American Economic Journal: Economic Policy 2023, 15(2): 292–330
https://doi.org/10.1257/pol.20200234

Can Forward Commodity Markets Improve Spot Market Performance? Evidence from Wholesale Electricity[†]

By Akshaya Jha and Frank A. Wolak*

Forward markets are believed to aggregate information about future spot prices and reduce the cost of producing the commodity. We develop a measure of the extent to which forward and spot prices agree in markets with transaction costs. Using this measure, we show that day-ahead prices better reflect real-time prices at all locations in California's electricity market after the introduction of financial trading. We then present evidence suggesting that operating costs and input fuel use fell after the introduction of financial trading on days when the nonconvexities inherent to the production and transmission of electricity are especially relevant. (*JEL* D23, D24, G13, L94, L98, Q48)

There is a growing empirical literature demonstrating that forward prices provide important information about future spot prices[1]. A number of scholars have argued that increasing forward market liquidity can also reduce the cost of producing the commodity (Working 1953; Gray 1964; Cox 1976). This is because the suppliers of goods that require incurring sunk costs to produce often use forward market outcomes to decide whether to incur these sunk costs.

For example, an iron ore mine might sign a forward contract with a steel manufacturer that finances the sunk cost of opening a new iron ore seam to serve this demand. Finding the least-cost source of additional iron ore for each steel manufacturer can be costly if there are many spatially distinct iron ore mines. Introducing purely financial participants into the forward market for iron ore can increase the likelihood that the least-cost mix of suppliers is found to serve all steel manufacturers. This is because financial players operating in the forward market earn the difference between the price at which they sell iron ore to a steel manufacturer and the forward price they pay to a mine owner for that raw material. Assuming there are many spatially distinct mines and steel manufacturers, increased competition in the

*Jha: H. John Heinz III College, Carnegie Mellon University (email: akshayaj@andrew.cmu.edu); Wolak: Program on Energy and Sustainable Development and Department of Economics, Stanford University (email: wolak@zia.stanford.edu). Lucas Davis was coeditor for this article. We would like to thank Severin Borenstein, Jim Bushnell, Meredith Fowlie, Bill Hogan, Han Hong, Ali Hortacsu, Paul Joskow, Lynne Kiesling, Erin Mansur, Ignacia Mercadal, Steve Puller, Mar Reguant, Peter Reiss, Andres Santos, and Ali Yurukoglu for their helpful comments on previous drafts.

† Go to https://doi.org/10.1257/pol.20200234 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

1 See Cheng and Xiong (2014) for a survey of this literature.

forward market for iron ore that operates months or even years in advance of actual production should result in the sunk costs necessary to produce being incurred by the combination of mines that yields the least-cost supply of iron ore to all of the steel manufacturers.

This example illustrates why empirical researchers have found it difficult to link increases in forward market liquidity to increases in the extent to which forward prices reflect future spot prices or reductions in production costs. Specifically, forward markets typically clear months or even years before the commodity is delivered, and many confounding changes can occur during this time interval. We overcome this empirical challenge by studying wholesale electricity supply where sunk cost commitments made in a day-ahead forward market impact the real-time cost of supplying locational demands throughout the transmission network. In the day-ahead and real-time markets, the same product, electrical energy at a specific location in the transmission network in a specific hour, is bought and sold. Similar to the case of iron ore mines, electricity suppliers use day-ahead forward market outcomes to determine whether to incur the sunk costs associated with starting up their units to serve demand in real time.

This paper studies the introduction of financial trading to California's wholesale electricity market. This allows purely financial participants that do not produce or consume electricity to trade day-ahead/real-time price differences at thousands of locations in the transmission network. Purely financial market participants can submit bids to buy or offers to sell energy in the day-ahead market at a location with the understanding that any energy bought or sold in the day-ahead market must be sold or purchased in the real-time market at the prevailing real-time price.

The California Independent System Operator (ISO) introduced purely financial participation in order to reduce differences between day-ahead and real-time prices at over 4,000 locations in the state's transmission network.[2] Analogous to the iron ore example, reducing differences between day-ahead and real-time prices was thought to lead to reductions in operating costs and input energy use. Trades submitted by purely financial participants at thousands of locations in the transmission network can result in forward market outcomes that better reflect real-time conditions and thus, a lower-cost combination of individual generation unit owners incurring sunk costs to produce to serve locational demands throughout the transmission network.

For the case of iron ore mines and steel manufacturers, the market efficiency benefits from financial trading are likely to be largest when there are many spatially distinct buyers and sellers because of the need to coordinate many sources of supply with many locations of demand. By the same token, we expect the market efficiency benefits from purely financial participation in wholesale electricity markets to be largest when there are many distinct buyer and seller locations in the electricity market. There are likely many distinct "submarkets" when many of the operating constraints inherent to electricity production and transmission bind. Namely, similar to the production process for iron ore as well as many other products, electricity generation units have sunk costs to start up, minimum safe

[2] See https://www.caiso.com/Documents/MSCFinalOpiniononConvergenceBidding.pdf.

operating levels significantly greater than zero, restrictions on how fast they can move to different operating levels, and restrictions on the minimum time they must operate or remain idle. There are also thousands of constraints on allowable flows of energy between locations in the transmission network, which implies the potential for substantial differences in real-time prices across these locations. We expect the benefits from introducing purely financial participation to US wholesale electricity markets to occur primarily when a significant fraction of these constraints bind.

To examine how differences between day-ahead and real-time prices changed after the implementation of financial trading, we introduce a measure of the extent to which the 24-dimensional vector of hourly average day-ahead prices at a location in the transmission network reflects the 24-dimensional vector of hourly average real-time prices at the same location. This measure is based on a model of the behavior of a purely financial participant with the ability to trade 24 assets corresponding to the average day-ahead/real-time price differences for each hour of the day.[3] This trader faces a per unit transaction cost associated with buying or selling one megawatt-hour (MWh) of any of these assets. The trader buys or sells the portfolio of day-ahead/real-time price differences that yields the highest expected profits after accounting for this transaction cost.[4]

Using this model of trading behavior, we compute two measures of implied transaction costs using hourly, location-specific data on day-ahead and real-time prices before and after financial trading was introduced at the vast majority of pricing locations in California (California ISO 2009–2012). We find that both of these measures fell substantially after California introduced purely financial participation.[5] In addition, both of these measures fell more at locations where trading was particularly restricted prior to the introduction of financial trading.

We also show that the volatility of the vector of day-ahead/real-time price differences and the volatility of the vector of real-time prices fell at the vast majority of pricing locations after the introduction of financial trading. This result is consistent with financial trading reducing the cost of serving demand because it indicates that the location-specific generation unit output levels that emerge from the day-ahead market are typically closer to their real-time operating levels. This implies fewer instances of real-time changes in operating levels for generation units relative to their day-ahead schedules, which implies fewer costly real-time starts and shutdowns of generation units and potentially lower average daily operating costs.

To explore this hypothesis, we compare daily market outcomes before versus after the introduction of financial trading on "high-complexity days" relative to

[3] This formulation is consistent with the market rules governing all US wholesale electricity markets. Specifically, purely financial trading in these markets occurs daily rather than hourly because financial participants simultaneously submit bids and offers to trade day-ahead/real-time price differences at each location in the transmission network for all 24 hours of the following day.

[4] The relatively small literature on the role of trading costs in explaining differences between forward and spot commodity prices focuses either on incorporating trading costs into theoretical models (see Williams 1987; Hirshleifer 1988; Dávila and Parlatore 2021) or quantifying the explicit transaction fees paid by traders in different markets, as discussed in Frazzini, Israel, and Moskowitz (2018).

[5] Consistent with our empirical results, a simulation study by Li, Svoboda, and Oren (2015) finds that the revenues earned from implementing their optimal strategy for trading day-ahead/real-time price differences fell significantly after California introduced financial trading.

"low-complexity days." We classify days as high complexity if many transmission and generation unit operating constraints bind, which can significantly increase the complexity of finding the least-cost combination of sunk cost commitments and operating levels for generation units to serve demands at the thousands of locations in the transmission network. The introduction of financial trading has the potential to deliver production cost savings on high-complexity days because the bids submitted by purely financial participants to the day-ahead market can result in a lower cost combination of day-ahead energy schedules for generation units to meet real-time demands across the network. On "low-complexity" days, few, if any, differences in real-time prices across locations in the state are expected because transmission network and other operating constraints are not likely to bind in the real-time market. On these days, the ability of retailers to alter their service territory–level demand bids in the day-ahead market and competition among suppliers to sell energy in an unconstrained transmission network are likely to yield day-ahead generation schedules that are close to least-cost real-time operating levels.

We employ an event study framework common to the finance literature to examine this hypothesis.[6] Specifically, we estimate the relationship between the market outcome and control variables, comparing residualized outcomes in the pre– versus post–financial trading sample periods for both high-complexity days and all other days. Our primary market outcomes are the log of the daily total fuel costs incurred by California's gas-fired fleet divided by the daily total output produced by these units and the log of the daily total fuel used by these plants divided by their daily total output.

We consider three measures of high-complexity days, all of which attempt to measure the extent to which a significant fraction of operating constraints are expected to bind in the real-time market. The three measures are the level of daily total demand, the daily standard deviation across locations and hours of the day of real-time prices, and daily total number of starts by gas-fired units. For all three measures, we find that daily fuel costs per MWh and daily input energy use per MWh fell on high-complexity days after financial trading was introduced. In contrast, we find no differences in either fuel costs per MWh or input energy use per MWh before versus after the introduction of financial trading on all other days in our sample. This is consistent with our view that there is little potential for financial traders to reduce operating costs or input fuel use during days when finding the least-cost mix of generation unit output levels is as straightforward as finding the intersection between the aggregate supply curve and the aggregate demand curve. We estimate that fuel costs per MWh (input fuel use per MWh) are 2 percent (1.5 percent) lower after the introduction of financial trading on days with demand above the seventy-fifth percentile of daily total demand. The annual fuel cost savings and annual reduction in carbon emissions implied by these estimates are roughly \$16.6 million and 160,635 tons of CO_2, respectively.

The actions of purely financial participants are not without controversy. Specifically, many argue that financial traders earn revenues primarily at the

[6] For more details on event study models, see MacKinlay (1997) and Eckbo (2008).

expense of producers and consumers of the commodity.[7] For wholesale electricity, some have argued that financial participants submit bids intended to profit from the physical realities inherent to electricity production and transmission, either by taking advantage of rules pertaining to starting up or ramping power plants (Parsons et al. 2015) or by inducing transmission congestion (Birge et al. 2018). On the other hand, previous work documents substantial differences between forward and spot prices in electricity markets without financial participation, due either to the exercise of market power (Borenstein et al. 2008; Ito and Reguant 2016) or risk preferences (Routledge, Seppi, and Spatt 2001; Bessembinder and Lemmon 2002, 2006; Longstaff and Wang 2004). Moreover, others have argued that increases in financial trading volumes decrease the exercise of unilateral market power (Saravia 2003; Mercadal 2022), decrease the volatility of electricity prices (Hadsell 2007), and increase grid reliability (Isemonger 2006). We contribute to this existing literature by providing evidence that the introduction of financial trading to California's wholesale electricity market led to a reduction in the implicit cost of trading day-ahead/real-time price differences, the volatility of these prices differences, and the volatility of real-time prices. We also show that operating costs and fuel use fell on high-complexity days after the introduction of purely financial participation for three different measures of complexity.

The remainder of the paper proceeds as follows. The next section describes how California and other US wholesale electricity markets operated before versus after the introduction of financial trading. Section II discusses several representative examples of how the attempts of purely financial participants to profit from differences between day-ahead and real-time prices at individual locations in the transmission network can reduce the cost of serving demand throughout the transmission network during high-complexity days. We present descriptive trends in day-ahead/real-time price differences for California's wholesale electricity market in Section III. This is followed by the derivation of our measure of how well forward prices reflect expected real-time prices in Section IV. Section V presents the results from applying this methodology to California's wholesale electricity market. Section VI provides evidence suggesting that the introduction of financial trading had market efficiency benefits on high-complexity days. Finally, we conclude in Section VII by exploring the implications of our findings for electricity market design.

I. Market Operation with and without Financial Trading

In this section, we first describe how day-ahead and real-time markets operated in California and other US wholesale electricity markets prior to the introduction of purely financial trading, termed "virtual bidding" by industry participants. This discussion emphasizes the computational complexity associated with finding the least-cost mix of generation unit output levels to meet real-time locational demands

[7] See "Traders Profit as Power Grid Is Overworked," *New York Times*, August 14, 2014 for the case of wholesale electricity markets. See "U.S. Suit Sees Manipulation of Oil Trades," *New York Times*, May 24, 2011 for the case of oil. See "Did Goldman Sachs Rig Commodities Markets?" *CNN Business*, November 20, 2014 for the case of aluminum.

when a significant fraction of the thousands of transmission network and other operating constraints that must be respected for reliable operation of the grid are likely to bind. The next subsection describes how purely financial trading is integrated into the day-ahead and real-time market-clearing processes. The final subsection discusses the limited ways in which physical market participants could trade day-ahead/real-time price differences prior to the introduction of financial trading.

A. *Locational Marginal Pricing in Multisettlement Markets*

In most markets, products are shipped directly from seller to buyer. In wholesale electricity markets, generation units inject electricity into the transmission network, and this electricity flows to buyers according to Kirchhoff's laws.[8] Thus, commitments between buyers and sellers of electricity constitute financial rather than physical arrangements. The buyer does not withdraw the actual energy injected into the transmission grid by the seller. Only the amount of energy injected to the grid by generation units and the amount of energy withdrawn from the grid by electricity demanders can be measured.[9]

All electricity supply industries have high-voltage transmission networks with finite transfer capacity between locations where energy is injected or withdrawn. For this reason, the system operator must sometimes satisfy demand at a given location using nearby higher-cost generation units rather lower-cost units located farther away. Because the extent to which these capacity constraints bind has grown over time, all US wholesale markets have adopted a dispatch and pricing mechanism that accounts for these operating constraints by setting potentially different prices at all nodes in the transmission network. This dispatch and pricing mechanism is called nodal pricing or locational marginal pricing (LMP).

The LMP algorithm sets a price at each node that reflects all relevant transmission network constraints, transmission losses, generation unit start-up and ramping constraints, and all other operating constraints relevant to withdrawing one more MWh of energy at that location.[10] Locational marginal prices in the day-ahead market are determined based on hourly offer curves submitted by suppliers and hourly bid curves submitted by electricity retailers.[11] Specifically, suppliers submit generation unit–level offer curves and retailers submit locational demand curves for each of the 24 hours of the following day. Market participants must submit all 24 of their hourly bid curves by 10 AM on day $t - 1$ for electricity to be delivered on day t.

Generation unit offer curves have three parts: a start-up cost offer, a minimum-load cost offer, and an energy supply curve. The start-up cost offer is a fixed dollar amount that must be paid to the generation unit owner if the unit is not generating electricity

[8] Schweppe et al. (2013) provides an accessible introduction to power system operation.

[9] As discussed in Schweppe et al. (2013), both the quantity of electricity injected by each generation unit as well as where this electricity is withdrawn depends on the level of demand at all locations in the transmission network, the output levels of all of the generation units, the configuration of the transmission network, and a host of other physical operating constraints.

[10] Bohn, Caramanis, and Schweppe (1984) summarizes the basics of the locational marginal pricing algorithm.

[11] We use the more familiar term "retailer" rather than the more technical term "load-serving entity" to refer to any entity that withdraws energy from the high-voltage transmission grid for its own consumption or sale to retail customers.

at the start of day t but is accepted to produce a positive amount of energy at some point during that day. The minimum-load cost offer is a fixed dollar payment that must be paid to the generation unit owner for each hour that the unit is producing energy at its minimum safe operating level.[12] Finally, the energy offer curve for hour h indicates how much additional electricity the supplier is willing to provide from the unit in hour h of day t as a function of the market-clearing day-ahead price at the unit's location. This energy offer curve is a nondecreasing step function, where each price-quantity step determines the minimum price that the generation unit owner must be paid in order to produce at most the quantity of energy associated with that step.[13] The sum of the quantity increments for each energy offer curve is restricted to be less than or equal to the capacity of the generation unit.

Retailers submit willingness-to-purchase bid curves in the day-ahead market that are nonincreasing in the price. This willingness-to-purchase function is composed of price-quantity pairs ordered from highest to lowest price. A retailer is willing to increase the amount of electricity it purchases by the offer quantity increment provided that the market-clearing price is at or below the corresponding offer price increment. In California, retailers submit willingness-to-purchase bid curves at the service territory level. The market operator then allocates shares of these bid curves to individual demand nodes in the retailer's service territory.[14] This allocation is based on the market operator's estimate of the fraction of the retailer's total demand that is withdrawn from each of the locations in its service territory.

California's Independent System Operator clears the day-ahead market by maximizing the sum of consumer and producer surplus for all 24 hours of the following day subject to a host of constraints. These include meeting the demand for energy and operating reserves at more than 4,000 locations in the transmission network during all 24 hours of the following day, respecting the ISO's best estimate of the configuration of the transmission network on the following day, as well as generation unit operating constraints.[15] The locational marginal price (LMP) at each location in the transmission network is equal to the increase in the maximized value of the objective function from the California ISO's optimization problem as a result of increasing the amount of energy withdrawn at that location by one MWh.

All market participants are notified of these LMPs as well as their day-ahead supply and demand obligations at 1 PM on the day before the delivery date. Day-ahead supply and demand obligations are firm financial commitments to sell or buy these quantities of energy. For example, if a supplier sold 50 MWh of electricity to be injected at a given location in the 6 PM hour of the following day at a price of \$40 per MWh, it is guaranteed to be paid \$2,000 ($= 50$ MWh $\times$ \$40/MWh) regardless of

[12] All generation units have a minimum safe operating level that is significantly greater than zero.

