Estimating the Congestion Benefits of Batteries on Electricity Grids when Network Connections are Unobserved

A. Justin Kirkpatrick *

January 13, 2025

Abstract

I estimate the grid impacts of 180MW of battery storage capacity operating in California during 2009-2016. I find that one megawatt of energy storage decreases evening peak prices by up to $1.87/MWh at the pricing node where the storage is installed. Off-peak prices increase only slightly, implying a benefit to ratepayers of $79,942 per year. I use a Double Pooled LASSO-based estimator to uncover the unobserved network structure and thereby estimate the price effects at other locations across the grid. Overall, I find early energy storage mandates in California are only partially justified by ratepayer benefits. These patterns likely hold in any grid setting with inelastic peak supply relative to off-peak supply.

*Michigan State University, Department of Economics, 110 Marshall-Adams Hall, East Lansing, MI 48824 (email: jkirk@msu.edu). I thank Lori Bennear, Steve Sexton, Chris Timmins, Bryan Bollinger, Frank Wolak, Marty Smith, Michael Caramanis, Jackson Dorsey, Catie Hausman, Soren Anderson, participants at Camp Resources and the 2022 AEA annual meetings for comments and suggestions, as well as Ray Hohenstein at AES for providing context. Generous support for this research was provided by the Alfred P. Sloan Foundation Pre-doctoral Fellowship on Energy Economics, awarded through the NBER.

Key words: electricity grid, battery storage, renewables

JEL classification numbers: Q41, Q48, L14, H23

1 Introduction

Since the time of Edison, the absence of economical electricity storage has raised the cost of reliable electricity supply. Compelled to have supply equal demand at all moments in time, utilities must invest in seldom-used generation capacity characterized by low capital costs and high marginal costs. These "peaker" plants are called up to meet daily and seasonal peak demand with some operating for just a few hours on a handful of days each year.

The inability to store electricity is becoming more problematic amid substantial investments in intermittent renewable generation capacity that are partly a response to federal and state policies intended to achieve environmental and energy security objectives. By 2021, there were more than 2.8 million household solar installations across the United States, 118 GW of utility-scale solar, and 138 GW of wind (Bloomberg Finance, 2023). This renewable capacity cannot be dispatched by generators or grid operators. It is, instead, governed by sunlight or wind, which vary within days and across seasons, both predictably and unpredictably. This intermittency imposes an additional burden upon the grid operator who must ensure dispatchable generation meets demand net of renewable supply. Moreover, as renewable penetration grows, it can be expected to crowd out dispatchable baseload coal and natural gas due to its near-zero marginal costs of generation. 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., 2021).

Though battery storage capacity grew by 3.6GW per year in 2021, reflecting $5B in annual

investment (Bloomberg Finance, 2023; GTM Research, 2017), 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 is expected to facilitate 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 because wholesale markets are commonly settled by uniform price auctions. Batteries also lower costs associated with grid congestion. Congestion induces sub-optimal production because electricity trades across areas are constrained by transmission capacity. 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 benefits are not appropriated by private storage operators, whose returns from arbitraging high and low prices are expected to decline in storage capacity.

This paper examines the effect of storage-induced congestion relief on wholesale electricity prices in California, one of the largest markets for energy storage in the world with over 180 MW of installed capacity as of 2016. Of this 180 MW, 120 MW is pumped hydro storage that unexpectedly went offline in 2015 and nearly 60 MW is battery storage. Specifically, I develop an empirical model of electricity prices that accounts for intermittent generation and congestion. I estimate price changes using the staggered installation of 17 unique 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-2017). 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 benefits 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-megawatt (MW) increase in storage capacity reduces afternoon peak prices by up to $1.87 in the day-ahead market, primarily in the late-afternoon and early evening, coinciding with the period of highest electricity net demand. Meanwhile, off-peak prices move only slightly. Thus, I estimate that the annual cost of serving load falls by $79,942 per 1 MW increase in storage capacity. This annual grid benefit from congestion relief 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). Thus, these grid benefits are approximately 38% to 161% 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. 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 wind, solar, and electric vehicle integration (Woo et al., 2011, 2016; Novan, 2015; Bushnell and Novan, 2021; Craig et al., 2018; Wolak, 2016, 2018; Liski and Vehviläinen, 2020; Reguant, 2019; Jha and Leslie, 2019; Cicala, 2022; Petersen et al., 2022). While previous papers (Butters et al., 2021; Karaduman, 2021) model the market equilibrium changes associated with the addition of grid-scale batteries, and (Carson and Novan, 2013) develops a model that

considers price changes, no previous paper has estimated the price effect of energy storage directly as in this paper, though price effects are integral components of nearly all models of the storage market. The engineering literature considers their market impacts using simulation methods, focusing almost entirely on the private returns of storage investments and the battery properties that drive potential profits (Bradbury et al., 2014; Fioravanti et al., 2013; Nottrott et al., 2013; Walawalkar et al., 2007; Hittinger et al., 2012). Only 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. Lueken and Apt (2014) estimate potential savings of $4 billion in the PJM grid that covers Mid-Atlantic and Northeastern states.

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 (Burlig et al., 2016; Mercadal, 2022; Cicala, 2022; Prest et al., 2023). 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.

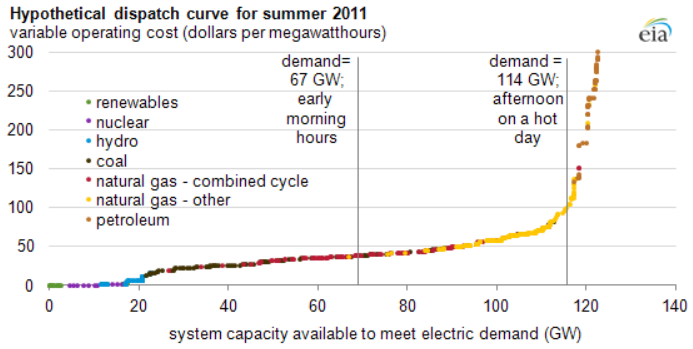
This paper proceeds as follows. Section 2 introduces a conceptual framework of spatial electricity price determination. Section 3 discusses data and empirical approach. Section 4 presents empirical results. Section 5 discusses results. Section 6 concludes.

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, as is shown in Figure 1. 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.

Figure 1: **Representative dispatch curve**. The merit-order dispatch curve is generated by ordering generators by cost. Source: Energy Information Administration



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.

In centrally dispatched energy markets such as those managed by the California Independent Systems Operator (CAISO) and by PJM in the northeast and mid-atlantic states, energy withdrawals are priced at the network node from which the withdrawals are made. A *node* is a physical substation or transmission point where generators may inject power, or load-serving entities (LSE's) may withdraw power. Nodal prices vary spatially and temporally (typically in 5 to 60 minute increments), and reflect the cost of withdrawing one additional unit of energy at that node, and injecting one unit of energy at a reference node.[1] Nodal prices determine the revenue that generators receive and the amount paid by LSEs. They are determined by the solutions to a constrained cost-minimization problem in which the grid operator must secure, via a uniform price auction, sufficient generation to meet node-level demand in every instant. The grid operator is constrained by transmission line capacity. Prices and generation change according to a non-linear Karush-Kuhn-Tucker (KKT) system of equations where slack conditions are on either generation constraints (e.g. prices change as higher-cost generation is required) or transmission constraints (e.g. prices change as lower-cost generation is unable to serve certain node-level demands) (Bohn et al., 1984).

Constrained optimization is a common tool in the economics literature, and is frequently used to generate equilibrium conditions and to define optimal production or consumption decisions. Here, the constrained optimization is not simply a convenient model for representing the grid. In the CAISO, network dispatch is literally determined by a numeric solution to the constrained optimization problem run hourly by CAISO staff. Rather than a represen-

[1] All withdrawals must have a concurrent injection to maintain the equality between supply and withdrawal. In practice, a withdrawal is offset by an increase in supply from a generator located on the grid, though not necessarily at the reference node. Both withdrawal and injection are priced relative to the reference node; the choice of reference node is arbitrary and does not affect nodal price levels.

tation of the optimal dispatch, a constrained optimization *is* the dispatch process. As such, I rely on the properties of the constrained optimization — specifically the shadow values of constraints that directly define the "locational marginal prices" or LMP's — to develop intuition about the effects of storage on prices.

In this section, I develop a model of electricity pricing first under a single convex supply curve, then under congestion in a multi-node network. The model draws a direct connection between the physical properties of electricity transmission that determine power flow and the constraints that determine nodal prices. I illustrate the price determination process and the interplay of these constraints and storage with numeric examples in Appendix C. 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. Nodal price is determined by the generators' supply bids and by transmission constraints that may preclude least-cost dispatch to serve some nodes. The effects of battery storage on nodal prices ultimately varies by congestion status.

2.2 Generation in a Single Node Market

To illustrate these price effects, consider first an uncongested market wherein a single, fixed set of generators serves demand. Let $p(Q) \geq 0$ be the marginal cost of energy at total quantity demanded Q. I assume the marginal cost of energy is increasing at an increasing rate in Q, i.e.

$$\frac{dp}{dQ} \geq 0 \tag{1}$$

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

Equation 2 is a consequence of merit-order dispatch and generator technologies, as shown

in Figure 1. Peaking plants used to serve periods of high demand are less capital intensive, but are less efficient and thus more expensive per unit generated.

