Estimating the Congestion Benefits of Batteries When Network Connections are Unobserved

A. Justin Kirkpatrick *

May 22, 2026

Abstract

I estimate the grid impacts of 126.3 MW of battery storage capacity operating in California during 2010-2017. I find that 1MW of energy storage decreases evening peak prices by $1.11/MWh at the pricing node where the storage is installed while off-peak prices do not increase. The unobserved network structure is revealed and the price effects are estimated at other locations across the grid using a double-pooled LASSO-based estimator that reflects the grid's constrained optimization price-determination process. Results imply a reduction in the cost of serving load of $123,700 per year per MW, an amount similar in magnitude to private revenue from arbitrage. Overall, I find early energy storage mandates in California are partially justified by reductions in the cost of serving load, and that effects are strongest at times in which increased penetration of zero-carbon solar generation impose higher costs on the grid. These patterns likely hold in any grid setting with inelastic peak supply relative to off-peak supply.

Key words: electricity grid, battery storage, renewables, decarbonization

JEL classification numbers: Q41, Q48, L14, H23

*Michigan State University, Department of Economics, 110 Marshall-Adams Hall, East Lansing, MI 48824 (email: jkirk@msu.edu).

1 Introduction

Since the time of Edison, the absence of economical electricity storage has raised the cost of reliable electricity supply, requiring costly generation that can follow net demand. The inability to store electricity is becoming more problematic amid substantial investments in intermittent renewable generation capacity that are a response to state and federal policies intended to achieve environmental and energy security objectives. Conventional power plants, therefore, increasingly sit idle during much of the day and undertake costly processes to ramp up generation to match fluctuations in renewable supply. The cost of conventional dispatch and, hence, the value of electricity storage are growing with renewable capacity (Joskow, 2011; Gowrisankaran et al., 2016; Bushnell and Novan, 2021; Butters et al., 2025).

From a policy perspective, electricity storage is necessary to address the costs of meeting environmental objectives through increased renewable penetration. To that end, in 2010, California passed AB2514 which mandated energy storage targets for publicly owned utilities with the express intent of "facilitating integration of renewables." Though battery storage capacity grew by 3.6GW per year in 2021 nationally, reflecting $5B in annual investment (Bloomberg Finance, 2023), the impacts of batteries on prices and costs of serving electricity load have not heretofore been empirically estimated in a manner robust to grid congestion, market power, and alternative battery operator objectives.

Energy storage facilitates integration of intermittent renewable generation by shifting supply from periods of abundant renewable generation to periods when it is scarce and dispatchable generation must be called up to meet the difference in load and renewables supply. Storage lowers peak prices by reducing the quantity of dispatchable generation demanded during peak periods. Because most electric grids are characterized by convex supply curves, the price reductions achieved by battery discharge at peak demand exceed the price increases induced by the recharging of storage capacity during off-peak periods. Such marginal price changes affect prices paid for every unit of generation be-

cause wholesale markets are commonly settled by uniform price auctions. Storage also lower costs associated with grid congestion. Congestion induces sub-optimal production because electricity trades across areas are constrained by transmission capacity, exacerbating price spikes resulting from increased renewable penetration. By increasing supply during peak periods, batteries relax these constraints, reducing congestion costs.

Reductions in costs of serving electricity demand, or load, are savings to load-serving entities (LSEs) that constitute transfers from non-marginal generators (Walawalkar et al., 2008). In states with cost-of-service regulation, such cost savings are expected to be passed on to ratepayers. These grid savings are not appropriated by private storage operators, whose returns from arbitraging high and low prices are expected to decline in storage capacity. This represents a misalignment between ratepayer savings and private storage investment – private arbitrage profits from storage are maximized when storage has no price effect; policy objectives are obtained when storage results in a significant price effect.

This paper examines the effect of storage-induced congestion relief on wholesale electricity prices in California, one of the largest early markets for energy storage in the world with over 126.3 megawatts (MW) of installed battery capacity as of 2017. I develop an empirical model of electricity prices that accounts for intermittent generation and congestion. I quantify price changes using the staggered installation of 19 utility-scale batteries during a period in which the battery market was maturing, but was largely driven by subsidies and mandates, rather than location-specific price spreads (pre-2018). I estimate price changes for the nodes at which storage is installed and for other nodes where prices are likely to respond because of grid spillovers. I aggregate these changes to calculate the magnitude of grid-level congestion-relief savings generated by marginal storage capacity. Because the architecture of the electric grid is unobserved due to national security considerations, I use machine learning to determine price dependencies across nodes of the grid.

I find that a 1 MW increase in storage capacity reduces afternoon peak prices by

up to \$1.11 in the day-ahead market, primarily in the late-afternoon and early evening, coinciding with the period of highest electricity demand net of non-dispatchable renewables. Meanwhile, off-peak prices remain largely unchanged. This results in a decrease in the annual cost of serving load of \$123,700 per 1 MW increase in storage capacity. This annual grid benefit partly justifies storage mandates like that of California, which were estimated to cost \$6,500,000 per MW of capacity (Penna et al., 2016), though costs are rapidly falling. Private benefits from arbitrage or ancillary services range from \$49,500 to \$208,500 per year (Penna et al., 2016).[1] Thus, these grid savings are approximately 59.0% to 248.6% of the private benefits that accrue to battery operators.

This is the first paper to flexibly and empirically estimate price effects of storage and its grid-level impacts in a market with congestion and high renewable penetration. As such, it contributes to two growing areas of literature. First, it contributes to literature that explores the impacts of new grid technology and policy on market outcomes and grid operations, focusing primarily on the price and welfare impacts of zero-carbon, non-dispatchable wind and solar (Woo et al., 2011; Kaffine et al., 2013; Novan, 2015; Fell and Kaffine, 2018; Bushnell and Novan, 2021; Fell et al., 2021; Craig et al., 2018; Wolak, 2016; Liski and Vehviläinen, 2020; Reguant, 2019; Jha and Leslie, 2025; LaRiviere and Lyu, 2022; Petersen et al., 2022). Although previous papers (Sioshansi, 2011; Butters et al., 2025; Karaduman, 2021; Liski and Vehviläinen, 2026) model the changes in market equilibrium associated with the addition of grid-scale batteries and Carson and Novan (2013) develop a model that considers price changes, no previous paper has estimated the price effect of energy storage directly as in this paper, although price effects are integral components of almost all models of the storage market. The engineering literature considers their market impacts using simulation methods, focusing mainly on the private returns of storage investments and the battery properties that drive potential profits

[1] In a modeling exercise using PJM observed day-ahead prices, Salles et al. (2021) find the 95th percentile of nodes in PJM would generate up to \$42,000 per year purely from arbitrage. I use the higher range in (Penna et al., 2016) as the setting is in California and reported revenue includes ancillary services as well as arbitrage. Deman et al. (2025) also note the use of storage in reserve markets which is not considered here.

(Bradbury et al., 2014; Fioravanti et al., 2013; Nottrott et al., 2013; Walawalkar et al., 2007; Hittinger et al., 2012; Salles et al., 2021). Sioshansi et al. (2009) considers the effect of storage on grid prices, primarily to consider whether arbitrage opportunities are dissipated when storage is used. However, the authors estimate the effect of energy storage on grid prices assuming a constant linear relationship between load and prices, and do so by estimating price responses to changes in demand, not storage capacity, limiting the robustness of results to grid characteristics like congestion.

Second, it contributes to the machine learning literature. It introduces a novel empirical approach leveraging LASSO estimation to recover the market-relevant network characteristics of the electricity grid when they are unreported by regulators or utilities. In doing so, it makes an important contribution to any endeavor concerned about network effects and congestion in energy markets, and contributes to the literature leveraging machine learning algorithms to provide prediction or aid in identification in energy contexts (Mercadal, 2022; Cicala, 2022). This is useful in the context of spatial energy markets where properties of the LASSO have real-world analogs, solving a persistent problem in the study of wholesale electricity markets.

This paper proceeds as follows. Section 2 develops a spatially explicit model of electricity pricing and discusses the channel through which storage operates. Section 3 presents the data and empirical methods. Sections 4 and 5 present and discuss the results, respectively. Section 6 concludes.

2 Conceptual model

2.1 Electricity markets, congestion, and prices

Electricity prices are governed by two phenomena. The first is an upward-sloping, convex supply curve defined by marginal costs that increase in quantity supplied. In an unconstrained competitive market, an inverse supply curve is constructed by ordering electricity generation capacity from lowest to highest marginal cost. This "merit order"

defines a supply curve that becomes steeper as quantity increases because of generation technology and the limited frequency at which "peaker plants" are called up. This increasing steepness leads to a large price decrease when storage is discharged during a peak period relative to a small price increase when storage is charged during an off-peak period. The net effect is to lower average prices even though total quantity of electricity demand is unchanged.

The second phenomenon governing wholesale electricity prices is congestion on the grid caused by transmission constraints that lead to out-of-merit order dispatch. In a market with binding congestion constraints, dispatch occurs over potentially many local supply curves, each characterized by the inability to access lower-cost generators. In doing so, congestion amplifies the convexity of the supply curve, where nodes with more binding congestion constraints face a steeper supply curve. The introduction of even a small amount of storage capacity at one of these nodes can affect dispatch, and thus price, at other nodes. Although storage arbitrages over time, it has price effects in space.

In this section, I establish a model of electricity pricing based on Bohn et al. (1984), as employed in the California ISO, incorporating both phenomenon. The model draws a direct connection between the physical properties of electricity transmission that determine power flow and the constrained optimization that determines nodal prices. I illustrate the price determination process and the interplay of these constraints and storage with numeric examples in Appendix E, while Appendix F details a full model of pricing. Throughout, electricity demand is assumed to be exogenous because consumers face retail rates that may vary with time of day but do not vary instantaneously with wholesale prices.

2.2 Grid Operator's Problem

The grid operator solves the following constrained optimization for each hour of each day:

$$\min_{\boldsymbol{G}} \quad \sum_{n=1}^{N} p_n G_n(p_n) \tag{1}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} D_n = \sum_{n=1}^{N} G_n(p_n) \tag{2}$$

$$|\boldsymbol{\kappa} \times [\mathbf{G} - \mathbf{D}]\,| \leq \bar{K} \tag{3}$$

Equation 1 is the total cost of serving load to be minimized subject to constraints in 2-3. Equation 2 states that the total amount generated must equal the total amount demanded, summed across all nodes. In Equation 3, $\bar{K}$ is the vector of bidirectional capacities of the E lines that form the transmission and distribution network.[2] $\mathbf{G} - \mathbf{D}$ is the generation (net of same-node demand) for each node n, and $\boldsymbol{\kappa}$ is the [E x N] *shift factor matrix* that summarizes power flow over lines e between nodes n (see Appendix E for power flow model definitions, details, and examples).

The grid operator chooses only the vector of generation to dispatch, $\boldsymbol{G}$. A generator at node n submits bids $G_n(p_n)$ for each time period stating the quantity and price at which they are willing to supply electricity at node n, which maps price p_n to generation G_n. The grid operator determines $\boldsymbol{G}$ via the vector of prices $\boldsymbol{P}$ by numerically solving the Lagrangian where λ is the shadow value of relaxing the system-wide power flow constraint, and μ is a vector of length E that contains the shadow value of relaxing the transmission constraint for each line e:

$$\mathcal{L} = -\sum_{n=1}^{N} G_n(p_n)p_n + \overbrace{\lambda(0 - \sum_{n=1}^{N} G_n(p_n) - \sum_{n=1}^{N} D_n)}^{\text{System-wide power flow constraint}} - \underbrace{\boldsymbol{\mu}(\bar{K} - |\boldsymbol{\kappa} \times [\boldsymbol{G} - \boldsymbol{D}]\,|)}_{\text{line flow (slack) constraints}}, \tag{4}$$

I rely on the properties of the constrained optimization — specifically the shadow

[2] In power flow models, connections between nodes over which power flows are called *edges*. I use the more familiar term *line* but notate lines using e for clarity.

values of constraints that directly define the "locational marginal prices" or LMP's — to develop intuition about the effects of storage on prices. Specifically, $\boldsymbol{\mu}$ are shadow values of congestion on each network line that are represented by Karush-Kuhn-Tucker slack constraints. In the optimal solution, they take the value of 0 when not binding, but are non-zero when constraints are binding and enter the LMP scaled by the power flow in $\boldsymbol{\kappa}$. A formal model is provided in Appendix F.

2.3 Prices in an uncongested grid

When no transmission constraints are binding, the solution implies values of 0 for every μ_e, the shadow value of transmission constraints. Then, locational marginal price p_n^* is determined by λ and is equal to the bid price of the marginal generator in the uncongested merit order.[3] Let $p(Q) \geq 0$ be the marginal cost of energy at total quantity demanded Q from the merit order. I assume the marginal cost of energy is increasing at an increasing rate in Q, i.e.

$$\frac{dp}{dQ} \geq 0 \tag{5}$$

$$\frac{dp^2}{d^2Q} \geq 0 \tag{6}$$

Equation 6 is a consequence of merit-order dispatch and generator technologies. Peaking plants used to serve periods of high demand are less capital intensive, but are less efficient and thus more expensive per unit generated.[4]

Let E be the amount of energy a storage system transacts from the grid in an hour, where a transaction is either a discharge to the grid or a withdrawal from the grid. Let Q^L

[3]For expositional simplicity, I abstract away from the formation of λ as an envelope of generator bids combined with individual capacity constraints. See Bohn et al. (1984) for the construction of λ.

[4]Liski and Vehviläinen (2026) show that supply may have concave regions. This may occur where dispatchable generators and zero-cost renewables are on the margin and dispatchable generators face startup costs resulting in reservation prices. The authors find that 33% of hours in CAISO in 2015-2020 exhibit concave supply. The assumption in 6 may be relaxed to assume only that $\frac{dP}{dQ}$ is greater in peak periods than in nadir periods. The assumption is not an empirical restriction in estimation.

and Q^H be (nighttime) nadir and (daytime) peak exogenous demand, respectively, such that $Q^L < Q^H$. The withdrawal of E from the grid during the off-peak period increases Q^L to $Q^L + E$ ($\frac{dQ^L}{dE} > 0$). The new price paid for all units of energy is $p(Q^L + E)$. Similarly, Q^H is reduced by E when energy storage discharges during the peak period ($\frac{dQ^H}{dE} < 0$), and $p(Q^H - E) \leq p(Q^H)$.[5] Not only is $p(Q^H) \geq p(Q^L)$, creating a potentially profitable private opportunity for arbitrage, but also:

$$\frac{dp(Q^H)}{dQ} \geq \frac{dp(Q^L)}{dQ} \tag{7}$$

due to convexity of the supply curve. For strictly convex supply, storage lowers marginal costs at peak more than it raises them at low demand. Furthermore, because infra-marginal generators receive the same prices as those on the margin, the larger reduction applies to a larger number of peak-period units, while the smaller increase applies to a smaller number of nadir-period units. This yields an equation for the *total change in the cost of serving load* from the introduction of storage E:

$$\Delta\text{Cost of Serving Load} \approxeq Q^H \frac{dp(Q^H)}{dQ^H}\frac{dQ^H}{dE} - Q^L \frac{dp(Q^L)}{dQ^L}\frac{dQ^H}{dE} < 0 \tag{8}$$

Thus, storage is expected to reduce the total cost of serving load via a transfer from infra-marginal generators to LSEs. Savings to LSEs are passed onto retail electricity customers under cost-of-service regulation.[6] Net revenues for the battery owners are equal to the quantity stored multiplied by the difference between high and low prices, and notably do not include the net savings captured by the LSEs.

[5] In operation, all storage suffers round-trip losses, where E withdrawn for storage returns ϕE, $\phi < 1$. I abstract away from round-trip losses for the sake of notational clarity.

[6] The share passed through to consumers is determined by the regulatory procedure not modeled here. This area remains ripe for further research.

2.4 Congestion and KKT Slack Conditions

In a congested market, one or more constraints in Equation 3 bind. When a constraint binds, the only two options available to the grid operator are to either impose a local blackout or to increase generation at another node whose path does not include the congested line, even when this generation is higher cost. In effect, congestion eliminates access to lower-cost generators, increasing the convexity of the local inverse supply curve at congested nodes. As a solution to the KKT system of equations with slack conditions, some values of $\boldsymbol{\mu} \neq 0$, reflecting the shadow cost of these constraints, and congestion prices are composed of the product of the shadow value of the constraint and the flow implied by the shift factor matrix $\boldsymbol{\mu\kappa}$. Nodes that share lines would similarly share congestion prices according to the flow that each node would cause over e.[7]

This relationship between price differences on the grid and transmission constraints is critical to the empirical strategy of this paper. As storage discharges, changes in net load at a node will affect prices at other nodes based on shift factors $\boldsymbol{\kappa}$ and transmission constraints $\bar{K}$. Estimation of price effects requires a specification that (1) accounts for the unobserved, underlying dependence on the shift matrix, and (2) admits the inequality constraints on transmission which frequently equal zero during uncongested times. In brief, the data generating process is a weighted sum of constraints where the constraints are non-linear and frequently zero, and the weights are unobserved. This estimation problem is, at its heart, one of variable selection. When $\mu_e \neq 0$ for some line e, storage at one node will affect *some* other nodes, but not all. A natural tool, then, is the LASSO estimator. This estimation strategy is discussed further in Section 3.1.

The effect of storage on prices is decomposed by Equation 4 into changes in λ and changes to elements of $\boldsymbol{\mu}$. The size of any storage relative to $\sum_{n=1}^{N} G_n$ is small as the latter is the sum of all demand in CAISO territory, and batteries are of limited size. A change in storage is unlikely to change the marginal generator and effects on λ would, in practice, be nearly impossible to detect. In contrast, slack conditions defining elements

[7] See Appendix E for example calculations of congestion price.

of $\boldsymbol{\mu}$ can switch between non-zero and 0 with even small storage additions, as illustrated in Appendix E. In the empirical application, I focus on the effect of storage on $\boldsymbol{\mu}$, which defines the change in congestion price, while controlling for the empirical analog of λ, the system price.

2.5 Other Determinants of Price

In Equation 4, I abstract away from other factors that enter the price determination process. Dynamic constraints on grid operation also restrict the ability of low-cost generation to serve load over short time periods. For instance, fixed startup costs or ramping constraints may preclude a lower-cost plant from being dispatched for a short period of time in favor of a more expensive plant that can be dispatched quickly (Bohn et al., 1984; Hogan and Ring, 2003). Locational market power may also lead to increased convexity of the supply curve (Hogan, 1997; Borenstein et al., 2000; Reguant, 2014; Mercadal, 2022; Ryan, 2021; Jha and Leslie, 2025). Storage may relieve these constraints by providing short-term supply, smoothing demand and allowing the grid optimization to select lower-cost plants. In this manner, storage may affect prices well outside of the periods in which it discharges or is expected to discharge.

Price decreases owing to storage are estimated in this paper, but no attempt is made to apportion the shares of estimated price decreases amongst relief of dynamic constraints or locational market power. This matter is left for subsequent work, noting that the overall total price effect of storage is the important Figure in the context of this paper.

3 Data and Methods

I estimate price effects of a marginal unit of storage capacity separately for the node at which the battery unit is installed and for all other nodes at which prices are affected by the battery unit. Cross-node effects depend upon unobserved network links; own node effects are defined by an assumed link between storage and the most proximal node. The

own-node effect is estimated in a panel regression that pools all storage units in order to maximize statistical power. Cross-node effects are estimated separately via LASSO, sacrificing power for flexible modeling of the network. It would be computationally infeasible to instead estimate cross-node effects in a pooled regression with each storage node and its interactions entering as independent variables. Doing so would ignore the sparse nature of network congestion prices that result from the KKT slack conditions presented in Section 2. The method is closest to that of Bushnell and Novan (2021), who estimate the effect of solar generation on the average hourly wholesale electricity price in CAISO, which primarily consists of the value of the total generation constraint λ in Equation 2, abstracting from constraints $\boldsymbol{\mu}$. In contrast, in addition to studying storage rather than solar, I study prices at 757 pricing nodes separately and disentangle own- and cross-node effects.

3.1 Own-node price effects

The per-MW price effect of energy storage is estimated at the node level within the CAISO network using hourly day-ahead prices. These prices are regressed on node-specific, time-varying energy storage capacity, yielding an estimate of the associated effect of storage capacity on prices. Unique coefficients are estimated by hour of the day and by season, admitting distinct charge and discharge behavior. Because electricity consumers do not face prices that instantaneously vary with marginal cost, demand is assumed to be exogenous.

Price effects at the local node are estimated by:

$$\lambda_{nt} = \sum_{s=1}^{4}\sum_{h=1}^{24} \beta_{hs}^{1} ES_{nt} \times HOUR_h \times SEASON_s + \beta_{nh}^{2}\lambda_{l(n)t}^{LAP} + \theta_{n}^{1}\mathbf{w}_{nt}^{1} + \theta_{nh}^{2}\mathbf{w}_{t}^{2} +$$
$$\sum_{n=1}^{N}\sum_{y=2010}^{2017}\sum_{s=1}^{4}\sum_{h=1}^{24}\sum_{w=0}^{1} \delta_{nhswy} \times NODE_n \times YEAR_y \times SEASON_s \times HOUR_h \times WEEKDAY_w + \varepsilon_{nt}, \tag{9}$$

where λ_{nt} is hourly (total) marginal price at node n and time t;[8] ES_{nt} is contemporaneous MW of installed storage capacity at the node; $\lambda_{l(n)t}^{LAP}$ is a weighted average of prices at nodes within the corresponding utility territory, $l(n)$, known as a Load Aggregation Point (LAP); $\mathbf{w}_{nt}^{1}$ consists of measures of local (nodal) solar generation and local relative temperature, both of which vary by hour of sample and node n; $\mathbf{w}_{t}^{2}$ consists of citygate natural gas price, an indicator for the Aliso Canyon blowout,[9] total available nuclear capacity, and the 12-month rolling average precipitation, which may differentially affect nodal prices through hydropower supply, all of which vary by day.[10] δ_{nhswy} are fixed effects for each node n, hour of the day h, quarter (season) of the year s, year of the sample y, and weekday/weekend w respectively. $\beta_{nh}^{2}\lambda_{l(n)t}^{LAP}$ controls for the shadow value of the total generation constraint in Equation 2, which focuses identification on congestion effects. The term also controls for any trends that may occur during the season in which storage is commissioned. Interest centers on β_{hs}^{1}, the season and hour-specific coefficients on energy storage capacity that reflects the change in node price due to a one MW change in storage capacity.