[13] In California, suppliers are permitted to submit generation unit–level offer curves with up to ten price-quantity pairs.

[14] Among the three major retailers in California, Pacific Gas and Electric has more than 1,500 nodes in its service territory, Southern California Edison has approximately 200 nodes in its territory, and San Diego Gas and Electric has approximately 300 nodes in its territory. Online Appendix Figure A.2 presents a map of the geographic territories served by each of these retailers.

[15] As discussed in Wolak (2019) and Buchsbaum et al. (2022), a collection of operating reserves are purchased by the ISO in order to ensure that supply equals demand at every instant in real time even in the event that one or more generation unit fails to produce electricity.

the actual production of energy from its generation units at that location during that hour of the following day. This is the sense in which all purchases and sales in the day-ahead market are financially binding.

Between the close of the day-ahead market and the start of real-time system operation, actual electricity demand at each location in the transmission network is realized. Some generation units must produce more or less than their day-ahead energy schedules in order to meet real-time demands at thousands of demand locations in California. At least 75 minutes in advance of each hour of real-time system operation, generation unit owners submit offer curves specifying their willingness to increase or decrease their output relative to their day-ahead schedules. Starting with midnight on the delivery date, these offer curves are used to clear the real-time market for each five-minute interval within the hour to meet the realized demand at each location in the transmission network subject to the real-time configuration of the transmission network and operating constraints on the output levels of all generation units. The configuration of the transmission network and the set of available generation units in real time can differ from the configuration of the transmission network and the set of available generation units used by the system operator to determine day-ahead market outcomes.

The real-time market-clearing process results in real-time prices and generation unit operating levels to serve real-time demands at all nodes. The five-minute real-time price at each location in the transmission network is equal to the increase in the maximized value of the sum of consumer and producer surplus associated with withdrawing an additional MWh at that location. The hourly real-time price is the average of the 12 five-minute real-time prices within that hour. The combination of a financially binding day-ahead forward market and a real-time spot market is called a two-settlement market because only real-time deviations from day-ahead generation and demand schedules are settled at the hourly real-time price.

Recall our previous example of a supplier that sold 50 MWh of energy in the day-ahead market at a price of $40 per MWh. If that supplier only produced 30 MWh of electricity between 6 PM and 7 PM on day *t*, it would have to purchase the remaining 20 MWh at the hourly real-time price for 6 PM at the same location in order to meet its forward market commitment. If the supplier's unit instead produced 55 MWh, then the additional 5 MWh beyond its day-ahead schedule of 50 MWh is sold at the hourly real-time price at that location.

B. *Purely Financial Trading of Wholesale Electricity*

All US wholesale electricity markets currently allow virtual bidding where every market participant has access to the following purely financial instrument: buy (sell) a specified quantity of electricity at a given location and hour of the day in the day-ahead market if the day-ahead price is below (above) the offer price, with the obligation to sell (buy) back the same quantity of electricity at the same location and hour of the day in the real-time market as a price-taker.[16] For example, if a

[16] See California ISO (2023) for a full list of the physical and financial participants that are licensed to submit virtual bids in California's wholesale electricity market as of March 2023.

virtual bidder sells five MWh of energy at a location in the day-ahead market, this market participant must purchase five MWh in the real-time market at the prevailing real-time price at that location. In short, any financial position taken in the day-ahead energy market by a virtual bidder must be closed out in the real-time market as a price-taker.

The market operator treats physical and virtual bids the same in the day-ahead market-clearing process. This implies that the actions of virtual bidders directly influence day-ahead and real-time market outcomes. For example, if a market participant expects the day-ahead price to exceed the real-time price at a location, she might place a virtual offer to supply energy to the day-ahead market at that location. If accepted, this supply offer to the day-ahead market has the potential to reduce the day-ahead price. Moreover, because virtual energy sold by the financial participant in the day-ahead market must be purchased back in the real-time market, this virtual bid increases the demand for energy at the location in the real-time market. This potentially increases the real-time price at that location. Placing a virtual supply bid at a location thus makes it less likely that the day-ahead price will be higher than the real-time price at that location.[17] By the same logic, an accepted virtual bid to purchase energy in the day-ahead market reduces real-time demand at that location, which can reduce the real-time price at that location.

Submitting virtual bids is not costless. First, the total value of virtual bids that each purely financial participant can submit must be less than the collateral it has posted with the California ISO. Market participants must also pay a monthly fee to the California ISO in order to be able to submit virtual bids. In addition, all US wholesale electricity markets charge transaction fees associated with submitting virtual bids as well as additional fees if these virtual bids are accepted in the day-ahead market. Finally, financial traders pay a significant share of total "uplift" charges. Uplift charges compensate generation unit owners for actions that they take at the request of the California ISO that are not recovered from selling energy and operating reserves in the day-ahead and real-time markets.

Although average uplift charges range from roughly 40–60 cents per MWh, there is substantial volatility in the hourly value of these charges. A large realization of uplift charges can easily make a purely financial trade that earns substantial revenues unprofitable. Online Appendix Section B provides further details on uplift charges and the transaction costs associated with financial trading in California's wholesale electricity market.

C. *Trading Day-Ahead/Real-Time Price Differences without Virtual Bidding*

Prior to the introduction of purely financial trading, a supplier that expects the day-ahead price to be higher (lower) than the real-time price at a location where it owns a generation unit might sell more (less) energy in the day-ahead market than it expects to produce in real-time. The energy offer curves submitted by suppliers must have a minimum offer quantity greater than the minimum safe operating

[17] Virtual bidding is also called convergence bidding precisely because it increases the likelihood of convergence between day-ahead and real-time prices.

level of the unit and a maximum offered quantity less than the unit's capacity. Consequently, prior to the introduction of financial trading, a generation unit owner could trade day-ahead/real-time price differences only at locations where it owns generation units and only for quantities of energy up to the capacity of the generation unit at that location.

As noted earlier, retailers must submit demand bids in the day-ahead market at the level of their service territory; the California ISO allocates each retailer's demand bid to individual locations in their service territory.[18] Consequently, the retailer cannot trade day-ahead/real-time price differences at a single location because adjusting its territory-level demand bid impacts the retailer's day-ahead positions at all locations in its service territory and thus, the retailer's trading profits at all locations in its service territory.

With the introduction of financial trading, any market participant can submit virtual bids at any location in the transmission network where virtual bidding is permitted.[19] This implies that each market participant now faces competition from all other market participants when trading day-ahead/real-time price differences at any location where virtual bidding is permitted. For this reason, we expect the introduction of financial trading to increase liquidity in the market for trading day-ahead/real-time price differences and thus, increase the extent of agreement between day-ahead and real-time prices at all locations in the transmission network.

We also expect day-ahead prices to better reflect real-time conditions at generation locations relative to demand locations in markets without purely financial participation. This is because suppliers can adjust their physical offers to trade day-ahead/real-time price differences at locations where they own generation units, while it is not possible for a retailer to adjust its service territory–level bids to trade price differences at an individual demand location. Consequently, introducing financial trading should result in a larger increase in the extent of agreement between day-ahead and real-time prices at demand locations relative to generation locations.

II. How FT Can Lower Production Costs

This section presents examples of how purely financial trading (henceforth denoted FT) at specific locations in the transmission network can lower system-wide production costs. FT aimed at profiting from differences between day-ahead and real-time prices at a location in the grid can produce lower aggregate production costs precisely because the market-clearing prices and output levels that emerge from both the day-ahead and real-time markets are solutions to nonconvex, mixed-integer

[18] The collection of pricing locations in each retailer's service territory is called its Load Aggregation Point (LAP). Each retailer is charged an hourly price equal to the weighted average of the nodal prices in their LAP, where the weights are the share of total demand in the LAP at that location during that pricing period. For example, if the ISO estimates that demand at each of the ten locations in a service area is the same, then the weights are equal to 1/10 for each location. In this case, the LAP is equal to the simple average of the ten locational prices in its service territory.

[19] US wholesale electricity markets typically restrict the set of locations at which virtual bidding is allowed. For example, California's ISO does not allow virtual bidding at locations it deems to be "electrically equivalent" to other locations where virtual bidding is allowed. During our sample period, virtual bidding was allowed at over 2,600 locations.

programming problems.[20] When many of the thousands of transmission and other operating constraints bind in the real-time market, it is unlikely that the solution to this nonconvex problem is a global optimum.[21] Consequently, there is the potential for FT at specific locations to both reduce the cost of serving real-time locational demands and earn a profit for the financial trader.

The use of forward markets to reduce the cost of short-term market operation in industries with nonconvexities in production is not unique to the electricity supply industry. Almost any industry that requires incurring a sunk cost to produce uses forward markets to reduce production costs. For example, airlines use information on forward market sales of travel between origin and destination pairs to assign planes and crews to routes in order to minimize the total cost of satisfying demand for air travel.[22] Gallamore and Meyer (2014) provide an example of how forward markets were used by US railroad companies in the aftermath of the Staggers Act of 1980 that deregulated rail rates. Namely, electric utilities signed long-term contracts with railroads to transport coal from the Powder River Basin to their power plants. These forward market sales allowed the railroad companies to incur the substantial sunk costs associated with infrastructures upgrades and track expansions that ultimately lowered the average cost per ton of delivering this coal.

The difference between wholesale electricity and these examples is the short time horizon between when a commitment to incur the sunk cost to start up a generation unit is made and real-time operation occurs. Therefore, our examples focus on instances where nonconvexities in the production and delivery of electricity combined with binding operating constraints create opportunities for FT to reduce production costs.

The first example involves a fossil fuel–fired generation unit with a start-up cost and long start-up time. Specifically, the unit cannot operate in real time unless it is scheduled to sell positive output in the day-ahead market because more than 75 minutes of advance notice is necessary to start the unit.[23] If the "long-start" unit does not start up, any increase in real-time demand at the unit's location must be satisfied by a fast-start unit that has a higher operating cost. Suppose that a financial trader recognizes that the day-ahead price at this location is persistently below the real-time price at this location for several hours of the day. Because of this, she submits a bid to purchase energy at this location in the day-ahead market for these hours, which she subsequently sells in real-time.

[20] For example, the decision to start up a generation unit is a binary variable and how much to operate each committed unit within its allowable operating range is a continuous variable. Burer and Letchford (2012) survey the available algorithms and software for solving mixed integer programs.

[21] If there are N constraints associated with turning a generation unit on or off and inequality constraints on energy flows between locations in the transmission network that must be respected in this optimization problem, then there are as many as 2^N possible binding constraint configurations. Because the real-time market must clear every five minutes, there is typically not sufficient time to exhaustively search through all of these possible constraint configurations.

[22] See Mercier, Cordeau, and Soumis (2005) for an example of this nonconvex optimization problem.

[23] During our sample period, roughly 20,000 MW of generation capacity in California required more than 75 minutes' advance notice to start.

The bid submitted by the purely financial participant increases day-ahead demand at the unit's location, which also increases day-ahead prices at that location. This day-ahead price increase makes it more likely that the long-start unit is scheduled to sell energy in the day-ahead market. If this low–variable cost, long-start unit is now available to produce electricity the following day, more expensive fast-start units are no longer required to meet demand in real time. Summarizing, in this case, FT increases day-ahead prices and reduces real-time prices, which also decreases the cost of serving real-time demand because the long-start, low-cost unit produces electricity in real time rather than fast-start, higher cost units.

We emphasize that the financial trader does not need to understand why day-ahead prices are persistently lower than real-time prices. She only needs to observe that fact and place the day-ahead market bid necessary to capture this locational price difference. Her virtual demand bid impacts the day-ahead schedule of generation at that location because purely financial bids are treated the same as physical bids in the day-ahead market-clearing process. In our example, the long-start unit is scheduled to produce in the day-ahead market as a consequence of the virtual demand bid. In real time, the virtual bidder reverses her position; no energy is actually produced or consumed by the bidder. Instead, the long-start, low–variable cost unit produces to meet real-time demand at the location. Consequently, the financial participant's bid results in day-ahead schedules for generation units that are closer to the least-cost mix of generation unit output levels necessary to meet real-time locational demands throughout the transmission network.

A second example demonstrates the role that FT can play in managing transmission constraints. Consider two locations in the transmission network, a generation-rich location A and a generation-deficient location B. Suppose that a financial trader notices that the price difference between locations A and B in the day-ahead market is systematically larger than the real-time price difference between A and B (i.e., $P_B(DA) - P_A(DA) > P_B(RT) - P_A(RT)$). The financial trader does not know that the reason that this is happening is because the market operator is systematically releasing more transmission capacity between locations in the transmission network in the real-time market than he is making available in the day-ahead market. This means that more energy can flow from A to B in real time than the market operator allows for when clearing the day-ahead market.

To exploit this difference across locations in day-ahead/real-time price spreads, the financial trader can submit a demand bid at A and a supply bid at B for the same number of MWh in the day-ahead market. These purely financial bids can allow more long-start, low-cost generation to be committed at A in the day-ahead market and less high-cost generation to be committed at B in the day-ahead market. In real time, the increased amount of energy produced at A can flow to be consumed at B because more transmission capacity between A and B is available than was scheduled when clearing the day-ahead market. Consequently, FT reduces the cost of serving demand in real time because the financial trader found systematic differences between day-ahead and real-time prices at A and B.

As these two examples illustrate, the potential for FT to reduce production costs is greatest when there are likely to be many binding operating constraints in real time. These binding constraints along with the nonconvexities associated with deciding which units to start up increase the complexity of the process of finding a global optimum to the real-time market optimization problem. On days when these nonconvexities are unlikely to be relevant, finding the global optimum solution to the day-ahead market is often as straightforward as finding the point of intersection between the system-wide supply curve and the system-wide demand curve for each hour of the day, which implies limited scope for FT to reduce the cost of serving real-time demand.

To understand the essential role that sunk start-up costs, significant start-up times, and other nonconvexities such as ramp rates, minimum safe operating levels, and transmission constraints play in creating the potential for FT to reduce production costs, consider the example of an electricity market where none of these costs or constraints are relevant. Assume that all firms have a constant marginal cost of producing energy, zero start-up costs, can produce at any level of output from zero to the maximum capacity of the unit instantaneously, and there is infinite transmission capacity between all locations in the transmission network. Under these conditions, a formal day-ahead market would have no market efficiency benefit. Day-ahead energy schedules would involve no sunk costs because suppliers can turn on and change their output levels instantaneously. Regardless of the day-ahead market sales by suppliers or purchases by retailers, in real time, the market operator would simply order the generation units in terms of their marginal cost from smallest to largest and set the price at the marginal cost of the highest-cost unit needed to serve demand. This would always yield the least-cost solution to serving demand in real time regardless of day-ahead market outcomes.

In this example market, bids and offers submitted in order to profit from expected differences between day-ahead and real-time prices would have no impact on the real-time output levels of units. For example, suppose that a supplier believes that the unconstrained day-ahead price will be higher than the unconstrained real-time price. He would sell more energy into the day-ahead market than he expects to produce in real time, profiting from this strategy if the realized day-ahead price is indeed larger than the real-time price. However, this has no impact on the final quantity of electricity produced in real time because there are no sunk costs to starting units or operating constraints that would limit the ability of his units to instantaneously produce as much as is needed to meet real-time demands across the grid at least cost. The same logic applies for a retailer that believes that the day-ahead price will be higher than the real-time price. In this case, the retailer can submit less demand at the service territory level into the day-ahead market and purchase the remaining demand in the real-time market. Once again, these actions would not impact the least-cost combination of generation units to meet real-time demand.

For this reason, we compare market outcomes before versus after the introduction of FT on two types of days, high-complexity days, when a significant fraction of operating constraints are likely to bind in real time, and all other days. Our proposed mechanism suggests that production costs should fall after the introduction of FT on high-complexity days but not all other days when the real-time market is adequately approximated by the simplified convex market discussed above.

III. Descriptive Analysis of Day-Ahead and Real-Time Prices

This section provides descriptive statistics on the first and second moments of hourly day-ahead and real-time prices at the more than 4,000 locations in California.[24] In the first subsection, we show that the average difference between day-ahead prices and real-time prices is smaller in absolute value after the introduction of FT. The second subsection documents differences in the sign of average day-ahead/real-time price differences across demand locations versus generation locations before versus after financial trading is introduced. This analysis yields evidence consistent with the introduction of FT limiting the ability of retailers to exercise unilateral market power in the day-ahead market. The final subsection demonstrates that the volatility of both the 24-dimensional vector of day-ahead/real-time price spreads and the 24-dimensional vector of real-time prices fell after the introduction of FT for the vast majority of locations in California.

A. *Descriptive Trends in Day-Ahead and Real-Time Prices*

Figure 1 plots the monthly average day-ahead price minus the monthly average real-time price. These averages are taken over both hours-of-sample in the month and locations in the California transmission network. The horizontal purple (green) line plots the overall average day-ahead/real-time price difference for the sample period before (after) FT was introduced.[25] This figure makes clear that day-ahead prices were lower than real-time prices on average prior to FT. After FT was introduced, the average day-ahead/real-time price spread fell considerably in absolute value.

It is important to note that a zero average difference between day-ahead and real-time prices does not imply the absence of a profitable FT strategy. For example, suppose that average price differences for the first 12 hours of the day are +\$10/MWh and average price differences for the remaining 12 hours of the day are −\$10/MWh. The average price difference would be zero, but a financial trader could earn considerable profits from selling energy in the day-ahead market during the first 12 hours of the day and buying energy in the day-ahead market during the second 12 hours of the day. Consequently, all of the elements of the 24-dimensional vector of hour-of-the-day-specific average differences between day-ahead and real-time prices must be small for a profitable trading strategy not to exist. This fact motivates our measure of the extent of agreement between day-ahead and real-time prices discussed in Section IV.