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 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 $Q^L + E$. 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, and $p(Q^H - E) \leq p(Q^H)$. Not only is $p(Q^H) \geq p(Q^L)$, creating a potentially profitable private opportunity for arbitrage (labeled C in Figure 2), but also:

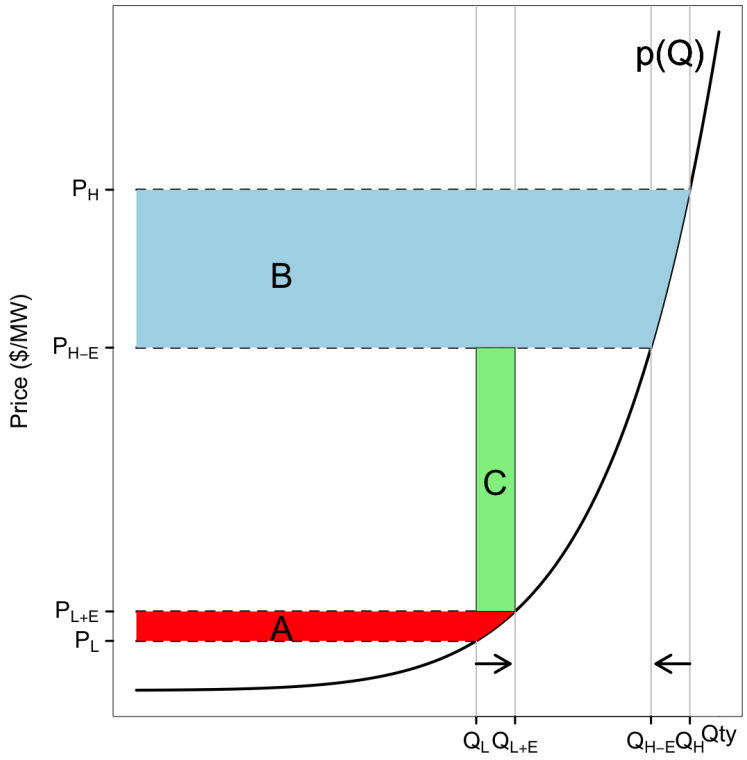
$$\frac{dp(Q^H)}{dQ} \geq \frac{dp(Q^L)}{dQ}, \tag{3}$$

due to convexity of the supply curve. Hence, the price changes from injection to and withdrawal from the grid may not be symmetric. For strictly convex supply, storage lowers marginal costs at peak more than it raises them at low demand.

While the energy storage operator profits per unit of energy stored, equal to the difference in $p(Q^H - E)$ and $p(Q^L + E)$, load serving entities (LSEs) face decreases in peak prices that are expected to outweigh increases in nadir prices. Because infra-marginal generators (e.g. baseload generators) receive the same prices as those on the margin, batteries reduce the price of every unit sold during peak periods and increase prices of every unit sold during the nighttime nadir. The gain to inframarginal generators (and cost to LSEs) from nadir price increases is depicted in area A in Figure 2, while the gain to LSEs from peak price reductions is depicted as region B. Importantly, because the reduction in peak prices is at least weakly greater than the increase in nadir prices and because $Q^H > Q^L$, peak cost savings to LSEs

dominate nadir cost increases, i.e., area B is greater than area A.

Figure 2: **Price effects of a discharge of energy from energy storage during daytime peak and nighttime nadir times**. Q is quantity of energy, and P is price. Area C is the private benefit accruing to the energy storage operator. Area B is the surplus gained by the LSE (load serving entity; utility) through the price effect during peak hour discharge. Area A is the cost to the LSE through the upward price effect during off-peak hour charging



These total net benefits to LSEs are given by:

$$\text{Total Benefits} \quad \approxeq \quad Q^H \frac{dp(Q^H)}{dQ} - Q^L \frac{dp(Q^L)}{dQ} > 0 \tag{4}$$

Thus, storage is expected to affect a net transfer from infra-marginal generators to LSEs. Such savings to LSEs are passed onto retail electricity customers under cost-of-service regulation.

Net revenues for the battery owners are equal to C, the quantity stored multiplied by the difference between high and low prices, and do not include the net benefits $B - A$.

2.3 Congestion

In a multi-node network, Kirchoff's Law determines energy flows over each transmission line between nodes, or "edge" in network terminology. Edges between nodes have two properties: *susceptance* is the ease with which energy flows over an edge, and *capacity* is the maximum amount of energy that can flow over an edge. Given two different edges between two nodes, for example, a generation node and a withdrawal node, electricity flows in proportion to the susceptance of the edges.[2] If one edge has a susceptance of 2, and another has a susceptance of 1, then $\frac{1}{3}$ of the flow between nodes will flow over the edge of lower susceptance, and $\frac{2}{3}$ will flow over the edge of higher susceptance. If the capacity of each edge is identical, one edge will reach capacity before the other line, inducing a transmission constraint. Even though one edge has not reached capacity, no additional power can flow between these two example nodes, nor between any other nodes whose flow would include the capacity-constrained edge, in contrast to e.g. a highway network where congestion on one highway causes motorists to reroute to another highway. Kirchoff's law precludes choosing where electrons flow — the flow across an electrical grids' lines is determined *entirely* by the net injections and withdrawals at each node, the susceptance of each transmission line, and the transmission line capacity. The grid operator does not have the ability to adjust or alter the constraints in the short term.

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 steepness and convexity of the local

[2]The inverse of susceptance, *resistance* is sometimes used in power flow modeling

inverse supply curve at congested nodes. The marginal cost differential between the lowest cost generator available to serve a given node and the lowest cost generator available at a reference node defines the *congestion cost* of electricity at the given node.

2.4 Generation and Congestion Constraints as KKT

The grid operator chooses only the vector of generation to dispatch, $\boldsymbol{G}$, which represents the generation authorized to operate over N grid nodes indexed by n. The operator chooses $\boldsymbol{G}$ to minimize the cost of serving load subject to transmission and generation constraints:

$$\min_{G_n} \quad \sum_{n=1}^{N} p_n(G_n)G_n \tag{5}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} D_n = \sum_{n=1}^{N} G_n \tag{6}$$

$$\boldsymbol{\kappa} \times [\mathbf{G} - \mathbf{D}] \leq \bar{K} \tag{7}$$

Here, $p_n(G_n)$ are bids placed by generators for each time period stating the quantity and price at which they are willing to supply electricity at node n. Equation 5 is the total cost of serving load. Equation 6 states that the total amount generated must equal the total amount demanded, summed across all nodes D_n. In equation 7, $\bar{K}$ is the vector of capacities of the E lines that form the transmission and distribution network. $\mathbf{G} - \mathbf{D}$ 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 edges e between nodes n (see Appendix C for power flow model definitions, details, and illustrative examples).

The grid operator determines $\boldsymbol{G}$ by solving the Lagrangian where λ is the shadow value of relaxing the system-wide power flow constraint by one unit, and μ is a vector of length E that contains the shadow value of relaxing the transmission constraint for each edge e:

$$\mathcal{L} = -\sum_{n=1}^{N} p_n(G_n)G_n + \overbrace{\lambda(0 - \sum_{n=1}^{N} G_n - \sum_{n=1}^{N} D_n)}^{\text{System-wide power flow constraint}} - \underbrace{\boldsymbol{\mu}(\bar{K} - \boldsymbol{\kappa} \times [\boldsymbol{G} - \boldsymbol{D}])}_{\text{line flow (slack) constraints}}, \tag{8}$$

$$\tag{9}$$

In an uncongested market when no transmission constraints are binding, the solution implies values of 0 for every μ_e. Then, price is determined by λ and is equal to the bid price of the marginal generator in the uncongested merit order.

In a congested market, one or more constraints in equation 7 bind. As a solution to the KKT system of equations with slack conditions, some values of $\boldsymbol{\mu} \neq 0$ 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 edges would similarly share congestion prices according to the flow that each node would cause over e.[3]

The effect of storage on prices is decomposed by equation 8 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 of $\boldsymbol{\mu}$ can switch between non-zero and 0 with even small storage additions, as illustrated in Appendix C. 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.

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 and transmission constraints.

[3]See Appendix C for example calculations of congestion price.

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 edge d, 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.

2.5 Other Determinants of Price

In equation 8, 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 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 (Hogan and Ring, 2003). Or, a lower-cost plant may not be capable (due to design limitations) of ramping quickly to meet demand (Bohn et al., 1984). Locational market power may also lead to increased convexity of the supply curve, and thus greater price decreases with the introduction of storage. Binding congestion constraints effectively "island" a generator into its own market and provide an incentive for a generator to shade bids upwards when the operator expects to face little or no competition due to transmission constraints (Mercadal, 2022). Dynamic constraints may also allow for exercises of market power (Reguant, 2014; Jha and Leslie, 2019). 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. This approach is motivated by the unobserved nature of the grid. 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 avarege hourly wholesale electricity price in CAISO, which primarily consists of the value of the total generation constraint λ in equation 6. 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.