Due to the saturation of fixed effects, identifying variation comes from node-level variation in storage capacity for a given node and hour within a specific 3-month interval

[8] I follow the energy literature and refer to λ_n as the nodal price. It is the sum of both the total generation constraint λ (common across all N nodes) in Equation 2 plus the sum of the transmission constraints that affect node n.

[9] The Southern California Gas-operated Aliso Canyon natural gas storage facility failed in Fall of 2015, limiting gas storage for Southern California. This supply chain interruption affects the marginal cost of all gas plants in the region, which in turn affects electricity prices in the region.

[10] θ_{n}^{1} represents node-specific relationships between node-hour varying controls $\mathbf{w}^{1}$. $\theta_{n}^{2}h$ represents node-by-hour-of-day relationships between time-varying controls $\mathbf{w}^{2}$ that are constant over nodes n.

(season). The identifying assumption of Equation 9 is that deviations from the node-hour-season-year average price, conditional on LAP prices, are uncorrelated with any omitted variables that correlate with the quantity of energy storage located at the node. Even node-specific trends, for instance, only confound if they introduce dynamics across three-month seasons during which storage is introduced.

This specification is akin to a difference-in-differences strategy to identify the effect of storage on own-node prices. I discuss the issue of selection as it pertains to the parallel trends assumption in section 3.5, below. Estimation of treatment effects with staggered treatment, as is the case here, can result in biased estimates when treatment effects are heterogeneous over time and adoption is staggered, resulting in the "forbidden comparison" (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Borusyak et al., 2024), which compares a newly-treated unit to an already-treated unit. Then, the estimator is a weighted average of treatment effects that can include negative weights on some estimates. However, negative weights do not occur when the number of observations of never-treated units in the period before the first treatment occurs exceeds the number of untreated observations for ever-treated units after the first period in which treatment occurs (see Proposition 4 of Borusyak et al. (2024)). In CAISO, there are 740 never-treated pricing nodes and 17 treated pricing nodes. The first storage commissioning occurs in late 2011 while the sample begins in January 2010, ensuring that this condition holds.[11]

Using this rich set of fixed effects limits identifying variation to a very fine window within the node-hour-season-year. With two exceptions, storage generally increases only once at each node. Therefore, coefficients that vary by season will only capture the effect within the hour-season-year of each increase in storage, while the fixed effects will absorb all variation in subsequent node-hour-season-years. While this improves internal validity, it comes at the cost of understanding the longer-run effects of storage.

[11] In the main estimation sample, there are 12,710,797 pre-treatment never-treated observations, far greater than the 1,156,306 ever-treated post-treatment observations.

3.2 Cross-Node Price Effects

Estimating the effect of energy storage on congestion pricing at other nodes requires knowledge about the characteristics of the network of generators and transmission lines. Network topology is not published by regulators, utilities, or grid operators because of grid security concerns. Hence, the network is unobservable to the econometrician. In order to recover market-relevant characteristics, I use a multi-step Double Pooled LASSO estimator that uncovers the structures of unobserved networks Manresa (2016).[12] If one takes any two nodes on the electric grid, then storage at one node will affect price at another if and only if (1) the two nodes share at least one non-zero shift matrix entry, and (2) the transmission constraint on that shared shift matrix entry has a non-zero shadow value (i.e. the constraint is binding) during some hour and season. Because the shift factor matrix and the transmission constraints of the electric grid are unobserved, and because the shadow values of transmission constraints (μ in Section 2) are, by definition, either zero or large, the LASSO estimator is an intuitive choice to uncover the relevant non-zero cross-node effects. In effect, the LASSO's "zeroing" property, resulting from the absolute value check function, captures the KKT slack conditions (non-zero when binding) that define shared constraints.

Cross-node effects are estimated by:

$$\lambda_{nt} = \sum_{s=1}^{4}\sum_{h=1}^{24} \beta^1_{nhs} ES_{nt} \times HR_h \times SEASON_s + \beta^2_{nh}\lambda^{LAP}_{l(n)t} + \theta^1_n \mathbf{w}^1_{nt} + \theta^2_{nh}\mathbf{w}^2_t + \sum_{i \neq n} \gamma_{nihs} ES_{it} \times HR_h \times SEASON_s + \delta_{nhswy} + \varepsilon_{nt}, \quad (10)$$

where the first four terms and δ_{nhswy} are as in 9. Parameters of interest in the cross-node equation are the non-zero entries in γ_{nihs} which represent cross-node price effects: the

[12] In Manresa (2016), the method is used to identify R&D linkages across technology firms to estimate spillover effects.

effect of 1 MW of storage at node i on prices at node n.[13] A LASSO algorithm is used to identify non-zero entries in γ_{nihs}.

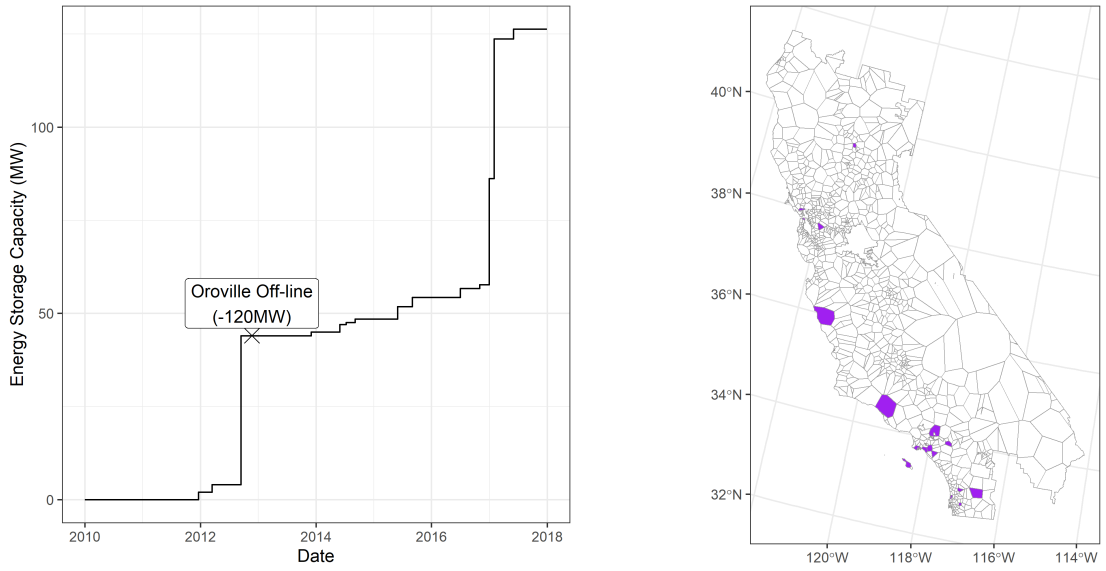
Estimating Equation 10 by LASSO presents an additional source of potential omitted-variables bias best characterized as *model selection omitted-variables bias* (Belloni et al., 2012, 2014). To obtain unbiased estimates of both the non-zero entries in γ_{nihs} and the magnitude of the cross-node price effect, I use the Double Pooled LASSO introduced in Manresa (2016), which leverages the "partialling out" method of Belloni et al. (2012) to address this bias. I discuss the issue further and detail the operationalization of the Double Pooled LASSO in Appendix G. In brief, the Double Pooled LASSO is estimated over three stages: the first stage generates values of λ_{nt} and w_{nt} that are orthogonal to ES_{it} in Equation 10. The second stage uses these versions of λ_{nt} and w_{nt} to estimate $\{\theta^1_n, \theta^2_{nh}\}$, then residualizes $\hat{\lambda}_{nt} = \lambda_{nt} - \hat{\theta}_n w_{nt}$. In the third stage, LASSO is used with Equation 10, replacing λ_{nt} with $\hat{\lambda}_{nt}$, resolving the potential for model selection omitted-variables bias. Finally, with non-zero entries in γ_{nihs} identified, OLS is used to estimate parameters. Estimation is done separately for each pricing node n, allowing separate linkages between pricing node n and storage node i to populate γ_{nihs}. I limit LASSO selection of storage nodes i to those within the same load aggregation area and are within 100km of pricing node i to limit selection of spurious linkages.

3.3 Data

Estimating the effects of energy storage on nodal prices requires three main datasets. The nodal prices themselves, the location and commissioning date of energy storage, and finally, data to control for potentially confounding factors that may affect nodal electricity prices, such as relative temperature, precipitation, which can constrain hydroelectric generation affecting nodal prices, and factors that influence the price of natural gas for generation. This section describes the nodal price and energy storage data sources

[13] I use n to index all pricing nodes and i to index storage nodes. All storage nodes are also pricing nodes.

Figure 1: (a) **Energy storage capacity added during study period**. Totals do not include storage introduced before January 2010. Note indicates date on which 120MW of Pumped Hydro was taken offline due to unexpected fire. (b) **Energy storage locations in CAISO territory**. Source: Dept. of Energy Global Energy Storage database.



(a) (b)

in detail. I construct measures of hourly nodal solar generation, hourly relative nodal temperatures, and a monthly 12-month rolling precipitation in Appendix B.

Commercial energy storage capacity in the CAISO area between 2010 and 2017 is observed in the U.S. Department of Energy's (DoE) Global Energy Storage (GS) Database (U.S. Department of Energy, 2016). During the study period, a total of 126.3 MW of energy storage was brought online,[14] equivalent to approximately .6% of the average peak daily load. The DoE GS database contains the commissioning date, storage size (MWh), discharge capacity (MW), owner, purpose, and geographic coordinates for every unit. I remove any unit whose purpose is related to emergency backup or other non-arbitrage applications, as these are unlikely to engage in regular discharges. I also exclude batteries 500kW of capacity and less. This yields 19 unique storage capacity additions to the CAISO service area as two storage sites receive multiple storage capacity additions. Figure 1 shows the timing and locations of storage installations in California.

[14] An additional 120MW of pumped storage at Oroville Dam that was built in 1969 was taken offline due to a fire during the study period. I include this change in the data.

I obtain Locational Marginal Prices (LMP's), λ_{nt}, for the day ahead market using the CAISO OASIS data portal for the period January 1, 2010 to December 31, 2017.[15] As shown in Figure 2, seasonal average LMPs exhibit expected patterns—daily peaks occur during the hours of 5-9pm, and some seasons also exhibit a smaller morning peak. These peaks show the potential for arbitrage. I also obtain each of the three Default Load Aggregation Point (LAP) prices, one for each major investor-owned LSE in California. Note that these prices differ from CAISO system price, as nodal prices include congestion whereas the CAISO system price does not. For exposition, I also calculate the deviation of each node from the LAP price for each observation. Prices and price deviations are shown in Figure 2. Appendix Figure A.1 shows spatial variation in nodal prices for select hours and select seasons.

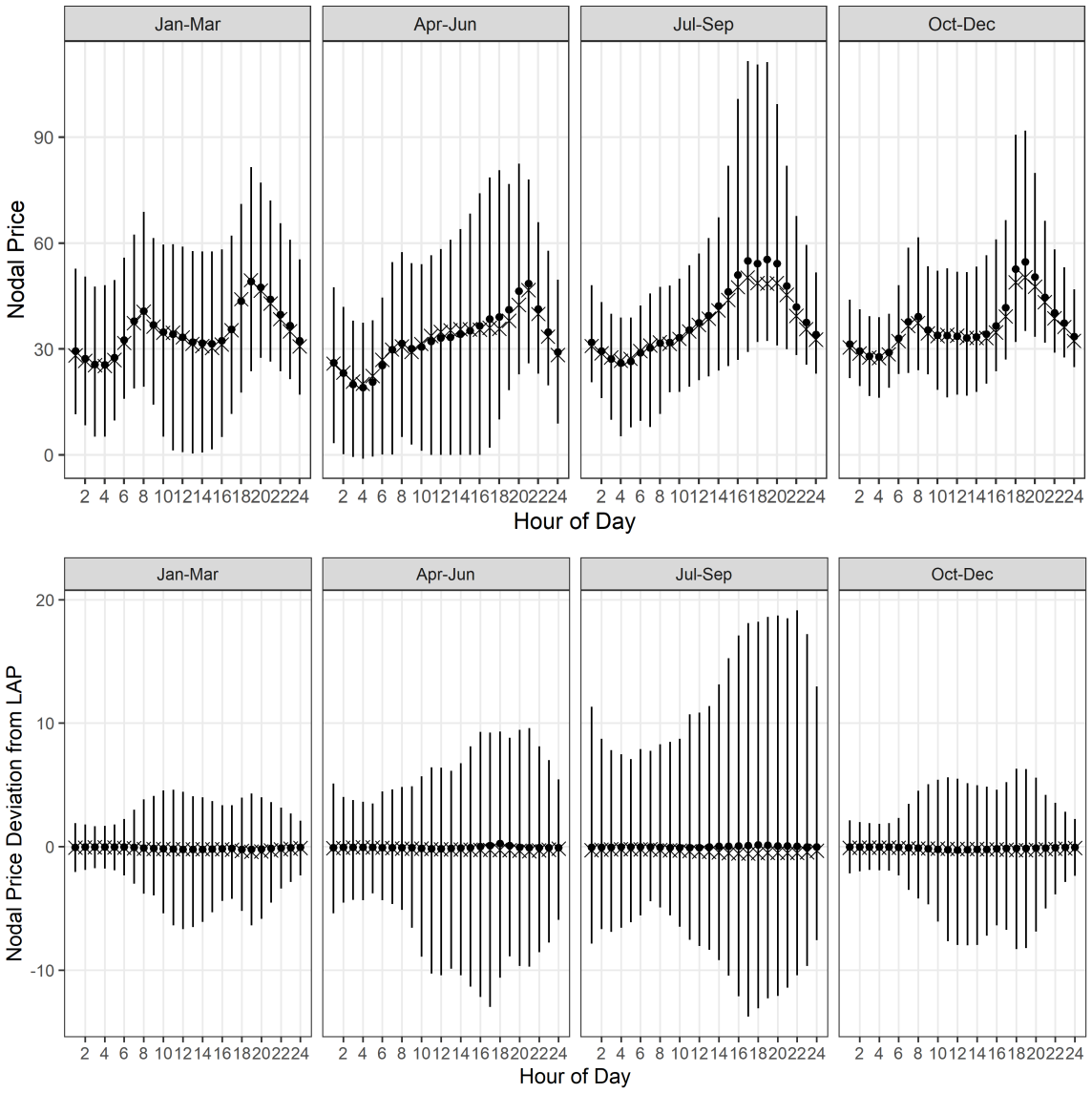
When necessary for calculating changes in cost of serving load, hourly energy withdrawals at each node are imputed as the product of hourly total load for each utility and respective, static Load Distribution Factors (LDFs) that report the share of utility load withdrawn at each node. Utility demand and LDFs are reported by CAISO.

I derive the physical location of each CAISO node from OASIS maps using a data scraping algorithm. The scraped location is not exact. Department of Homeland Security standards for information disclosure of "critical infrastructure" generally forbid disseminating information on electricity grid infrastructure. While reported node locations are not precise, they nevertheless appear to correspond to locations at which Google satellite imagery depicts large electricity substations. I assign each energy storage unit in the GS database to the nearest node and LAP. Error in the assignment may introduce bias through measurement error, biasing estimates towards zero. The magnitude of this attenuation is likely to be small, however, because nearby nodal prices are highly correlated. Nodes are clustered by name and proximity to form aggregate clusters of nodes and energy storage.[16] Proximity is determined by both physical distance and price path

[15] http://oasis.caiso.com. In 2024, CAISO restructured OASIS and removed the 2010-2017 data used here. I have archived the data as it was obtained in the online Appendix.

[16] Plants consisting of multiple generators will often have sequential naming e.g. ALAMT_1, ALAMT_2, etc., which are grouped here

Figure 2: **Hourly Nodal Prices and Deviations from LAP 2010-2017**. (Top) Points represent hour-season mean nodal price (λ_{nt}), $\times$ represents median nodal price, and vertical bars represent the 95% interval. (Bottom) Average, median, and 95% interval of nodal deviations from LAP price. Prices have positive skew in late afternoon hours during Apr-Jun and Jul-Sep. All prices are in dollars per-MW.



similarity. Clustering reduces the count of nodes substantially, from over 2,000 to 757. These 757 clusters are called "nodes" throughout this paper.

3.4 Summary Statistics

Table 1 shows summary statistics for the variables used in estimation. The weighted mean column shows the LDF-weighted mean for each variable.

Table 1: **Summary Statistics**. Statistics for $N = 757$ nodes in CAISO, hourly from 2010 through 2017. Weighted Mean is calculated by weighting each node by its load distribution factor (LDF), which represents the proportion of total load served at each point. (*) Includes Oroville pumped hydro. Weighted mean without Oroville is .35, unweighted mean is .06.

	N	Weighted Mean	Mean	Std. dev	Min	Max
LMP	51331382	36.30	35.85	17.91	−1240.47	4699.36
LAP	51331382	36.37	36.08	15.80	−71.16	806.26
LMP deviations	51331382	−0.07	−0.23	8.50	−1255.90	4640.90
Storage (MW)*	51331382	0.37	0.11	3.01	0.00	120.00
Solar Generation Index	51331382	10.97	12.07	108.19	0.00	8905.55
Relative Temperature	51331382	0.27	0.00	3.50	−24.26	18.81
Natural Gas Price (citygate)	51331382	3.99	3.99	1.06	1.79	23.58
1(Aliso Canyon blowout)	51331382	0.09	0.03	0.16	0.00	1.00
Nuclear Cap Available (GW)	51331382	2.51	2.51	0.95	0.00	4.39
Precipitation (12 mo)	51331382	21.49	21.52	6.67	7.28	35.65

3.5 Selection of Storage Sites

While an ideal experiment would place storage at randomly selected nodes and measure the change in prices with a causal interpretation, it is not feasible to allocate storage randomly. The DoE GS database consists entirely of batteries that are sited according to unobserved objectives — in the early stage of the battery storage industry studied here, arbitrage profits are not the sole objective maximized. Rather, storage owners

may also have considered the accessibility of the storage for monitoring, the cost and space constraints for adding large (semi-truck trailer sized) equipment at a node, or other considerations. The post-2017 boom in storage installed on the grid marked the beginning of the mature phase of energy storage in California, while pre-2018 storage marked a time when storage capacity installed was geared towards proof-of-scalability rather than towards optimal location. By limiting the study period to periods prior to the end of 2017, I leverage a set of energy storage installations that were sited according to constraints that are largely orthogonal to confounders. Of specific concern would be siting decisions that are correlated with unobserved, time-varying price trends.

The main concern is that storage operators chose sites that are correlated with either the outcome or time-varying trends in the outcome. The rich set of fixed effects in Equation 9 controls for unobserved, time-persistent differences in hourly nodal prices. However, time-varying unobservables that both (1) make a node more attractive to storage operators, and (2) influence the nodal price over time conditional on the fixed effects would still present an endogeneity issue. To the extent that the siting decision is correlated with prices, it is most likely that the price *spread*, which determines the potential for arbitrage, is the main concern. In particular, storage may have been sited at nodes with price spreads that were likely to increase through higher peak LMPs. If storage were sited at nodes that would have had higher peak prices and higher spreads absent storage, then the estimates of the price effect of storage are biased (upward) towards zero.

Tests for selection on observable intra-day price patterns and trends are shown in Appendix C. Intra-day price spreads prior to the study period are not sigificant correlates to selection as a storage site, suggesting that sites selected in the data are not sites with particularly large price spreads. Taking a dynamic approach, I test for trends in daily price spread prior to installation of storage and find no difference in the pre-installation price spread trend for storage sites relative to non-storage sites.

Both the selection of storage sites using price trends and the selection of storage sites using unobserved price response would result in an average treatment effect that is

upwardly biased toward zero. Therefore, I interpret the results as an upper-bound (or a lower-bound in magnitude) on the effect of storage.

4 Effects of Energy Storage on Nodal Prices

4.1 Own-node Effects

Point estimates of β^1_{hs} in Equation 9 and their 95% confidence intervals are plotted in Figure 3.[17] Driscoll-Kraay standard errors (Driscoll and Kraay, 1998) are two-way clustered at the node and hour-of-sample. Point estimates for effects in spring (April-June) are largely negative during peak afternoon and evening periods and are statistically significant. During the 6pm-10pm shoulder hours when prices are highest (see Figure 2), effects range from -$1.11 to -$0.76. These are precisely the hours in which declining solar generation coincides with increases in energy demand, forming the net demand "duck curve." Effects for all hours in spring are negative and nearly all significant while theory predicts weakly increasing prices during off-peak times. Lower prices may be a result of dynamics in the dispatch process where high-cost peak generation is avoided during peak times, allowing lower-cost generation to serve from on-peak into off-peak. Bushnell and Novan (2021) find that renewables shift generation in non-daylight "shoulder" hours away from low-cost dispatchable generation to higher marginal cost, more flexible generators. The estimated declines in non-daylight prices are consistent with storage facilitating the "integration of renewables" (the policy goal of AB2514) by enabling the use of lower marginal cost generators throughout the day and into the "shoulder" hours.