Figure 2 plots the average day-ahead price minus the average real-time price for each hour of the day. These averages are taken across days and pricing locations. The left panel focuses on the sample period before financial trading, while the right panel focuses on the sample period after FT. For each hourly average

[24] These data are downloaded from the OASIS API administered by California's Independent System Operator (California ISO 2009–2012).

[25] The before-FT sample period is April 1, 2009 to January 31, 2011, and the after-FT sample period is February 1, 2011 to November 30, 2012.

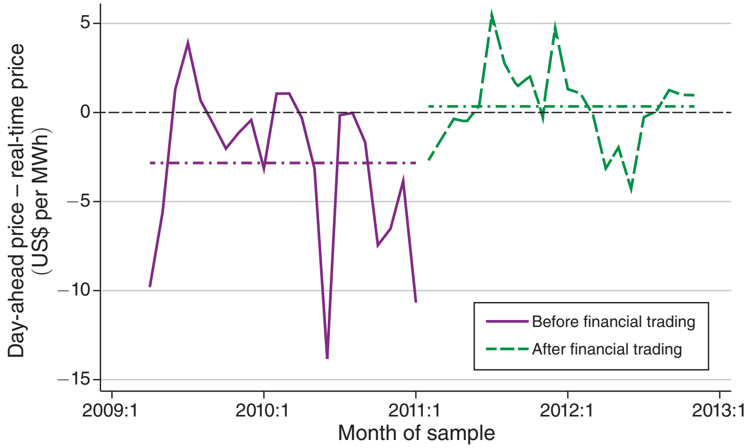


Figure 1. Monthly Average Day-Ahead Price Minus Real-Time Price

Notes: This figure plots the monthly average day-ahead price minus the monthly average real-time price. These averages are taken over all hours-of-sample in the month and pricing locations in California. The sample period considered is April 2009 to November 2012. Monthly averages for the sample period before the introduction of FT are plotted with a solid purple line; monthly averages for the sample period after FT are plotted with a dashed green line. The horizontal dashed purple (green) line plots the overall average day-ahead/real-time price difference for the sample period before (after) FT was introduced.

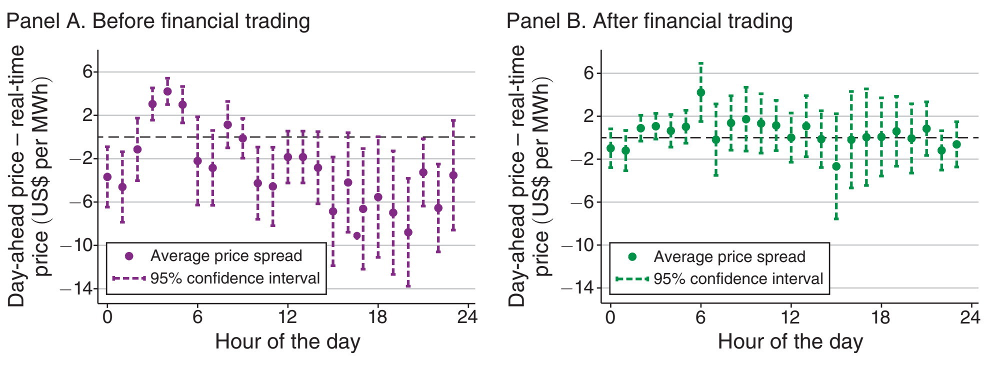


Figure 2. Hourly Average Day-Ahead Price Minus Hourly Average Real-Time Price with 95 Percent Confidence Intervals

Notes: This figure plots the average day-ahead price minus the average real-time price for each hour of the day. For the left panel (right panel), we utilize the sample period before (after) financial trading was introduced. For each hourly average day-ahead/real-time price spread, we include 95 percent confidence intervals based on standard errors two-way clustered by location and week-of-sample.

day-ahead/real-time price spread, we include 95 percent confidence intervals based on standard errors two-way clustered by location and week-of-sample. Before FT, there are many hours of the day with average price spreads that are statistically different from zero (see the left panel of Figure 2). In contrast, only the average price spread for 6 AM is statistically significant for the post-FT sample period (see the right panel of Figure 2). Nonzero average day-ahead/real-time price spreads do not imply that financial participants can profit from trading these price differences. This is because of the substantial transaction costs associated with buying or selling virtual energy in the day-ahead market, an issue we address when developing our model of an expected profit-maximizing purely financial trader in Section IV.

B. *Price Spreads for Generation versus Demand Locations*

Ito and Reguant (2016) argue that a persistent difference between day-ahead prices and real-time prices can be indicative of the exercise of unilateral market power. As shown in Figures 1 and 2, average day-ahead prices were lower than average real-time prices for most hours of the day prior to the introduction of financial trading in California. As discussed in Borenstein et al. (2008), this suggests that large retailers were able to exercise unilateral market power by withholding demand from the day-ahead market and driving down prices in the day-ahead market.

In contrast, day-ahead prices are not persistently higher or lower than real-time prices on average after the introduction of FT (see Figure 1 and panel B of Figure 2). All but one of the pointwise 95 percent confidence intervals for the hour-of-the-day-specific average price spreads contain 0 after FT. This suggests that the actions of financial participants to exploit any persistent day-ahead/real-time price differences prevented large buyers from exercising unilateral market power by withholding demand from the day-ahead market.

To explore this hypothesis more formally, we compare the signs of hourly average day-ahead/real-time price spreads at generation locations versus demand locations before versus after the introduction of FT. For each location in California, we first compute the mean of day-ahead/real-time price differences for each hour of the day for both the pre-FT and post-FT sample periods. For each location and each sample period, we then count the number of hours in the day with a positive average day-ahead/real-time price spread.

Figure 3 plots the resulting empirical distribution function (EDF) of this count across generation locations and across demand locations for both the pre-FT and post-FT samples. After FT, there is a substantial increase in the number of hours of day with a positive average day-ahead/real-time price difference for both generation and demand locations. The median across locations of the number of hours of day with a positive average price spread increases from 6 in the pre-FT period to 17 in the post-FT period. This rightward shift in EDFs after the introduction of FT is consistent with the argument advanced by Saravia (2003) and Mercadal (2022) that purely financial participants increase competition in the market for day-ahead/real-time price spreads. This increased competition reduces the ability of large buyers to withhold demand from the day-ahead market in order to lower day-ahead prices relative to real-time prices.

For both the pre-FT and post-FT sample periods, we also note that the EDF for demand locations is slightly shifted to the right relative to the EDF for generation locations. For virtually all k between 0 and 24, a larger fraction of demand locations have k or more hours of the day with positive average price spreads relative

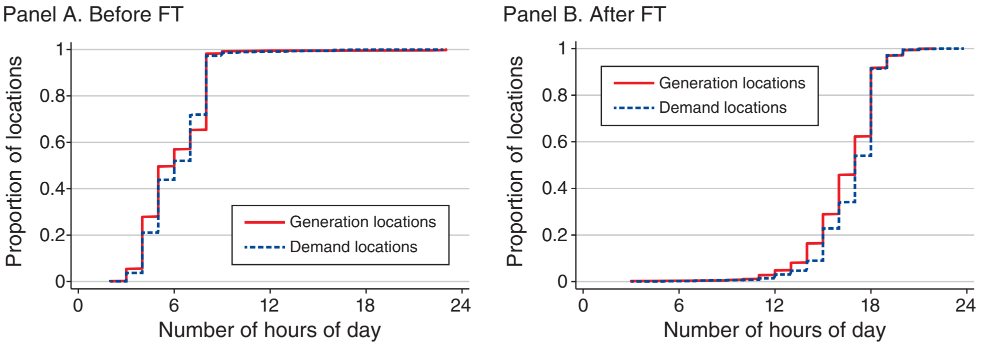


Figure 3. Number of Hours with Positive Average Price Spreads: Generation Locations versus Demand Locations before versus after FT

Notes: This figure plots the proportion of locations with at most the number of hours of day with positive average day-ahead/real-time price spreads listed on the x-axis. Separate empirical distribution functions are plotted for generation locations versus demand locations before versus after the introduction of FT. To calculate an observation underlying these empirical distribution functions, we compute the average day-ahead/real-time price spread across days-of-sample for each location and each hour of the day for both the pre-FT and post-FT sample periods. Then, we count the number of hours of day that the average price spread is positive for each location in each of the two sample periods. Finally, we plot the empirical distribution functions associated with these location-/sample period–level counts separately for generation locations and demand locations before versus after FT.

to generation locations.[26] This result is consistent with two features of California's multisettlement market. First, retailers must submit territory-level bid curves into the day-ahead market, limiting their ability to exercise market power in the day-ahead market at specific demand nodes. Second, with few exceptions, only suppliers are able to influence real-time prices because only they can submit location-specific offer curves to the real-time market.[27] The results in Figure 3 imply that electricity suppliers were successful at raising real-time prices above day-ahead prices for more hours of the day at locations where they owned generation units relative to demand locations both before and after FT.

C. *Volatility in Prices before versus after Financial Trading*

If day-ahead prices at all locations are closer to real-time prices at all locations, then the real-time output levels of generation units are also likely to be closer to the day-ahead schedules of these generation units. This implies less need for large deviations between day-ahead scheduled output and actual output in the real-time market, which leads to a lower variance in real-time prices. Consequently, FT should reduce both the variance of day-ahead/real-time price spreads and the variance of real-time prices. To test this hypothesis, we compare estimates of the covariance

[26] In online Appendix Section C.2, we show that, for both the pre-FT and post-FT sample periods, we fail to reject the null hypothesis that the EDF of demand locations first-order stochastic dominates the EDF of generation locations. Moreover, we reject the null hypothesis that the EDF of generation locations first-order stochastic dominates the EDF of demand locations for both the pre-FT and post-FT sample periods.

[27] Real-time prices are determined using the real-time offer curves of generation units and a small amount of dispatchable demand to meet actual locational demands throughout the transmission network.

matrices of day-ahead/real-time price spreads and real-time prices before versus after FT.

Let $X_{h,d} \equiv P^{\mathrm{DA}}_{h,d} - P^{\mathrm{RT}}_{h,d}$ be the difference between day-ahead and real-time prices in hour of the day h in day-of-sample d at a location. Suppressing the subscripts associated with location and sample period, let $\mathbf{X}_d = (X_{d,1}, X_{d,2}, \ldots, X_{d,24})'$ be the 24×1 vector composed of realized day-ahead/real-time price spreads for day-of-sample d. Each day, this vector is assumed to be drawn from a distribution with contemporaneous covariance matrix Λ^{pre} or Λ^{post} depending on whether the sample period considered is before or after the introduction of FT.

The variance of price spreads is larger "pre-FT" relative to "post-FT" if the difference between the covariance matrices for the pre-FT versus post-FT sample periods is a positive semidefinite matrix (i.e., $\Lambda^{pre} - \Lambda^{post} \geq 0$). We construct a statistical test of this null hypothesis by finding the eigenvalues $\{\hat{\omega}_j\}_{j=1}^{24}$ associated with $\hat{\Lambda}^{\mathrm{diff}} \equiv \hat{\Lambda}^{pre} - \hat{\Lambda}^{post}$. We test the joint null hypothesis that all of these eigenvalues are greater than or equal to zero using the methodology developed by Wolak (1989).[28]

We also calculate the test statistic associated with the null hypothesis that price spreads are more volatile post-FT relative to pre-FT (i.e., $\Lambda^{post} - \Lambda^{pre}$ is a positive semidefinite matrix). Finally, we perform tests of the same two null hypotheses focusing on the variance of real-time prices before versus after FT rather than the variance of day-ahead/real-time price spreads.

We conduct both sets of tests for 762 generation locations and 3,861 demand locations. Table 1 presents the proportion of locations for which we fail to reject a 0.05 size test of each of our null hypotheses, separately for locations associated with generation units ("Generation") versus locations not associated with generation units ("Demand"). The results of this table indicate that there are no discernible differences between generation versus demand nodes in the proportion of locations for which we reject any specific null hypothesis.

Thus, focusing on the rows corresponding to total proportions ("Total"), we fail to reject the null hypothesis that the volatility of price spreads is lower (higher) after FT is introduced for 97 percent (0.6 percent) of locations. Combined, these results provide evidence that the volatility of day-ahead/real-time price spreads fell at both generation and demand locations after the introduction of FT. This is consistent with the intuition that allowing purely financial participation results in day-ahead generation schedules that more closely resemble real-time operating levels.

The last column of Table 1 reports the results of these tests for the vector of real-time prices. We fail to reject the null hypothesis that real-time prices are more volatile pre-FT relative to post-FT for roughly 97 percent of nodes and reject the

[28] Our test statistic is $TS = \min_{\{z \geq 0\}} (\hat{\Omega} - z)' [\widehat{\mathrm{var}(\hat{\Omega})}]^{-1} (\hat{\Omega} - z)$, where $\hat{\Omega}$ is the 24×1 vector containing the eigenvalues $\{\hat{\omega}_j\}_{j=1}^{24}$. The covariance matrix $\mathrm{var}(\hat{\Omega})$ is estimated using a moving block bootstrap procedure. We construct L moving block resamples separately for the sample periods before versus after FT. For each resample $b \in \{1, 2, \ldots, L\}$, we estimate the contemporaneous covariance matrices associated with day-ahead/real-time price spreads in each sample period (i.e., $\hat{\Lambda}_b^{pre}$ and $\hat{\Lambda}_b^{post}$). This allows us to compute $\hat{\Lambda}_b^{\mathrm{diff}} \equiv \hat{\Lambda}_b^{pre} - \hat{\Lambda}_b^{post}$ as well as the 24 eigenvalues associated with $\hat{\Lambda}_b^{\mathrm{diff}}$. We denote the 24×1 vector of these eigenvalues $\hat{\Omega}_b$. Finally, our estimate of the covariance matrix associated with $\hat{\Omega}$ is $(1/L)\sum_{b=1}^{L} (\hat{\Omega}_b - \hat{\Omega})(\hat{\Omega}_b - \hat{\Omega})'$, where $\hat{\Omega}$ is estimated using data from the entire sample period.

Table 1—Proportion of Locations Where Volatility Test Is Not Rejected

	Location type	Price spread	Real-time price
$H_0: \Lambda^{pre} - \Lambda^{post} \geq 0$	Generation	0.971	0.963
	Demand	0.974	0.966
	Total	0.974	0.966
$H_0: \Lambda^{post} - \Lambda^{pre} \geq 0$	Generation	0.005	0.001
	Demand	0.006	0.003
	Total	0.006	0.003

Notes: This table reports the proportion of locations for which we fail to reject different null hypotheses regarding the volatility of day-ahead/real-time price spreads as well as the volatility of real-time prices before versus after the introduction of FT. We report these proportions separately for locations associated with generation units ("Generation") versus locations not associated with generation units ("Demand"). We also report the total proportions aggregated across all pricing locations ("Total").

null hypothesis that real-time prices are more volatile after FT for almost all nodes. This constitutes additional evidence that the introduction of FT reduced the number of costly adjustments to output relative to day-ahead schedules necessary to meet real-time demands at all locations.

IV. Expected Profit-Maximizing Financial Trading with Transactions Costs

In this section, we develop a model of expected profit-maximizing trading of the 24×1 vector of hourly day-ahead/real-time price differences accounting for the presence of a per MWh trading charge. We use this model to construct two indices that quantify how well the vector of day-ahead prices reflects the vector of real-time prices. We then compute these two indices before and after FT.

Before presenting our model, we note that trading strategies based on the first lag of the vector of day-ahead/real-time price differences are not feasible because market participants submit their offers to the day-ahead market for day t before the vector of day-ahead/real-time price differences for day $t-1$ is realized.[29] Market participants can thus only condition their trading strategies on the vector of realized price differences from two or more days prior.

In online Appendix Section C.3, we formulate and implement a test of the null hypothesis that the elements of the autocovariance matrices between the current vector of day-ahead/real-time price spreads and the second through tenth lags of this vector are jointly zero. We reject a test of this null hypothesis for over 40 percent of generation and demand locations before FT but fail to reject this same null hypothesis for close to 95 percent of generation and demand locations after FT. This suggests that purely financial traders cannot earn significantly more profits by conditioning their trading strategies on realized price spreads from two or more days prior to the current day.[30] For this reason, we consider expected profit-maximizing trading strategies that do not condition on lags of the vector of daily price differences.

[29] Bids to the day-ahead market for day t must be submitted by 10 AM of day $t-1$.

[30] See online Appendix Tables C.4 and C.5 for the results when applying this statistical test to the price spreads faced by California's three large retailers and location-specific price spreads, respectively.

A. *The Trader's Problem*

In this framework, a financial participant is assumed to buy or sell hourly positions a_h associated with day-ahead/real-time price spreads $X_{h,d}$ for each hour h in day d. Because the trader can condition her positions on the hour of the day, a_h can take on a different value for each $h \in \{1, 2, \ldots, 24\}$. Let $a = (a_1, a_2, \ldots, a_{24})'$ denote the 24×1 vector of hourly positions. Consistent with market rules, the trader chooses positions for all hours of the day simultaneously.

A positive (negative) value of a_h implies selling (buying) energy in the day-ahead market and buying (selling) it back in the real-time market. Holding a positive (negative) position earns revenues if and only if the day-ahead price for hour h of day d is higher (lower) than the real-time price in hour h of day d. In other words, a trader earns positive revenue if and only if her position a_h has the same sign as the realized price spread $X_{d,h}$. We assume that this trader is small relative to the market so that her purely financial bids do not affect day-ahead or real-time prices.