Specifically, price effects at the local node are estimated by:

$$\lambda_{nt} = \beta_1 \lambda_t^{LAP} + \sum_{s=1}^{4}\sum_{h=1}^{24} \beta_{hs} ES_{nt} \times HOUR_h \times SEASON_s + \theta_n^1 w_{nt}^1 + \theta_{nh}^2 w_{nt}^2 + \sum_{n=1}^{N}\sum_{y=2009}^{2017}\sum_{s=1}^{4}\sum_{h=1}^{24}\sum_{w=0}^{1} \delta_{nhsyw} \times NODE_n \times YEAR_y \times SEASON_s \times HOUR_h \times WEEKDAY_w + \varepsilon_{nt}, \tag{10}$$

where λ_{nt} is hourly (total) marginal price at node n and time t; ES_{nt} is contemporaneous MW of installed storage capacity at the node; w_{nt}^1 consists of measures of local solar generation and local relative temperature,[4] both of which vary by day and by hour; w_{nt}^2 consists of an indicator for the Aliso Canyon blowout,[5] and the 12-month rolling average precipitation, which may differentially affect nodal prices through hydropower supply. δ_{nhsyw} 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; and λ_t^{LAP} is a weighted average of prices at nodes other than n within the corresponding utility territory known as a Load Aggregation Point (LAP).[6] This term controls for the shadow value of the total generation constraint in equation 6, which focuses identification on congestion effects. The term also controls for any trends

[4]Local solar generation is measured as the product of solar capacity multiplied by the hourly solar irradiance in the territory of node n; local relative temperature is measured as the difference between the hourly temperature in node n and the leave-one-out average temperature

[5]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.

[6]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 6 plus the sum of the transmission constraints that affect node n.

that may occur during the season in which storage is commissioned. Interest centers on β_{hs}, 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. The identifying assumption of equation 10 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.

Using this rich set of fixed effects limits identifying variation to a very fine window within the node-hour-season-year. With few 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.

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 borrow an identification strategy from the Research and Development (R&D) literature that leverages a multi-step Double Pooled LASSO estimator

to uncover the structures of unobserved networks Manresa (2016).[7] 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:

$$\lambda_{nt} = \lambda_t^{LAP} + \sum_{s=1}^{4}\sum_{h=1}^{24} \beta_{hs} ES_{nt} \times HR_t \times SEASON_s + \theta_n^1 w_{nt}^1 + \theta_{nh}^2 w_{nt}^2 \\ \sum_{i \neq n}^{N_{ES}} \gamma_{inhs} ES_{nt} \times HR_h \times SEASON_s + \delta_{nhs} + \eta_y + \varepsilon_{nt}, \quad (11)$$

where λ_{nt} is the nodal price, λ_t^{LAP} is the utility-level price as before, and $\{\theta_n^1, \theta_{nh}^2\}$ is the marginal price effect of $w_{nt} = \{w_{nt}^1, w_{nt}^2\}$, which is comprised as in 10 of temperature and precipitation covariates, local (within node) hourly solar generation, and a dummy indicator for the Aliso Canyon blowout. Variables in w_{nt} are potential confounders that are included to avoid omitted variables bias and improve precision. Specifically, if storage is located in an area characterized by more frequent temperature spikes, then omitting these variables will attribute their effect to storage, biasing the coefficient of interest. Parameters of interest are the non-zero entries in γ_{inhs} which represent cross-node price effects: the effect of 1MW

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

of storage at node n on prices at node i.[8] A LASSO algorithm is used to identify non-zero entries in γ_{inhs}.

Estimating equation 11 by LASSO presents an additional source of potential omitted variables bias best characterized as *model selection omitted variables bias* (Belloni et al., 2012, 2014a,b). To obtain unbiased estimates of both the non-zero entries in γ_{inhs} 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 E. 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_{nt} in Equation 11. The second stage uses these versions of λ_{nt} and w_{nt} to estimate θ_n, then residualizes $\hat{\lambda}_{nt} = \lambda_{nt} - \hat{\theta}_n w_{nt}$. In the third stage, LASSO is used with equation 11, replacing λ_{nt} with $\hat{\lambda}_{nt}$, resolving the potential for model selection omitted variables bias. Finally, with non-zero entries in γ_{inhs} identified, OLS is used to estimate parameters.

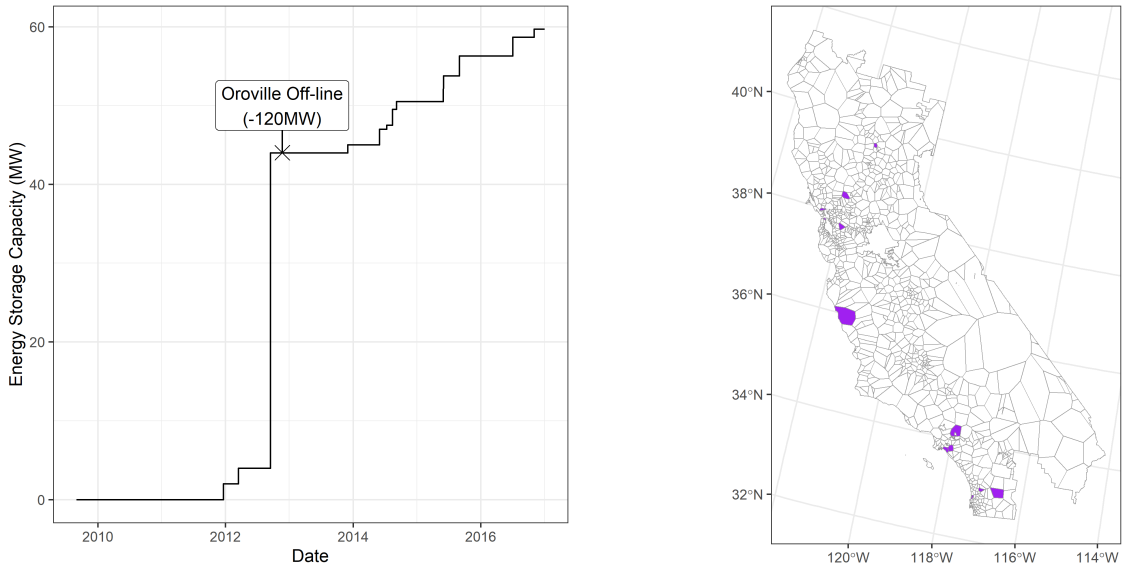
3.3 Data

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 temperature, precipitation, or the price of natural gas. This section describes the data sources in detail.

Commercial energy storage capacity in the CAISO area between 2009 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 180MW of energy storage was brought online or went offline, equivalent to approximately .4% of the average peak daily

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

Figure 3: (a) **Energy storage capacity added during study period**. Totals do not include storage introduced before September 2009. 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.



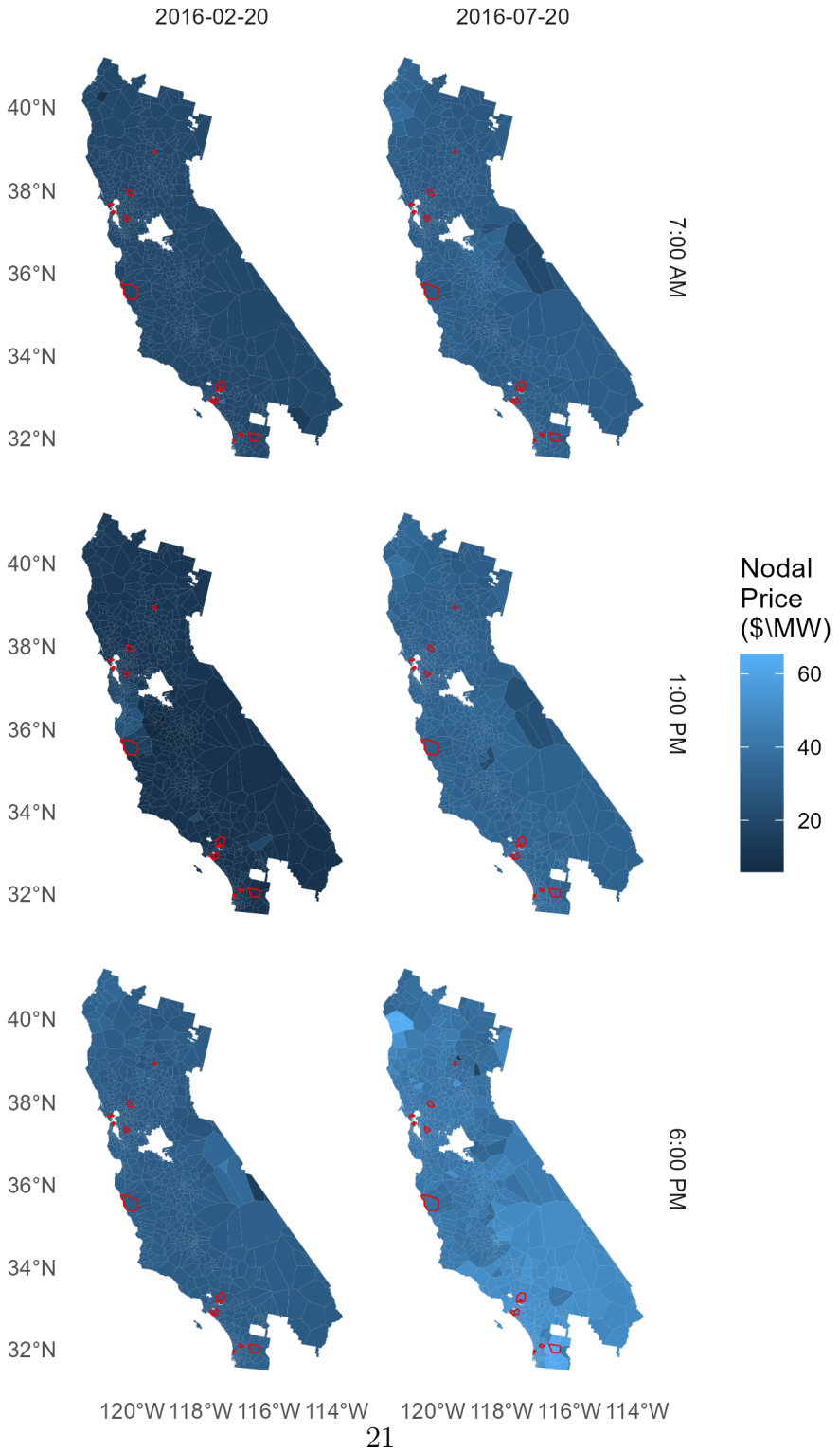
(a) (b)