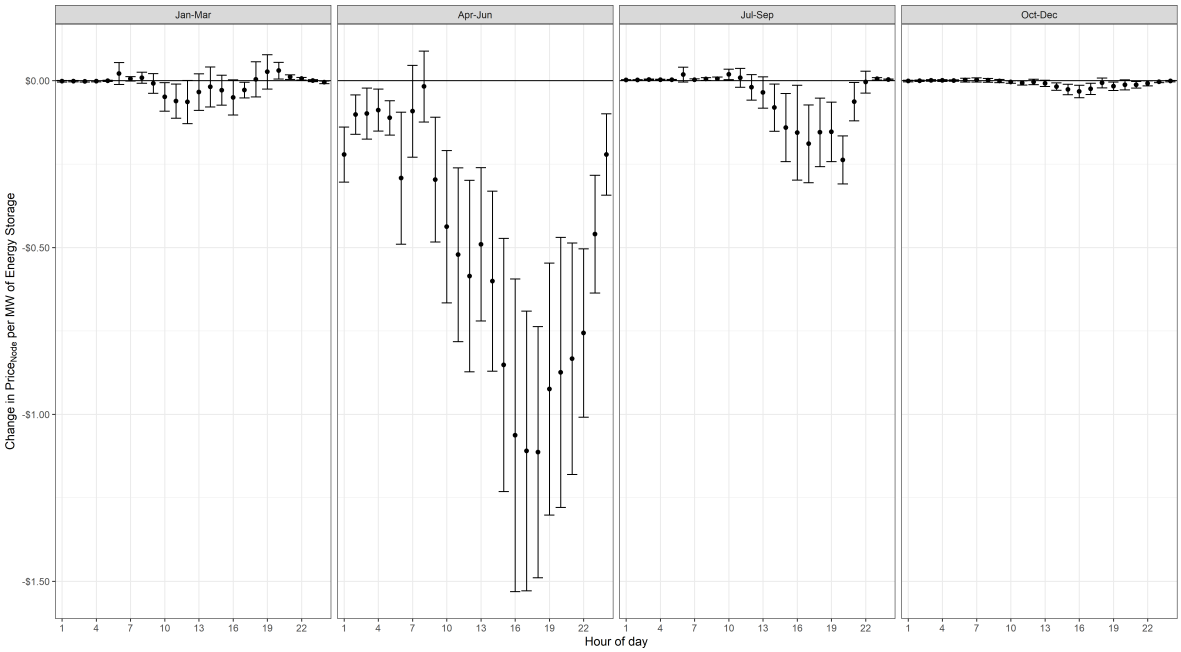
In summer (July-September), price effects are smaller, with a maximum price decline of -$0.24 for the 8pm hour, which nearly perfectly coincides with the highest-priced hour during summer (7pm) as shown in Figure 2. There is a clear negative and significant afternoon and evening trend in summer, and overnight prices are precisely estimated zeros. Results for summer follow the theory in Section 2 well. The relative magnitude of

[17] A full table of results is in Table A.1

results for July-September relative to April-June may be the result of increased demand during July-Septermber months. The power flow model indicates that prices decrease when storage alleviates a binding constraint. Doing so is less likely (or requires a larger amount of storage) when demand is higher, as in July-September.[18]

The winter January-March effects are small and mostly insignificant. In October-December, storage shows nearly no effect with the exception of very small afternoon negative and significant estimates.

Figure 3: **Coefficient Estimates**: Equation 9 with 95% CI, by Season and Hour. Includes node x season x hour x weekend FE. Driscoll-Kraay standard errors shown.



Lower hourly nodal prices benefit LSEs and, ultimately, ratepayers by reducing the cost of serving load. Reductions in price at a single node may lead to only small overall impacts. Results are robust to sample specification and reduced controls (see Appendix A). In Appendix A.1, I re-estimate Equation 9 but weight by node LDF in order to rule out biased estimates due to large capacity at low-demand nodes. We now turn to

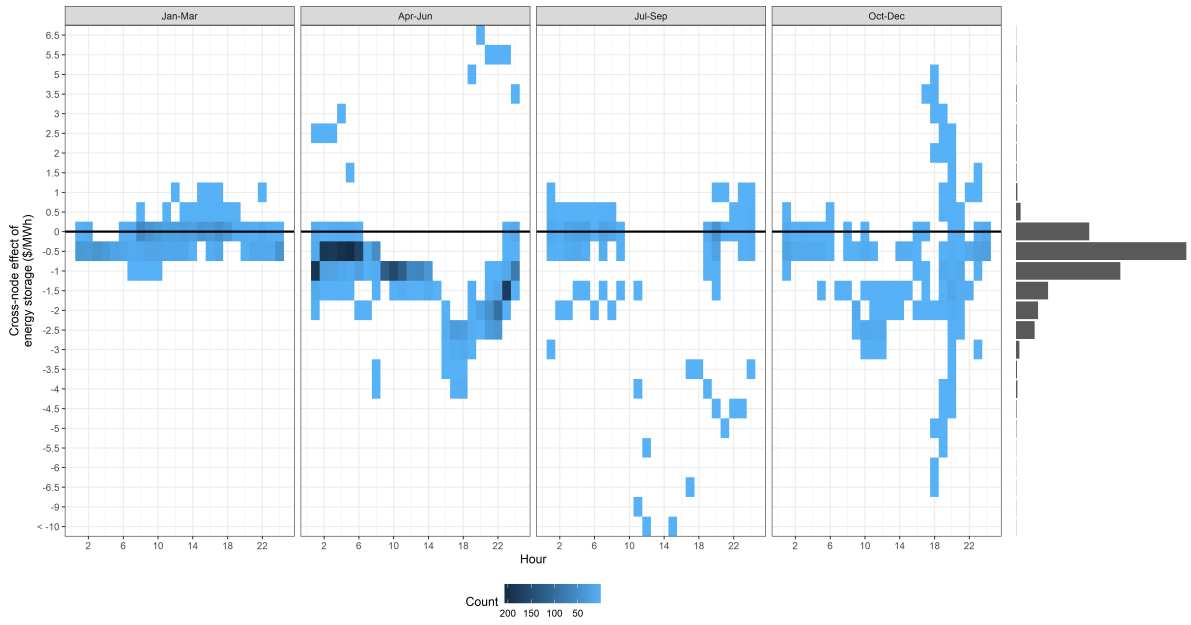
[18] In Appendix A.4, I estimate Equation 9 splitting the sample by high and low demand days, stratifying by season. Results are consistent with higher demand generating smaller price impacts, even within season.

cross-node effects.

4.2 Cross-node Effects

Results from estimating the cross-node equation are plotted in Figure 4. Each cell represents a count of non-zero coefficient from Γ_{hs}, a cross-node price response to a MW of storage at another node linked through the network identified by the LASSO estimator. Confidence intervals are not included for visual clarity, however, only statistically significant coefficients are included. Most estimates are significant at the 5% level, a result not surprising given the LASSO method used to select non-zero effects. In the final stage of estimation, each node is estimated separately. This sacrifices some precision that would be gained from pooling across nodes, but gains flexibility in estimating Γ_{hs}.

Figure 4: **Cross-node effects**. Each cell count represents a statistically significant (at 5%) entry in Γ_{hs}, and is the coefficient representing the change in one node's price at hour h, season s, resulting from a 1MW increase in storage at another node. Darker colors indicate greater mass of coefficients of that value. Aggregated counts for each price coefficient are shown in the histogram to the right.



While the overall mass of points is negative, a considerable number of positive points appear. This is an expected ramification of the nature of congestion in the network.

When congestion binds, the congested region has higher prices as it cannot access lower-cost generation. However, on the non-congested side, prices are lower as fewer areas can be served by generation in the uncongested area.[19] Reducing congestion allows prices to equilibrate, lowering them on the congested side but increasing them in the uncongested side, lowering the overall cost of serving load. Most coefficients show price effects smaller than the estimated own-node effect and the average cross-node effect (conditional on being non-zero) is \$-0.53/MW. If storage reduces locational market power, changes in generator bidding strategies in response may affect prices away from the storage node more than in the node itself.

The spring season (April-June) and the summer season (July-September), hypothesized to be the periods when congestion costs would be highest and thus storage would exhibit the greatest effect, contain the most non-zero coefficients. In both, a mass of negative coefficients are observed in the evening hours (6-10pm). Furthermore, a mass of coefficients are negative in April-June during the morning hours, when ramping for the morning peak would begin. The positive effects in some hours represent a smaller positive change (on a lower quantity) than the afternoon on-peak effects. This is consistent with the effects hypothesized in Section 2. Some negative effects with positive offsets are observed during evening hours in October-December. Note that Figure 2 shows a second daily peak occurring during the morning (8 - 10am) in October-December. This morning price peak occurs as households rise to prepare for their day during the colder, darker fall mornings.

Nodes designated by the LASSO estimation as linked during some hour and season ($\gamma_{nihs} \in \mathcal{L}_{LASSO}$) may not necessarily be spatially contiguous as transmission and distribution infrastructure is not a perfect lattice. However, linked nodes should be close. Select linked nodes are depicted in Figure 5. As these Figures show, linked nodes are generally nearby, lending validity to the LASSO selection procedure. However, distant nodes near the 100km limit are not entirely eliminated. Appendix Figure A.6 plots the

[19] In terms of the model in section 2, the shift factor matrix for some nodes may have negative entries, which translates to a lower price for energy withdrawn when that withdrawal relaxes the constraint.

density of all selected storage node effects (Γ_{rs}) by distance between pricing and storage nodes. Although some storage nodes lying far from the pricing node are selected, the larger mass of points selected lies within 75 km, and distant nodes tend to exhibit variance in sign, suggesting selection is subject to noise.

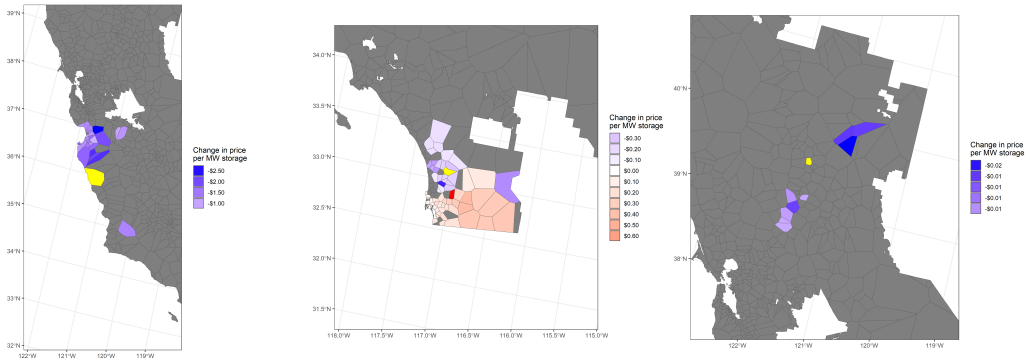


Figure 5: **Examples of cross-node effects**. Polygon in yellow represents site of storage node. (Left) effect of storage in Jolon, CA, hour ending at 8pm in Q4. (Middle) effect of storage in San Diego, CA, hour ending at 8pm in Q3. (Right) storage in Oroville, CA, hour ending 10am in Q4.

These cross-node results indicate that price effects of energy storage extend beyond the capacity node to other nodes in the network. Cross-node effects are heterogeneous across the 19 storage installations. In some areas, effects are concentrated in the area around the location of storage as in Figure 5 (left). In others, effects appear to be widespread across a region as in Figure 5 (middle and right). In some cases, the gradient of congestion relief can be seen in the results – Figure 5 (middle) shows that storage reduces prices to the north and east, but increases (by relieving a constraint) prices to the south. This heterogeneity highlights the importance of considering the local network topology and congestion patterns in calculating the reduction in the cost of serving load.

4.3 Changes in the Cost of Serving Load

To estimate the magnitude of the change in the total cost of serving load (ΔCSL) from a MW storage capacity, I multiply the own-node point estimates in Figure 3 and the cross-

node results in Figure 4 with the observed price at each node in each hour of sample, the total load withdrawn at that node, and the total amount of energy storage affecting price at the node:

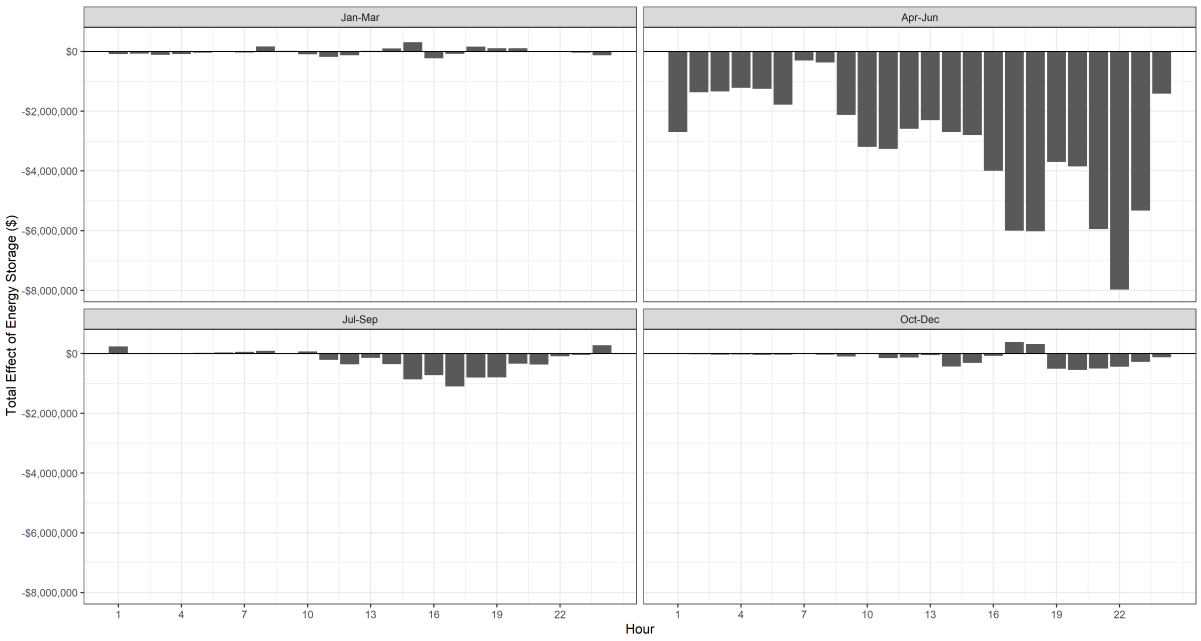
$$\Delta CSL = \sum_{t=1}^{T}\sum_{h=1}^{H}\sum_{n=1}^{N}\left(\beta_{hs}^{1} \times ES_{nt} + \sum_{i \neq n}\gamma_{nihs} \times ES_{it}\right) \times \left(LOAD_{l(n)t} \times LDF_{n}\right), \quad (11)$$

where β_{hs}^{1} is the price change estimated in Equation 9, γ_{nihs} is the cross-node effect in Equation 10, t is the hour-of-sample, $LOAD_{l(n)t}$ is the reported total load for the LAP that contains node n, and LDF_n is the load distribution factor. Calculations for total change in cost of serving load are described in Appendix F and defined in Equation 16.

The total estimated changes in the cost of serving load for the entire study period (2010-2017) is -$82.6 million (representing savings in the cost of serving load), of which -$37.7 million, or 45.6% of the total reduction, is from cross-node effects. Season-by-hour total changes (2010-2017) calculated using 11 are shown in Figure 6. Small positive amounts are shown in the first quarter and at a handful of other times, while the vast majority of savings occur in the second and third quarters, representing the spring and summer months – not only because point estimates are largest, but also because load is greatest during these hours and seasons.

Because each energy storage installation is introduced on a different day of the sample (see Figure 1a), I calculate an average total savings per hour, per megawatt of storage ("capacity-hour"). I divide the total savings from energy storage by the total number of MW of capacity-hours on the grid (5.82 million) to generate an average savings of $14.12 per MW of capacity per hour on grid. Over one year (8,760 hours), the sagings from 1MW of energy storage is $\$14.12 \times 8,760 = \$123,700$. For perspective, the most common behind-the-meter energy storage unit, has a capacity of 2kW, or .002MW, yielding savings from own-node changes of $247.40 per year.

Figure 6: **Aggregated own- and cross-node effects of all storage over study period**. Change in total cost of serving load over all estimate own- and cross-node effects. Totals reflect estimated cost of serving load savings starting from observed commissioning date in sample.



5 Discussion

Storage reduces the cost of serving load through own-node and cross-node effects. These savings represent transfers from inframarginal generators, and do not necessarily represent increases in social surplus. However, since the costs of storage are mandated by law, ratepayers ultimately pay for storage through rate regulation. Thus, it is important to consider the magnitude of the transfer on a distributional basis, asking "do savings recover in transfers an amount that justifies the cost?"

Savings in the cost of serving load are estimated to be $123,700 per MW per year. While the price of battery installations is falling rapidly, estimates from 2016 for the full installation and connection cost in California are around $6.5M per installed MW of storage capacity. Based purely on savings, the costs of mandated storage can only be justified under infeasibly long battery life expectancies. Environmentally-motivated storage mandates such as AB2514 can potentially still be justified by electricity market distortions (Schmalensee, 2022), learning spillovers, or first-mover advantages.

Standing on their own or in tandem, the total savings and private benefits from energy storage only partially justify the estimated cost of storage during the study period. This paper estimates only the savings in the cost of serving load, draws assumptions on private benefits and storage cost from published reports, and does not seek to quantify other tangible economic benefits. This is largely because costs of batteries are declining rapidly, with large-scale installations falling below $1,000,000 per MW in 2022, less than one-sixth of the reported 2016 price (Lambert, 2018; Viswanathan et al.).

5.1 Savings in the Cost of Serving Load Relative to Private Benefits

It is important to consider the economic incentives present in a low-cost storage world. If storage acts as any other generation asset, then investment in storage will be driven by the *private* marginal savings. Private investors will not consider the savings in the cost of serving load that cannot be captured. In fact, higher savings will *decrease* private revenues as the magnitude of the diurnal arbitrage opportunities (peak prices) are reduced. Therefore, a fiscal public goods problem arises. To this end, I calculate the savings as a fraction of the total private benefits from arbitrage. If savings are large relative to private benefits, then policy interventions that coordinate savings and private arbitrage profits may be justified.

An exact determination of private revenues from arbitrage requires information about daily battery charge and discharge, which is not publicly available data. However, private revenues for one pilot storage project in California are reported in Penna et al. (2016), which estimates private revenues between $49,500 and $208,500 per MW of storage per year. Hence, the storage savings accruing to ratepayers (and not captured by the operator) of $123,700 amount to 59.3-249.9% of these private revenues.

Savings in the cost of serving load represents a considerable portion of total overall benefits from storage operation, possibly even exceeding private benefits. These gains are

distributional, as they come as a transfer from incumbent inframarginal generators who would otherwise have received a higher market-clearing price. These reductions in the cost of serving load, not captured by private storage operators, justify policy on distributional grounds, but not economic grounds, if they are greater than the cost of storage. In short, if the transfer savings in the cost of serving load plus the private arbitrage profits do not exceed the cost of storage, then even a distributional policy preference does not justify a mandate.

The magnitude of savings relative to private benefits highlight the importance of rate regulation as well. Under pass-through utility regulation, savings in the cost of serving load are passed on to ratepayers. This paper does not examine the underlying assumption, but acknowledges that the intent of a policy and the implementation of that policy may differ dramatically. When storage is necessary to avoid or delay a local distribution line upgrade (i.e. to lower costs associated with local congestion costs), the utility-borne costs may be included in a ratepayer's billed distribution charge (Bierman, 2018). Under these conditions, it is important to understand the ratepayer savings, as the ratepayer is bearing the capital costs.

5.2 Relevance to Other Grids

California's experience with energy storage provides insight for other states and grid operators that may consider the role of storage in grid operations. The effects measured in this paper pertain to local nodal prices relative to a system-wide average LAP; therefore, grids with little congestion may be unlikely to see similar effects, even with a proportionally equivalent amount of storage. Historically, California has seen higher than average congestion costs, with the New York ISO (NYISO) having the highest historical average cost of congestion (Lesieutre and Eto, 2003). Comparison across grids is hampered by variation in how congestion is priced and measured. However, California's congestion costs are driven in part by internal, branch group (e.g. intra-distribution level) congestion, or congestion resulting from insufficient capacity on the smaller distribution grid,

rather than exclusively by constraints on larger high-voltage regional transmission lines (DOE, 2015). Results here are relevant to grids with similar qualities.

6 Conclusion

This research provides the first empirical estimates of the price effect of energy storage. Panel data estimates show a significant downward effect on nodal prices during peak hours and during peak summer months, suggesting that energy storage has a measurable effect on grid operations. Spring and summer afternoon peaks coincide with times of highest demand for expensive "peaking" generation, and are the costliest demand periods for which grid operators must plan. These price effects generate an average $123,700 in cost savings per year, per MW to load-serving entities that are typically passed on to consumers through rate regulation. The 126.3 MW of energy storage capacity in California is estimated to have reduced energy costs across the grid by $82.6M, which represents a transfer from inframarginal generators. However, even on a distributional basis ("do ratepayers receive a transfer equal to the cost of energy storage?"), rather than efficiency basis ("is overall welfare increased?"), this amount is not sufficient to fully justify mandates. Battery operators are not compensated for these cost savings to utilities and their rate payers. These savings are transfers from inframarginal generation, constitute a substantial fraction of the private benefits from energy price arbitrage, and point to a public interest in storage capacity expansion in the future.

References

Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012. ISSN 1468-0262.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2): 608–650, 2014. ISSN 0034-6527.

Evan M. Bierman. Direct Testimony of Evan M. Bierman On Behalf of San Diego Gas and Electric Company Before the Public Utilities Commission of the State of California, 2018.

LP Bloomberg Finance. Sustainable energy in america factbook. *The Business Council for Sustainable Energy. Smart thermostat facts on*, page 112, 2023.

Roger E Bohn, Michael C Caramanis, and Fred C Schweppe. Optimal pricing in electrical networks over space and time. *The Rand Journal of Economics*, pages 360–376, 1984. ISSN 0741-6261.

Severin Borenstein, James Bushnell, and Steven Stoft. The competitive effects of transmission capacity in a deregulated electricity industry. *RAND Journal of Economics*, 31(2):294–325, 2000.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024.

Kyle Bradbury, Lincoln Pratson, and Dalia Patiño-Echeverri. Economic viability of energy storage systems based on price arbitrage potential in real-time U.S. electricity markets. *Applied Energy*, 114:512–519, 2014. ISSN 03062619. doi: 10.1016/j.apenergy.2013.10.010. URL http://www.sciencedirect.com/science/article/pii/S0306261913008301.