Let $\mu_h \equiv E(X_{h,d}) = E(P_{h,d}^{\text{DA}}) - E(P_{h,d}^{\text{RT}})$ be the unconditional expectation of the day-ahead/real-time price spread for hour h. Define μ to be the 24×1 vector composed of $(\mu_1, \mu_2, \ldots, \mu_{24})'$. The trader's expected profit-maximization problem is

$$\max_{a \in R^{24}} \underbrace{a'\mu - c\sum_{i=1}^{24}|a_i|}_{\text{Expected Profits}} \text{ subject to } \underbrace{\sum_{i=1}^{24}|a_i| = 1}_{\text{Absolute Position Constraint}}, \tag{1}$$

where c is the dollar per MWh transaction cost associated with buying or selling 1 MWh of any combination of these 24 assets. The vector of positions $a^*(\mu) \in R^{24}$ denotes the solution to the constrained optimization problem described in equation (1).

The trader pays the same per unit trading cost c regardless of whether she buys or sells the asset. This is why overall trading costs are calculated based on the sum of the absolute values of the portfolio weights (i.e., trading costs are $c\sum_{i=1}^{24}|a_i|$). The trader's expected revenue from solving equation (1) is

$$\phi(\mu) \equiv a^*(\mu)'\mu = \max_{h \in \{1,\ldots,24\}} |\mu_h|. \tag{2}$$

In words, the trader simply buys or sells one MWh of the asset with the highest expected payoff in absolute value.

B. *Measures of Agreement between Day-Ahead and Real-Time Prices*

Our two measures of the extent of agreement between the vector of day-ahead prices and real-time prices are derived from tests of two null hypotheses: (i) the null hypothesis that a profitable trading strategy exists (H_0: $a^*(\mu)'\mu - c > 0$) and (ii) the null hypothesis that no profitable trading strategies exist (H_0: $a^*(\mu)'\mu - c \leq 0$). The first measure, c_{lower}, is the smallest value of c that would cause rejection of a size 0.05

test of the first null hypothesis. The second measure, c_{upper}, is the largest value of c that results in rejection of a size 0.05 test of the second null hypothesis.

We estimate these two measures for each pricing location in California for the sample periods before and after the introduction of FT. Our estimate of the unconditional expectation of day-ahead/real-time price spreads for each hour of the day is simply the sample average: $\hat{\mu} \equiv \bar{X} = (1/N)\sum_{d=1}^{N}\mathbf{X}_d$. Our estimate of the revenue from the trader's optimal strategy, presented in equation (2), is the element of $\bar{X}$ that is largest in absolute value, $\phi(\bar{X}) \equiv \max_{h \in \{1,\ldots,24\}}|\bar{X}_h|$. Both of our test statistics are based on the difference between $\phi(\bar{X})$ and the per unit trading cost c.

We cannot use the Delta Method to derive the asymptotic distribution of $\phi(\bar{X})$ because the maximum operator is not differentiable. Instead, we use the method developed by Fang and Santos (2019) for testing hypotheses involving directionally differentiable functions of a regular parameter estimate. This method is applicable because $\phi(\mu)$ is a directionally differentiable function of the parameter vector μ and the sample average $\bar{X}$ is a regular estimator of population average μ_0.[31] Fang and Santos (2019) proposes a modified bootstrap estimator for the asymptotic distribution of $\sqrt{N}(\phi(\bar{X}) - \phi(\mu_0))$.

To implement this estimator, we simulate the distribution of $\phi(\bar{X})$ using a procedure based on numerical derivatives developed by Hong and Li (2018). For this procedure, we first compute moving blocks bootstrap resamples of $\bar{X}$ with block size equal to the largest integer less than or equal to $N^{1/3}$, as recommended by Kunsch (1989).[32] Let the sample average calculated from the b^{th} bootstrap resample be denoted $\bar{X}^b$. We next construct

$$Z^b = \frac{\phi\left(\bar{X} + \sqrt{N}\left(\bar{X}^b - \bar{X}\right)\epsilon\right) - \phi(\bar{X})}{\epsilon} \tag{3}$$

for $b = 1, 2, \ldots, B$. Hong and Li (2018) demonstrates that the asymptotic distribution of $\sqrt{N}(\phi(\bar{X}) - \phi(\mu_0))$ can be approximated by the bootstrap distribution of Z^b provided that, as sample size N goes to infinity, ϵ tends to zero but $\sqrt{N}\epsilon$ tends to infinity. To satisfy these conditions, we set $\epsilon = N^{-1/3}$, which is the value recommended by Hong and Li (2018).

This estimate of the asymptotic distribution of $\sqrt{N}(\phi(\bar{X}) - \phi(\mu_0))$ allows us to compute two test statistics related to the existence of profitable trading strategies given a per unit trading cost of c. To do this, we estimate the distribution of $\phi(\bar{X})$ using the moving blocks bootstrap. In particular, the b^{th} resample gives us

$$\phi(\bar{X})^b = \phi(\bar{X}) + \frac{Z^b}{\sqrt{N}}. \tag{4}$$

[31] The sample average $\bar{X}$ is a regular estimator of population average μ_0 because $\sqrt{N}(\bar{X} - \mu_0)$ is asymptotically normally distributed.

[32] Given a sample $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$, each moving blocks bootstrap resample $b \in \{1, 2, \ldots, B\}$ is constructed as follows. First, we partition the data into K nonoverlapping blocks of size M: $\{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K\} \equiv \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M\}$, $\{\mathbf{X}_{M+1}, \ldots, \mathbf{X}_{2M}\}, \ldots, \{\mathbf{X}_{M(K-1)+1}, \ldots, \mathbf{X}_{KM}\}$. Next, let S be a discrete uniform variable over the integers $\{1, 2, \ldots, K\}$; we construct the b^{th} bootstrap resample by drawing K integers from S independently and identically and merging together blocks based on these draws. For example, if we draw $\{2, 5, K, \ldots, 5\}$, then the bootstrap sample would be $\{\mathbf{B}_2, \mathbf{B}_5, \mathbf{B}_K, \ldots, \mathbf{B}_5\}$. When implementing this procedure, we set $M = \text{floor}(N^{1/3})$ and $K \equiv \text{floor}(N/M)$, where $\text{floor}(Y)$ is equal to Y rounded down to the nearest integer.

We use this bootstrap distribution to compute our two measures of the extent of agreement between the 24×1 vectors of the daily averages of day-ahead and real-time prices. The first, c_{lower}, is equal to the fifth percentile of the bootstrap distribution of $\phi(\bar{X})$. The second magnitude, c_{upper}, is equal to the ninetieth percentile of the distribution of $\phi(\bar{X})$.

We call these two summary measures of the extent of agreement between day-ahead prices and real-time prices "implied trading costs." A larger value of either c_{lower} or c_{upper} corresponds to less agreement between day-ahead and real-time prices. We estimate c_{lower} and c_{upper} for the day-ahead/real-time prices associated with each pricing location for sample periods before and after the introduction of FT.

V. Estimates of Implied Trading Costs

This section is split into two parts. In the first subsection, we present the distribution across locations of implied trading costs before versus after the introduction of FT. The next subsection describes how implied trading costs change at locations with and without electricity generation units after FT is introduced. We find that implied trading costs fell after FT was introduced, with significantly larger decreases at demand locations where it was not possible to trade day-ahead/real-time price differences prior to the introduction of FT.

A. *Spatial and Temporal Trends in Implied Trading Costs*

Figure 4 plots various percentiles of the across-location distributions of c_{lower} and c_{upper}, separately for locations associated with generation units ("Generation locations") and locations not associated with generation units ("Demand locations"). Figure 4 shows that the distributions of c_{lower} and c_{upper} shift downward after FT is introduced for both generation locations and demand locations. This indicates greater agreement between day-ahead and real-time prices at each location after the introduction of FT.[33]

Next, we compute the bootstrap distribution of $c_{pre} - c_{post}$ for each of the more than 4,000 locations that existed both before and after the introduction of FT. The first row of Table 2 reports the proportion of locations for which we reject a size 0.05 test of the null hypothesis that implied trading costs increased after FT (i.e., $H_0\colon\ c_{pre} \leq c_{post}$). We report these proportions separately for generation locations and demand locations. The second row of Table 2 reports the proportion of locations for which we reject a size 0.05 test of the null hypothesis that implied trading costs decreased after the introduction of FT (i.e., $H_0\colon\ c_{pre} \geq c_{post}$), again separately for generation locations and demand locations.

We reject the null hypothesis that implied trading costs increased after the introduction of FT for roughly 39 percent of locations. In contrast, we reject the null hypothesis that implied trading costs fell after the introduction of FT for less than

[33] In online Appendix Section A.3.1, we compute implied trading costs using data on the territory-level day-ahead and real-time prices paid by each of California's three major retailers. Both measures of implied trading costs fell after FT was introduced for all three retailers.

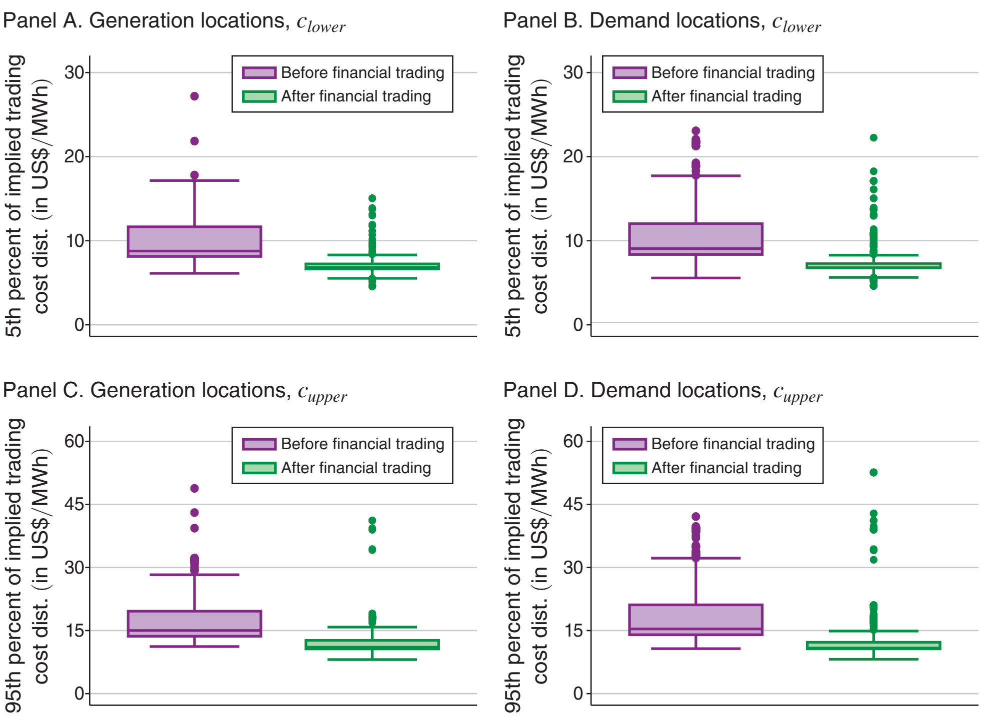


Figure 4. Nodal-Level Distribution of Implied Trading Costs: Before and after Financial Trading

Notes: This figure plots percentiles of the across-location distribution of c_{lower} and c_{upper}. These implied trading costs are estimated separately for each location for the sample period before financial trading and the sample period after FT. We plot the percentiles of the across-location distributions of c_{lower} and c_{upper} separately for locations associated with generation units versus demand locations. The box portion of this box and whiskers plot contains all locations within the twenty-fifth through seventy-fifth percentiles of the distribution of implied trading costs. The bottom (top) whisker is defined by the smallest (largest) value that is within 1.5 × *IQR* of the twenty-fifth (seventy-fifth) percentile of the distribution of implied trading costs, where IQR (interquartile range) is the distance between the twenty-fifth and seventy-fifth percentiles of the distribution. The remaining points are outliers.

Table 2—Proportion of Locations That Reject $c_{pre} \leq c_{post}$ or $c_{pre} \geq c_{post}$

	Total	Generation locations	Demand locations
$H_0: c_{pre} \leq c_{post}$	0.389	0.364	0.394
$H_0: c_{pre} \geq c_{post}$	0.009	0.012	0.008
Number of Locations	4,623	762	3,861

Notes: The first row of this table reports the proportion of locations for which we can reject a size 0.05 test of the null hypothesis that implied trading costs increase after financial trading was introduced, separately for locations associated with generation units versus demand locations. The second row of this table reports the proportion of locations for which we can reject a size 0.05 test of the null hypothesis that implied trading costs fell after FT, again separately for generation locations versus demand locations.

1 percent of locations. A rejection frequency of 1 percent is consistent with the null hypothesis being true for all locations because the size of each hypothesis test is $\alpha = 0.05$. Combined, the results suggest that day-ahead prices better reflect real-time prices at all locations after the introduction of FT.

B. *Heterogeneous Impacts of FT on Implied Trading Costs*

As noted in Section IC, prior to the introduction of FT, each supplier could implicitly trade differences between day-ahead and real-time prices by adjusting their physical bids; this strategy was only feasible at locations where the supplier owned generation units. In contrast, retailers could only submit bids at the service territory level, making it impossible for them to trade day-ahead/real-time price differences at specific locations. Consequently, in the absence of purely financial participation, we would expect implied trading costs to be higher at demand locations relative to generation locations because no market participant can adjust its physical bids to profit from an expected day-ahead/real-time price difference at an individual demand location. Because the introduction of FT allows any market participant to take a financial position in the market for day-ahead/real-time price differences at virtually any location, the reduction in implied trading costs after FT should be larger for demand locations relative to generation locations.

To test these two hypotheses, we regress our estimate of the implied trading cost c_{lower} at each location before and after FT on a constant, an indicator variable that is equal to one if the location is associated with a generation unit ("Generation"), an indicator variable that is equal to one if the implied trading cost is estimated using the post-FT sample period ("Post FT"), and an indicator variable that is equal to one if the observation corresponds to a generation location during the post-FT sample period ("Post FT $\times$ Generation"). The unit of observation for this regression is thus a location in the pre-FT or post-FT sample period. White (1980) standard errors are reported in parentheses below the coefficient estimates. We run the same regression considering our estimate of c_{upper} for each location before and after FT as the dependent variable.

Table 3 presents the estimates from these difference-in-difference regressions. Focusing first on columns 1 and 5, we see that the coefficient estimate on "Post FT" is negative for both c_{lower} and c_{upper}. This indicates that the average level of implied trading costs across locations fell after FT. Moreover, the coefficient estimates corresponding to "Generation" indicate that both c_{lower} and c_{upper} are significantly lower for generation locations relative to demand locations prior to FT.[34] This difference in implied trading costs across generation versus demand locations is essentially eliminated after FT. Specifically, we fail to reject the null hypothesis that the sum of the coefficients corresponding to the variables "Generation" and "Post-FT $\times$ Generation" is zero for both c_{lower} and c_{upper}. Put another way, as we expected, the reduction in implied trading costs after FT is larger for demand locations relative to generation locations.

These results are robust to a variety of alternative specifications. For example, columns 2 and 6 of Table 3 demonstrate that the results continue to hold if we

[34] In online Appendix Table A.3, we split generation locations into two categories: "baseload" locations, where the amount of electricity injected at that location is greater than 0 in at least 75 percent of hours-of-sample, versus "peaker" locations, where the amount of electricity injected at that location was greater than 0 in less than 75 percent of hours-of-sample. The results presented in online Appendix Table A.3 indicate that implied trading costs before FT are lower for baseload generation locations relative to peaker generation locations. This is consistent with the logic that baseload units that operate more frequently find it less costly to adjust their day-ahead schedules relative to their real-time output in order to trade day-ahead/real-time price differences than peaker units that operate less frequently.

Table 3—Implied Trading Costs before versus after FT for Generation versus Demand Locations

	Dependent variable: c_{lower}				Dependent variable: c_{upper}			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post FT × Gen Node	0.508	0.524	0.645	0.166	1.619	1.496	1.655	1.834
	(0.114)	(0.106)	(0.085)	(0.210)	(0.258)	(0.226)	(0.227)	(0.500)
Gen Node	−0.536	−0.482	−0.624	−0.136	−1.363	−1.200	−1.374	−0.218
	(0.107)	(0.099)	(0.072)	(0.199)	(0.219)	(0.208)	(0.182)	(0.342)
Post FT	−3.494	−3.329	−1.247	−9.495	−6.659	−6.577	−3.555	−17.182
	(0.051)	(0.046)	(0.040)	(0.083)	(0.115)	(0.102)	(0.103)	(0.162)
Constant	10.519	10.351	9.124	19.902	18.577	18.306	15.970	36.248
	(0.048)	(0.044)	(0.036)	(0.079)	(0.102)	(0.096)	(0.088)	(0.131)
Residualized	No	No	Yes	No	No	No	Yes	No
Trim top and bottom 1%	No	Yes	No	No	No	Yes	No	No
Full sample period	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Six months	No	No	No	Yes	No	No	No	Yes
Mean of Dep. Var.	8.686	8.614	8.439	15.012	15.083	14.893	14.066	27.532
Std. Dev. of Dep. Var.	2.770	2.770	1.821	5.956	5.920	5.920	4.757	11.314
R^2	0.382	0.405	0.106	0.631	0.296	0.349	0.125	0.558
Number of observations	9,486	9,302	9,485	9,440	9,486	9,298	9,485	9,440

Notes: This table reports estimates from difference-in-difference regressions comparing implied trading costs before versus after the introduction of financial trading for generation locations versus demand locations. The unit of observation is a location before versus after the introduction of FT. We report White (1980) standard errors in parentheses. We trim the top and bottom 1 percent of the dependent variable before estimating the regressions for columns 2 and 6. In columns 3 and 7, implied trading costs are estimated using residualized day-ahead/real-time price spreads; we residualize price spreads by regressing them on natural gas prices at the PG&E and SCG citygates; hourly electricity demand; hourly net electricity imports; and three separate controls for the hourly total production from nuclear, hydro, and renewable sources, respectively. In columns 4 and 8, we estimate implied trading costs using only data from the six months before and after the introduction of FT.

trim observations corresponding to the top 1 percent and bottom 1 percent of the distribution of the relevant dependent variable before estimating the regressions. Moreover, in columns 3 and 7, we residualize hourly day-ahead/real-time prices spreads at each location before estimating c_{lower} and c_{upper}. Specifically, we regress price spreads on natural gas prices at the Pacific Gas and Electricity and Southern California Gas citygates; hourly electricity demand; hourly net electricity imports; and three separate controls for the hourly total production from nuclear, hydro, and renewable sources, respectively. The results based on estimating c_{lower} and c_{upper} using the residuals from this regression are similar to the results from our primary specification in columns 1 and 5. This assuages concerns that our findings are driven by coincident changes over time in factors such as investments in renewables or falling natural gas prices.