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 20 unique storage capacity additions to the CAISO service area as two storage sites receive multiple storage capacity additions. Figure 3a shows the timing and locations of storage installations in California. Storage installations are concentrated in urban areas, likely to exploit local congestion-related price spikes. Capacity also exists in rural areas, and some capacity is located near utility-scale renewables.

I obtain Locational Marginal Prices (LMPs), λ_{nt}, for the day ahead market using the CAISO OASIS data portal for the period September 1, 2009 to December 31, 2016[9]. As

[9] http://oasis.caiso.com

Figure 4: **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.



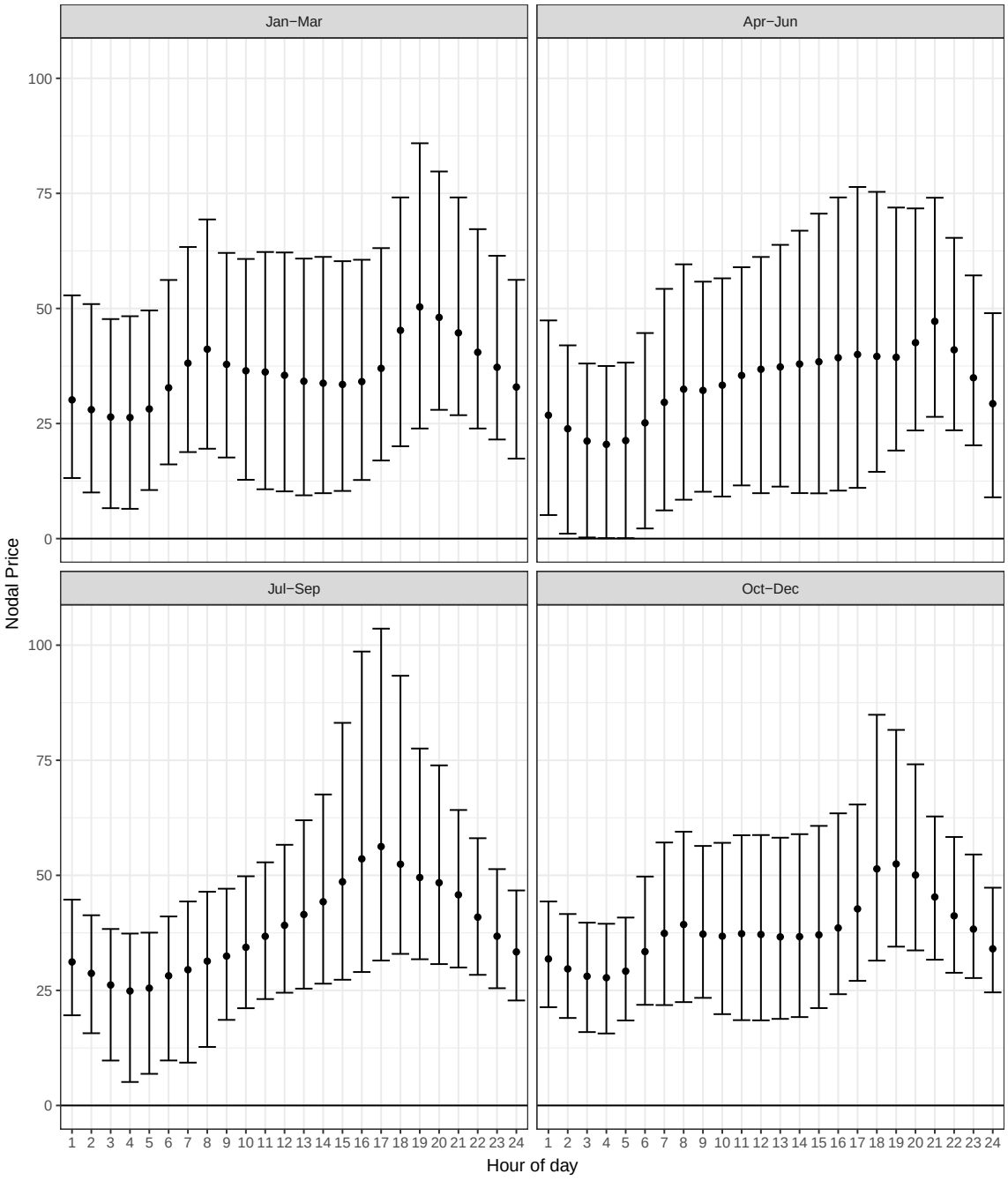
shown in Figure 5, seasonal average LMPs exhibit expected patterns—daily peaks occur during the hours of 5-9pm, and some seasons also exhibit a smaller morning peak. These peaks show the potential for arbitrage using energy storage. I also obtain each of the three Default Load Aggregation Point (DLAP) prices, one for each major investor-owned LSE in California. Figure 4 shows spatial variation in nodal prices for select hours and select seasons. Note that these prices differ from CAISO system price, as nodal prices include congestion whereas the CAISO system price does not.

When necessary for calculating changes in cost of serving load, hourly energy withdrawals at each node are imputed as the product of hourly total demand for each utility and respective, static Load Distribution Factors (LDFs) that report the share of utility demand 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 DLAP. 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 [10] and proximity to form aggregate clusters of nodes and energy storage. Proximity is determined by both physical distance and price path similarity. These clusters are called "nodes" throughout this paper. Polygons are generated around each node containing all spatial points that are closest in proximity to that node.

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

Figure 5: **Hourly Nodal Price** ($), λ_n, by Season; 95% CI



I then construct a proxy of local solar generation. 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). Conditional on solar capacity, solar generation is proportional to GHI.

To create a node-level measure 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 subregion 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 within-node hourly solar generation.

3.4 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-2017 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 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 10 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. Notably, the vast majority of storage in the GS dataset is privately owned and operated. 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. Because nighttime prices are largely uncongested and follow a single grid-wide price, an increase in the price spread at a particular node can only follow from an increase in peak congestion at that node. 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 B. 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