James Bushnell and Kevin Novan. Setting with the sun: The impacts of renewable energy on conventional generation. *Journal of the Association of Environmental and Resource Economists*, 8(4):759–796, 2021.

R. Andrew Butters, Jackson Dorsey, and Gautam Gowrisankaran. Soaking up the Sun: Battery Investment, Renewable Energy, and Market Equilibrium. *Econometrica*, 93(3):891–927, 2025. ISSN 0012-9682. doi: 10.3982/ECTA20411.

California Solar Initiative. CSI Database, 2018. URL https://www.californiasolarstatistics.ca.gov/archived_working_data_files/.

Richard T. Carson and Kevin Novan. The private and social economics of bulk electricity storage. *Journal of Environmental Economics and Management*, 66(3):404–423, 11 2013. ISSN 0095-0696. doi: 10.1016/J.JEEM.2013.06.002. URL https://www.sciencedirect.com/science/article/pii/S0095069613000417.

Jiangzhuo Chen, Matthew Macauley, and Achla Marathe. Network topology and locational market power. *Computational Economics*, 34(1):21–35, 2009. ISSN 0927-7099.

Steve Cicala. Imperfect markets versus imperfect regulation in us electricity generation. *American Economic Review*, 112(2):409–41, February 2022. doi: 10.1257/aer.20172034. URL https://www.aeaweb.org/articles?id=10.1257/aer.20172034.

Michael T. Craig, Paulina Jaramillo, Bri-Mathias Hodge, Nathaniel J. Williams, and Edson Severnini. A retrospective analysis of the market price response to distributed photovoltaic generation in California. *Energy Policy*, 121:394–403, 10 2018. ISSN 0301-4215. doi: 10.1016/J.ENPOL.2018.05.061. URL https://www.sciencedirect.com/science/article/pii/S0301421518303781#.

Clément de Chaisemartin and Xavier D'Haultfœuille. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996, September 2020. ISSN 0002-8282. doi: 10.1257/aer.20181169.

Laureen Deman, Afzal S Siddiqui, Cédric Clastres, and Quentin Boucher. Day-ahead and reserve prices in a renewable-based power system: adapting electricity-market design for energy storage. *The Energy Journal*, page 01956574241309557, 2025.

U S DOE. National electric transmission congestion study, 2015.

John C Driscoll and Aart C Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, 80(4):549–560, 1998.

Harrison Fell and Daniel T. Kaffine. The Fall of Coal: Joint Impacts of Fuel Prices and Renewables on Generation and Emissions. *American Economic Journal: Economic Policy*, 10(2):90–116, May 2018. ISSN 1945-7731. doi: 10.1257/pol.20150321.

Harrison Fell, Daniel T. Kaffine, and Kevin Novan. Emissions, Transmission, and the Environmental Value of Renewable Energy. *American Economic Journal: Economic Policy*, 13(2): 241–272, May 2021. ISSN 1945-7731. doi: 10.1257/pol.20190258.

Richard Fioravanti, Michael Kleinber, Warren Katzenstein, Sudipta Lahiri, Nellie Ton, Alicia Abrams, Jessica Harrison, and Charlie Vartanian. Energy Storage Cost-effectiveness Methodology and Results. Technical report, DNV KEMA Energy and Sustainability, Oakland, CA, 2013.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, December 2021. ISSN 0304-4076. doi: 10.1016/j.jeconom.2021.03.014.

Gautam Gowrisankaran, Stanley S Reynolds, and Mario Samano. Intermittency and the value of renewable energy. *Journal of Political Economy*, 124(4):1187–1234, 2016.

Eric Hittinger, J F Whitacre, and Jay Apt. What properties of grid energy storage are most valuable? *Journal of Power Sources*, 206:436–449, 2012. ISSN 0378-7753.

William W Hogan. A market power model with strategic interaction in electricity networks. *The Energy Journal*, pages 107–141, 1997. ISSN 0195-6574.

William W Hogan and Brendan J Ring. On minimum-uplift pricing for electricity markets. *Electricity Policy Group*, pages 1–30, 2003.

Akshaya Jha and Gordon Leslie. Start-up costs and market power: Lessons from the renewable energy transition. *American Economic Review*, 115(2):690–726, 2025.

Paul L Joskow. Comparing the costs of intermittent and dispatchable electricity generating technologies. *American Economic Review*, 101(3):238–241, 2011.

Daniel T Kaffine, Brannin J McBee, and Jozef Lieskovsky. Emissions savings from wind power generation in texas. *The Energy Journal*, 34(1), 2013.

Ömer Karaduman. Economics of grid-scale energy storage in wholesale electricity markets. Technical report, 2021.

Frederic Lambert. Tesla's massive Powerpack battery in Australia cost $66 million and already made up to ˜$17 million. *Electrek*, 9 2018.

Jacob LaRiviere and Xueying Lyu. Transmission constraints, intermittent renewables and welfare. *Journal of Environmental Economics and Management*, 112:102618, March 2022. ISSN 0095-0696. doi: 10.1016/j.jeem.2022.102618.

Bernard C Lesieutre and Joseph H Eto. Electricity transmission congestion costs: A review of recent reports. 2003.

Matti Liski and Iivo Vehviläinen. Gone with the wind? an empirical analysis of the equilibrium impact of renewable energy. *Journal of the Association of Environmental and Resource Economists*, 7(5):873–900, 2020.

Matti Liski and Iivo Vehviläinen. Redistribution Through Efficiency: Theory and Evidence from Three Electricity Markets. *The RAND Journal of Economics*, 57(1):225–244, 2026. URL https://onlinelibrary.wiley.com/doi/10.1111/1756-2171.70048.

Elena Manresa. Estimating the structure of social interactions using panel data. *Unpublished Manuscript. MIT Sloan*, 2016.

Ignacia Mercadal. Dynamic competition and arbitrage in electricity markets: The role of financial players. *American Economic Journal: Microeconomics*, 14(3):665–99, 2022.

National Renewable Energy Laboratory. National Solar Radiation Database (NSRDB), 2018. URL https://nsrdb.nrel.gov/nsrdb-viewer.

A Nottrott, J Kleissl, and B Washom. Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems. *Renewable Energy*, 55: 230–240, 2013. ISSN 0960-1481.

Kevin Novan. Valuing the wind: renewable energy policies and air pollution avoided. *American Economic Journal: Economic Policy*, 7(3):291–326, 2015. ISSN 1945-7731.

Mike Della Penna, Manho Yeung, and David Fribush. Electric Program Investment Charge (EPIC) - Energy Storage for Market Operations. Technical report, Pacific Gas and Electric, Electric Asset Management Department, San Francisco, CA, 2016. URL https://www.pge.com/pge_global/common/pdfs/about-pge/environment/what-we-are-doing/electric-program-investment-charge/PGE-EPIC-Project-1.01.pdf.

Claire Petersen, Mar Reguant, and Lola Segura. Measuring the impact of wind power and intermittency. *Available at SSRN 4291672*, 2022.

Mar Reguant. Complementary bidding mechanisms and startup costs in electricity markets. *The Review of Economic Studies*, 81(4):1708–1742, 2014.

Mar Reguant. The efficiency and sectoral distributional impacts of large-scale renewable energy policies. *Journal of the Association of Environmental and Resource Economists*, 6(S1):S129–S168, 2019.

Nicholas Ryan. The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market. *American Economic Journal: Microeconomics*, 13(2):202–242, May 2021. ISSN 1945-7669, 1945-7685. doi: 10.1257/mic.20180191.

Maurício BC Salles, Taina N Gadotti, Michael J Aziz, and William W Hogan. Potential revenue and breakeven of energy storage systems in pjm energy markets. *Environmental Science and Pollution Research*, 28:12357–12368, 2021.

Richard Schmalensee. Competitive energy storage and the duck curve. *The Energy Journal*, 43 (2):1–16, 2022.

Ramteen Sioshansi. Increasing the value of wind with energy storage. *The Energy Journal*, 32 (2):1–30, 2011.

Ramteen Sioshansi, Paul Denholm, Thomas Jenkin, and Jurgen Weiss. Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects. *Energy economics*, 31(2): 269–277, 2009. ISSN 0140-9883.

U.S. Department of Energy. Global Energy Storage Database, 2016. URL http://www.energystorageexchange.org/projects.

Vilayanur Viswanathan, Kendall Mongird, Ryan Franks, Xiaolin Li, and Vincent Sprenkle. 2022 Grid Energy Storage Technology Cost and Performance Assessment.

Rahul Walawalkar, Jay Apt, and Rick Mancini. Economics of electric energy storage for energy arbitrage and regulation in New York. *Energy Policy*, 35(4):2558–2568, 2007. ISSN 0301-4215.

Rahul Walawalkar, Seth Blumsack, Jay Apt, and Stephen Fernands. An economic welfare analysis of demand response in the PJM electricity market. *Energy Policy*, 36(10):3692–3702, 2008. ISSN 0301-4215.

Frank A Wolak. Level versus Variability Trade-offs in Wind and Solar Generation Investments: The Case of California. Technical report, 2016.

Chi-Keung Woo, Ira Horowitz, Jack Moore, and Andres Pacheco. The impact of wind generation on the electricity spot-market price level and variance: The Texas experience. *Energy Policy*, 39(7):3939–3944, 2011. ISSN 0301-4215.

Acknowledgements

I thank Lori Bennear, Steve Sexton, Chris Timmins, Bryan Bollinger, Frank Wolak, Marty Smith, Michael Caramanis, Jackson Dorsey, Catie Hausman, Soren Anderson, participants at 2018 CEnREP Camp Resources and the 2022 AEA annual meetings for comments and suggestions, as well as Ray Hohenstein at AES for providing context.

Declaration of Conflicting Interests

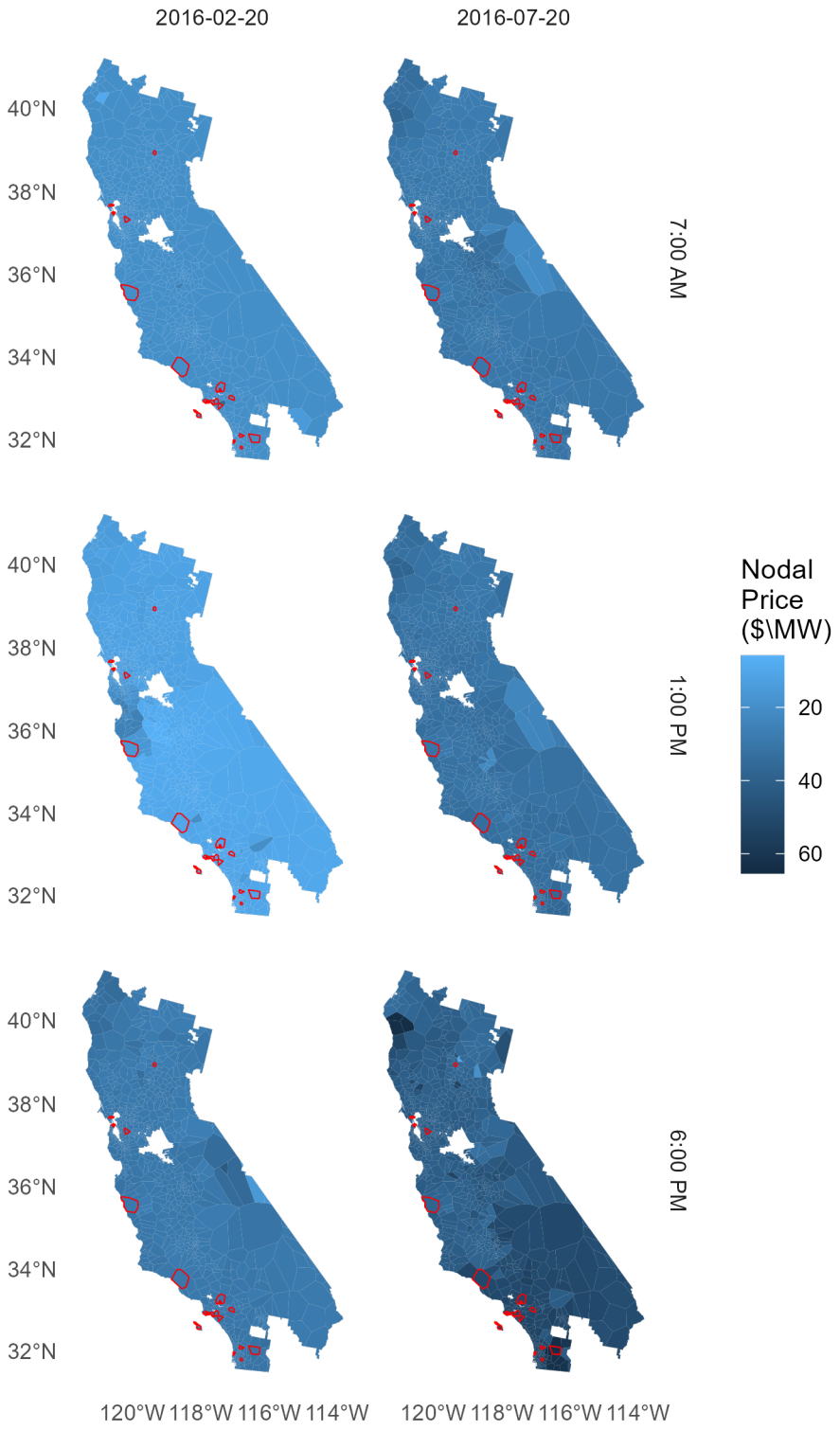
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) acknowledge financial support from the Alfred P. Sloan Foundation Predoctoral Fellowship on Energy Economics, awarded through the NBER.

Appendix A Additional Figures and Tables

Figure A.1: **Example of spatio-temporal variation in hourly nodal price ($/MW),** λ_n. Note the "hot-spot" of congestion in the Northwest portion of the state which is greatest during the lower-right panel, 6pm on July 20, 2016. Red outlined polygons indicate storage nodes.



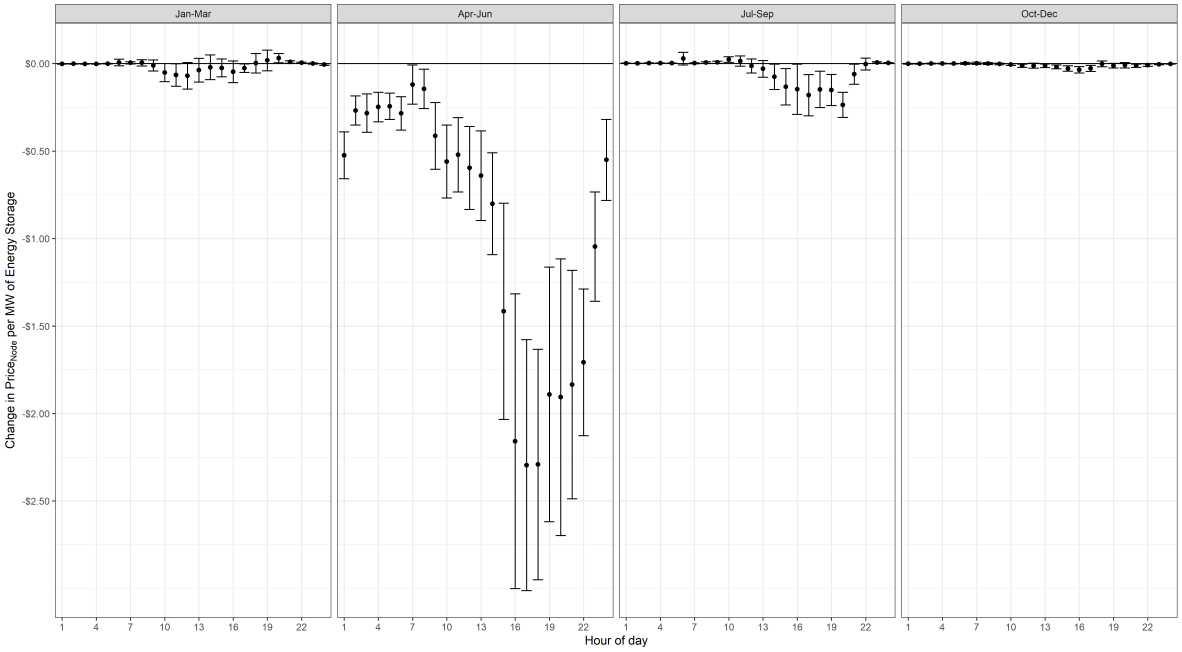
A.1 Main Results Weighting by LDF

In estimating the own-node effect, the capacity of storage relative to load may be important in understanding the effect of storage. Of particular concern may be a case where storage is located at a low-demand node, and thus has an outsize impact on that node's price. The LMP determination process illustrated in Appendix E details how prices at the node of installation are intertwined with (potentially higher-demand) connected nodes, and the estimation of cross-node effects accounts for changes in prices at other nodes that may be of higher or lower demand.

As an alternative that illustrates the relationship between storage capacity and node demand would be to estimate the effect of storage controlling for (or relative to) node-level demand. Hourly node-level demand is not available from CAISO. Instead, I re-estimate Equation 9 but weight each node by it's Load Distribution Factor (LDF), which is a static measure of overall relative load share. Results therefore show the effect of storage at higher-demand withdrawal nodes. If the main estimates of the effect of storage are influenced by the location of storage at low-LDF nodes (and thus have very high capacity-to-demand ratios), then results of this weighted estimation will be significantly different.

This is not the case. Figure A.2 shows that results with weighting are nearly identical to the main results in qualitative form. The effects of storage are *more negative* during the evening "shoulder" hours in April-June, double the magnitude as the main results. July-September results are slightly attenuated relative to the main results. These results are consistent with the model of LMPs in which the storage capacity at one node affects all nodes connected by a shared capacity constraint, and the magnitude of the effect is determined by the capacity relative to the constraint, rather than capacity relative to the node at which it is installed. This highlights the importance of estimating the cross-node effects in order to form a full picture of the price effect of energy storage.

Figure A.2: **Estimation of main results weighted by node LDF**. Results show estimation of Equation 9 weighted by the node's LDF. This places emphasis on the capacity of storage relative to the node's size, emphasizing storage nodes with higher demand.



A.2 Main Results Using Only Treated Quarters

As an alternative specification to the results in Figure 3, I estimate Equation 9 using only those quarter-nodes in which storage is commissioned. Since all identification is within-quarter, this provides an alternative estimation that does not use prices a non-storage nodes, and does not use prices during non-storage times.

Figure A.3 shows that results are qualitatively similar to the main results, though with some differences. In these results, price effects are somewhat attenuated, especially in Q2 in the evening, where the main results showed the largest effect. In the hours following the largest effects, estimates are sometimes positive, potentially indicating a time-shift for the price peak (e.g. later in the evening) rather than a smoothing of the price. Both Q2 and Q3 still exhibit late-afternoon and early-evening price decreases, consistent with theory.

Figure A.3: **Coefficient Estimates**: Equation 9 with 95% CI, by Season and Hour, estimated using (gray) full sample (main results), (yellow) only quarters in which one node has storage introduced, and (blue) only those nodes and quarters in which storage is introduced. Includes node x season x hour x weekend FE. Driscoll-Kraay standard errors shown.

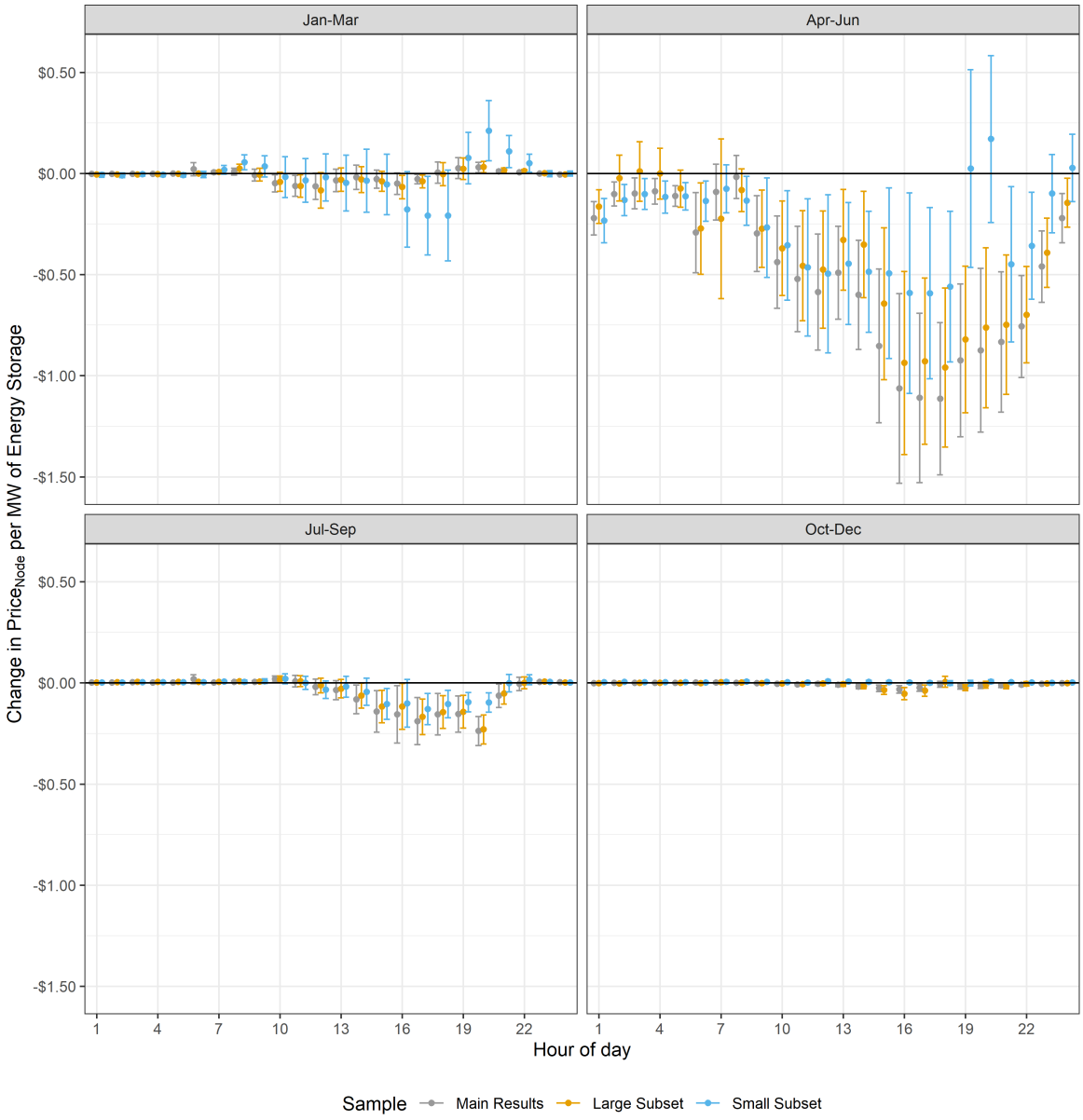


Table A.1: **Coefficient Estimates**: Equation 9 with 95% CI, by Season and Hour. Includes node x season x hour x weekend FE. Column (1) is the preferred specification, (2) uses only quarters in which storage is commissioned, and (3) uses only quarters and nodes in which storage is commissioned. Driscoll-Kraay standard errors shown.