As further evidence against this concern, columns 4 and 8 of Table 3 demonstrate that the qualitative conclusions drawn from the analysis remain the same if c_{lower} and c_{upper} are estimated using only the six months before and after the introduction of FT. Longer-run trends in market factors such as changes in natural gas prices and investment in renewables are unlikely to drive the results estimated for this shorter 13-month sample window.[35]

[35] Online Appendix Figures A.7 and D.1 show that production from renewables and natural gas prices did not change substantially in the 13-month window around February 1, 2011. Moreover, this shorter window excludes the closing of the San Onofre nuclear plant in February 2012.

Summarizing, Table 3 provides statistical evidence consistent with all three of our hypotheses: (i) implied trading costs are lower for generation locations relative to demand locations prior to FT, (ii) implied trading costs fell after FT, and (iii) the reduction in implied trading costs after the introduction of FT was smaller for generation locations relative to demand locations.

VI. Suggestive Evidence of the Efficiency Benefits of Financial Trading

Section II argues that the introduction of financial trading has the potential to reduce production costs on days when the solution to the optimization problem required to clear the real-time market involves many binding operating constraints, which we call "high-complexity" days. This section provides evidence in support of this hypothesis.

The section is split into four subsections. The first presents descriptive trends in each of our outcome and control variables to explore concerns that our results may be due to changes in economic conditions coincident with the introduction of FT. The second subsection provides evidence that fuel costs per MWh of gas-fired electricity production fell after FT was introduced on high-complexity days but not low-complexity days. This is consistent with the intuition that there is limited scope for FT to reduce production costs on days when clearing the real-time market involves few, if any, binding operating constraints.

The third subsection presents estimates from a difference-in-difference framework comparing market outcomes before versus after FT on high-complexity days versus low-complexity days. We close by focusing on a specific mechanism by which FT can result in lower production costs on high-complexity days, changes in the number and type of generation units that start up to meet locational demands throughout California. Consistent with the two examples in Section II, our results suggest that quick-start, high-cost units were forced to start up fewer times after the introduction of FT, leading to lower operating costs on high-complexity days after FT.

A. *Data Sources and Descriptive Trends*

Our analysis uses hourly generation unit–level data on input fuel measured in millions of British Thermal Units (MMBTU) and electricity output in MWh from the Continuous Emissions Monitoring Systems database administered by the United States Environmental Protection Agency (US EPA 2009–2012). Our primary analysis focuses on the sample period April 1, 2009 to November 30, 2012. We focus on gas-fired generation units because coal- and oil-fired sources combined accounted for only 1.8 percent (0.8 percent) of the electricity produced in California in 2009 (2012).[36]

The monthly average natural gas price paid by each gas-fired unit is calculated using transaction-level data on the fuel purchased by US power plants from the

[36] Fifty-six and 61 percent of in-state electricity production in California came from gas-fired sources in 2009 and 2012, respectively.

Energy Information Administration (EIA 2009–2012). Specifically, we calculate the monthly average natural gas price paid by plants supplied by Pacific Gas and Electricity (PG&E) versus Southern California Gas (SCG). Gas-fired units located in the PG&E service territory are assigned the gas price corresponding to PG&E, while gas-fired units located in the Southern California Edison (SCE) or San Diego Gas and Electricity (SDG&E) service territories are assigned the gas price for SCG. Online Appendix Section D.2 discusses the construction of natural gas prices and the assignment of these prices to gas-fired units in more detail.

Each unit's hourly total input fuel use is multiplied by the relevant natural gas price to obtain its hourly total fuel cost. We sum the fuel costs and heat energy utilized across units and hours of the day in order to calculate system-wide daily total fuel costs and fuel use. This allows us to construct our two outcome variables: the log of daily total fuel cost divided by daily total gas-fired output in MWh and the log of daily total fuel use per MWh of gas-fired output.[37]

Online Appendix Figure A.10 presents the monthly averages of each of our outcome variables for our sample period. Neither of the market outcomes display clear upward or downward trends over this time period.[38] This provides some comfort that comparing market outcomes before versus after the introduction of financial trading will not be confounded by secular time trends in either outcome variable. Nevertheless, both outcomes exhibit substantial seasonality. The variability induced by this seasonality can obfuscate comparisons of the outcomes across the sample periods before versus after FT. We include fixed effects to account for this seasonality in the model specifications discussed below.

The two market outcomes are also likely to depend on other economic factors that change over time. We therefore control for the monthly average natural gas prices paid by power plants in California. Online Appendix Figure D.1 plots the monthly average natural gas prices paid by plants supplied by PG&E versus SCG over our sample period. This figure documents that natural gas prices did not fluctuate much in the six months before and after the introduction of FT, suggesting that gas prices did not respond to this event. Nevertheless, we believe it is still important to control for natural gas prices given the substantial increase in gas prices between April 2009 and January 2010 as well as the sizable decline in these prices from July 2011 through March 2012.

Our specifications also control for the logs of daily total electricity demand and daily total net electricity imports. These data come from the OASIS API operated by California ISO (California ISO 2009–2012). In addition, we control for the monthly total output from: (i) nuclear plants, (ii) renewable sources, and (iii) hydroelectric

[37] In online Appendix Section E.1, we explore changes in the costs of procuring the operating reserves necessary to ensure that electricity supply equals electricity demand at every instant even in the face of unanticipated changes in physical conditions, such as generation unit outages or transmission outages (Wolak 2019; Buchsbaum et al. 2022).

[38] Online Appendix Table A.4 presents p-values from tests of the null hypothesis that the daily time series of each outcome variable contains a unit root and finds evidence against this null hypothesis for both variables. This suggests that both outcome variables are stationary time series.

units.[39] These control variables are constructed using data on monthly plant-level output from the Energy Information Administration (EIA 2009–2012).

Online Appendix Figure A.7 plots monthly total output by type of generation source, monthly total demand, and monthly total net imports. The decline in output from nuclear sources beginning in March 2012 is due to the shutdown of the San Onofre nuclear power plant discussed in Davis and Hausman (2016). In addition, this figure documents that output from gas-fired units declined between June 2010 and May 2011. This decline is due to reductions in demand coupled with increases in output from hydroelectric sources. These trends emphasize the importance of controlling for system demand, monthly total output from nuclear plants, and monthly total output from hydroelectric sources when exploring changes in fuel costs per MWh of gas-fired output before versus after the introduction of FT.

Online Appendix Figure A.9 plots annual total electricity generating capacity by type. We see from this figure that there were no major investments in generating capacity during our sample period. However, this figure documents a steady increase in the installation of renewable capacity during the latter half of our sample period as well as the aforementioned retirement of the San Onofre nuclear power plant. Controlling flexibly for monthly total output from nuclear plants and renewables is important given these trends. Finally, there were no major investments in electricity transmission capacity in the territory served by California ISO during our sample period (California ISO 2010, 2011, 2012a, 2013).[40]

B. *Empirical Strategy*

Our empirical strategy compares outcomes before versus after financial trading on days when the real-time market-clearing optimization problem is more versus less complex to solve. We consider three indicators of complexity. Our first measure is the standard deviation across pricing locations and hours of the day of real-time prices for each day of the sample period. The intuition behind this measure is that physical operating constraints such as transmission congestion lead to larger deviations in prices across locations on the grid. Days with a larger standard deviation in real-time prices are thus more likely to be days in which more system operating constraints are binding. Our second and third measures of complexity are the level of daily total real-time demand and the daily total number of times that gas-fired units started up. The intuition behind these two measures is that transmission and other operating constraints are more likely to bind during high-demand days when more units have to start up or ramp up to satisfy demand at locations across California.

This subsection and the next subsection present estimates based on measuring complexity using either daily total demand or the daily standard deviation in real-time prices. We discuss results based on daily total starts in the final subsection.

[39] The classification "renewables" includes wind, solar, and geothermal sources as well as hydro sources with capacity less than 30 MW.

[40] The California ISO provides a list of transmission upgrades in each year in its annual report. The list for 2009 is on page 5.26 (California ISO 2010), for 2010 is on pages 125–26 (California ISO 2011), and for 2011 is on page 146 (California ISO 2012a). There were no notable transmission upgrades in 2012 (California ISO 2013).

In online Appendix Section F.1, we show that our three indicators of complexity are positively correlated with each other and are positively correlated with fuel costs per MWh. Moreover, online Appendix Section F.2 provides suggestive evidence that the aggregate marginal cost curve is far steeper at high levels of the residual demand to be served by the gas-fired fleet. Combined, the descriptive evidence presented in online Appendix Sections F.1 and F.2 suggests that suppliers are forced to deploy significantly higher marginal cost units when our three indicators of complexity are large. This supports our logic that the potential for production cost savings from location-specific FT is greatest on high-complexity days.

That being said, there are many concerns with comparing market outcomes before versus after the introduction of financial trading across high- versus low-complexity days. First, the introduction of financial trading is inherently a market-wide shock. There is thus no natural "control group" that is completely unaffected by the change in policy. Consequently, a comparison across high- versus low-complexity days could understate the full benefits of financial trading to the extent that purely financial participation also has benefits on low-complexity days.[41] In addition, this comparison is potentially confounded by unobserved factors that vary over time and differentially impact high-complexity days versus low-complexity days. Indeed, the descriptive trends discussed in the previous subsection indicate that hydroelectric production was unusually high in the first part of 2011, while electricity demand was unusually low during this period.

To explore the importance of these concerns, we compare residualized market outcomes before versus after the introduction of FT. This approach is similar to the event study models commonly employed in finance (MacKinlay 1997; Eckbo 2008). One of the primary considerations in this literature is the length of the window around the "event." Informed by the descriptive trends discussed above, we residualize market outcomes using regressions estimated on the full sample period. This allows us to better account for seasonality as well as idiosyncratic changes in economic conditions such as the unusually high levels of hydroelectric production in the first part of 2011.

We consider the following regression model:

$$
\begin{aligned}
(5) \qquad Y_t = {} & \alpha_{m,\mathrm{HIGH}} + \theta_w + \gamma_{y,m} \\
& + \sum_{s=1}^{4} \sum_{k=1}^{K} \left[\left(X_{k,t} - \bar{X}_k \right)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} \mathbf{1}\left[X_{k,t} \in \mathrm{BIN}_{k,b} \right] \right] + u_t,
\end{aligned}
$$

where Y_t is the outcome variable for day-of-sample t in calendar month m of year y. Depending on specification, either daily total demand or the daily standard deviation in real-time prices is used to categorize days as high versus low complexity. For a given measure, a day is deemed to be high complexity if and only if the value of the measure for the day is above the seventy-fifth percentile of the daily distribution of this measure across days of the sample period.

[41] That being said, we hypothesize that the operating cost reductions from introducing FT are likely to be small on low-complexity days for the reasons noted in Section II.

Equation (5) includes an indicator variable for whether the day of sample is a weekday or weekend (θ_w), month-of-sample fixed effects $(\gamma_{y,m})$, as well as separate sets of calendar month fixed effects for days categorized as "high" versus "low" complexity days $(\alpha_{m,\text{HIGH}})$. In addition, we control for the variables in X_t: the log of daily total electricity demand, the log of daily net electricity imports, the log of monthly average natural gas prices paid by power plants in California ISO, as well as logs of total monthly production from (i) renewables, (ii) nuclear sources, and (iii) hydro sources. Specifically, we center each variable in X_t at its sample mean. For each centered variable x in X_t, equation (5) includes x, x^2, x^3, x^4 and ten separate indicators defined using the deciles of the distribution of x.[42] This allows us to control flexibly for potential nonlinear relationships between the outcome variable and economic factors, which may be especially important given the unusually high levels of hydroelectric production and unusually low levels of demand in the first portion of 2011. In online Appendix Figure E.4, we show that the trends in residualized outcomes from estimating equation (5) remain similar if we only control linearly for the variables in X_t.

Figure 5 presents the monthly average of the residuals calculated from estimating equation (5). The outcome considered in the two left panels of this figure is the log of fuel costs per MWh of gas-fired output, while the two right panels focus on the log of input energy use per MWh produced. The top two panels measure complexity using daily total demand, while the bottom two panels use the daily standard deviation of real-time prices across locations and hours of the day. The vertical black dashed line denotes the introduction of FT. The solid red horizontal lines plot the overall averages of residuals for low-complexity days taken separately over the pre-FT and post-FT sample periods. Similarly, the dashed blue horizontal lines plot overall averages for high-complexity days in the pre-FT and post-FT sample periods.

For both measures of complexity, Figure 5 indicates that the overall averages of residualized fuel costs per MWh and residualized input energy per MWh fell after the introduction of FT on high-complexity days. In contrast, we do not see any significant difference in the overall averages of residualized outcomes before versus after FT on low-complexity days. This is consistent with the mechanism described in Section II of when we should expect purely financial participation to result in lower production costs.

Figure 5 also shows that the decrease in residualized outcomes on high-complexity days after FT does not appear to be driven by preexisting trends in residualized outcomes prior to financial trading being introduced. That being said, one might be concerned that the post-FT reduction in residualized outcomes on high-complexity days is driven by low values of the residuals in the first months after February 2011. To assuage this concern, online Appendix Figure E.3 plots the monthly average residualized outcomes dropping the six months before and after February 2011. Even after dropping these months, we see that monthly average residualized outcomes for high-complexity

[42] Controlling for deciles of the relevant economic factors is similar in spirit to the specification considered in Davis and Hausman (2016).

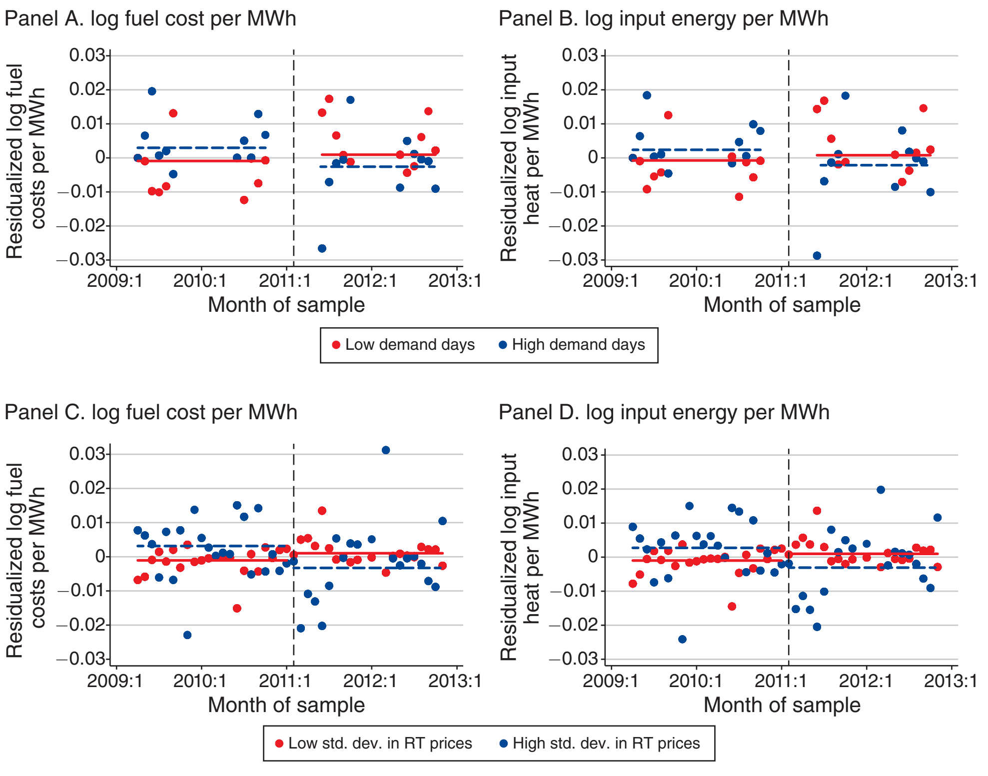


FIGURE 5. MONTHLY AVERAGE RESIDUALIZED OUTCOMES BEFORE VERSUS AFTER FT

Notes: This figure plots the monthly averages of the residualized outcome for high-complexity days versus low-complexity days. We plot only months with both high-complexity days and low-complexity days. We measure complexity using daily total demand for the top two panels and the daily standard deviation over locations and hours of real-time prices for the bottom two panels. For a given measure of complexity, a day is defined as being "high complexity" if the value of the measure on the day is above the seventy-fifth percentile of the distribution of this measure across our sample period. Residuals are calculated based on a daily level regression of the relevant outcome on separate sets of calendar month fixed effects for high- versus low-complexity days, month-of-sample fixed effects, an indicator for weekday versus weekend, and the control variables X_t described in Section VIB.

days fall after FT, while average residualized outcomes for low-complexity days are not statistically different across the two time periods.