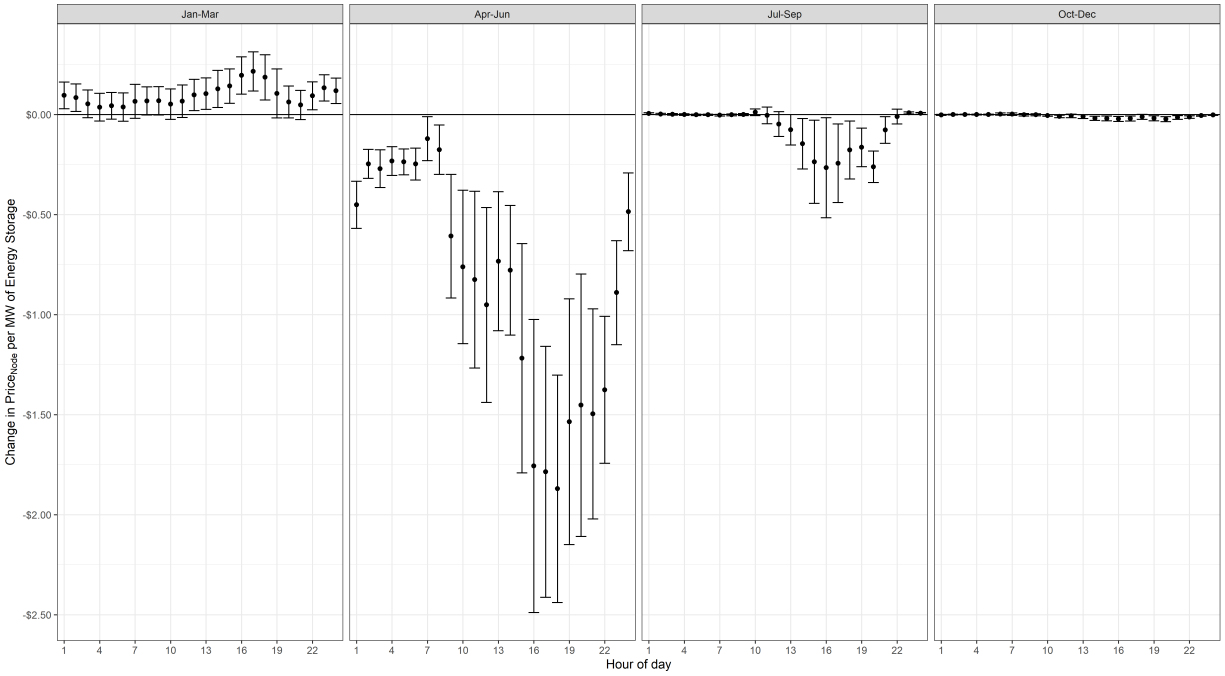
4.1 Own-node Effects

Point estimates of β_{hs} in equation (10) and their 95% confidence intervals are plotted in Figure 6. Driscoll-Kraay standard errors (Driscoll and Kraay, 1998) are two-way clustered at the node and hour-of-sample. Results for Spring (April-June) are noisy, but point estimates are largely negative during peak (afternoon) periods. When afternoon demand is highest on the CAISO grid, storage effects are statistically significant and reach as much as -$1.87 at the 6pm hour. During the 6pm-10pm shoulder hours, effects range from -$1.87 to -$1.37 at 10pm. These are precisely the hours in which declining solar generation coincides with increases in energy demand, forming the "duck curve." Effects for all hours are negative and nearly all significant. While theory predicts weakly increasing prices during off-peak times, the magnitudes of the off-peak effects are small. 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. In summer (July-September), when theory posits the price effect will be most pronounced because of high demand and the convexity of the supply curve, price effects are smaller, with a maximum price decline of -$0.26 for the 4pm and 8pm hours, which coincide with the highest-priced hours during summer as shown in Figure 5.

The January-March effects are largely small but positive and significant. Average price and quantity demanded are smaller during this season, though the overall effect is positive. In October-December, storage shows nearly no effect. Overall, effects are focused on afternoons in spring and summer, broadly consistent with the theory in section 2.

Coefficients for hours usually considered "off-peak" are close to zero or statistically insignificant, consistent with markets clearing at a flat and uncongested region of supply curves. Point estimates for the early morning nadir hours are small but positive in January-March and October-December, in line with the theory presented in Section 2. However, negative effects in the early morning hours in April-June contradict the theory model. While it is feasible that changes in dispatch resulting from the presence of energy storage on a node may alter dispatch at other hours, it is an unlikely explanation for these results. While estimates may be somewhat noisy, the afternoon pattern of the storage price effect holds, even if only relative to the early-morning price effects seen in April-June.

Figure 6: **Coefficient Estimates**: Equation 10 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. To estimate the magnitude of these total savings from a MW storage capacity, denoted TS, I multiply the point estimates in Figure 6 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 located at the node:

$$TS = \sum_{d=1}^{D} \sum_{n=1}^{N} \sum_{h=1}^{H} \beta_{hs} \times ES_{dn} \times \lambda_n \times LOAD_{hd}.$$

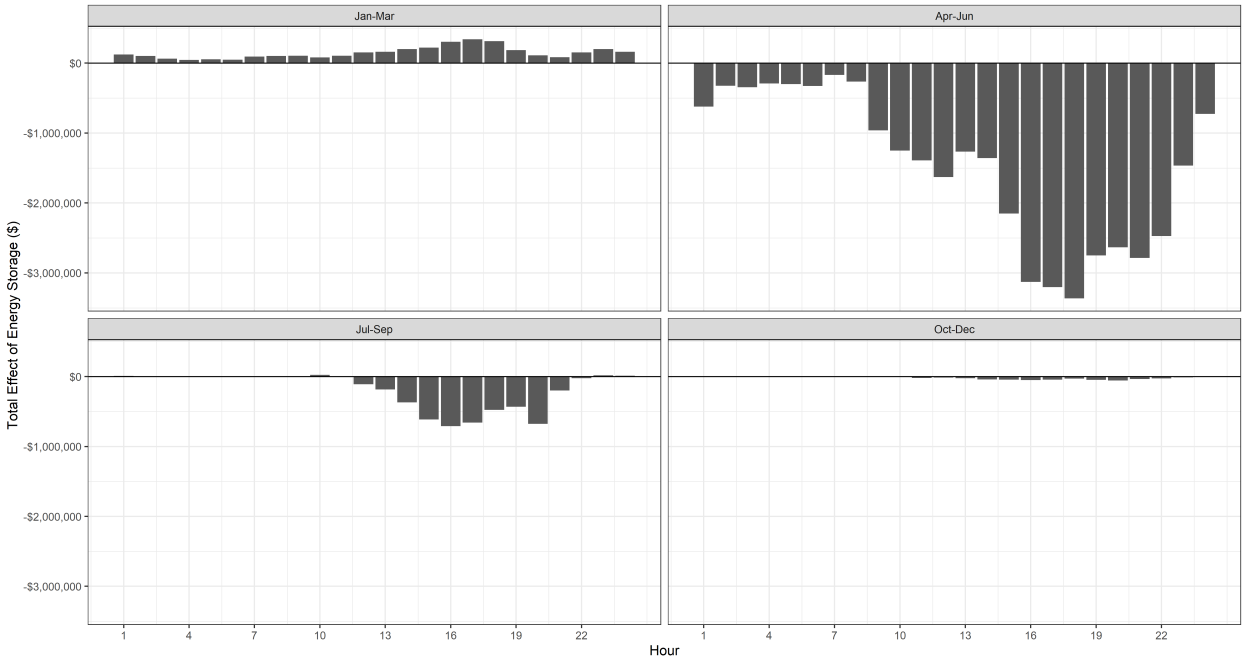
where β_{hs} is the price change estimated in equation 10, and d is the day-of-sample. Total load withdrawn at each node is not reported, though total grid load is reported. To calculate node-level load, I multiply total grid load $LOAD_{hd}$ by load distribution factors λ_n (the share of total load served, on average, by each node) to calculate hourly load at each node.

Calculations for total change in cost of serving load are described in appendix D and defined in equation 16.

The total own-node savings for the entire study period (2009-2016) is $36.5 million. The effects of storage on load-serving costs are greatest in the second and third quarters of the year—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 3a), I calculate an average total savings per hour, per megawatt of storage ("capacity-hour"). I divide the total benefits of energy storage by the total number of MW of capacity-hours on the grid to generate a public benefit (reduced cost of serving load) of $7.02 per MW of capacity per hour on grid.

Over one year (8,760 hours), the annual own-node benefit of 1MW of energy storage is $\$7.02 \times 8,760 = \$61,514$. For perspective, a Tesla Powerwall, the most widely-known brand of behind-the-meter energy storage, has a capacity of 2kW, or .002MW, yielding cost savings from own-node changes of $123.02 per year.

Figure 7: **Aggregated own-node inframarginal effects of all storage over study period**.



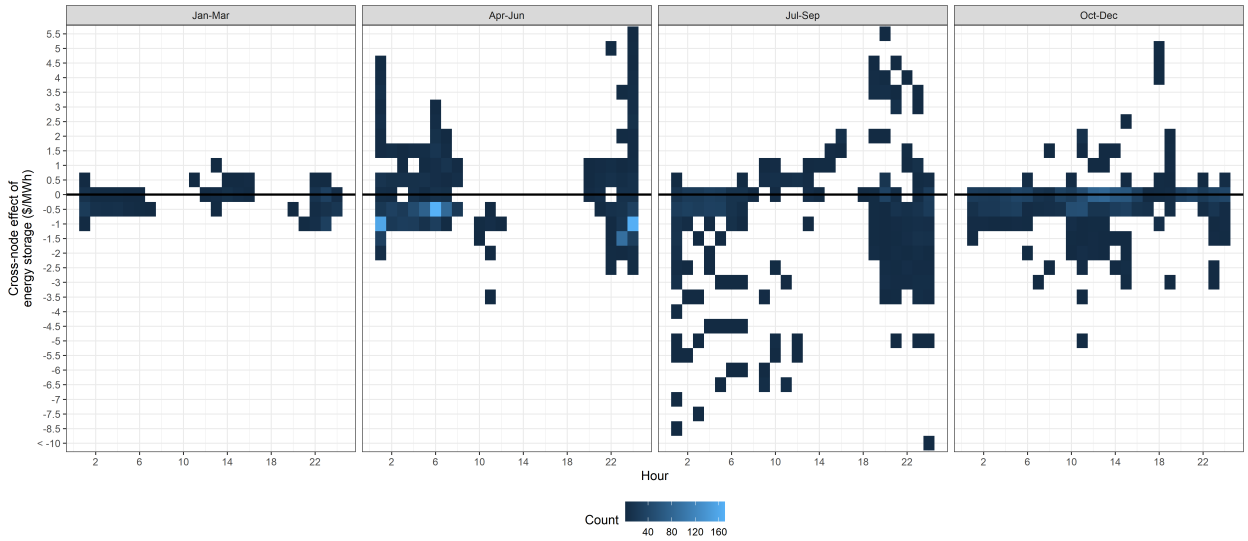
Another way to compare ratepayer benefits is to state them as a fraction of the private benefits that can be gained from daily arbitrage of prices. This is of particular importance for understanding the magnitude of ratepayer benefits not captured by private operators. A high amount of unappropriated (public) benefit justifies mandates or subsidies, as are used in California. Arbitrage opportunities decline in storage capacity though ratepayer benefits of storage do not. Hence, internalizing the effect may become increasingly important.