Model:	(1)	(2)	(3)
Variables			
Storage × Hour 1 × Season Q1	-0.002 (0.001)	-0.005** (0.002)	-0.007 (0.006)
Storage × Hour 2 × Season Q1	-0.001 (0.001)	-0.005*** (0.002)	-0.009 (0.006)
Storage × Hour 3 × Season Q1	-0.002* (0.0010)	-0.005*** (0.002)	-0.004 (0.004)
Storage × Hour 4 × Season Q1	-0.002* (0.0009)	-0.004*** (0.001)	-0.006 (0.005)
Storage × Hour 5 × Season Q1	-0.0003 (0.001)	-0.002 (0.001)	-0.007 (0.005)
Storage × Hour 6 × Season Q1	0.02 (0.02)	-7.4×10^{-5} (0.002)	-0.005 (0.008)
Storage × Hour 7 × Season Q1	0.006* (0.003)	0.008** (0.004)	0.02 (0.01)
Storage × Hour 8 × Season Q1	0.009 (0.008)	0.02** (0.01)	0.06*** (0.02)
Storage × Hour 9 × Season Q1	-0.008 (0.01)	-0.006 (0.02)	0.04 (0.03)
Storage × Hour 10 × Season Q1	-0.05** (0.02)	-0.04* (0.02)	-0.02 (0.05)
Storage × Hour 11 × Season Q1	-0.06** (0.03)	-0.06** (0.03)	-0.03 (0.06)
Storage × Hour 12 × Season Q1	-0.06* (0.03)	-0.08* (0.04)	-0.02 (0.06)
Storage × Hour 13 × Season Q1	-0.03 (0.03)	-0.03 (0.03)	-0.05 (0.07)
Storage × Hour 14 × Season Q1	-0.02 (0.03)	-0.03 (0.03)	-0.04 (0.08)
Storage × Hour 15 × Season Q1	-0.03 (0.02)	-0.04 (0.02)	-0.05 (0.08)
Storage × Hour 16 × Season Q1	-0.05* (0.03)	-0.07** (0.03)	-0.18* (0.10)
Storage × Hour 17 × Season Q1	-0.03** (0.01)	-0.04** (0.02)	-0.21** (0.10)
Storage × Hour 18 × Season Q1	0.004 (0.03)	-0.003 (0.03)	-0.21* (0.11)
Storage × Hour 19 × Season Q1	0.03 (0.03)	0.02 (0.03)	0.08 (0.06)
Storage × Hour 20 × Season Q1	0.03** (0.01)	0.03** (0.01)	0.21*** (0.08)
Storage × Hour 21 × Season Q1	0.01*** (0.004)	0.02*** (0.006)	0.11*** (0.04)
Storage × Hour 22 × Season Q1	0.006*** (0.002)	0.01*** (0.004)	0.05** (0.02)
Storage × Hour 23 × Season Q1	0.0003 (0.002)	0.001 (0.002)	-0.0003 (0.007)
Storage × Hour 24 × Season Q1	-0.005** (0.002)	-0.005** (0.002)	-0.0001 (0.006)
Storage × Hour 1 × Season Q2	-0.22*** (0.04)	-0.16*** (0.04)	-0.23*** (0.06)
Storage × Hour 2 × Season Q2	-0.10*** (0.03)	-0.02 (0.06)	-0.13*** (0.04)
Storage × Hour 3 × Season Q2	-0.10** (0.04)	0.01 (0.07)	-0.10*** (0.04)
Storage × Hour 4 × Season Q2	-0.09*** (0.03)	-0.0007 (0.06)	-0.12*** (0.04)
Storage × Hour 5 × Season Q2	-0.11*** (0.03)	-0.07 (0.05)	-0.11*** (0.04)
Storage × Hour 6 × Season Q2	-0.29*** (0.10)	-0.27** (0.12)	-0.14*** (0.05)
Storage × Hour 7 × Season Q2	-0.09 (0.07)	-0.22 (0.20)	-0.07 (0.06)
Storage × Hour 8 × Season Q2	-0.02 (0.05)	-0.08 (0.05)	-0.13** (0.06)
Storage × Hour 9 × Season Q2	-0.30*** (0.10)	-0.27*** (0.10)	-0.27** (0.13)
Storage × Hour 10 × Season Q2	-0.44*** (0.12)	-0.37*** (0.12)	-0.35** (0.14)
Storage × Hour 11 × Season Q2	-0.52*** (0.13)	-0.46*** (0.14)	-0.46*** (0.17)
Storage × Hour 12 × Season Q2	-0.59*** (0.15)	-0.48*** (0.15)	-0.50** (0.20)
Storage × Hour 13 × Season Q2	-0.49*** (0.12)	-0.33*** (0.13)	-0.45*** (0.15)
Storage × Hour 14 × Season Q2	-0.60*** (0.14)	-0.35*** (0.13)	-0.49*** (0.15)
Storage × Hour 15 × Season Q2	-0.85*** (0.19)	-0.64*** (0.19)	-0.49** (0.22)
Storage × Hour 16 × Season Q2	-1.1*** (0.24)	-0.94*** (0.23)	-0.59** (0.25)
Storage × Hour 17 × Season Q2	-1.1*** (0.21)	-0.93*** (0.21)	-0.59*** (0.22)
Storage × Hour 18 × Season Q2	-1.1*** (0.19)	-0.96*** (0.20)	-0.56*** (0.19)
Storage × Hour 19 × Season Q2	-0.92*** (0.19)	-0.82*** (0.18)	0.02 (0.25)
Storage × Hour 20 × Season Q2	-0.87*** (0.21)	-0.76*** (0.20)	0.17 (0.21)
Storage × Hour 21 × Season Q2	-0.83*** (0.18)	-0.75*** (0.18)	-0.45** (0.20)
Storage × Hour 22 × Season Q2	-0.76*** (0.13)	-0.70*** (0.12)	-0.36*** (0.13)
Storage × Hour 23 × Season Q2	-0.46*** (0.09)	-0.39*** (0.09)	-0.10 (0.10)
Storage × Hour 24 × Season Q2	-0.22*** (0.06)	-0.14** (0.06)	0.03 (0.08)
Storage × Hour 1 × Season Q3	0.002*** (0.0007)	0.003*** (0.001)	0.002 (0.002)
Storage × Hour 2 × Season Q3	0.002*** (0.0007)	0.004*** (0.001)	0.003 (0.002)
Storage × Hour 3 × Season Q3	0.003*** (0.0007)	0.006*** (0.001)	0.003 (0.003)
Storage × Hour 4 × Season Q3	0.003*** (0.0007)	0.005*** (0.001)	0.003 (0.002)
Storage × Hour 5 × Season Q3	0.003*** (0.0007)	0.005*** (0.001)	0.003 (0.003)
Storage × Hour 6 × Season Q3	0.02 (0.01)	0.006*** (0.002)	0.004 (0.003)
Storage × Hour 7 × Season Q3	0.003** (0.001)	0.006*** (0.002)	0.006* (0.004)
Storage × Hour 8 × Season Q3	0.005** (0.002)	0.008*** (0.003)	0.006 (0.005)
Storage × Hour 9 × Season Q3	0.006** (0.003)	0.007** (0.003)	0.009 (0.006)
Storage × Hour 10 × Season Q3	0.02** (0.008)	0.02** (0.008)	0.02 (0.01)
Storage × Hour 11 × Season Q3	0.009 (0.01)	0.009 (0.01)	0.0004 (0.02)
Storage × Hour 12 × Season Q3	-0.02 (0.02)	-0.01 (0.02)	-0.03 (0.02)
Storage × Hour 13 × Season Q3	-0.04 (0.02)	-0.03 (0.02)	-0.02 (0.03)
Storage × Hour 14 × Season Q3	-0.08** (0.04)	-0.06** (0.03)	-0.04 (0.03)
Storage × Hour 15 × Season Q3	-0.14*** (0.05)	-0.12*** (0.04)	-0.10*** (0.04)
Storage × Hour 16 × Season Q3	-0.16** (0.07)	-0.12** (0.06)	-0.10* (0.06)
Storage × Hour 17 × Season Q3	-0.19*** (0.06)	-0.17*** (0.04)	-0.13*** (0.04)
Storage × Hour 18 × Season Q3	-0.15*** (0.05)	-0.14*** (0.04)	-0.10*** (0.04)
Storage × Hour 19 × Season Q3	-0.15*** (0.04)	-0.14*** (0.04)	-0.10*** (0.02)
Storage × Hour 20 × Season Q3	-0.24*** (0.04)	-0.23*** (0.04)	-0.10*** (0.02)
Storage × Hour 21 × Season Q3	-0.06** (0.03)	-0.05* (0.03)	-0.0005 (0.02)
Storage × Hour 22 × Season Q3	-0.004 (0.02)	0.0004 (0.01)	0.02 (0.01)
Storage × Hour 23 × Season Q3	0.006*** (0.002)	0.006*** (0.002)	0.005* (0.003)
Storage × Hour 24 × Season Q3	0.003*** (0.0009)	0.003* (0.002)	0.002 (0.002)
Storage × Hour 1 × Season Q4	-0.0006 (0.001)	-0.001 (0.002)	0.003*** (0.001)
Storage × Hour 2 × Season Q4	-5.8×10^{-5} (0.001)	-0.002 (0.002)	0.005*** (0.001)
Storage × Hour 3 × Season Q4	0.0008 (0.001)	1.9×10^{-5} (0.002)	0.004*** (0.001)
Storage × Hour 4 × Season Q4	0.0008 (0.001)	0.0004 (0.002)	0.004*** (0.001)
Storage × Hour 5 × Season Q4	0.0005 (0.001)	-6.2×10^{-5} (0.002)	0.004*** (0.001)
Storage × Hour 6 × Season Q4	0.002 (0.003)	-0.0005 (0.003)	0.002 (0.002)
Storage × Hour 7 × Season Q4	0.002 (0.003)	0.003 (0.004)	0.006* (0.003)
Storage × Hour 8 × Season Q4	0.001 (0.003)	0.003 (0.003)	0.007* (0.004)
Storage × Hour 9 × Season Q4	-0.0007 (0.002)	-0.0003 (0.003)	0.006* (0.003)
Storage × Hour 10 × Season Q4	-0.004** (0.002)	-0.003 (0.002)	0.004 (0.003)
Storage × Hour 11 × Season Q4	-0.007** (0.003)	-0.005 (0.004)	0.002 (0.003)
Storage × Hour 12 × Season Q4	-0.004 (0.004)	-0.003 (0.004)	0.008* (0.004)
Storage × Hour 13 × Season Q4	-0.008* (0.005)	-0.006 (0.005)	0.007** (0.004)
Storage × Hour 14 × Season Q4	-0.02*** (0.006)	-0.02*** (0.006)	0.005 (0.004)
Storage × Hour 15 × Season Q4	-0.03*** (0.008)	-0.04*** (0.01)	0.004 (0.004)
Storage × Hour 16 × Season Q4	-0.03*** (0.009)	-0.05*** (0.01)	0.002 (0.004)
Storage × Hour 17 × Season Q4	-0.02*** (0.009)	-0.04*** (0.01)	0.0002 (0.004)
Storage × Hour 18 × Season Q4	-0.006 (0.007)	0.006 (0.01)	-0.0006 (0.006)
Storage × Hour 19 × Season Q4	-0.02** (0.006)	-0.02*** (0.008)	-0.0006 (0.007)
Storage × Hour 20 × Season Q4	-0.01 (0.008)	-0.008 (0.009)	0.007 (0.005)
Storage × Hour 21 × Season Q4	-0.01** (0.005)	-0.02** (0.006)	0.003 (0.004)
Storage × Hour 22 × Season Q4	-0.008** (0.004)	-0.004 (0.006)	0.003* (0.002)
Storage × Hour 23 × Season Q4	-0.003** (0.001)	-0.002 (0.002)	0.001 (0.001)
Storage × Hour 24 × Season Q4	-0.001 (0.0010)	-0.001 (0.002)	0.002** (0.001)
Fixed-effects			
Hour-Node-season-Year-weekend	Yes	Yes	Yes
Varying Slopes			
Solar gen index (Node)	Yes	Yes	Yes
Relative temp (Node)	Yes	Yes	Yes
Aliso Canyon blowout (Hour x Node)	Yes	Yes	Yes
12 month rolling pcpt (Hour x Node)	Yes	Yes	Yes
Distribution Load Agg. Price (Hour x Node)	Yes	Yes	Yes
CAISO natural gas price (Hour x Node)	Yes	Yes	Yes
Total nuclear avail (Hour x Node)	Yes	Yes	Yes
Fit statistics			
Observations	51,331,382	20,726,325	38,339
R^2	0.81706	0.71812	0.96826

A.3 Stepwise Addition of Covariates

Equation 9 includes flexible controls for each node or node-by-hour for the controls described in the main text. Here I show the coefficient estimates in a stepwise addition of fixed effects and controls. Columns (1)-(3) show the addition of fixed effects. Columns (4)-(10) show the addition of controls. (4) adds node-level solar generation, (5) adds node-level hourly relative temperature, (6) adds the Distributed Load Aggregation Point (DLAP) price, (7) adds the 12 month rolling precipitation measure, (8) adds the citygate natural gas price, (9) adds the indicator variable for the Aliso Canyon blowout, and (10) adds the total nuclear capacity available.

Large changes in coefficient estimates occur first when adding season x year x hour x node x weekend fixed effects (column (3)) and in Column (6), when I add the DLAP price, identifying the effect of storage on it's own node from deviations in the DLAP price. The large change illustrates the role of DLAP price in controlling for other drivers of nodal price – natural gas prices, demand, etc. Adding additional controls from Column (7)-(10) changes results only slightly. Column (10) contains the preferred specification and is the main result.

Table A.2: **Results: Stepwise Addition of Controls**. This table shows storage coefficient results as controls are added stepwise. Distributed Load Aggregation Price (DLAP) x node x hour controls account for regional price differences. Thus, DLAP absorbs natural gas price, the Aliso Canyon blowout, and nuclear capacity. Adding these controls beyond DLAP does not change point estimates or statistical inference in any meaningful way, indicating the appropriateness of DLAP controls.