C. *Difference-in-Difference Framework*

Figure 5 provides evidence that residualized fuel costs per MWh and residualized fuel use per MWh fell after FT was introduced on high-complexity days but not on low-complexity days. Further, this figure suggests that the trends in residualized outcomes across high- versus low-complexity days are similar before the introduction of FT. We test this "common trends" assumption formally by regressing the first difference of each outcome on an indicator for high-complexity days. The results of this analysis, presented in online Appendix Table E.1, indicate that there is no statistical difference in the time trends of the two outcome variables on high- versus low-complexity days prior to the introduction of FT.

With this evidence in hand, we quantify the difference in each market outcome on high-complexity days relative to low-complexity days after relative to before the introduction of FT using the following regression:

$$(6) \qquad Y_t = \alpha_{m,\mathrm{HIGH}} + \theta_w + \gamma_{y,m} + \delta_{DD}(\mathrm{HIGH}_t \times \mathrm{POSTFT}_t)$$
$$+ \sum_{s=1}^{4} \sum_{k=1}^{K} \left[(X_{k,t} - \bar{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} \mathbf{1}[X_{k,t} \in \mathrm{BIN}_{k,b}] \right] + u_t,$$

where we define HIGH_t to be an indicator that is equal to one if and only if the relevant indicator of complexity on day-of-sample t is above the seventy-fifth percentile of the distribution of this measure across our sample period. The independent variable of interest is $\mathrm{HIGH}_t \times \mathrm{POSTFT}_t$. This variable captures the change in outcome after the introduction of FT on high-complexity days relative to low-complexity days.

We control for the set of variables X_t in the same way as in equation (5). As before, we include separate sets of calendar year fixed effects for high-complexity days and low-complexity days ($\alpha_{m,\mathrm{HIGH}}$), weekend versus weekday fixed effects (θ_w), and month-of-sample fixed effects ($\gamma_{y,m}$). Standard errors are clustered by week-of-sample.

Table 4 presents the estimates from the difference-in-difference framework. For the first four columns of this table, we define high-complexity days using daily total demand; the last four columns present results based on defining high-complexity days using the daily standard deviation in real-time prices. Column 1 of the top panel of Table 4 indicates that average fuel costs per MWh fell by 2 percent after the introduction of FT on relatively high-demand days. Aggregating across all of the power plants in our sample, this 2 percent decrease in fuel costs per MWh corresponds to a \$16.6 million reduction in the annual fuel costs incurred on high-demand days. Similarly, we find a 1.5 percent reduction in average input heat energy per MWh on high-demand days after FT was introduced (see column 1 of the bottom panel of Table 4). A 1.5 percent decrease in thermal energy per MWh translates into an annual reduction in CO_2 emissions of 160,635 tons on high-demand days.

Columns 5–8 of Table 4 present estimates based on defining high-complexity days as days above the seventy-fifth percentile of the distribution of the daily standard deviation in real-time prices. The estimated reductions in fuel costs per MWh and input energy use per MWh remain precisely estimated for this alternative indicator of complexity. In addition, columns 2 and 5 of Table 4 demonstrate that the results remain similar if we trim observations corresponding to the top and bottom 1 percent of the distribution of the outcome variable. Moreover, columns 3 and 6 indicate that the estimates remain precisely estimated if we trim observations corresponding to the top and bottom 1 percent of the distribution of the residualized outcome, where outcomes are residualized using equation (5). Combined, columns 2, 3, 5, and 6 assuage concerns that the estimates are driven by outliers either in the outcome or the residualized outcome. Finally, for columns 4 and 8, we exclude

Table 4—Difference-in-Differences: Changes after FT on Relatively High-Complexity Days

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Dep. var.: log fuel cost per MWh								
$\text{HIGH}_t \times \text{POSTFT}_t$	−0.020	−0.019	−0.015	−0.018	−0.013	−0.012	−0.010	−0.015
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
R^2	0.963	0.963	0.971	0.957	0.961	0.962	0.970	0.955
Mean of dep. var.	3.680	3.679	3.677	3.680	3.680	3.679	3.677	3.680
Number of observations	1,340	1,313	1,313	1,340	1,340	1,313	1,312	1,340
Measure: Total demand	Y	Y	Y	Y	N	N	N	N
Measure: SD RT price	N	N	N	N	Y	Y	Y	Y
Trimmed dep. var.?	N	Y	N	N	N	Y	N	N
Trimmed res. dep. var.?	N	N	Y	N	N	N	Y	N
No nonlinear controls	N	N	N	Y	N	N	N	Y
Panel B. Dep. var.: log input energy use per MWh								
$\text{HIGH}_t \times \text{POSTFT}_t$	−0.015	−0.014	−0.012	−0.014	−0.013	−0.012	−0.009	−0.014
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
R^2	0.729	0.714	0.775	0.688	0.722	0.707	0.773	0.680
Mean of dep. var.	2.051	2.050	2.049	2.051	2.051	2.050	2.049	2.051
Number of observations	1,340	1,314	1,314	1,340	1,340	1,314	1,312	1,340
Measure: Total demand	Y	Y	Y	Y	N	N	N	N
Measure: SD RT price	N	N	N	N	Y	Y	Y	Y
Trimmed dep. var.?	N	Y	N	N	N	Y	N	N
Trimmed res. dep. var.?	N	N	Y	N	N	N	Y	N
No nonlinear controls	N	N	N	Y	N	N	N	Y

Notes: This table presents the difference-in-difference estimates of the change in outcome after the introduction of financial trading on high-complexity days relative to low-complexity days. The unit of observation for these regressions is day-of-sample. The dependent variable considered in the top (bottom) panel of this table is the log of fuel costs per MWh (the log of input energy per MWh). For the first four columns of each panel, the indicator variable HIGH_t is equal to one if and only if the daily total demand on day t is higher than the seventy-fifth percentile of the distribution of daily total demand across our sample period. For the last four columns, HIGH_t is equal to one if the daily standard deviation of real-time prices over locations and hours in day t is greater than the seventy-fifth percentile of the distribution of daily standard deviations. All specifications include separate sets of calendar month fixed effects for high- versus low-complexity days, month-of-sample fixed effects, an indicator for weekday versus weekend, and the control variables X_t described in Section VIB. In columns 2 and 6 (columns 3 and 7), we trim the top 1 percent and bottom 1 percent of observations based on the dependent variable (dependent variable residualized using equation (5)). In column 8, we exclude the nonlinear functions of the variables in X_t from the regression. Standard errors are clustered by week-of-sample and are reported in parentheses.

the quadratic, cubic, and quartic terms for each (centered) control variable in X_t as well as the ten separate indicators defined based on the deciles of the variable. The estimates presented in columns 4 and 8 are similar to those presented in the other columns, suggesting further that our results are not driven by nonlinear relationships between the control variables and the outcome.

In online Appendix Tables E.2 and E.3, we define high-complexity days based on the fiftieth, sixtieth, seventieth, eightieth, or ninetieth percentiles of the distribution of the relevant indicator of complexity. The estimated reductions in fuel cost per MWh and input energy use remain precisely estimated regardless of the indicator of complexity considered or the cutoff used to define high-complexity days. These estimated reductions generally increase in absolute value as the cutoff increases, providing suggestive evidence that the per MWh benefits from purely financial participation are larger on days when it is more difficult to find the least-cost combination of real-time output levels of all generation units.

D. *Changes in Total Starts as a Measure of Complexity*

In the first example in Section II, purely financial trades resulted in real-time demand being satisfied by a long-start unit with lower marginal costs rather than a quick-start unit with higher marginal costs. Purely financial trades are more likely to lead to changes in which units are dispatched when a significant number of operating constraints bind in real time. Based on this intuition, we hypothesize that the number of units that must start up to meet daily total demand on high-complexity days falls after the introduction of FT. Moreover, we expect these reductions in starts to be more pronounced when focusing on units with shorter start-up times but higher marginal costs rather than units with longer start-up times but smaller marginal costs.

Exploring our first hypothesis, online Appendix Table F.2 documents how the number of starts by gas-fired units per MWh of gas-fired output changes after FT on days when clearing real-time markets is relatively complex. We consider two different indicators of complexity: (i) daily total demand and (ii) the daily standard deviation across locations and hours in real-time prices. For both measures, we find that the number of starts per MWh decreases after FT was introduced on high-complexity days.

As discussed above, FT potentially influences not just the number of units that start up but the type of units that start up. As evidence of this, online Appendix Table F.3 documents that the number of starts by high–variable cost units falls relative to the number of starts by low–variable cost units on high-complexity days after FT is introduced.[43] This result holds for both of the indicators of complexity discussed above. This suggests that purely financial participation potentially impacts production costs both through reductions in starts and changes in the type of units that start up.

We isolate how production costs adjust to changes in the type of units that start up after FT using a difference-in-difference framework similar to the one specified in equation (6). Specifically, we estimate how production costs change after FT on days with a larger versus smaller number of starts. Days are defined to have a high number of starts if the daily total number of starts by gas-fired units is above the X^{th} percentile of the daily distribution of starts; we consider the fiftieth, sixtieth, seventieth, eightieth, and ninetieth percentiles of daily total starts as potential cutoffs. The results from this specification are presented in online Appendix Table F.1. We find that fuel costs per MWh and input energy use per MWh fall by roughly 1 percent after FT on days with a high number of starts relative to days with a low number of starts. Combined, these results suggest that (i) purely financial trades lead suppliers to shift toward utilizing lower-cost units rather than starting up high-cost units to meet real-time locational demand and (ii) this shift is associated with meaningful reductions in aggregate production costs.

[43] A unit is categorized as high (low) variable cost if its aggregate fuel cost divided by its aggregate output is above (below) the median of this magnitude across units.

VII. Implications for Electricity Market Design

In this paper, we formulate a new statistical test of the null hypothesis that expected profits can be earned by trading differences between forward and spot commodity prices that accounts for the presence of transaction costs. Inverting this test, we provide evidence that day-ahead prices better reflect real-time prices after the introduction of financial trading. This post-FT increase in the agreement between day-ahead and real-time prices is more pronounced at demand locations relative to generation locations. This outcome is consistent with the fact that suppliers had the limited ability to trade day-ahead/real-time price differences at locations where they owned generation units before FT. In contrast, retailers could only submit service territory–level bids, which made it impossible to trade day-ahead/real-time price differences at specific demand locations prior to FT.

The second half of the paper explored the extent to which daily total production costs and input fuel use changed after the introduction of FT. Our empirical strategy is based on the hypothesis that purely financial participation is likely to result in reductions in production cost when a significant fraction of transmission network and system operating constraints are likely to bind in the real-time market. This is because generation unit owners must make irreversible sunk commitments to start up their units in the day-ahead market in order for these units to be available to produce the following day. The process of finding the least-cost combination of units to start up is significantly more complex when many operating constraints bind in real time. Under these conditions, financial participants trading day-ahead/real-time price differences at thousands of locations in the transmission network have the potential to yield lower cost solutions.

We provide suggestive evidence within an event study framework consistent with this hypothesis. Specifically, our results indicate that residualized fuel cost per MWh of gas-fired output and input fuel use per MWh of gas-fired output fell after financial trading was introduced on days when the complexity of the real-time market-clearing problem was high but not on days when complexity was low. This is true regardless of whether complexity is measured using daily total demand, the daily standard deviation across locations and hours in real-time prices, or the total number of daily starts. Finally, using a difference-in-difference framework, we estimate that fuel costs per MWh (input energy use per MWh) fell by 2.0 percent (1.5 percent) after the introduction of financial trading on high-demand days relative to low-demand days. Decreasing fuel costs per MWh by 2 percent on high-demand days would result in a $16.6 million reduction in annual total fuel costs on average.

Our results have important implications for the design of wholesale electricity markets with large shares of intermittent renewable resources. As the share of electricity demand met by intermittent renewables in a region grows, system operators are likely to be required to impose more operating constraints on day-ahead and real-time markets in order to maintain supply and demand balance throughout the day.[44] Controllable generation resources with positive start-up costs are also likely

[44] For example, the California ISO implemented a flexible ramping product constraint to ensure that there is sufficient available output from controllable generation capacity when output from solar resources declines in the

to start up and shut down more frequently because intermittent renewables can start and stop producing with very little advance notice (Schill, Pahle, and Gambardella 2017). An increasing fraction of transmission network constraints are also likely to bind in real time as the percentage of electricity production from wind and solar resources distributed across the grid increases. The location-specific bids and offers submitted by purely financial participants can reduce day-ahead/real-time price differences caused by day-ahead uncertainty in intermittent renewable output. These purely financial bids and offers can also ensure that the least-cost combination of controllable resources are available to compensate for fluctuations in real-time output from intermittent renewables on high-complexity days.

It is generally acknowledged that investment in renewables must be coupled with investments in electricity storage in order to substantially reduce the carbon emissions from burning fossil fuels to produce electricity. Storing excess renewable energy and withdrawing it when little renewable energy is being produced can help ensure that electricity demands across the grid are satisfied in settings with a large share of intermittent renewable generation.[45] However, it is unclear precisely how to utilize a portfolio of storage resources to reduce the cost of serving demands at all locations in the transmission network. The actions of purely financial participants have the potential to result in real-time injection and withdrawal actions by storage resources that lower the cost of serving real-time locational demands. The intuition discussed in this paragraph and the previous one suggest that it is increasingly important to foster purely financial participation in multisettlement locational marginal pricing markets as regions across the United States and around the world take steps to transition away from fossil fuel–fired production in favor of low-carbon intermittent renewable resources.

The degree of financial participation in the day-ahead market is tied to the per unit costs of trading in the market. With this motivation in mind, we use our estimates to provide a back-of-the-envelope calculation of the benefits from reducing the per unit costs of trading in California's wholesale electricity market. Specifically, our estimates of implied trading costs fell by \$3.41 per MWh on average after financial trading was introduced.[46] We also estimate that fuel costs per MWh decrease by 76 cents on high-demand days relative to low-demand days after purely financial participation was allowed. Taken at face value, our estimates thus imply that a 5 cent reduction in transaction costs corresponds to a roughly 1.11 $(= 5 \times (0.76/3.41))$ cent reduction in fuel costs per MWh on high-demand days on average. However, there is substantial heterogeneity across locations and days in our estimates of the costs and benefits from financial trading. This suggests that the efficiency gains from reducing transaction fees would likely be even larger if these fees were allowed to

early evening. For more information, see http://www.caiso.com/informed/Pages/StakeholderProcesses/CompletedClosedStakeholderInitiatives/FlexibleRampingProduct.aspx.

[45] For example, solar resources produce only during the day when the sun is shining. In contrast, wind resources in California typically produce the most energy during the early morning and late evening.

[46] The average across all nodes of the fifth percentile of the bootstrapped distribution of the maximum over hours of absolute average day-ahead/real-time price spreads is 10.43 (7.02) for the pre-FT sample period (post-FT sample period).

vary by location and time in order to reflect differences in the expected benefits from financial trading across locations and days.

REFERENCES

Bessembinder, Hendrik, and Michael L. Lemmon. 2002. "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets." *Journal of Finance* 57 (3): 1347–82.

Bessembinder, Hendrik, and Michael L. Lemmon. 2006. "Gains from Trade under Uncertainty: The Case of Electric Power Markets." *Journal of Business* 79 (4): 1755–82.

Birge, John R., Ali Hortaçsu, Ignacia Mercadal, and J. Michael Pavlin. 2018. "Limits to Arbitrage in Electricity Markets: A Case Study of MISO." *Energy Economics* 75: 518–33.

Bohn, Roger E., Michael C. Caramanis, and Fred C. Schweppe. 1984. "Optimal Pricing in Electrical Networks over Space and Time." *Rand Journal of Economics* 15 (3): 360–76.

Borenstein, Severin, James Bushnell, Christopher R. Knittel, and Catherine Wolfram. 2008. "Inefficiencies and Market Power in Financial Arbitrage: A Study of California's Electricity Markets." *Journal of Industrial Economics* 56 (2): 347–78.

Buchsbaum, Jesse F., Catherine Hausman, Johanna L. Mathieu, and Jing Peng. 2022. "Spillovers from Ancillary Services to Wholesale Energy Markets." NBER Working Paper 28027.

Burer, Samuel, and Adam N. Letchford. 2012. "Nonconvex Mixed-Integer Nonlinear Programming: A Survey." *Surveys in Operations Research and Management Science* 17 (2): 97–106.

California ISO-Department of Market Monitoring. 2010. *2009 Annual Report on Market Issues and Performance*. Folsom, CA: Department of Market Monitoring-California ISO.

California ISO-Department of Market Monitoring. 2011. *2010 Annual Report on Market Issues and Performance*. Folsom, CA: Department of Market Monitoring-California ISO.

California ISO-Department of Market Monitoring. 2012a. *2011 Annual Report on Market Issues and Performance*. Folsom, CA: Department of Market Monitoring-California ISO.

California ISO-Department of Market Monitoring. 2012b. *Q3 2012 Report on Market Issues and Performance*. https://www.caiso.com/Documents/2012ThirdQuarterReport-MarketIssues-Performance-Nov2012.pdf.

California ISO-Department of Market Monitoring. 2013. *2012 Annual Report on Market Issues and Performance*. Folsom, CA: Department of Market Monitoring-California ISO.

California Independent System Operator. 2009–2012. "Open Access Same-time Information System." http://oasis.caiso.com/.

California Independent System Operator. 2015. "Convergence Bidding: Settlement Implications and Convergence Bidding Fees." https://www.caiso.com/Documents/ConvergenceBiddingSession7-ConvergenceBiddingFees.pdf.