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 benefits accruing to ratepayers (and not captured by the operator) of $61,514 amount to 29-124% of these private revenues.

4.2 Cross-node Effects

Results from estimating Equation 17 are plotted in Figure 8. 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 8: **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.



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 gen-

eration. However, on the non-congested side, prices are lower as fewer areas can be served by generation in the uncongested area.[11] 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.

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 (8-11pm). 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 are observed during morning hours in October-December, and during the evening hours during January-March. Note that Figure 5 shows a second daily peak occurring during the morning 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 ($\gamma_{inhs} \in \mathcal{L}_{LASSO}$) may not necessarily be spatially contiguous as transmission and distribution infrastructure is not a perfect lattice. However, linked nodes should be close. For example, it is unlikely that a node in San Diego influences prices at the northern border of California. Select linked nodes are depicted in Figure 9. As these figures show, linked nodes are generally nearby, lending validity to the LASSO selection procedure. However, distant nodes are not entirely eliminated. Appendix Figure A.1 plots the 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

[11] 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.

lies within 100 km, and distant nodes tend to exhibit variance in sign, suggesting selection is subject to noise.

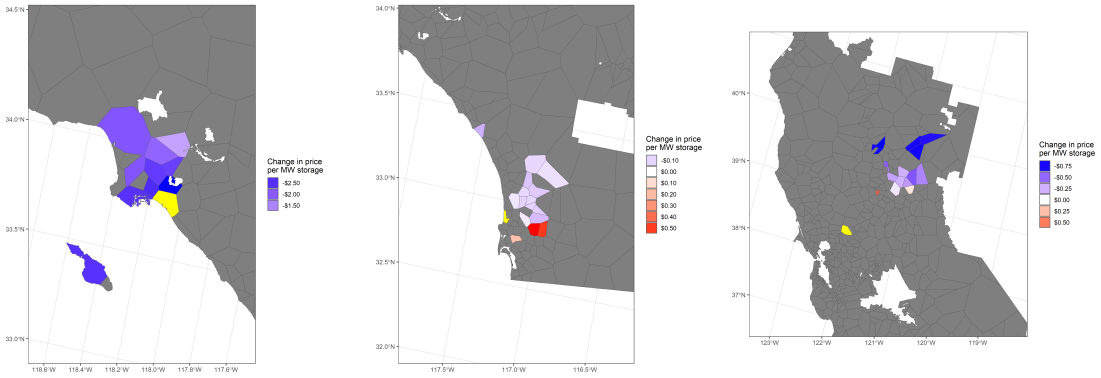


Figure 9: Examples of cross-node effects. Polygon in yellow represents site of storage node. (Left) storage in Long Beach, CA. (Middle) storage in Torey Pines, CA. (Right) storage in Vacaville, CA.

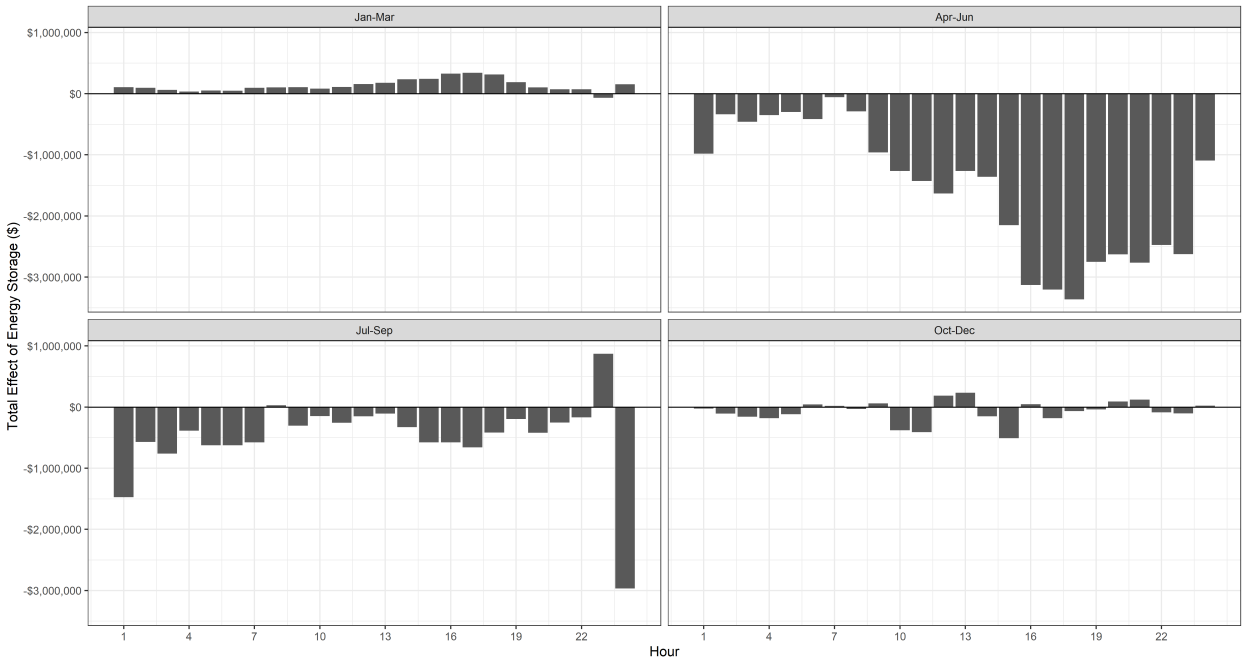
These cross-node results indicate that price effects of energy storage extend beyond the capacity node to other nodes in the network. The sign of these effects is predominantly negative. When aggregated, cross-node effects generate total savings (e.g. decreases in cost of serving load) of approximately $10,937,431, or approximately 23% of the total benefits realized at the own-node. This is approximately 9% to 37% of private revenues. Cross-node effects are heterogeneous across the 18 storage installations. In some areas, effects are concentrated in the area around the location of storage as in Figure 9 (left and middle). In others, effects appear to be widespread across a region as in Figure 9 (right). This heterogeneity highlights the importance of considering the local network topology and congestion patterns in calculating the benefits of energy storage to ratepayers.

4.3 Total Cost of Serving Load

Figure 10 shows the season-by-hour aggregated effects. Values represent the change in the total cost of serving load over the entire study period 2009-2016. Small positive amounts are

show in the first quarter, while the vast majority of savings occurs in the second and third quarters, representing the spring and summer months.

Figure 10: Aggregated own- and cross-node effects of all storage over study period



5 Discussion

Ratepayers benefit from storage when the cost of serving load decreases, as is the case with storage own-node and cross-node effects. These ratepayer benefits 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 ratepayers recover in transfers an amount that justifies the cost?"

Ratepayer benefits from own-node price responses are estimated to be $61,514/MW per year, and cross-node benefits are 23% of that amount, $18,428/MW per year, totalling

$79,942 per MW of capacity, per year. The average ratepayer benefit is $9.13 per hour. 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. Arbitraging energy prices in CAISO energy or participating in the ancillary services market generates *private* revenues of $45,500 to $208,500 per year, per MW (Penna et al., 2016).

Standing on their own or in tandem, the total public and private benefits from energy storage only partially justify the estimated cost of storage in 2016. Mandates such as AB 2514 can potentially still be justified by learning spillovers and first-mover advantage, and California has historically been a first-mover in renewable and low-carbon energy policy. This paper estimates only the ratepayer benefits, 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 Ratepayer Benefits 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 benefits. Private investors will not consider the ratepayer benefits that cannot be captured. In fact, higher ratepayer benefits will *decrease* private revenues as the magnitude of the diurnal arbitrage opportunities (peak prices) are reduced. Therefore, a fiscal public goods problem arises, and a policy intervention is necessary to ensure that the ratepayer benefits are accounted for in the investment decision.

To this end, I calculate the ratepayer benefits as a fraction of the total private benefits from arbitrage. If ratepayer benefits are large relative to private benefits, then policy in-

tervention may be justified. If the benefits are small relative to private benefits, then the unsubsidized outcome will result only in storage that can be justified by private benefits, and no policy is necessary.

Results show that ratepayer benefits are a significant portion of benefits from storage. Own-node effects total 29-124% of the total private benefits from arbitrage, and cross-node effects total another 9-37%, summing to 38%-161%. Therefore, ratepayers gain 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 reduction 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 to ratepayers 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 public ratepayer benefits 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. In California, the California Public Utilities Commission examines all rate change requests, and receives input from ratepayer advocates in determining its approval of a rate change. 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 benefits, as the ratepayer is bearing the capital costs.