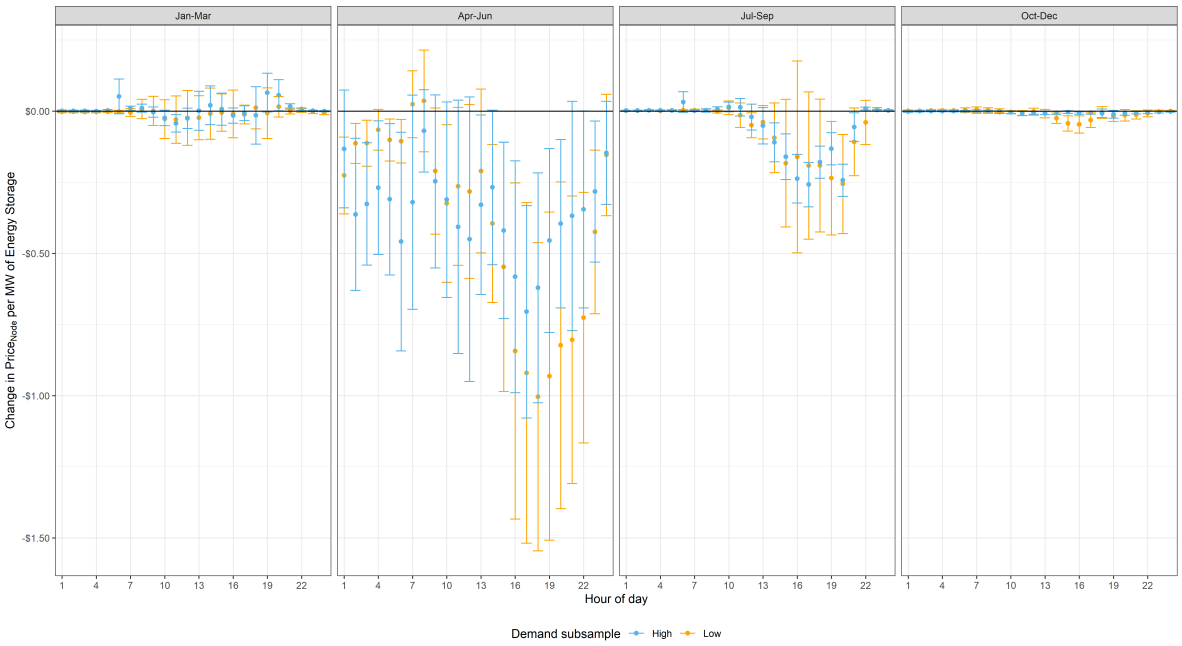
Season		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Q1	Hour 1	−0.106*** (0.005)	−0.038*** (0.004)	−0.350*** (0.035)	−0.350*** (0.035)	−0.374*** (0.035)	−0.002 (0.001)	−0.001 (0.001)	−0.002 (0.001)	−0.002 (0.001)	−0.002 (0.001)
	Hour 2	−0.134*** (0.005)	−0.044*** (0.004)	−0.422*** (0.038)	−0.422*** (0.038)	−0.452*** (0.039)	−0.002 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)
	Hour 3	−0.155*** (0.005)	−0.046*** (0.004)	−0.448*** (0.040)	−0.448*** (0.040)	−0.478*** (0.042)	−0.002* (0.001)	−0.002+ (0.001)	−0.002+ (0.001)	−0.002+ (0.001)	−0.002+ (0.001)
	Hour 4	−0.157*** (0.005)	−0.045*** (0.004)	−0.419*** (0.038)	−0.419*** (0.038)	−0.446*** (0.039)	−0.002+ (0.001)	−0.001 (0.001)	−0.002 (0.001)	−0.002 (0.001)	−0.002+ (0.001)
	Hour 5	−0.128*** (0.005)	−0.037*** (0.004)	−0.309*** (0.033)	−0.309*** (0.033)	−0.330*** (0.036)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
	Hour 6	−0.065*** (0.004)	−0.028*** (0.004)	−0.134*** (0.028)	−0.134*** (0.028)	−0.156*** (0.030)	−0.033 (0.024)	−0.011 (0.011)	0.022 (0.017)	0.022 (0.017)	0.022 (0.017)
	Hour 7	0.000 (0.005)	−0.025*** (0.004)	−0.055 (0.040)	−0.054 (0.040)	−0.087* (0.042)	0.005 (0.003)	0.006+ (0.003)	0.006 (0.003)	0.006 (0.003)	0.006+ (0.003)
	Hour 8	0.030*** (0.006)	−0.030*** (0.004)	−0.241*** (0.059)	−0.233*** (0.059)	−0.259*** (0.060)	−0.003 (0.009)	0.003 (0.009)	0.008 (0.008)	0.008 (0.008)	0.009 (0.009)
	Hour 9	−0.014** (0.005)	−0.028*** (0.005)	−0.540*** (0.063)	−0.529*** (0.064)	−0.543*** (0.065)	−0.015 (0.016)	−0.011 (0.015)	−0.009 (0.015)	−0.009 (0.015)	−0.008 (0.015)
	Hour 10	−0.040*** (0.006)	−0.032*** (0.005)	−0.752*** (0.073)	−0.736*** (0.073)	−0.753*** (0.073)	−0.060** (0.022)	−0.050* (0.022)	−0.048* (0.022)	−0.048* (0.022)	−0.048* (0.022)
	Hour 11	−0.046*** (0.007)	−0.032*** (0.005)	−0.767*** (0.073)	−0.744*** (0.072)	−0.762*** (0.070)	−0.077** (0.026)	−0.065* (0.026)	−0.060* (0.026)	−0.060* (0.026)	−0.061* (0.026)
	Hour 12	−0.056*** (0.007)	−0.034*** (0.005)	−0.709*** (0.077)	−0.682*** (0.077)	−0.693*** (0.076)	−0.082* (0.034)	−0.070* (0.033)	−0.061+ (0.033)	−0.061+ (0.033)	−0.064+ (0.033)
	Hour 13	−0.070*** (0.007)	−0.034*** (0.005)	−0.617*** (0.075)	−0.587*** (0.076)	−0.587*** (0.074)	−0.051+ (0.028)	−0.040 (0.028)	−0.033 (0.028)	−0.033 (0.028)	−0.034 (0.028)
	Hour 14	−0.072*** (0.007)	−0.030*** (0.005)	−0.562*** (0.084)	−0.531*** (0.085)	−0.544*** (0.082)	−0.036 (0.030)	−0.025 (0.030)	−0.018 (0.030)	−0.018 (0.030)	−0.019 (0.031)
	Hour 15	−0.075*** (0.007)	−0.032*** (0.005)	−0.634*** (0.076)	−0.603*** (0.077)	−0.605*** (0.075)	−0.053* (0.023)	−0.038+ (0.023)	−0.027 (0.023)	−0.027 (0.023)	−0.028 (0.023)
	Hour 16	−0.067*** (0.006)	−0.034*** (0.005)	−0.696*** (0.079)	−0.668*** (0.080)	−0.670*** (0.079)	−0.076** (0.027)	−0.061* (0.027)	−0.049+ (0.027)	−0.049+ (0.027)	−0.050+ (0.027)
	Hour 17	−0.034*** (0.005)	−0.039*** (0.004)	−0.660*** (0.067)	−0.632*** (0.067)	−0.645*** (0.064)	−0.046*** (0.012)	−0.040** (0.013)	−0.027* (0.012)	−0.027* (0.012)	−0.028* (0.012)
	Hour 18	0.062*** (0.007)	−0.040*** (0.005)	−0.710*** (0.083)	−0.697*** (0.082)	−0.715*** (0.083)	0.001 (0.026)	−0.006 (0.028)	0.005 (0.027)	0.005 (0.027)	0.004 (0.027)
	Hour 19	0.128*** (0.008)	−0.039*** (0.005)	−0.145* (0.066)	−0.142* (0.065)	−0.181** (0.068)	0.023 (0.026)	0.022 (0.026)	0.027 (0.026)	0.027 (0.026)	0.027 (0.026)
	Hour 20	0.109*** (0.007)	−0.035*** (0.005)	0.119* (0.054)	0.119* (0.054)	0.087 (0.054)	0.029* (0.012)	0.028* (0.013)	0.031* (0.013)	0.031* (0.013)	0.030* (0.013)
	Hour 21	0.067*** (0.006)	−0.037*** (0.004)	−0.027 (0.033)	−0.027 (0.033)	−0.064+ (0.035)	0.004 (0.003)	0.008* (0.003)	0.012** (0.004)	0.012** (0.004)	0.011** (0.004)
	Hour 22	0.017*** (0.004)	−0.035*** (0.004)	−0.103*** (0.027)	−0.103*** (0.027)	−0.120*** (0.027)	0.003+ (0.001)	0.004* (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.006** (0.002)
	Hour 23	−0.020*** (0.004)	−0.034*** (0.004)	−0.213*** (0.020)	−0.213*** (0.020)	−0.242*** (0.022)	−0.001 (0.002)	−0.001 (0.002)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
	Hour 24	−0.070*** (0.004)	−0.036*** (0.004)	−0.285*** (0.025)	−0.285*** (0.025)	−0.319*** (0.026)	−0.008* (0.004)	−0.006* (0.003)	−0.005* (0.002)	−0.005* (0.002)	−0.005* (0.002)
Q2	Hour 1	−0.157*** (0.004)	−0.044*** (0.005)	0.330 (0.249)	0.330 (0.249)	−0.308 (0.238)	−0.217*** (0.041)	−0.218*** (0.041)	−0.221*** (0.042)	−0.222*** (0.042)	−0.221*** (0.042)
	Hour 2	−0.192*** (0.004)	−0.045*** (0.005)	0.333 (0.230)	0.333 (0.230)	−0.300 (0.218)	−0.102*** (0.029)	−0.103*** (0.029)	−0.101*** (0.030)	−0.101*** (0.030)	−0.102*** (0.030)
	Hour 3	−0.229*** (0.005)	−0.044*** (0.005)	0.565* (0.248)	0.565* (0.248)	−0.078 (0.237)	−0.099** (0.037)	−0.099** (0.037)	−0.097* (0.039)	−0.098* (0.039)	−0.099* (0.039)
	Hour 4	−0.240*** (0.005)	−0.040*** (0.005)	0.324 (0.279)	0.324 (0.280)	−0.297 (0.271)	−0.087** (0.030)	−0.088** (0.030)	−0.086** (0.033)	−0.087** (0.033)	−0.088** (0.032)
	Hour 5	−0.218*** (0.004)	−0.033*** (0.005)	−0.145 (0.217)	−0.145 (0.217)	−0.747*** (0.213)	−0.108*** (0.025)	−0.110*** (0.025)	−0.109*** (0.026)	−0.110*** (0.026)	−0.111*** (0.026)
	Hour 6	−0.159*** (0.005)	−0.023*** (0.006)	−1.675*** (0.208)	−1.675*** (0.208)	−2.357*** (0.222)	−0.248** (0.078)	−0.260** (0.085)	−0.291** (0.101)	−0.292** (0.101)	−0.292** (0.101)
	Hour 7	−0.106*** (0.006)	−0.019** (0.006)	−3.392*** (0.326)	−3.283*** (0.322)	−4.008*** (0.340)	−0.093 (0.062)	−0.097 (0.064)	−0.091 (0.070)	−0.092 (0.070)	−0.091 (0.070)
	Hour 8	−0.087*** (0.006)	−0.024*** (0.005)	−2.952*** (0.321)	−2.961*** (0.321)	−3.732*** (0.338)	−0.033 (0.047)	−0.036 (0.048)	−0.016 (0.054)	−0.017 (0.054)	−0.017 (0.054)
	Hour 9	−0.106*** (0.006)	−0.037*** (0.004)	−1.116*** (0.306)	−1.209*** (0.294)	−2.078*** (0.294)	−0.293** (0.093)	−0.298** (0.093)	−0.293** (0.095)	−0.293** (0.095)	−0.296** (0.095)
	Hour 10	−0.100*** (0.006)	−0.039*** (0.005)	0.107 (0.377)	−0.043 (0.369)	−0.912** (0.352)	−0.416*** (0.114)	−0.425*** (0.114)	−0.429*** (0.117)	−0.430*** (0.117)	−0.437*** (0.117)
	Hour 11	−0.082*** (0.006)	−0.038*** (0.005)	1.229** (0.474)	1.090* (0.474)	0.188 (0.449)	−0.466*** (0.131)	−0.477*** (0.131)	−0.510*** (0.133)	−0.513*** (0.133)	−0.521*** (0.133)
	Hour 12	−0.070*** (0.007)	−0.038*** (0.005)	2.356*** (0.550)	2.207*** (0.546)	1.274* (0.512)	−0.523*** (0.145)	−0.534*** (0.145)	−0.575*** (0.146)	−0.579*** (0.146)	−0.585*** (0.146)
	Hour 13	−0.070*** (0.007)	−0.044*** (0.005)	3.145*** (0.600)	3.142*** (0.604)	2.209*** (0.562)	−0.428*** (0.116)	−0.437*** (0.117)	−0.477*** (0.117)	−0.482*** (0.117)	−0.490*** (0.117)
	Hour 14	−0.063*** (0.007)	−0.047*** (0.005)	3.871*** (0.668)	3.865*** (0.669)	2.919*** (0.626)	−0.535*** (0.133)	−0.543*** (0.133)	−0.589*** (0.138)	−0.594*** (0.138)	−0.600*** (0.138)
	Hour 15	−0.051*** (0.007)	−0.046*** (0.006)	4.357*** (0.734)	4.358*** (0.734)	3.438*** (0.686)	−0.787*** (0.190)	−0.794*** (0.190)	−0.841*** (0.194)	−0.844*** (0.194)	−0.852*** (0.194)
	Hour 16	−0.037*** (0.008)	−0.046*** (0.007)	4.758*** (0.875)	4.799*** (0.876)	3.890*** (0.824)	−1.006*** (0.236)	−1.011*** (0.236)	−1.054*** (0.240)	−1.055*** (0.240)	−1.063*** (0.239)
	Hour 17	−0.017+ (0.009)	−0.042*** (0.008)	5.880*** (1.425)	5.902*** (1.433)	5.016*** (1.379)	−1.072*** (0.211)	−1.078*** (0.211)	−1.103*** (0.214)	−1.104*** (0.214)	−1.109*** (0.214)
	Hour 18	−0.012 (0.009)	−0.046*** (0.008)	4.931** (1.575)	4.995** (1.579)	4.129** (1.522)	−1.085*** (0.189)	−1.091*** (0.189)	−1.108*** (0.192)	−1.109*** (0.192)	−1.113*** (0.192)
	Hour 19	0.026* (0.012)	−0.016 (0.011)	4.231* (2.017)	4.342* (2.016)	3.508+ (1.963)	−0.895*** (0.190)	−0.894*** (0.190)	−0.906*** (0.193)	−0.907*** (0.193)	−0.924*** (0.193)
	Hour 20	0.106*** (0.018)	0.017 (0.016)	2.321 (3.075)	2.416 (3.071)	1.635 (3.026)	−0.846*** (0.203)	−0.844*** (0.203)	−0.858*** (0.207)	−0.859*** (0.207)	−0.874*** (0.206)
	Hour 21	0.128*** (0.010)	−0.003 (0.007)	−1.873+ (1.006)	−1.873+ (1.006)	−2.633** (0.969)	−0.817*** (0.174)	−0.817*** (0.174)	−0.822*** (0.177)	−0.823*** (0.177)	−0.833*** (0.177)
	Hour 22	0.036*** (0.006)	−0.018** (0.006)	−1.116*** (0.273)	−1.116*** (0.273)	−1.812*** (0.261)	−0.739*** (0.126)	−0.741*** (0.126)	−0.753*** (0.129)	−0.754*** (0.129)	−0.756*** (0.129)
	Hour 23	−0.045*** (0.004)	−0.027*** (0.005)	−0.774** (0.236)	−0.774** (0.236)	−1.489*** (0.224)	−0.454*** (0.088)	−0.455*** (0.088)	−0.460*** (0.090)	−0.461*** (0.090)	−0.460*** (0.090)
	Hour 24	−0.119*** (0.004)	−0.042*** (0.005)	0.443* (0.212)	0.443* (0.212)	−0.240 (0.191)	−0.220*** (0.061)	−0.220*** (0.061)	−0.221*** (0.062)	−0.222*** (0.062)	−0.221*** (0.062)
Q3	Hour 1	−0.083*** (0.003)	−0.035*** (0.003)	0.027** (0.009)	0.027** (0.009)	0.048*** (0.011)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)
	Hour 2	−0.113*** (0.003)	−0.038*** (0.003)	0.039*** (0.010)	0.039*** (0.010)	0.063*** (0.011)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)	0.002** (0.001)
	Hour 3	−0.141*** (0.003)	−0.039*** (0.003)	0.086*** (0.016)	0.086*** (0.016)	0.120*** (0.017)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
	Hour 4	−0.157*** (0.003)	−0.041*** (0.003)	0.106*** (0.020)	0.106*** (0.020)	0.132*** (0.021)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
	Hour 5	−0.150*** (0.003)	−0.040*** (0.002)	0.104*** (0.017)	0.104*** (0.017)	0.133*** (0.019)	0.002*** (0.001)	0.002*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
	Hour 6	−0.119*** (0.003)	−0.038*** (0.003)	0.091*** (0.015)	0.091*** (0.015)	0.126*** (0.018)	0.028 (0.018)	0.027 (0.018)	0.019 (0.011)	0.019 (0.011)	0.019 (0.011)
	Hour 7	−0.103*** (0.004)	−0.039*** (0.003)	0.139*** (0.015)	0.138*** (0.015)	0.175*** (0.017)	0.003+ (0.001)	0.003+ (0.001)	0.003* (0.001)	0.003* (0.001)	0.003* (0.001)
	Hour 8	−0.084*** (0.004)	−0.035*** (0.003)	0.118*** (0.013)	0.116*** (0.013)	0.153*** (0.016)	0.007** (0.002)	0.007** (0.002)	0.005* (0.002)	0.005* (0.002)	0.005* (0.002)
	Hour 9	−0.079*** (0.003)	−0.033*** (0.002)	0.102*** (0.015)	0.102*** (0.015)	0.131*** (0.019)	0.008** (0.003)	0.008** (0.003)	0.006* (0.003)	0.006* (0.003)	0.006* (0.003)
	Hour 10	−0.060*** (0.003)	−0.032*** (0.003)	0.101*** (0.022)	0.100*** (0.022)	0.136*** (0.026)	0.023** (0.008)	0.023** (0.008)	0.019* (0.008)	0.019* (0.008)	0.019* (0.008)
	Hour 11	−0.036*** (0.004)	−0.035*** (0.003)	0.114** (0.042)	0.114** (0.042)	0.142** (0.046)	0.014 (0.015)	0.014 (0.015)	0.009 (0.014)	0.009 (0.014)	0.009 (0.014)
	Hour 12	−0.012*** (0.004)	−0.039*** (0.003)	0.134** (0.052)	0.133* (0.052)	0.158** (0.057)	−0.015 (0.021)	−0.015 (0.021)	−0.020 (0.020)	−0.020 (0.020)	−0.020 (0.020)
	Hour 13	0.010** (0.004)	−0.044*** (0.004)	0.188* (0.075)	0.187* (0.075)	0.210* (0.082)	−0.032 (0.025)	−0.032 (0.025)	−0.036 (0.024)	−0.036 (0.024)	−0.035 (0.024)
	Hour 14	0.051*** (0.005)	−0.035*** (0.005)	0.252* (0.118)	0.251* (0.118)	0.271* (0.121)	−0.079* (0.038)	−0.079* (0.038)	−0.081* (0.036)	−0.081* (0.036)	−0.080* (0.036)
	Hour 15	0.111*** (0.009)	−0.025** (0.009)	0.256 (0.174)	0.255 (0.174)	0.276 (0.180)	−0.137* (0.054)	−0.137* (0.054)	−0.141** (0.052)	−0.141** (0.052)	−0.140** (0.052)
	Hour 16	0.194*** (0.015)	0.003 (0.014)	0.311 (0.209)	0.309 (0.209)	0.321 (0.210)	−0.151* (0.074)	−0.151* (0.074)	−0.156* (0.073)	−0.156* (0.073)	−0.156* (0.072)
	Hour 17	0.255*** (0.018)	0.025 (0.016)	0.165 (0.201)	0.162 (0.201)	0.163 (0.203)	−0.183** (0.061)	−0.183** (0.061)	−0.189** (0.059)	−0.189** (0.059)	−0.189** (0.059)
	Hour 18	0.241*** (0.021)	0.025 (0.017)	0.081 (0.153)	0.078 (0.153)	0.102 (0.155)	−0.148** (0.053)	−0.149** (0.053)	−0.155** (0.052)	−0.155** (0.052)	−0.155** (0.052)
	Hour 19	0.262*** (0.030)	0.046* (0.019)	0.122+ (0.064)	0.119+ (0.064)	0.138* (0.065)	−0.152*** (0.045)	−0.152*** (0.045)	−0.153*** (0.045)	−0.153*** (0.045)	−0.153*** (0.045)
	Hour 20	0.242*** (0.024)	0.039* (0.016)	0.378*** (0.056)	0.378*** (0.056)	0.401*** (0.058)	−0.237*** (0.037)	−0.237*** (0.037)	−0.238*** (0.037)	−0.238*** (0.037)	−0.237*** (0.037)
	Hour 21	0.147*** (0.012)	0.012 (0.011)	0.207*** (0.058)	0.207*** (0.058)	0.235*** (0.060)	−0.062* (0.030)	−0.062* (0.030)	−0.063* (0.029)	−0.063* (0.029)	−0.063* (0.029)
	Hour 22	0.059*** (0.007)	−0.006 (0.008)	0.137*** (0.031)	0.137*** (0.031)	0.157*** (0.031)	−0.004 (0.017)	−0.004 (0.017)	−0.004 (0.017)	−0.004 (0.017)	−0.004 (0.017)
	Hour 23	−0.013*** (0.003)	−0.026*** (0.004)	0.054*** (0.016)	0.054*** (0.016)	0.070*** (0.017)	0.006** (0.002)	0.006** (0.002)	0.006** (0.002)	0.006** (0.002)	0.006** (0.002)
	Hour 24	−0.058*** (0.003)	−0.037*** (0.003)	0.042*** (0.009)	0.042*** (0.009)	0.062*** (0.010)	0.004*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
Q4	Hour 1	−0.075*** (0.003)	−0.012*** (0.002)	0.022*** (0.004)	0.022*** (0.004)	0.013** (0.004)	0.002** (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)
	Hour 2	−0.102*** (0.003)	−0.014*** (0.002)	0.021*** (0.004)	0.021*** (0.004)	0.012** (0.004)	0.003*** (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
	Hour 3	−0.122*** (0.003)	−0.017*** (0.002)	0.018*** (0.004)	0.018*** (0.004)	0.009* (0.004)	0.003*** (0.001)	0.001 (0.001)	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)
	Hour 4	−0.126*** (0.003)	−0.018*** (0.002)	0.017*** (0.005)	0.017*** (0.005)	0.008 (0.005)	0.002** (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
	Hour 5	−0.109*** (0.003)	−0.016*** (0.002)	0.019*** (0.005)	0.019*** (0.005)	0.011+ (0.006)	0.002*** (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
	Hour 6	−0.054*** (0.003)	−0.015*** (0.003)	0.017*** (0.004)	0.017*** (0.004)	0.010* (0.005)	0.006+ (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.002 (0.003)
	Hour 7	0.012+ (0.006)	−0.010* (0.005)	0.015* (0.007)	0.015* (0.007)	0.007 (0.007)	0.007* (0.003)	0.002 (0.003)	0.001 (0.003)	0.001 (0.003)	0.002 (0.003)
	Hour 8	0.029*** (0.006)	−0.013** (0.005)	0.021* (0.008)	0.021* (0.008)	0.014 (0.009)	0.007** (0.003)	0.000 (0.003)	0.000 (0.003)	0.000 (0.003)	0.001 (0.003)
	Hour 9	−0.017*** (0.004)	−0.010** (0.003)	0.009 (0.006)	0.009 (0.006)	0.001 (0.007)	0.004+ (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)
	Hour 10	−0.030*** (0.004)	−0.004 (0.003)	0.019*** (0.006)	0.018** (0.006)	0.011+ (0.006)	0.002 (0.002)	−0.004* (0.002)	−0.004* (0.002)	−0.004* (0.002)	−0.004* (0.002)
	Hour 11	−0.026*** (0.004)	0.001 (0.004)	0.037*** (0.007)	0.036*** (0.007)	0.030*** (0.007)	−0.001 (0.003)	−0.007* (0.003)	−0.007* (0.003)	−0.007* (0.003)	−0.007* (0.003)
	Hour 12	−0.027*** (0.005)	0.002 (0.004)	0.041*** (0.007)	0.039*** (0.007)	0.034*** (0.007)	0.003 (0.004)	−0.005 (0.004)	−0.004 (0.004)	−0.004 (0.004)	−0.004 (0.004)
	Hour 13	−0.034*** (0.005)	0.002 (0.004)	0.049*** (0.007)	0.047*** (0.007)	0.042*** (0.007)	−0.001 (0.005)	−0.009+ (0.005)	−0.008+ (0.005)	−0.008+ (0.005)	−0.008+ (0.005)
	Hour 14	−0.029*** (0.005)	0.003 (0.004)	0.049*** (0.008)	0.047*** (0.008)	0.042*** (0.008)	−0.011+ (0.006)	−0.018** (0.006)	−0.018** (0.006)	−0.018** (0.006)	−0.018** (0.006)
	Hour 15	−0.020*** (0.006)	0.003 (0.004)	0.065*** (0.011)	0.063*** (0.011)	0.058*** (0.011)	−0.020* (0.009)	−0.025** (0.008)	−0.026** (0.008)	−0.026** (0.008)	−0.026** (0.008)
	Hour 16	0.009 (0.007)	0.001 (0.005)	0.071*** (0.016)	0.070*** (0.016)	0.064*** (0.016)	−0.028** (0.010)	−0.031*** (0.009)	−0.032*** (0.009)	−0.032*** (0.009)	−0.032*** (0.009)
	Hour 17	0.077*** (0.009)	0.003 (0.007)	0.064** (0.021)	0.063** (0.021)	0.057** (0.020)	−0.020* (0.009)	−0.025** (0.009)	−0.025** (0.009)	−0.025** (0.009)	−0.024** (0.009)
	Hour 18	0.232*** (0.018)	0.013 (0.011)	−0.010 (0.015)	−0.011 (0.015)	−0.020 (0.015)	−0.002 (0.007)	−0.009 (0.007)	−0.008 (0.007)	−0.008 (0.007)	−0.006 (0.007)
	Hour 19	0.258*** (0.023)	0.017 (0.013)	−0.001 (0.011)	−0.001 (0.011)	−0.013 (0.011)	−0.006 (0.006)	−0.019** (0.006)	−0.018** (0.006)	−0.018** (0.006)	−0.017* (0.006)
	Hour 20	0.189*** (0.014)	0.002 (0.008)	0.032*** (0.009)	0.032*** (0.009)	0.019* (0.009)	0.000 (0.007)	−0.015* (0.007)	−0.015+ (0.007)	−0.015+ (0.007)	−0.012 (0.008)
	Hour 21	0.108*** (0.008)	−0.005 (0.005)	0.007 (0.006)	0.007 (0.006)	−0.005 (0.006)	−0.004 (0.005)	−0.013** (0.005)	−0.013** (0.005)	−0.013** (0.005)	−0.012* (0.005)
	Hour 22	0.049*** (0.006)	−0.008+ (0.004)	0.008 (0.006)	0.008 (0.006)	−0.003 (0.006)	−0.005 (0.003)	−0.010** (0.003)	−0.009* (0.004)	−0.009* (0.004)	−0.008* (0.004)
	Hour 23	0.008* (0.004)	−0.008** (0.003)	0.013** (0.005)	0.013** (0.005)	0.003 (0.006)	0.000 (0.001)	−0.005*** (0.001)	−0.004** (0.001)	−0.004** (0.001)	−0.003* (0.001)
	Hour 24	−0.048*** (0.003)	−0.015*** (0.003)	0.013*** (0.004)	0.013*** (0.004)	0.003 (0.004)	0.001+ (0.001)	−0.002* (0.001)	−0.002+ (0.001)	−0.002+ (0.001)	−0.001 (0.001)
	Num.Obs.	51 501 655	51 501 655	51 501 655	51 458 534	51 458 534	51 458 534	51 458 534	51 331 382	51 331 382	51 331 382
	R2	0.000	0.239	0.495	0.496	0.499	0.817	0.817	0.817	0.817	0.817
	R2 Adj.	0.000	0.236	0.484	0.484	0.488	0.812	0.812	0.813	0.812	0.812
	R2 Within		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Season x Weekend x Hour x Node		X								
	Season x Year x Weekend x Hour x Node			X	X	X	X	X	X	X	X
	Solar gen index x Node				X	X	X	X	X	X	X
	Relative temp x Node					X	X	X	X	X	X
	Distribution Load Agg. Price x Hour x Node						X	X	X	X	X
	12 month rolling pcpt x Hour x Node							X	X	X	X
	CAISO natural gas price x Hour x Node								X	X	X
	Aliso Canyon blowout x Hour x Node									X	X
	Total nuclear avail x Hour x Node										X

Figure A.4: **Coefficient Estimates**: Equation 9 with 95% CI, by Season and Hour, estimated using stepwise controls and fixed effects (listed at bottom of table). Driscoll-Kraay standard errors shown.