California Independent System Operator. 2023. "List of SCs, CRR Holders, Convergence Bidding Entities." http://www.caiso.com/PublishedDocuments/ListofSchedulingCoordinatorsCRRHoldersandConvergenceBiddingEntities.pdf.

Cheng, Ing-Haw, and Wei Xiong. 2014. "Financialization of Commodity Markets." *Annual Review of Financial Economics* 6 (1): 419–41.

Cicala, Steve. 2015. "When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation." *American Economic Review* 105 (1): 411–44.

Cox, Charles C. 1976. "Futures Trading and Market Information." *Journal of Political Economy* 84 (6): 1215–37.

Dávila, Eduardo, and Cecilia Parlatore. 2021. "Trading Costs and Informational Efficiency." *Journal of Finance* 76 (3): 1471–1539.

Davis, Lucas, and Catherine Hausman. 2016. "Market Impacts of a Nuclear Power Plant Closure." *American Economic Journal: Applied Economics* 8 (2): 92–122.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74 (366a): 427–31.

Eckbo, B. Espen. 2008. *Handbook of Empirical Corporate Finance*. Amsterdam: Elsevier.

Fang, Zheng, and Andres Santos. 2019. "Inference on Directionally Differentiable Functions." *Review of Economic Studies* 86 (1): 377–412.

Federal Energy Regulatory Commission. 2014. "Staff Analysis of Uplift in RTO and ISO Markets." https://www.ferc.gov/legal/staff-reports/2014/08-13-14-uplift.pdf.

Federal Energy Regulatory Commission. 2015. "Energy Primer: A Handbook of Energy Market Basics." https://www.ferc.gov/sites/default/files/2020-05/energy-primer.pdf.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz. 2018. "Trading costs." Unpublished.

Gallamore, Robert E., and John Robert Meyer. 2014. *American Railroads*. Cambridge, MA: Harvard University Press.

Gray, Roger W. 1964. "The Attack upon Potato Futures Trading in the United States." *Food Research Institute Studies* 4: 97–121.

Hadsell, Lester. 2007. "The Impact of Virtual Bidding on Price Volatility in New York's Wholesale Electricity Market." *Economics Letters* 95 (1): 66–72.

Hirshleifer, David. 1988. "Residual Risk, Trading Costs, and Commodity Futures Risk Premia." *Review of Financial Studies* 1 (2): 173–93.

Hong, Han, and Jessie Li. 2018. "The Numerical Delta Method." *Journal of Econometrics* 206 (2): 379–94.

Isemonger, Alan G. 2006. "The Benefits and Risks of Virtual Bidding in Multisettlement Markets." *Electricity Journal* 19 (9): 26–36.

Ito, Koichiro, and Mar Reguant. 2016. "Sequential Markets, Market Power, and Arbitrage." *American Economic Review* 106 (7): 1921–57.

Jha, Akshaya, and Frank Wolak. 2023. "Replication data for: Can Forward Commodity Markets Improve Spot Market Performance? Evidence from Wholesale Electricity." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E141521V1.

Kunsch, Hans R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics* 17 (3): 1217–41.

Kurlinski, Ryan E. 2013. "Real-time Revenue Imbalance in CAISO Markets." http://www.caiso.com/documents/discussionpaper-real-timerevenueimbalance_californiaiso_markets.pdf.

Li, Ruoyang, Alva J. Svoboda, and Shmuel S. Oren. 2015. "Efficiency Impact of Convergence Bidding in the California Electricity Market." *Journal of Regulatory Economics* 48 (3): 245–84.

Longstaff, Francis A., and Ashley W. Wang. 2004. "Electricity Forward Prices: A High-Frequency Empirical Analysis." *Journal of Finance* 59 (4): 1877–1900.

MacKinlay, A. Craig. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature* 35 (1): 13–39.

MacKinnon, James G. 1994. "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests." *Journal of Business & Economic Statistics* 12 (2): 167–76.

Mercadal, Ignacia. 2022. "Dynamic Competition and Arbitrage in Electricity Markets: The Role of Financial Players." *American Economic Journal: Microeconomics* 14 (3): 665–99.

Mercier, Anne, Jean-François Cordeau, and François Soumis. 2005. "A Computational Study of Benders Decomposition for the Integrated Aircraft Routing and Crew Scheduling Problem." *Computers & Operations Research* 32 (6): 1451–76.

Parsons, John E., Cathleen Colbert, Jeremy Larrieu, Taylor Martin, and Erin Mastrangelo. 2015. "Financial Arbitrage and Efficient Dispatch in Wholesale Electricity Markets." Unpublished.

Phillips, Peter C. B., and Pierre Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika* 75 (2): 335–46.

Routledge, Bryan R., Duane J. Seppi, and Chester S. Spatt. 2001. "The "Spark Spread": An Equilibrium Model of Cross-Commodity Price Relationships in Electricity." Unpublished.

S&P Global Platts. 2009–2012. "Citygate Spot Natural Gas Prices." Data retrieved January 24, 2020 from Bloomberg terminal.

Saravia, Celeste. 2003. "Speculative Trading and Market Performance: The Effect of Arbitrageurs on Efficiency and Market Power in the New York Electricity Market." Center for the Study of Energy Markets Working Paper 121.

Schill, Wolf-Peter, Michael Pahle, and Christian Gambardella. 2017. "Start-up Costs of Thermal Power Plants in Markets with Increasing Shares of Variable Renewable Generation." *Nature Energy* 2 (6): 1–6.

Schmid, Friedrich, and Mark Trede. 1996. "Testing for First-Order Stochastic Dominance: A New Distribution-Free Test." *Journal of the Royal Statistical Society: Series D* (*The Statistician*) 45 (3): 371–80.

Schweppe, Fred C., Michael C. Caramanis, Richard D. Tabors, and Roger E. Bohn. 2013. *Spot Pricing of Electricity*. Boston: Kluwer Academic Publishers.

US Energy Information Administration. 2009–2012. "Form EIA-923: Data on fuel purchases and electricity output for U.S. power plants." https://www.eia.gov/electricity/data/eia923/.

US Environmental Protection Agency. 1996–2012. "eGrid Database." https://www.epa.gov/egrid/download-data.

US Environmental Protection Agency. 2009–2012. "Continuous Emissions Monitoring Database." https://gaftp.epa.gov/DMDnLoad/emissions/hourly/monthly/.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.

Williams, Jeffrey. 1987. "Futures Markets: A Consequence of Risk Aversion or Transactions Costs?" *Journal of Political Economy* 95 (5): 1000–1023.

Wolak, Frank A. 1989. "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models." *Econometric Theory* 5 (1): 1–35.

Wolak, Frank A. 2019. "The Role of Efficient Pricing in Enabling a Low-Carbon Electricity Sector." *Economics of Energy & Environmental Policy* 8 (2).

Working, Holbrook. 1953. "Hedging Reconsidered." *Journal of Farm Economics* 35 (4): 544–61.

Online Appendix

Can Forward Commodity Markets Improve Spot Market Performance? Evidence from Wholesale Electricity

by Akshaya Jha and Frank A. Wolak

A Additional Tables and Figures

A.1 Additional Tables and Figures: Financial Participation

Appendix Table A.1 documents that physical participants account for only roughly 4% of financial trading (FT) volumes and less than 0.5% of net revenues from financial trades in 2012.[47] It is thus unlikely that physical participants have better information than financial participants as it relates to profiting from expected day-ahead/real-time price differences. The fact that physical participants represent such a low percentage of trading volumes also makes it unlikely that they use purely financial bids to hedge against day-ahead price and demand uncertainty.

The left panel of Appendix Figure A.1 plots the monthly average hourly volume of purely financial trades submitted and cleared in the day-ahead market over the period October 2011 to December 2012. The right panel plots the average for each hour of the day of trading volumes submitted and cleared for this same time period. These panels document that the absolute net volume of financial trades submitted and cleared is larger during the summer months and in the evening, both time periods when generation unit and system operating constraints are more likely to bind in the real-time market. That being said, the changes in financial trading volumes across months and hours documented in Appendix Figure A.1 are relatively small, especially when compared to the large increase in forward market liquidity due to the introduction of financial trading.

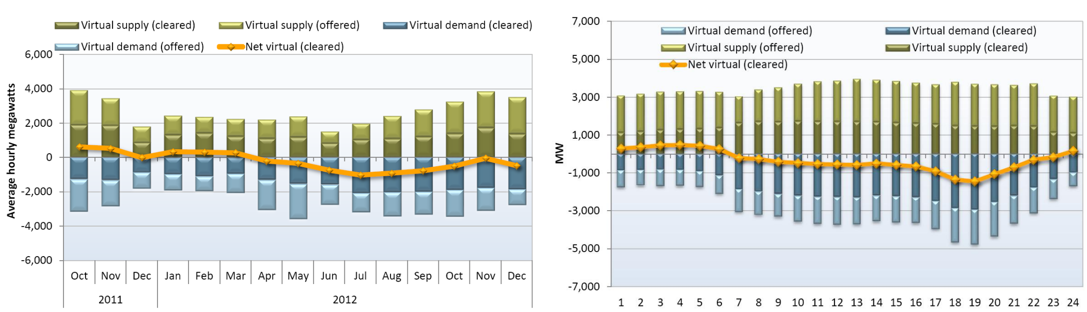
[47] This table is reproduced from CAISO's 2012 Annual Report (CAISO (2012a)).

Table A.1: Financial Trading Volumes and Revenues by Participant Type in 2012

Average Hourly Megawatts			
Trading Entities	**Virtual Demand**	**Virtual Supply**	**Total**
Financial	1,049	757	1,807
Marketer	467	374	841
Physical Generation	61	70	131
Physical Load	8	36	45
Revenues (Million Dollars)			
Trading Entities	**Virtual Demand**	**Virtual Supply**	**Total**
Financial	31.2	18.7	49.9
Marketer	6.8	-0.3	6.5
Physical Generation	1.8	0.0	1.8
Physical Load	-1.1	-0.5	-1.6

Notes: This is Table E.1 from CAISO's 2012 Annual Report (CAISO (2012a)). Financial entities are defined as "participants who control no physical power, do not serve any load, and participate in only the convergence bidding and congestion revenue rights markets." In contrast, generation unit owners are in the "Physical Generation" category while electricity retailers are in the "Physical Load" category.

Figure A.1: Monthly and Hourly Averages of Trading Volumes



(a) Monthly Averages (b) Hourly Averages

Notes: The left panel of this figure plots the monthly average of the hourly volume of trades submitted and cleared in the day-ahead market over the period October 2011 to December 2012. Trades are split by whether the offer corresponded to buying electricity (virtual demand) or selling electricity (virtual supply) in the day-ahead market. The right panel of this figure plots the average for each hour of the day of trading volumes submitted and cleared, once again split out by virtual supply versus virtual demand. These figures are from page 103 of CAISO (2012a).

A.2 Day-Ahead and Real-time Prices by Service Territory

California is home to three major investor-owned utilities: Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E). Appendix Figure A.2 presents a map of the territories served by each of California's investor-owned utilities.

Appendix Figure A.3 presents monthly average day-ahead and real-time prices paid by each of California's three major investor-owned utilities.[48] Specifically, the top left panel, the top right panel, and the bottom left panel plot the quantity-weighted average of prices over locations in the territories served by PG&E, SCE, and SDG&E, respectively. The bottom right panel of Appendix Figure A.3 plots the monthly average day-ahead price minus the monthly average real-time price for each of the three utilities. A vertical dashed black line is placed at February 2011 to indicate that financial trading was introduced in California's wholesale electricity market on February 1, 2011. It is immediately apparent from this figure that: (1) before FT, day-ahead prices are consistently below real-time prices on average and (2) the average day-ahead/real-time price spread is smaller in absolute value after February 1, 2011.

Appendix Figure A.4 presents daily average day-ahead/real-time price spreads for each of the 24 hours of the day along with their pointwise 95% confidence intervals. As before, we focus on PG&E, SCE, and SDG&E. There are separate plots for the sample periods before versus after FT is introduced.

Appendix Figure A.4 demonstrates that day-ahead/real-time price spreads are larger in absolute value before the introduction of FT for all three of the utilities. For example, before FT, day-ahead prices for PG&E are much lower than real-time prices on average for the hours of 8PM to 12AM. Indeed, prior to FT, the 95% confidence interval around average price spreads does not include zero for many hours of the day for all three utilities. In contrast, after FT, the 95% confidence interval covers zero

[48] As noted in Section I.C, these prices are quantity-weighted averages of the locational prices in each utility's service territory. Hourly day-ahead and real-time prices for each utility can be downloaded from the OASIS API administered by California's Independent System Operator (CAISO, 2009-2012).

Figure A.2: Territories Served by California's Three Major Investor-Owned Utilities

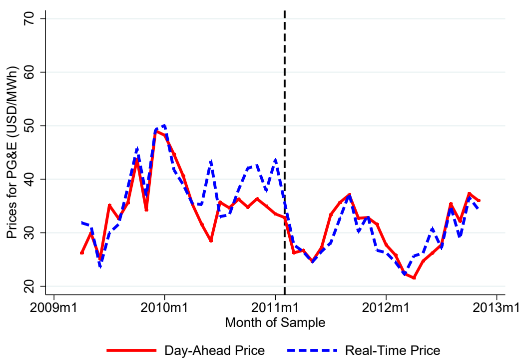
Notes: This is a map of the territories served by each of the three major investor-owned electric utilities in California. These three utilities are Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E). This map is reproduced from FERC (2015).

for the vast majority of hours of the day for each of the utilities.

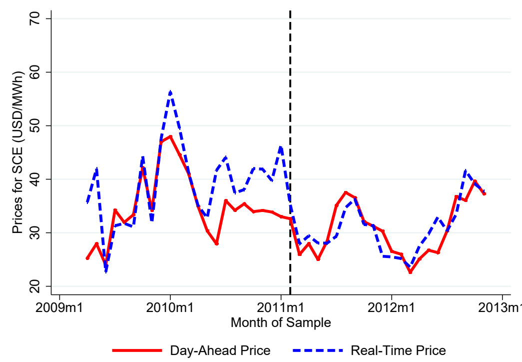
These plots also demonstrate that day-ahead prices are lower than real-time prices on average for the majority of hours of the day for all three utilities prior to FT. This is consistent with the results in Borenstein et al. (2008), which argues that large retailers in California withheld demand from the day-ahead market in order to lower day-ahead prices prior to FT. This strategy was likely to increase the utility's profits because it purchased the bulk of its energy from the day-ahead market. Day-ahead/real-time price spreads do not seem to be persistently negative or persistently positive after FT.

In Appendix Section C.1, we demonstrate that the post-FT reduction in average day-ahead/real-time price differences is statistically different from zero. As shown in Section III.C, the volatility of both day-ahead/real-time price spreads and real-time prices fell after the introduction of FT. The reduction in both the mean and volatility of price spreads after February 1st 2011 is consistent with day-ahead prices better reflecting real-time prices after FT was introduced.

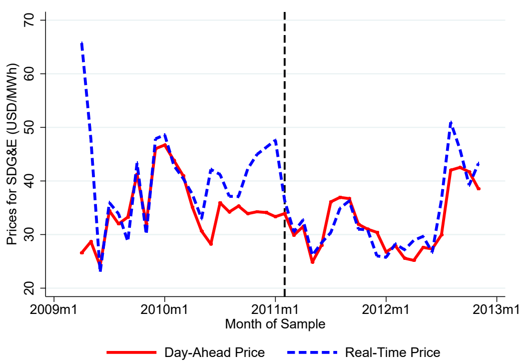
Figure A.3: Monthly Average Day-Ahead and Real-Time Prices By Service Territory



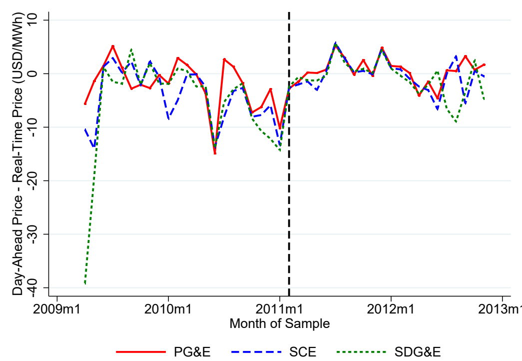
(a) Pacific Gas and Electric



(b) Southern California Edison



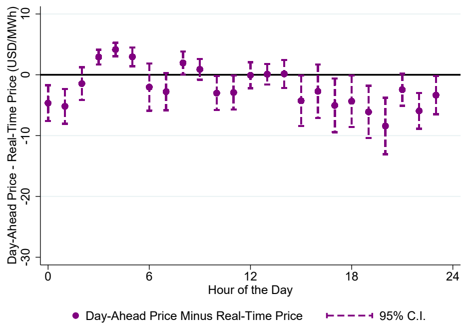
(c) San Diego Gas and Electric



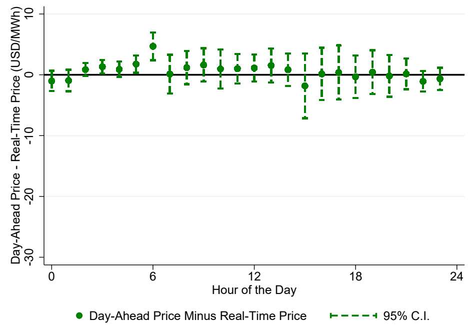
(d) Day-Ahead Price − Real-Time Price

Notes: The top left, top right, and bottom left panels of this figure present the monthly average day-ahead price and the monthly average real-time price paid by PG&E, SCE, and SDG&E respectively. The bottom right panel presents the monthly average day-ahead price minus the monthly average real-time price for each of the three aforementioned electric utilities.