5.2 Relevance to Other Grids

California has historically been a first-mover in renewable and low-carbon energy policy, and retains this position in mandating storage under AB2514. 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 average ratepayer 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). For example, the spatial distribution of "hot spots" in Figure 4 illustrates this issue in the Northwest corner of the state during summer peak hours. Figure 4 also illustrates the larger transmission-level congestion constraints — the lower-right panel shows a clear gradient between the Southern and Central regions around 34°N. This transmission-level congestion has persisted for years in the CAISO. Note that, in this figure, the local "hot-spot" in the Northwest corner is of greater magnitude than the longstanding South-Central gradient, although the Northwest "hot-spot" is in far fewer nodes, and covers nodes with much lower total volume. For 2016, the CAISO-estimated total cost of congestion was $99 million (CAISO, 2016). This stands in contrast to ERCOT, the Texas grid, where transmission constraints between wind-generation rich areas in the Northwest and load centers in the East drive congestion costs (LaRiviere and Lu, 2017).

5.3 Does Storage Facilitate Renewables?

This paper demonstrates the role of storage in "facilitating integration of renewables," the goal of AB2514. Results indicate that storage reduces local nodal prices at the times in which local solar generally imposes high costs, i.e., late-afternoon and evening hours during spring and summer. Solar generation has two observable price effects throughout the day. First, it decreases the grid-level cost of electricity during the late afternoon hours as generation is high relative to demand, which has the effect of depressing prices due to low net load (demand net of zero-marginal-cost renewables). Second, during both the morning ramp-up of solar generation and, especially, the evening ramp-down as the sun sets, more expensive fossil-fuel generation is necessary to accommodate the fast ramp. Bushnell and Novan (2021) quantify these effects, and identify more expensive gas turbine (GT) generation as responding to the evening ramp in lieu of the more efficient combined cycle gas turbine fleet (CCGT). Therefore, over the day, solar reduces total costs, but accommodating the morning and evening ramp requires more expensive generation, increasing prices during those hours.

Energy storage reduces local congestion costs during these critical hours, as shown in Figures 6 and 8. Thus, storage appears to facilitate integration of solar on the grid. By reducing local congestion, storage makes quickly-dispatched, lower cost plants available to follow changes in solar generation, alleviating system-wide cost increases associated with solar capacity (Bushnell and Novan 2021; Bierman 2018). This follows the theory in Section 2, where price effects are timed to peak prices, which in turn are a result of net demand and line congestion.

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. 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. The effect during spring evening hours reaches as much as -$1.87 (-2.2%) per hour, per MW, while summer effects are smaller at up to -$0.26. These price effects generate an average $79,942 in cost savings per year, per MW to load-serving entities that are typically passed on to consumers through rate regulation. The 180 MW of energy storage capacity in California is estimated to have reduced energy costs across the grid by $47.45M, 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 public ratepayer 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

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

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

A. Belloni, V. Chernozhukov, and C. Hansen. High-dimensional methods and inference on

structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50, 2014b. ISSN 0895-3309.

E. 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.

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

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

K. Bradbury, L. Pratson, and D. 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.

F. Burlig, C. Knittel, D. Rapson, M. Reguant, and C. Wolfram. Learning from schools about energy efficiency. Technical report, 2016.

J. Bushnell and K. 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. A. Butters, J. Dorsey, and G. Gowrisankaran. Soaking up the sun: Battery investment, renewable energy, and market equilibrium. Technical report, National Bureau of Economic Research, 2021.

CAISO. Annual Report on Market Issues and Performance. 2016. URL http://www.caiso.com/Documents/2014AnnualReport_MarketIssues_Performance.pdf.

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

R. T. Carson and K. 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.

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

S. 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.

M. T. Craig, P. Jaramillo, B.-M. Hodge, N. J. Williams, and E. 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#.

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

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

R. Fioravanti, M. Kleinber, W. Katzenstein, S. Lahiri, N. Ton, A. Abrams, J. Harrison, and C. Vartanian. Energy Storage Cost-effectiveness Methodology and Results. Technical report, DNV KEMA Energy and Sustainability, Oakland, CA, 2013.

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

GTM Research. US Energy Storage Monitor: 2016 Year in Review and Q1 2017 Executive Summary. Technical report, Energy Storage Association, 2017.

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

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

A. Jha and G. Leslie. Dynamic costs and market power: The rooftop solar transition in western australia. Technical report, working paper, 2019.

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

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

F. Lambert. Teslas massive Powerpack battery in Australia cost $66 million and already made up to ˜$17 million. *Electrek*, 9 2018.

J. LaRiviere and X. Lu. Transmission Constraints, Intermittent Renewables and Welfare. 2017.

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

M. Liski and I. 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.

R. Lueken and J. Apt. The effects of bulk electricity storage on the PJM market. *Energy Systems*, 5(4):677–704, 2014. ISSN 1868-3967.

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

I. 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.

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

M. D. Penna, M. Yeung, and D. 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.

C. Petersen, M. Reguant, and L. Segura. Measuring the impact of wind power and intermittency. *Available at SSRN 4291672*, 2022.

B. C. Prest, C. J. Wichman, and K. Palmer. Rcts against the machine: Can machine learning prediction methods recover experimental treatment effects? *Journal of the Association of Environmental and Resource Economists*, 10(5):1231–1264, 2023.

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

M. 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.

R. Sioshansi, P. Denholm, T. Jenkin, and J. 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.

V. Viswanathan, K. Mongird, R. Franks, X. Li, and V. Sprenkle. 2022 Grid Energy Storage Technology Cost and Performance Assessment.

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

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

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

F. A. Wolak. The Evidence from California on the Economic Impact of Inefficient Distribution Network Pricing. Technical report, 2018.

C.-K. Woo, I. Horowitz, J. Moore, and A. 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.

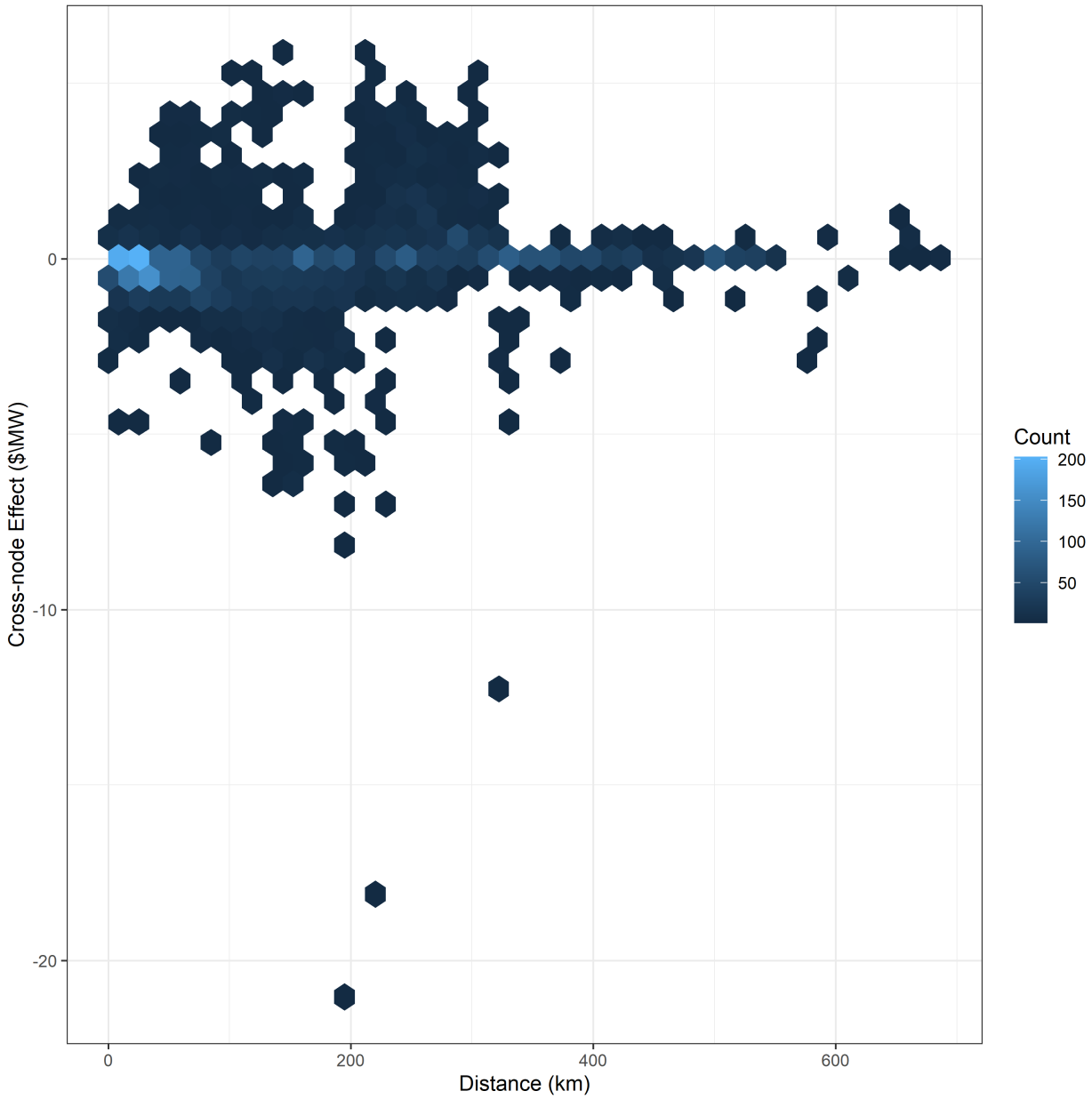
C.-K. Woo, J. Moore, B. Schneiderman, T. Ho, A. Olson, L. Alagappan, K. Chawla, N. Toyama, and J. Zarnikau. Merit-order effects of renewable energy and price divergence in Californias day-ahead and real-time electricity markets. *Energy Policy*, 92:299–312, 2016. ISSN 0301-4215.

