar	19.2%	19.2%	0	27.4%	27.4%	0
	Wind	15.6%	15.6%	0	14.1%	14.1%	0

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

Table 10: Simulation model results: Effect of California's cap-and-trade program on emissions and leakage, 2013 and 2016

2013	No cap	Carbon cap
Local emissions in CA	35.49	34.94
Emissions of CA imports	54.50	42.48
Regulated emissions in CA	89.99	77.42
Emissions in WECC-NonCA	258.91	259.05
- NW	108.80	108.79
- RoW	88.99	89.10
- SW	61.13	61.17
Total emissions in WECC	294.40	293.99
Change in CA local emissions (E_1)		-0.55
Change in CA import emissions (E_2)		-12.02
Change in CA regulated emissions ($E_3 = E_1 + E_2$)		-12.56
Change in WECC-NonCA emissions (E_4)		0.14
Change in WECC emissions (E_5)		-0.41
Leakage $[(1 - E_5/E_3) \times 100\%]$		*96.8%*

2016	No cap	Carbon cap
Local emissions in CA	40.33	32.64
Emissions of CA imports	33.27	25.40
Regulated emissions in CA	73.60	58.03
Emissions in WECC-NonCA	232.33	241.64
- NW	102.70	106.84
- RoW	81.96	82.38
- SW	47.68	52.43
Total emissions in WECC	272.66	274.28
Change in CA local emissions (E_1)		-7.69
Change in CA import emissions (E_2)		-7.88
Change in CA regulated emissions ($E_3 = E_1 + E_2$)		-15.57
Change in WECC-NonCA emissions (E_4)		9.31
Change in WECC emissions (E_5)		1.62
Leakage $[(1 - E_5/E_3) \times 100\%]$		*110.4%*

Note: Emissions are in million metric tons of CO_2 per year.

Table 11: Comparison of emissions and leakage results, 2013 and 2016

2013	Simulation model (JHSMINE)	Econometric model Robust 95% CI, $M = 0$		Econometric model Robust 95% CI, $M > 0$	
		LB	UB	LB	UB
Change in CA local emissions (E_1)	−0.22	−4.31	0.38	−4.16	0.62
Change in CA import emissions (E_2)	−12.02	−5.34	−5.34	−5.34	−5.34
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	−12.23	−9.65	−4.96	−9.50	−4.72
Change in WECC-NonCA emissions (E_4)	0.06	4.69	11.74	5.95	22.00
Change in WECC emissions ($E_5 = E_1 + E_4$)	−0.15	0.38	12.12	1.80	22.62
Leakage $[(1 - E_5/E_3) \times 100\%]$	*98.7%*	*104.0%*	*344.4%*	*118.9%*	*579.2%*

2016	Simulation model (JHSMINE)	Econometric model Robust 95% CI, $M = 0$		Econometric model Robust 95% CI, $M > 0$	
		LB	UB	LB	UB
Change in CA local emissions (E_1)	−7.32	−5.33	−1.33	−6.44	−1.33
Change in CA import emissions (E_2)	−7.88	−5.91	−5.91	−5.91	−5.91
Change in CA regulated emissions ($E_3 = E_1 + E_2$)	−15.20	−11.23	−7.23	−12.35	−7.24
Change in WECC-NonCA emissions (E_4)	2.42	3.06	21.34	2.37	29.40
Change in WECC emissions ($E_5 = E_1 + E_4$)	−4.90	−2.26	20.02	−4.07	28.07
Leakage $[(1 - E_5/E_3) \times 100\%]$	*67.8%*	*79.9%*	*376.7%*	*67.0%*	*487.8%*

Note: Changes in emissions are in million metric tons of CO_2 per year. For both models, changes in CA local emissions refer to NGCC plants, and changes in WECC-NonCA emissions refer to coal-fired plants in the NW and RoW leaker regions. LB indicates the lower bound of the robust confidence interval, while UB refers to its upper bound.