A.4 Heterogeneity in Own-Node Effects by Total Grid Demand

I re-estimate Equation 9 splitting the sample with in each season at the median daily forecast demand. Thus, I take the highest (lowest) 50% of all Spring days, highest (lowest) 50% of all Summer days, etc. and estimate on each sample separately. Results are shown in Figure A.5.

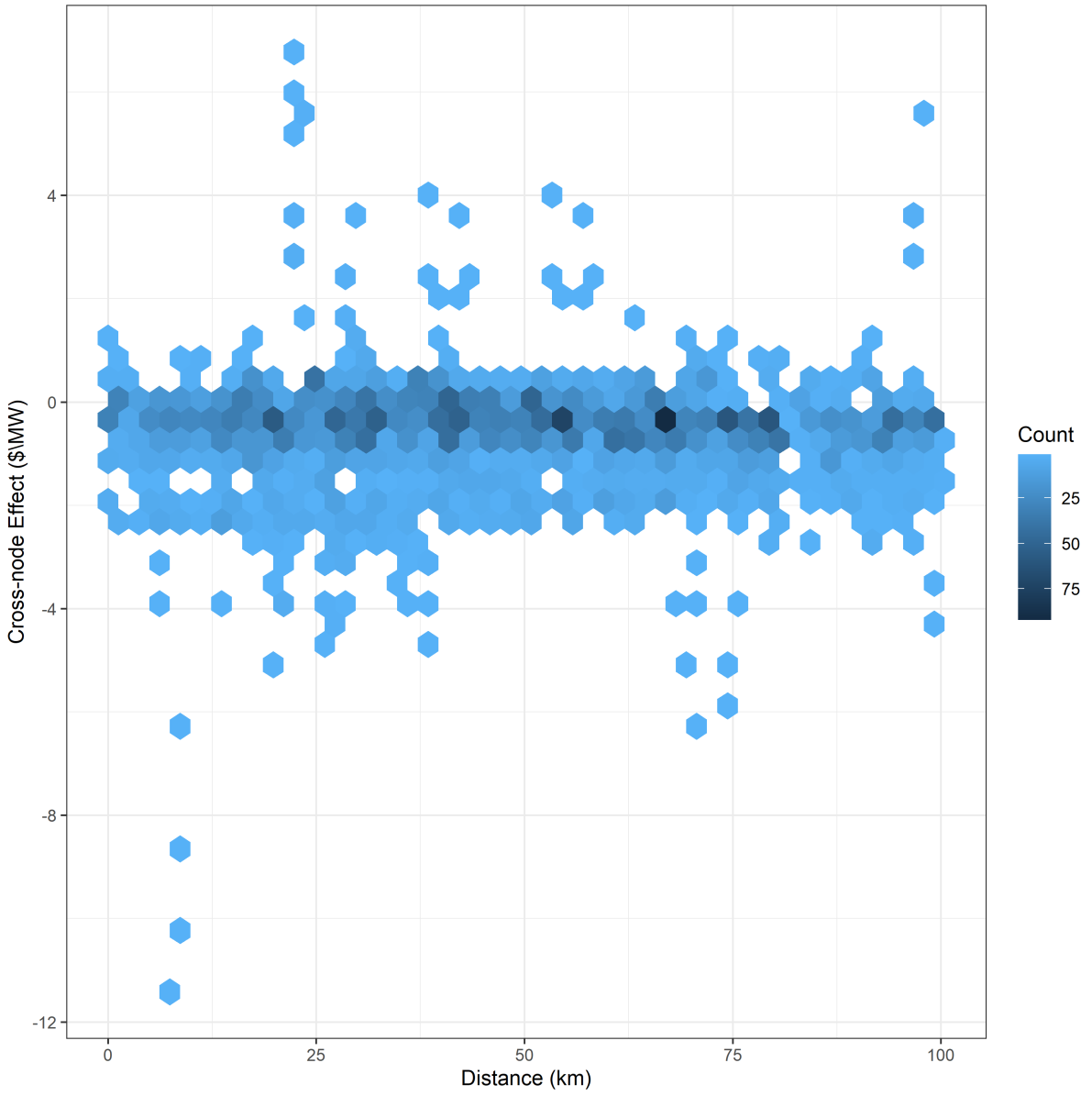
Figure A.5: **Own-Node Results Splitting Sample by Daily Forecast Demand**. The upper panel shows coefficient results for the lower 50% of demand days in the sample. The lower panel shows the same but fo the highest-demand dates.



Results are consistent with the effect of storage scaling inversely with the daily grid demand. In Apr-Jun, the sample consisting of low-demand spring days yields more negative results that the spring high-demand days. The high-demand days still show negative afternoon price effects, but of much smaller magnitude and over fewer hours. Similarly Jul-Sep results show low-demand days with small but negative point estimates, while high-demand summer days are attenuated towards zero in the evening "shoulder" hours. While higher-demand periods face more convex supply, the level of storage necessary to alleviate constraints also increases (see Appendix E for an example).

A.5 Cross-node effects of energy storage by distance

Figure A.6: **Cross-node effects by distance**. This figure shows the density of distance between a LASSO-selected pricing node and the energy storage node estimated to have a price effect. The y-axis shows the magnitude and sign of the effect. A mass of negative effects lies below 80km, though identified effects persist out to 400km and beyond.



Appendix B Data Construction

In this section, I detail the construction of the hourly solar generation index, the hourly relative temperature measure, total available nuclear capacity, the rolling 12-month average precipitation measure, CAISO-reported citygate natural gas prices, and the Aliso

Canyon Blowout variable.

B.1 Solar Generation Index

The hourly solar generation index controls for the price effect of local solar generation, which is a function of both capacity, which changes daily, and irradiance, which is measured hourly. First, I obtain hourly Global Horizontal Irradiance (GHI) measurements for the centroid of each polygon from the National Renewable Energy Laboratory's (NREL) National Solar Radiation Database (NSRDB) (National Renewable Energy Laboratory, 2018). Combined with a physical solar model such as NREL's System Advisory Model (SAM) and data on a system's tilt, azimuth, and inverter, it is possible to convert GHI to electricity generated as in Craig et al. (2018). However, I do not directly convert to generation. Instead, I note that, conditional on solar capacity, solar generation is proportional to GHI.

To create a node-level index of local solar generation, daily solar capacity within each polygon is combined with each polygon's GHI measure and standardized. Locations and installation dates of household solar are extracted from the California Solar Initiative (CSI) database (California Solar Initiative, 2018) and interpolated across each zip code as exact coordinates are not available in the CSI. Utility-scale solar generation dominates distributed generation in California by more than 2:1. I use the EPA eGrid database to find the location, capacity, and commission date of each utility-scale solar facility in the CAMX eGrid sub-region that contains the CAISO territory, and match each to the nearest node. An hourly time series of total installed solar capacity is generated for each polygon and multiplied by hourly GHI. The resulting standardized measure represents a within-node hourly solar generation index proportional to actual generation.

B.2 Relative Temperature

Local temperature at node n will affect demand, and thus congestion, at node n. While controls for the LAP-level price are included in all specifications, local deviations in temperature relative to other nodes in the LAP may cause deviations from this relationship. To improve precision, I generate a measure of nodal temperature relative to other nodal temperatures.

To construct this measure, I obtain hourly temperature measures for the centroid of each node from the NSRDB, as described above. For each node n and hour t, I calculate the leave-one-out mean LAP temperature. I then subtract this mean from each node's hourly temperature, which represents that node's temperature deviation from the rest-of-LAP hourly mean.

B.3 Total Available Nuclear Capacity

I merge daily nuclear capacity availability for the two nuclear power plants in CAISO territory during the study period – San Onofre and Diablo Canyon. This data, obtained from the EIA,[20] shows the daily amount of capacity available. In general, all nuclear

[20] https://www.eia.gov/opendata/browser/nuclear-outages/facility-nuclear-outages?frequency=daily&data=capacity;outage;percentOutage;&facets=facility;&facility=360;

capacity not undergoing maintenance or otherwise reported as unavailable is used for generation. Although hourly nuclear generation is not reported in archived data, capacity availability is a reliable proxy. San Onofre was decommissioned during the study period at which time all of its capacity was set to zero. Data is merged at the daily level, and thus reflects a grid-wide availability of nuclear generation.

B.4 12-month Precipitation

I obtain monthly measures of statewide precipitation for California from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) Climate-at-a-glance (CAG) website.[21]

B.5 Citygate Gas Prices

I obtain citygate natural gas prices from CAISO's OASIS platform for daily natural gas prices reported for the CISO region. This price reflects both national price trends and local supply constraints or local demand. The intent of controlling for natural gas prices is to allow for local generation prices to reflect fuel prices and the local fuel mix. Thus, local prices, such as citygate price, are the preferred measure of local natural gas prices. Eight dates do not report a CISO price on CAISO, and are dropped from analysis. None of the missing data occur in the days surrounding the addition of any energy storage, making statistical issues with missing data unlikely.

B.6 Aliso Canyon Blowout

The failure of the Aliso Canyon natural gas storage field caused a local spike in natural gas prices that largely affected Southern California generators. Since this departure from both national and CAISO-wide pricing was locally concentrated, I generate an Aliso Canyon Blowout indicator variable based on the date of the explosion (October 23, 2015) and proximity to the blowout (Southern California Edison and San Diego Gas and Electric territories).

Appendix C Selection of Storage Sites

I first test for any relationship between a node's daily price spread and the selection of the node for installation of energy storage. While this relationship would not threaten identification due to the rich set of fixed effects, rejecting the null hypothesis of "no relationship between storage and price spread" would indicate that arbitrage was a factor in siting decisions, giving rise to concerns about time-varying unobservable trends. Table C.1 shows the results of a regression of a binary indicator for ever having storage on measures of the price spread at the node. To generate a relevant index for the price spread, I first calculate the daily price spread (maximum daily hourly price - minimum daily

6099;&start=2009-01-01&end=2018-12-31&sortColumn=period;&sortDirection=desc

[21] see https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series

hourly price) for every node in the data and take the node-level median for the time period prior to any storage installation. Additionally, I calculate the daily difference between the 90th and 10th percentile of hourly prices for each day, and also calculate the node-level median. These approximate the cross-sectional arbitrage potential for each node prior to the addition of storage. I estimate this using both OLS and a logit specification. Results in table C.1 indicate that the installation of storage is not significantly correlated with greater price spreads. These results are consistent with storage location decisions in the pre-2018 period being more about proof-of-concept and less about active price arbitrage attempts.

To test for selection on trends in spread, I use an event study regression specification to show the difference in the daily spread for storage and non-storage nodes during the periods prior to the installation of storage. Figure C.1 shows the results from this event study. Daily spreads for storage nodes show no differential trend prior to the installation of storage, indicating that trends in arbitrage opportunities are likely not driving selection of nodes for storage.

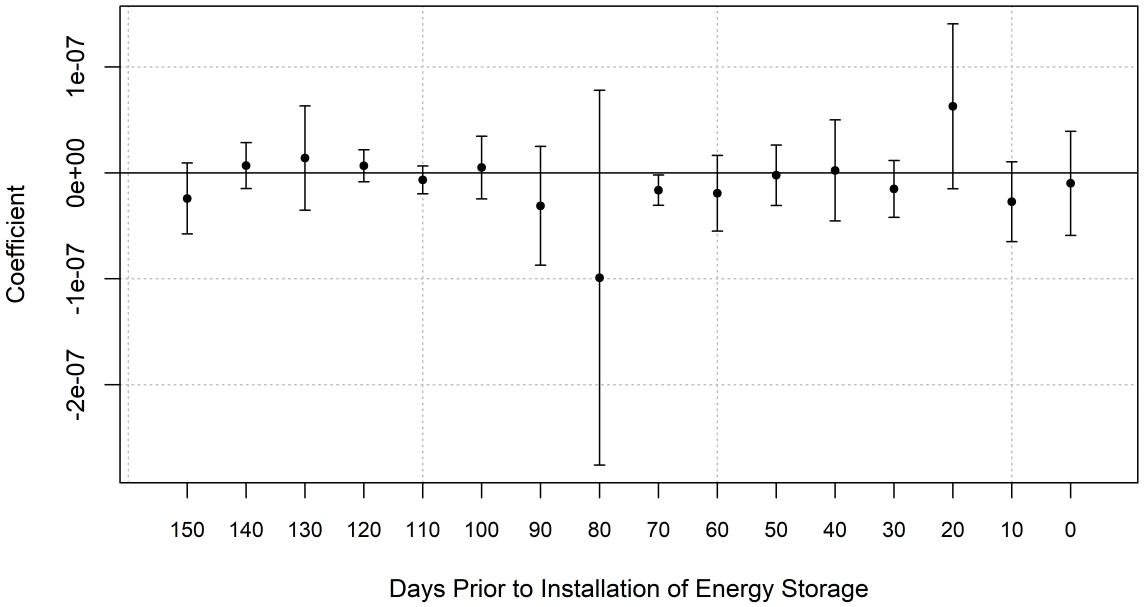
An additional threat to identification would arise from the decision to install storage being correlated with low price response. Because storage operating at an LMP may change the peak price at that node, a profit-maximizing strategy for siting would be to choose nodes with large spreads, but that will not respond in price strongly to the additional supply. As noted in the above section, the price determination process embedded in the power flow model makes sudden, significant changes in price possible – optimal siting of storage would consider this effect, assuming perfect knowledge of them. If this were the case, and conditional on observing little correlation between price spread and storage, I interpret these results to be a lower-bound on the (negative) price effect as any incentive to locate at nodes with low price response would bias estimates upward towards zero.

Table C.1: **Tests for selection on price spread**. Results from regression of binary indicator for installation of energy storage on node-level price spreads. HAC-robust errors in parentheses.

Dependent Variable:	Has storage			
Model:	(1) OLS	(2) OLS	(3) Logit	(4) Logit
Variables				
Median Within-day Max Spread	0.0027		0.1170	
	(0.0032)		(0.1285)	
Median $90^{th} - 10^{th}$ decile		0.0055		0.2354
		(0.0042)		(0.1663)
Fit statistics				
Pseudo R^2	-0.00141	-0.00317	0.00620	0.01391
Observations	757	757	757	757

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Figure C.1: **Test for trend in price spread**. Event study coefficient plot showing daily price spread differences between storage and non-storage nodes for up to 150 days prior to the installation of storage. Includes node and day two-way fixed effects. Errors clustered at the node.



Given the evidence that siting of storage during the study period was not geared towards arbitrage opportunities, the effects estimated are best interpreted as average treatment effects on the treated, or ATT. Siting of future storage based on price spreads, price responsiveness, or a combination of the two (siting for large arbitrage opportunities in areas where storage will not dissipate spread) may yield an ambiguous effect on the overall cost of serving load. On average, the effect of future storage should approach the effect estimated here. Thus, the ATT estimated here serves a useful purpose in understanding the overall impact of energy storage on the cost of serving load borne by ratepayers relative to the costs associated with mandates.

Appendix D Wind Generation

Correlation between energy storage at a pricing node and wind generation within California may bias estimates of the price effect of storage. Identification of the price effect relies on within-season variation in storage which is largely monotonically increasing[22], and may be confounded by wind or solar generation within a season if such generation is monotonically increasing due to seasonal patterns. I control for solar directly in main

[22]The exception is Oroville, where 120MW of energy storage was taken offline due to a control room fire, a genuinely exogenous event.

results. However, wind generation forecasts were not established until mid-2012, and CAISO archived data reports forecasts for day-ahead starting only in March 2015.

To address this issue, I establish that wind generation is uncorrelated with energy storage in California during the period from March 2015 to December 2017, when CAISO wind forecast data overlaps with my study period. I estimate the following equation:

$$ES_{nt} = \beta_0 + \beta_1 Wind_t + \delta_{nhswy}$$

where δ_{nhswy} is a node-hour-season-year-weekday fixed effect as in Equation 9. I show results estimated both with and without the fixed effect in Table D.1.

Table D.1: **Regression of node-level storage on CAISO day-ahead wind generation.** This regression shows that within-season wind generation is not correlated with energy storage at the node level. Note that the sample is truncated relative to main results, with wind data available only for March 2015 onward.

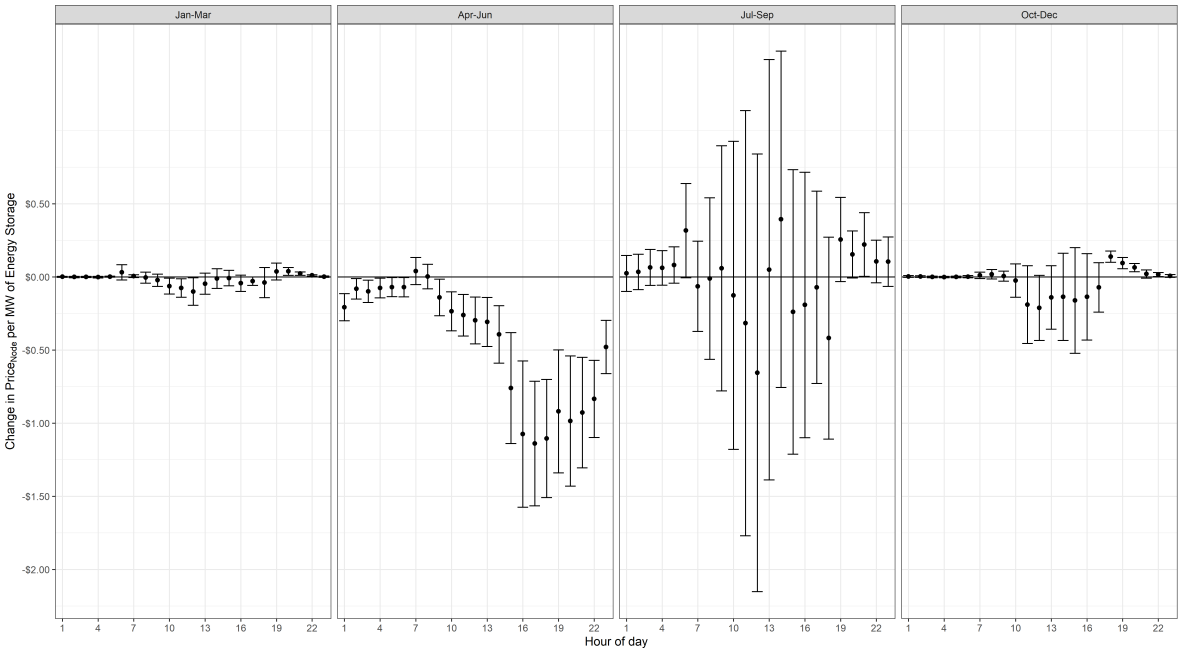
Dependent Variable:	Total Storage (MW)	
Model:	(1)	(2)
Variables		
Total Grid Wind (DAM Forecast)	$1.69 \times 10^{-6*}$	-1.96×10^{-8}
	(9.73×10^{-7})	(1.97×10^{-7})
Fixed-effects		
Hour-Node-Season-Year-Weekday		Yes
Fit statistics		
Observations	16,881,444	16,881,444
R^2	6.79×10^{-7}	0.99315
Within R^2		9.06×10^{-9}

Driscoll-Kraay (L=12) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Column (1) (without fixed effects) show significant correlation at the 10% level. However, conditional on the same fixed effects used in the main results, wind generation is not correlated with the presence of storage during the period of March 2015 to December 2017.

I also re-estimate Equation 9 and include hour-by-node-by-season controls for wind generation. Doing so limits the sample size to observations after March 2015, which comes at the expense of variation in energy storage, particularly in Q3, as 89.5% of all Q3 storage occurs before March 2015. Other quarters remain relatively proportionally reduced. Results in Figure D.1 show nearly identical results in Q2, and small changes in results for Q1 and Q4. Results in Q3 are noisy and point estimates do not follow the pattern shown in the main results. This is due to the loss of nearly all identifying variation for this quarter.

Figure D.1: **Coefficient Estimates**: Equation 9 with 95% CI, by Season and Hour. Includes node x season x hour x weekend FE, controls from Equation 9, and adds controls for node x forecast wind generation. Note that the sample is limited by wind data availability to March 2015 and later. Driscoll-Kraay standard errors shown.



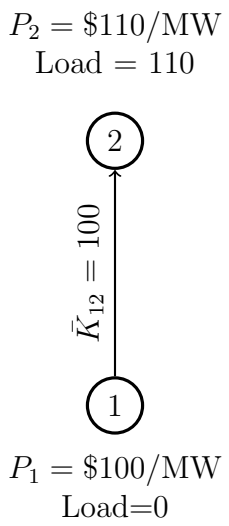
Appendix E Examples of Price Determination with Power Flow Models

In this Appendix, I present numerical examples that illustrate the role of transmission constraints in determining nodal prices.

E.1 Price determination in two-node grid

For illustrative purposes, consider the two-node grid depicted in Figure E.1. The grid consists of two nodes, 1 and 2. There is a generator at each node and a transmission line (edge) connecting them. The generator at Node 1, $G1$, has lower marginal cost, and all load is located at Node 2. Suppose load is 110MW, and each generator has sufficient capacity to serve the full load. In an unconstrained scenario, the lowest-cost generator, $G1$, would serve all load. If transmission line capacity is $\bar{K}_{12} = 100$MW, then it is clear that the more costly $G2$ generator must be dispatched. Thus, the clearing price at Node 2, P_2 is the price of the expensive generator, while the price at Node 1, P_1, remains the price of the less expensive generator. The congestion cost at Node 2 is the price difference.