figuresection tablesection

Appendix A Additional Figures

A.1 Cross-node effects of energy storage by distance

Figure A.1: **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 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 B.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 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 B.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 B.2 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

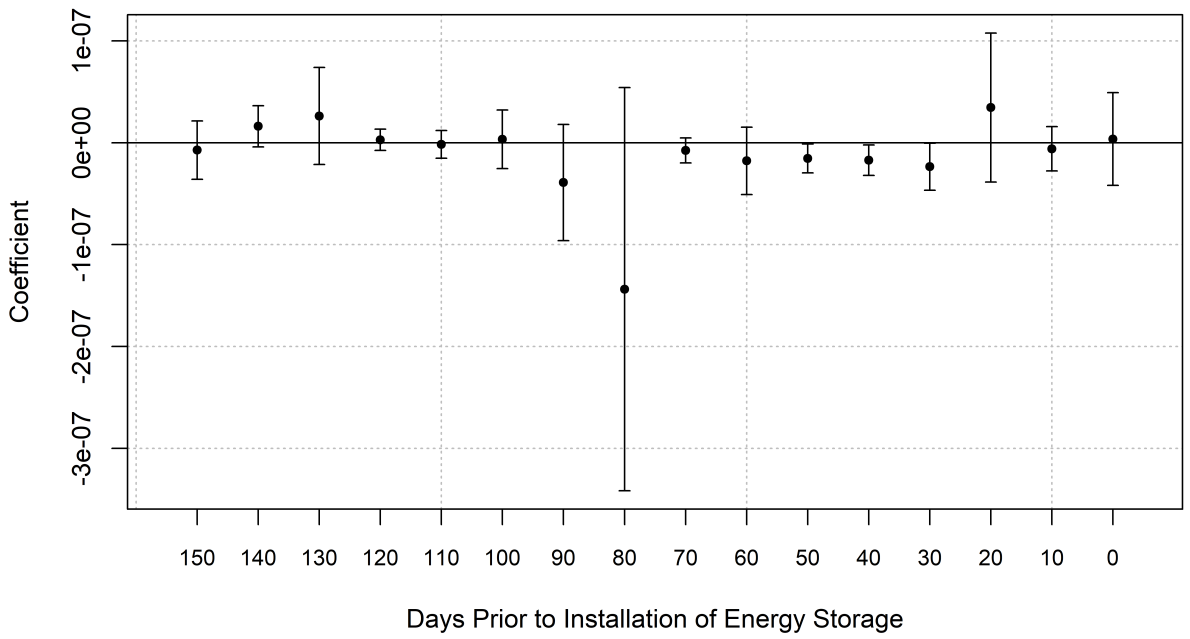
the peak price at that node, a profit-maximizing strategy for siting would be to choose nodes with large spreads, but that will not respond in price strongly to the additional supply. As noted in the above section, the price determination process embedded in the power flow model makes sudden, significant 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 B.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	-0.0019		-0.1168	
	(0.0030)		(0.2006)	
Median $90^{th} - 10^{th}$ decile		-0.0002		-0.0109
		(0.0039)		(0.2321)
Fit statistics				
Pseudo R^2	-0.00064	-3.5×10^{-6}	0.00475	2.51×10^{-5}
Observations	757	757	757	757

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

Figure B.2: **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.

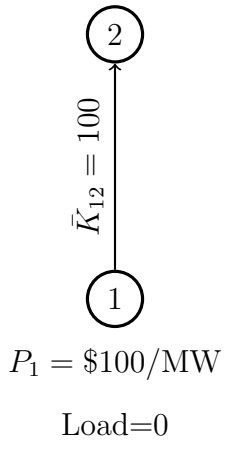
Appendix C Examples of Price Determination with Power Flow Models

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

C.1 Price determination in two-node grid

For illustrative purposes, consider the two-node grid depicted in Figure C.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 C.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 edge in the network, and one column for each node. Column n of $\boldsymbol{\kappa}$ represents the flow over each edge 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,[12] 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 $\boldsymbol{P}$. 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

[12] 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

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.

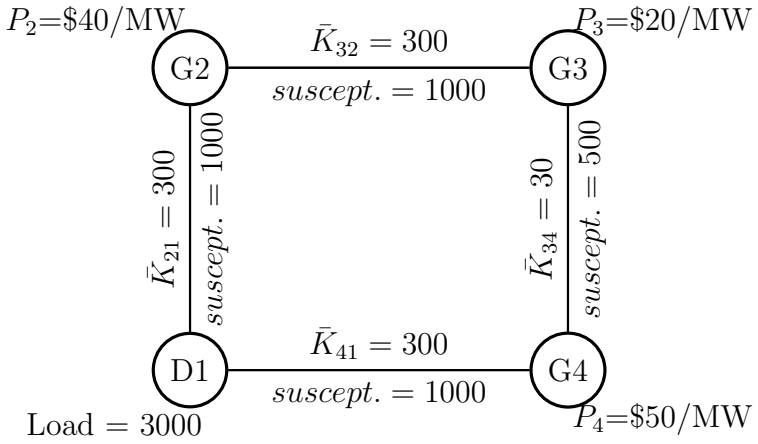
If storage is located at Node 2 which becomes congested during peak demand (e.g., in order to defer a transmission upgrade), then transmission constraints bind less frequently with the storage capacity, and the nodal price will be weakly closer to the grid price, even on days where grid-wide demand is 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.

C.2 Price determination in 4-node grid

A final example demonstrates price dependencies amid congestion. Consider the 4-node network shown in Figure C.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[13].

Figure C.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 edge K_{34}, and higher priced generation will be dispatched to meet demand at $D1$.



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 edges of the network.[14] 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 edge resulting from injecting at the reference node and withdrawing at node $n \in$

[13]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 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)

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

$\{D1, G2, G3, G4\}$. 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 ther 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 prices and net demand vector $\boldsymbol{D} = [300, 0, 0, 0]$, 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:

$$\boldsymbol{G} = \begin{bmatrix} 0 \\ 750 \\ 1000 \\ 1250 \end{bmatrix}, \boldsymbol{P} = \begin{bmatrix} 45 \\ 40 \\ 35 \\ 50 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 25 \\ 0 \end{bmatrix}$$

The least expensive generator is not serving load at all nodes. In fact, the generator with greatest supply in the solution is the most expensive. The constraint, μ times the matrix $\boldsymbol{\kappa}$ determines the price difference between the reference node, $D1$ and the other nodes. For instance, at $G3$, $0 + 0 + (25 \times 0.4) + 0 =$\$10 less at $G3$ relative to $D1$. *If storage at any node were to eliminate the congestion on e_{34}, the price at nodes $G2$, $G3$, and $G4$ would decrease.*

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, 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 D 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 edges 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 edge (transmission line) in the network. The $\{1, \ldots, E\}$ elements of $\boldsymbol{\kappa}_n$ represent the energy flow across each edge 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 3— 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, $\boldsymbol{Q}$:

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

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 edge 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:

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

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 benefits of a marginal unit of storage capacity are calculated as the sum of price changes over *all* hours of the day:

$$\text{Total Benefits} = s \times \left(\sum_{t=1}^{24} \frac{d\boldsymbol{\lambda}_t(\boldsymbol{Q_t})}{dES} \boldsymbol{Q_t} \right). \tag{16}$$

Appendix E 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, 2014a,b). 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., 2014a). 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_{inhs}. 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_{inhs} 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:

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

where π is the LASSO penalty term, ϕ_{in} is the penalty loading, and γ_{inhs} 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".[15] π, the penalty term, is calculated via a simulation process.

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

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. (2014b)):

$$\lambda_{nt} = \sum_{s=1}^{4}\sum_{h=1}^{24} \beta_{hs} ES_{nt} \times HR_t \times SEASON_t + \theta_n w_{nt} + \sum_{i \neq n}^{N_{ES}} \mathbb{1}\left(\gamma_{inhs} \in \mathcal{L}_{LASSO}\right) \gamma_{inhs} ES_{nt} \times HR_t \times SEASON_t + \delta_{nhs} + \eta_y + \varepsilon_{nt} \quad (17)$$

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