Figure E.1: **Example of 2-node grid with one line and one potential line constraint**. Load is 110MW at Node 2, generation at both Node 1 and Node 2. Note that the lowest-cost electricity at Node 1 will be constrained by the capacity of the line between Node 1 and Node 2, and higher priced generation will be dispatched to meet demand (load) at Node 2.



The optimal power flow in this example can be solved using a Karush-Kuhn-Tucker generalization of the lagrangian to minimize cost of generation subject to total generation and line flow (inequality) constraints. It is assumed for this example that generator constraints do not bind, leaving only two constraints — that the sum of generation must equal demand, and that flow over the line be less than or equal to capacity. Let nodes be denoted as $n \in N$, generation as G_n, load as D_n, and capacity as $\bar{K}$. Then the cost minimization problem is given by:

$$\mathcal{L} = -\sum_{n=1}^{N} p_n(G_n)G_n + \overbrace{\lambda(0 - \sum_{n=1}^{N} G_n - \sum_{n=1}^{N} D_n)}^{\text{System-wide power flow constraint}} - \underbrace{\mu(\bar{K} - |\boldsymbol{\kappa} \times [\boldsymbol{G} - \boldsymbol{D}]\,|)}_{\text{line flow (slack) constraints}},$$

where $\boldsymbol{G}$ is the $[2 \times 1]$ vector of generation, $\boldsymbol{D}$ is the $[2 \times 1]$ vector of demand, and $\boldsymbol{\kappa}$ is the *shift factor* determined by the susceptances of the lines in the network. $\boldsymbol{G}$ is the grid operator's choice variable. λ is the familiar KKT (lagrangian) multiplier on the total energy constraint, and μ is the KKT multiplier on the line constraint. $\boldsymbol{\kappa}$ contains one row for each line in the network, and one column for each node. Column n of $\boldsymbol{\kappa}$ represents the flow over each line that will result from the injection of 1 unit of energy at the reference node and the withdrawal of 1 unit of energy from node n. If Node 1 is the reference node,[23] then $\boldsymbol{\kappa} = [0, 1]$. That is, injecting 1 unit of energy at Node 1 and withdrawing it at Node 2 will increase flow on the line between them by 1 unit. The solution takes the form of optimal (cost minimizing) generator dispatch and associated prices, denoted $\boldsymbol{P}$.

[23] A reference node is used to provide the model-implied reference location, but is chosen arbitrarily without affecting outcomes in a manner similar to the choice of a base or omitted category in a regression including fixed effects

It is:

$$\begin{aligned} \boldsymbol{G} &= [100, 10] \\ \boldsymbol{P} &= [10, 20]. \end{aligned}$$

Because the objective function minimizes total cost, and because each transmission capacity is included as a constraint, the KKT multiplier μ is the incremental reduction in the cost of serving load that would result from relaxing the transmission line constraint by 1 unit. When the constraint is relaxed, it allows one unit of energy to flow from the less expensive generator to the more expensive node, displacing costlier generation there, and lowering the total cost. The shadow value of this constraint is $\mu = P_1 - P_2 = -10$. If demand were to decrease to 100MW, or capacity were to increase to 110 MW, the constraint would become "slack", and μ would be equal to zero.

In the simple example, the weight on the transmission constraint is 1, as a unit of energy withdrawn at Node 2 (and injected at Node 1) resulted in an increase in flow on the line of 1. In a more complex network, multiple paths will exist between the withdrawal node and the injection node. Each transmission line will have a constraint which, if binding, will have a corresponding non-zero KKT multiplier. The differential between prices at the withdrawal node and the injection node will be the sum of these corresponding KKT multipliers, weighted by the amount of flow that passes over each line.

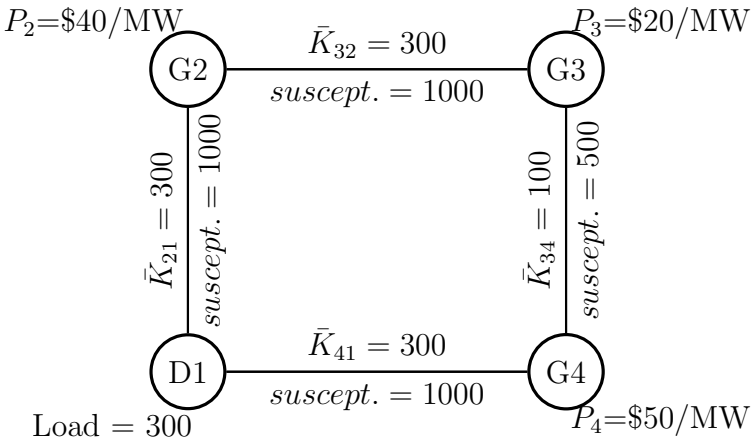
If storage is located at Node 2 which becomes congested during peak demand (e.g., in order to defer a transmission upgrade), then transmission constraints bind less frequently with the storage capacity, and the nodal price will be weakly closer to the grid price, even on days where grid-wide demand is high. In this case, storage at the congested Node 2 does not influence the price at Node 1, and prices at Node 1 would be unchanged. If there were a *third* node located in the congested part of the grid near Node 2, served also by the same single, congestable transmission line, then prices at Node 3 would also respond to the introduction of storage capacity assuming lines between Nodes 2 and 3 are uncongested. Prices at the two nodes in the congested portion of the grid would decline while prices at Node 1 would be unchanged. Because the network is not a perfect lattice where all physical neighbors are connected, these congested parts of the grid could be physically disconnected, but linked in the network. Without knowledge of the physical network structure (as in Chen et al. (2009)), estimating price effects is difficult.

E.2 Price determination in 4-node grid

A final example demonstrates price dependencies amid congestion. Consider the 4-node network shown in Figure E.2. Demand is located at Node D1, and generation is located at nodes G2, G3, and G4, with prices as shown. Capacities are denoted $\bar{K}$. The susceptance of each line determines the flow when power is injected at one node and withdrawn at another. With some algebraic manipulation, the susceptances form the shift factor matrix shown in Equation 12, with Node D1 chosen as the reference node[24].

[24]The choice of which node is the "reference" node is arbitrary. In practice, an increase in withdrawal at one node is met with an increase in supply at another node, or many other nodes, not necessarily the reference node. In this case, the change in the power flow equations is the sum of the effect of withdrawing

Figure E.2: **Example of 4-node grid with line constraints**. Note that the lowest-cost electricity at $G3$ will be constrained by the low capacity of the line K_{34}, and higher priced generation will be dispatched to meet demand at $D1$. All constraints are bi-directional.



The shift factor matrix for this grid, $\boldsymbol{\kappa}$ is derived using a power flow model with Node D1 as the reference node. The power flow model is an algebraic manipulation of the susceptances and lines of the network.[25] Based on the susceptances, the shift factor matrix is:

$$\boldsymbol{\kappa} = \begin{bmatrix} 0 & 0.8 & 0.6 & 0.2 \\ 0 & -0.2 & 0.6 & 0.2 \\ 0 & 0.2 & 0.4 & -0.2 \\ 0 & 0.2 & 0.4 & 0.8 \end{bmatrix} \tag{12}$$

Each column in $\boldsymbol{\kappa}$ represents a node, while each row represents a line. I index lines between nodes by $e \in E$ where $E = \{21, 32, 34, 41\}$. Each column reports the flow of energy across each line resulting from injecting at the reference node and withdrawing at node $n \in \{D1, G2, G3, G4\}$. For instance, 1MW injected at $D1$ and withdrawn at $G3$ will increase the flow over e_{21} by 0.6, e_{32} by 0.6, e_{34} by 0.4, and e_{41} by 0.4 (all units in MW). Similarly, injecting at $G3$ and withdrawing at $D1$ yields the opposite — e_{21} by -0.6, e_{32} by -0.6, e_{34} by -0.4, and e_{41} by -0.4. The lower relative susceptance on e_{34} causes the lower flow relative to other lines. The shift factors in column 1, the reference node, are zero because injecting 1MW and withdrawing 1MW at the reference node does not increase flow anywhere on the network.

Given the generator prices $P_1, \cdots, P_3$ and demand at $D1 = 300$, one can solve a constrained optimization for the generation vector that results in the lowest-cost power flow subject to the line constraints. The solution is:

one unit of energy at the withdrawal node and injecting at the reference node, and withdrawing one unit at the reference node and injecting at the generation node(s)

[25] I abstract away from reactive power and angles to illustrate the nature of congestion pricing in a simplified manner.

$$\boldsymbol{G} = \begin{bmatrix} 0 \\ 0 \\ 266.67 \\ 33.33 \end{bmatrix}, \boldsymbol{P} = \begin{bmatrix} 40 \\ 30 \\ 20 \\ 50 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 50 \\ 0 \end{bmatrix}$$

The least expensive generator is not serving all load. In fact, the most expensive generator serves an equal amount of load. The constraint, μ times the matrix $\boldsymbol{\kappa}$ determines the price difference between the reference node, $D1$ and the other nodes. For instance, at $G3$, $0 + 0 + (50 \times 0.4) + 0 =$\$20 less at $G3$ relative to $D1$. If storage were placed at node $D1$ of sufficient size (50MW), the constraint would not bind and the relevant entry in $\boldsymbol{\mu}$ would equal zero. Line e_{34} constraint would not bind, all load would be served by the lowest-cost generator at $G3$, and the price at nodes $G2$ and $G4$ would decrease to \$20. In this example, no other constraint is binding after the addition of storage. However, in other examples, additional constraints that may not bind under the optimal solution without storage may bind with storage. Alleviating one constraint does not alleviate all constraints.

Entries in the shift factor matrix $\boldsymbol{\kappa}$ may be negative, as is the case here — an increase in injections and withdrawals at some pair of nodes may *reduce* flow across some lines. This can be particularly beneficial when this relieves a congested line. Because the shadow value on a constraint is always positive for some lines, the negative entry in the shift factor matrix is the source of negative congestion prices. Negative congestion prices signal that withdrawals at the negative-priced node help to relieve a constraint, allowing additional (less expensive) power to flow to other nodes, and, thereby reducing the cost of serving load (i.e. the objective function).

Appendix F Formal Model of Grid Pricing

Let the congestion portion of the nodal energy constraint be denoted as $\boldsymbol{\lambda}^c$, with elements λ_n^c, and the total nodal energy constraint (the nodal price) be denoted as $\boldsymbol{\lambda}$ with elements λ_n. The KKT shadow value for transmission congestion is $\boldsymbol{\mu}$, with elements μ_e. For a network with E lines and N nodes, $\boldsymbol{\mu} \in \mathbb{R}^E$ and $\boldsymbol{\lambda}^c \in \mathbb{R}^N$:

$$\begin{aligned} \lambda_n &= \lambda^e + \lambda_n^c \\ &= \lambda^e + \boldsymbol{\mu}' \boldsymbol{\kappa}_n, \end{aligned} \tag{13}$$

where λ^e is the marginal cost of energy at the reference node, λ^c is the congestion component, and $\boldsymbol{\kappa}_n$ is the nth column of the shift factor matrix derived directly from the physical resistance of each transmission line (edge) in the network. The $\{1, \ldots, E\}$ elements of $\boldsymbol{\kappa}_n$ represent the energy flow across each line e from the reference node to node n. They are not constrained to positive numbers, though the value of the elements of $\boldsymbol{\mu}$ must be greater than zero. Therefore, it is possible for $\lambda_n^c < 0$. This can occur at a node that has low-cost energy accessible to it, but congestion keeps that low-cost energy from flowing to the reference node. If one increases the withdrawal at a negative congestion cost node ($\lambda_n^c < 0$), and increases the injection at the reference node (which was unable to

be served by the low-cost generator due to congestion), then (1) the total cost of serving demand has decreased as the cheaper generator is being used, and (2) the net flow on the congested line has decreased.

Congestion is priced at every node in a network but can be affected by withdrawals or injections at *any* node, especially those that are grid "neighbors". As a result, the effect of energy storage does not follow a congestion analog to Equation 7— the discharge of energy at peak times may *increase* the nodal price, or may cause another previously uncongested transmission line to become the binding constraint, raising nodal prices elsewhere. However, decreases in demand during peak periods will, over the entire grid, strictly decrease total congestion costs. The magnitude and location of price changes due to energy storage are determined by network topology, which is not publicly available, and, therefore, unobserved to the econometrician. Changes in network congestion costs affect transfers to LSEs and ratepayers.

The vector $\boldsymbol{\lambda^c}$ represents the congestion component of the cost of injecting one unit of energy at the reference node and withdrawing one additional unit of electricity at any node n. The price paid for *all* units of energy at node n includes λ_n^c. Interest centers on the effect of a change in nodal demand on congestion costs: $\frac{d\boldsymbol{\lambda^c}}{dQ_n}$. A change in demand at node n, however, may change congestion costs at *all* nodes by changing the transmission constraint shadow value vector, $\boldsymbol{\mu}$. Furthermore, the change in $\boldsymbol{\lambda^c}$ is itself dependent on the vector of net demand, $\boldsymbol{Q}$:

$$\begin{aligned}
\frac{d\boldsymbol{\lambda}}{dQ_n} &= \frac{d\lambda^e}{dQ_n} + \frac{d\boldsymbol{\lambda^c(Q)}}{dQ_n} \\
&= \frac{d\lambda^e}{dQ_n} + \boldsymbol{\kappa}' \left(\frac{d\boldsymbol{\mu}(\boldsymbol{Q})}{dQ_n} \right) \\
&= \frac{d\lambda^e}{dQ_n} + \left[\sum_{e=1}^{E} \frac{d\mu_e(\boldsymbol{Q})}{dQ_n} \kappa'_{1e}, \sum_{e=1}^{E} \frac{d\mu_e(\boldsymbol{Q})}{dQ_n} \kappa'_{2e}, \ldots, \sum_{e=1}^{E} \frac{d\mu_e(\boldsymbol{Q})}{dQ_n} \kappa'_{Ne} \right]' \qquad (14)
\end{aligned}$$

where the first equality follows from the definition in Equation 13 and that $\boldsymbol{\kappa}$ is a constant. The second equality states that the change in each node's congestion cost, λ_n^c, is the sum of the changes in each of the E transmission constraints, μ_e, weighted by the Kirchoff's Law-derived flows between the reference node and n. The KKT slack conditions on $\boldsymbol{\mu}$ provide intuition for the difference between peak period and nadir period price effects. When a constraint is not binding on line e, slack conditions require $\mu_e = 0$. If energy storage at node n is charged during the nadir, and transmission constraints are likely to be slack at this time, then the probability distribution of price effects will have mass at zero (e.g. charging during off-peak times will not increase congestion costs).

Amid congestion, interest centers on storage effects on both *own-node* congestion prices, $\frac{d\lambda_n^c}{dQ_n}$ and *cross-node* congestion prices, $\frac{d\boldsymbol{\lambda}^c}{dQ_n}$. These determine the total congestion price effect of discharging stored energy at node n during high demand $\boldsymbol{Q}^H$ recharging storage capacity during low demand, $\boldsymbol{Q}^L$. A change in Q_n will result in a change in generation at *some unknown node(s)*, further complicating estimation of the total effect. The total change in congestion cost per unit of energy storage at node n may be written as:

$$\text{Total Change} \approx \left(-s \times \frac{d\boldsymbol{\lambda^c(Q^H)}}{dQ_n} \times \boldsymbol{Q^H}\right) + \left(s \times \frac{d\boldsymbol{\lambda^c(Q^L)}}{dQ_n} \times \boldsymbol{Q^L}\right) \tag{15}$$

The first term in parentheses is the energy storage transaction $-s$ at node n multiplied by the price effect at each node from a change in Q_n, multiplied by the total demand at each node during the peak period. The second term is identical, but for an increase in withdrawal of s during the nadir.

In the previous sections, peak and nadir times are summarized by $\{H, L\}$. Operation of energy storage generally extends beyond a single hour. Storage is charged (or discharged) over 1-4-hour periods. Storage operated to shave peak demand or engage in energy arbitrage will discharge during a node's peak period. Because nodal prices are unique to each node and determined as solutions to a complex non-linear optimization problem, the peak and nadir periods may occur at different times for different nodes. Therefore, storage price effects may be measured over multiple hours that differ from node to node. Battery operations are not observed, so total daily savings of a marginal unit of storage capacity are calculated as the sum of price changes over *all* hours of the day:

$$\Delta\text{Cost of Serving Load} = s \times \left(\sum_{t=1}^{24} \frac{d\boldsymbol{\lambda}_t(\boldsymbol{Q_t})}{dES} \boldsymbol{Q_t}\right). \tag{16}$$

Appendix G LASSO Bias and Methods

The source of this bias is rooted in the machine learning nature of the parameter selection. This bias is generated as follows: If two independent variables are correlated with each other, and each has an associated effect on the dependent variable, a LASSO algorithm will select only one, zeroing out the coefficient of the other, despite their correlation. This is because the LASSO objective is parsimonious *prediction* and not inference. If the dependent variable can be explained with one variable instead of two, then the penalty term results in the selection of the more parsimonious specification. In this case, an OLS specification would partition out the explained variance, keeping both variables, provided the two are not perfectly correlated.

By selecting only one of two explanatory variables, LASSO introduces the possibility of an omitted variables bias Belloni et al. (2012, 2014). The solution requires either including any variables in the variable selection stage that also explain the variable of interest (e.g. the treatment variable) derived from a first-stage LASSO procedure, or orthoganolizing the variables of interest and potential covariates (Belloni et al., 2014). Following Manresa (2016), I do the latter. In practice, this ensures that covariates in w_{nt} do not confound the LASSO selection for non-zero entries in γ_{nihs}. Although ES_{nt} is independent of within-season covariates in w_{nt}, spurious correlations would result in mis-selection of non-zero entries. Results are robust to either method, and are little changed when following a "naive" single-stage approach.

The Double Pooled LASSO is introduced in Manresa (2016). In this application, the

objective of the procedure is to estimate an unbiased sparse matrix of cross-node effects of energy storage for every hour-season combination, Γ_{hs}, where each entry γ_{nihs} is the price effect at node j of 1 MW of energy storage at node i in hour h and season s. It addresses the model selection omitted variables bias via the "partialling out" method of Belloni et al. (2012), which

The Double-Pooled LASSO is estimated in the following three stages:

The first stage regresses w_{it} and λ_{it} on ES_{nt}:

$$\begin{aligned} w_{it} &= \rho_i ES_{nt} + \epsilon_\rho \\ \lambda_{it} &= \phi_i ES_{nt} + \epsilon_\phi \end{aligned}$$

and generates residuals $\tilde{w}_{it}$ and $\tilde{\lambda}_{it}$, orthogonal to ES_{nt}:

$$\begin{aligned} \tilde{w}_{nt} &= w_{nt} - \hat{\rho}_n ES_{nt} \\ \tilde{\lambda}_{nt} &= \lambda_{nt} - \hat{\phi}_n ES_{nt} \end{aligned}$$

In the second stage, a consistent estimate of θ_n, $\hat{\theta}_n$ is generated using $\tilde{w}_{nt}$ and $\tilde{\lambda}_{nt}$:

$$\tilde{\lambda}_{nt} = \theta_n \tilde{w}_{nt} + \epsilon_{\theta_n}$$

and $\hat{\lambda}_{it}$, orthogonal to $\tilde{w}_{it}$ is generated:

$$\hat{\lambda}_{nt} = \lambda_{nt} - \hat{\theta}_n \tilde{w}_{nt}$$

$\hat{\lambda}_{nt}$ is a consistent estimate of prices that is orthogonal to the variation in w_{it} that is uncorrelated with ES_{nt}. Thus, the effect of w_{it} is accounted for, but because the effect of w_{nt} that is removed is orthogonal to ES_{nt}, LASSO is not subject to model selection omitted variables bias — the covariance between ES_{nt} and λ_{nt} remains intact.

Finally, the LASSO estimator is used to estimate Γ_{hs}, the $n \times i$ matrix of cross-node effects on price at node n from storage at node i for hour h and season s. The LASSO procedure estimates each row, i, of the matrix separately:

$$\underset{\boldsymbol{\gamma}_{nhs}}{\arg\min} \sum_{t=1}^{T} \left(\hat{\lambda}_{nt} - \sum_{i}^{N_{ES}} \gamma_{nihs} ES_{it} \right)^2 + \pi \sum_{i}^{N_{ES}} |\gamma_{nihs}| \phi_{in},$$

where π is the LASSO penalty term, ϕ_{in} is the penalty loading, and γ_{nihs} is the hour-season specific effect of node i on node n, which is assumed to be sparse.

Penalty loadings, ϕ_{ij} are selected through an iterated estimation process described in Belloni et al. (2012). The penalty weights allow for sharp properties of the LASSO estimator, even in the presence of dependence in errors in a time-series setting (Manresa,

2016). The LASSO is estimated by adapting Christian Hansen's "shooting algorithm".[26] π, the penalty term, is calculated via a simulation process.

To remove the shrinkage bias inherent in LASSO feature selection (owing to the penalty term), a final OLS estimation is done for each n using nodes i identified as having $\gamma_{nihs} > 0$, setting aside all storage nodes designated by LASSO as being zero. Let this set be denoted $\mathcal{L}_{LASSO}$. This *Double Pooled LASSO* estimator eliminates shrinkage bias (see Manresa (2016); Belloni et al. (2014)) in estimating 10:

$$\lambda_{nt} = \sum_{s=1}^{4}\sum_{h=1}^{24} \beta^1_{nhs} ES_{nt} \times HR_h \times SEASON_s + \beta^2_{nh}\lambda^{LAP}_{l(n)t} + \theta^1_n w^1_{nt} + \theta^2_{nh} w^2_{nt} + \sum_{i\neq n}^{N_{ES}} \mathbb{1}\left(\gamma_{nihs} \in \mathcal{L}_{LASSO}\right) \gamma_{nihs} ES_{it} \times HR_h \times SEASON_t + \delta_{nhswy} + \varepsilon_{nt} \quad (17)$$

Proofs and methods for estimation are provided by Manresa (2016).

[26] http://faculty.chicagobooth.edu/christian.hansen/research/#Code