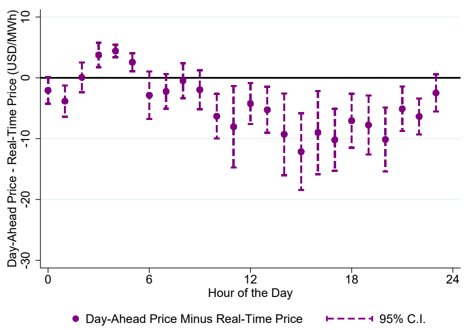
Figure A.4: Hourly Average Day-Ahead/Real-Time Price Spreads: Before and After Financial Trading



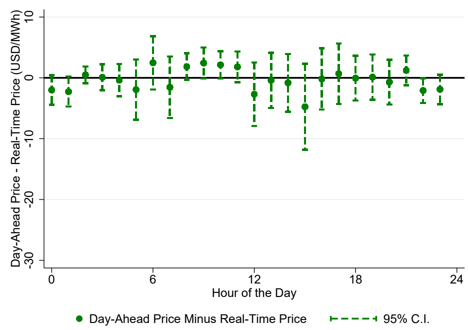
(a) PG&E, Before Financial Trading



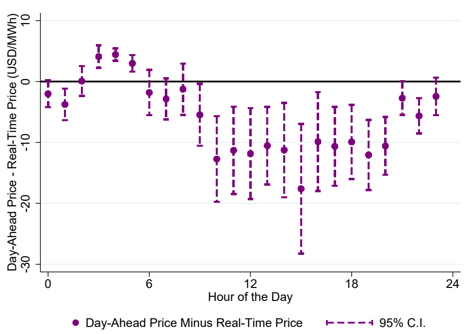
(b) PG&E, After Financial Trading



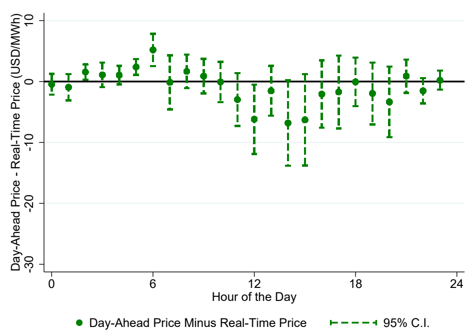
(c) SCE, Before Financial Trading



(d) SCE, After Financial Trading



(e) SDG&E, Before Financial Trading



(f) SDG&E, After Financial Trading

Notes: This figure presents the hourly average day-ahead price minus the hourly average real-time price for the following three electric utilities: Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E). We plot hourly average day-ahead/real-time price spreads separately for the sample periods before versus after financial trading was introduced. This figure also includes the pointwise 95% confidence interval associated with the average day-ahead/real-time price spread for each hour of the day.

A.3 Additional Tables and Figures: Implied Trading Costs

A.3.1 Results By Service Territory

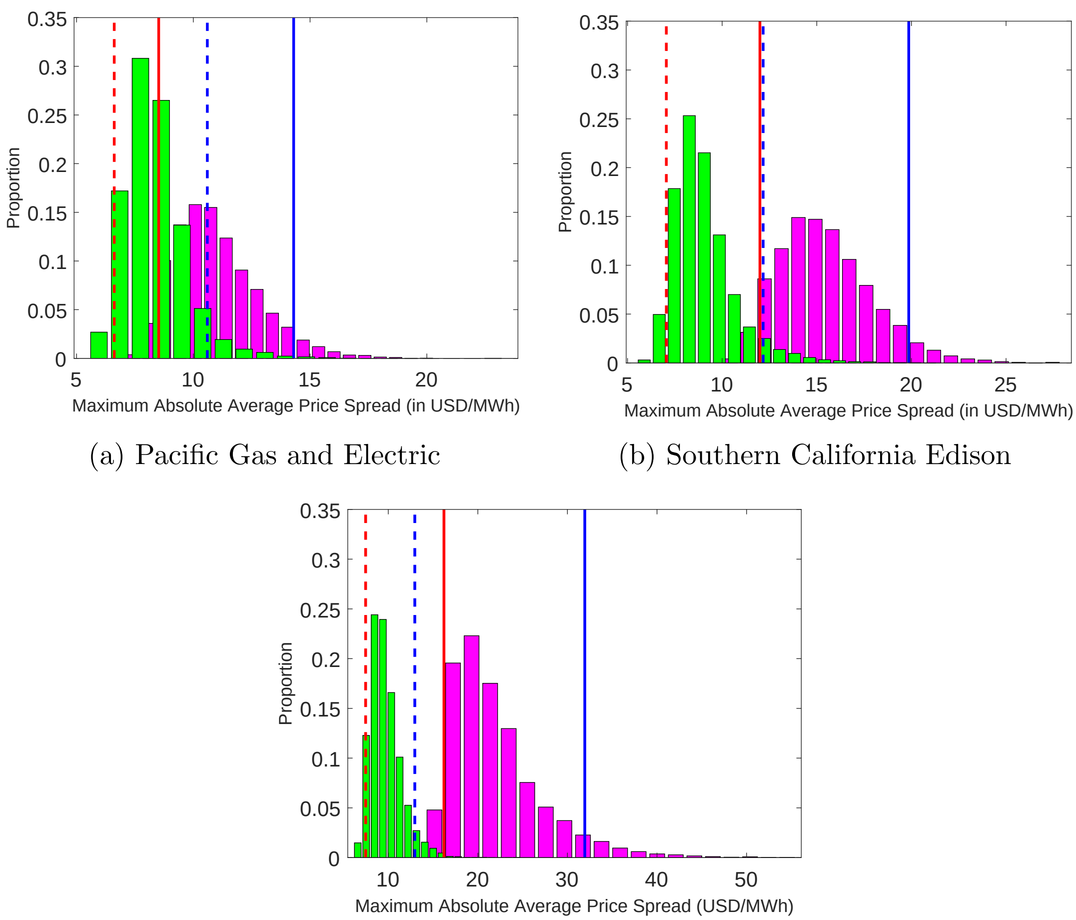
Appendix Table A.2 reports estimates of our two measures of implied trading costs before and after the implementation of FT for the day-ahead/real-time price spreads corresponding to the territories served by PG&E, SCE, and SDG&E. Recall that, as discussed in Section IV.B, c_{lower} is the smallest value of per-unit trading cost for which we can reject the null hypothesis that a profitable strategy exists while c_{upper} is the largest value of trading cost for which we can reject the null hypothesis that no profitable trading strategy exists. Appendix Table A.2 demonstrates that our estimates of c_{lower} and c_{upper} are substantially lower after the introduction of financial trading for all three utilities.

The top left panel, the top right panel, and the bottom middle panel of Appendix Figure A.5 plots the bootstrap distributions of implied trading costs corresponding to the service-territory-level day-ahead and real-time prices paid by PG&E, SCE and SDG&E respectively. We plot separate distributions for the pre-FT sample period in purple and the post-FT sample period in green. The solid vertical lines on each graph in this figure denote our estimated values for c_{lower} (in red) and c_{upper} (in blue) for the pre-FT sample period while the dashed vertical lines denote our estimated values for c_{lower} and c_{upper} for the post-FT sample.

All three panels of Appendix Figure A.5 indicate that both c_{lower} and c_{upper} fell substantially after the introduction of financial trading. That being said, Appendix Figure A.6 presents results from a formal test of the null hypothesis that c_{lower} and c_{upper} remained the same after financial trading was introduced.

Specifically, Appendix Figure A.6 plots the bootstrap distribution of the *difference* in implied trading costs for each utility before versus after financial trading. The left vertical line in this figure is the 10th percentile of the distribution of $c_{pre} - c_{post}$ and the right vertical line is the 90th percentile of this distribution. If the 10th percentile

Figure A.5: Bootstrap Distribution of Implied Trading Costs For Each Service Territory: Pre-FT in Purple and Post-FT in Green



(a) Pacific Gas and Electric

(b) Southern California Edison

(c) San Diego Gas and Electric

Notes: This figure plots the bootstrap distributions of implied trading costs for sample periods before versus after the introduction of financial trading ("FT") in purple and green respectively. The top left panel, the top right panel, and the bottom middle panel of this figure focus on the implied trading costs associated with the day-ahead/real-time price spreads faced by PG&E, SCE, and SDG&E respectively. The solid vertical lines on each graph in this figure denote our estimated values for c_{lower} (in red) and c_{upper} (in blue) for the pre-FT sample period while the dashed vertical lines denote our estimated values for c_{lower} and c_{upper} for the post-FT sample. Implied trading costs c_{lower} and c_{upper} are defined in Section IV.B.

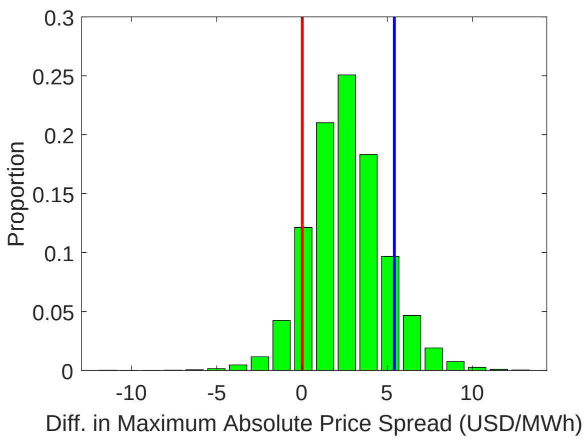
Table A.2: Implied Trading Costs by Territory (in USD/MWh)

	Utility	Before FT	After FT
Lower 5% C.I.	PG&E	8.518	6.614
(c_{lower})	SCE	11.995	7.050
	SDG&E	16.217	7.471
Upper 5% C.I.	PG&E	14.297	10.600
(c_{upper})	SCE	19.858	12.166
	SDG&E	31.939	12.961

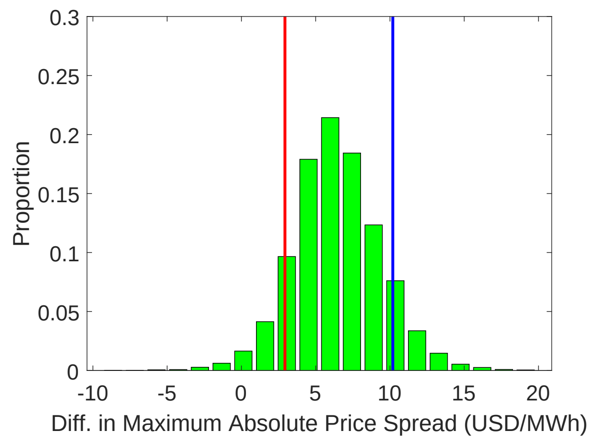
Notes: This table presents the implied trading costs estimated using the modeling framework discussed in Section IV. We estimate implied trading costs separately for each utility service territory for the sample periods before versus after the introduction of FT. The three service territories considered in this table correspond to California's three major electric utiilties: PG&E, SCE, and SDG&E.

of this distribution is greater than zero, then we can reject the null hypothesis that $c_{pre} \leq c_{post}$ at a 10% significance level. Similarly, we can reject the null hypothesis that $c_{pre} \geq c_{post}$ at a 10% significance level if the 90th percentile of the bootstrap distribution of $c_{pre} - c_{post}$ is less than zero. For all three utilities, we reject the null hypothesis that implied trading costs are higher post-FT relative to pre-FT, but fail to reject the null hypothesis that implied trading costs are higher pre-FT relative to post-FT.

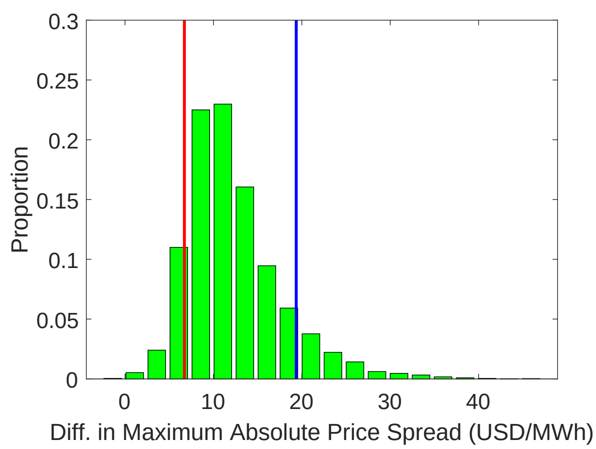
Figure A.6: Bootstrap Distribution of the Difference in Implied Trading Costs



(a) Pacific Gas and Electric



(b) Southern California Edison



(c) San Diego Gas and Electric

Notes: This figure plots the bootstrap distribution of the difference in "implied trading costs" (i.e.: $c_{pre} - c_{post}$), where "pre" indicates the sample period before the introduction of financial trading ("FT") and "post" indicates the sample period after FT. We plot this bootstrap distribution separately for the day-ahead/real-time price spreads paid by each of California's three major investor-owned distribution utilities: PG&E, SCE, and SDG&E. The left vertical line on the graph in red is the 10th percentile of the distribution of $c_{pre} - c_{post}$ and the right vertical line in blue is the 90th percentile of this distribution.

A.3.2 Additional Heterogeneity in Implied Trading Costs By Location

Appendix Table A.3 presents estimates for how implied trading costs changed before versus after FT across three types of locations: (1) "baseload" locations where the amount of electricity injected into the location was greater than zero in at least 75% of hours in our sample period, (2) "peaker" locations where the amount of electricity injected into the location was greater than zero in less than 75% of hours-of-sample, and (3) "demand" locations not associated with a generation unit. The unit of observation for the regressions presented in Appendix Table A.3 is a location in one of two sample periods, before FT and after FT.

Columns 1-2 (Columns 3-4) focus on c_{lower} (c_{upper}): the 5th (95th) percentile of the bootstrap distribution of the maximum over hours of the day of the absolute value of the 24×1 vector of hourly average day-ahead/real-time price spreads. For Columns 2 and 4, we trim observations corresponding to the top 1% and bottom 1% of the distribution of the outcome variable before estimating the regression. White (1980) standard errors are provided in parentheses.

Appendix Table A.3 tests the intuition that some types of units find it more costly to adjust their day-ahead schedules relative to their real-time output to profit from expected differences between day-ahead and real-time prices. Specifically, we hypothesize that units that operate less frequently find it more costly to inject more electricity than expected in real-time because these units are typically not needed to serve demand. This limits the extent to which the owners of these units can adjust their physical bids to profit from expected day-ahead/real-time price spreads. In contrast, owners of units that frequently operate can easily adjust how much of their expected real-time output to sell in the day-ahead versus real-time markets depending on their expectations about the day-ahead/real-time price spread.

Consistent with this logic, the results presented in Appendix Table A.3 indicate that implied trading costs prior to the introduction of financial trading were smallest for baseload locations, followed by peaker locations, with demand locations exhibiting

Table A.3: Implied Trading Costs Before vs. After Financial Trading For Baseload versus Peaker versus Demand Locations

Dep. Var.	c_{lower}		c_{upper}	
	(1)	(2)	(3)	(4)
Post FT × Gen Node × Baseload	0.640	0.493	1.358	1.579
	(0.196)	(0.186)	(0.446)	(0.402)
Post FT × Gen Node	0.237	0.311	1.046	0.844
	(0.152)	(0.139)	(0.342)	(0.291)
Gen Node × Baseload	-0.817	-0.630	-1.647	-1.468
	(0.183)	(0.174)	(0.372)	(0.359)
Gen Node	-0.198	-0.218	-0.681	-0.589
	(0.142)	(0.129)	(0.292)	(0.272)
Post FT	-3.494	-3.329	-6.659	-6.577
	(0.051)	(0.046)	(0.115)	(0.102)
Constant	10.519	10.351	18.577	18.306
	(0.048)	(0.044)	(0.102)	(0.096)
Residualized	Yes	Yes	Yes	Yes
Trim Top and Bottom 1%	No	No	No	No
Mean of Dep. Var.	8.686	8.614	15.083	14.893
Std. Dev. of Dep. Var.	2.770	2.770	5.920	5.920
R^2	0.383	0.407	0.297	0.351
Number of Obs.	9,486	9,302	9,486	9,298

Notes: This table reports the results from our difference-in-differences specification comparing implied trading costs before versus after the introduction of financial trading ("FT") for pricing locations associated with generation units ("Generation") versus not associated with generation units. We consider two types of Generation Locations: locations associated with generation units that produced in over 75% of hours-of-sample ("Baseload") versus locations associated with generation units that produced in less than 75% of hours-of-sample ("Peaker"). The unit of observation for these regressions is a location in the sample period before FT versus after FT. We report White (1980) standard errors in parentheses. We consider two dependent variables: c_{lower} in the first two columns and c_{upper} in the last two columns. For Columns 2 and 4, we trim observations corresponding to the top 1% and bottom 1% of the distribution of the outcome variable before estimating the regression.

Variable Definitions: Post FT is an indicator variable that is equal to one if the observation corresponds to the sample period after FT. Generation is an indicator variable that is equal to one if the location is associated with a generation unit. Baseload is an indicator variable that is equal to one if the amount of electricity injected into the location was greater than zero in at least 75% of hours-of-sample.

the largest implied trading costs. The coefficient estimates also suggest that implied trading costs are the same across baseload, peaker, and demand locations after FT.

Combined, the results presented in Appendix Table A.3 are consistent with the intuition that, prior to financial trading, the implied trading costs associated with adjusting real-time output to trade day-ahead/real-time price spreads are smaller for units that operate more frequently. After FT, all market participants can trade day-ahead/real-time price spreads at most locations. Therefore, we no longer find systematic differences in implied trading costs across baseload, peaker, and demand locations after FT.

A.4 Additional Tables and Figures: Generation and Capacity

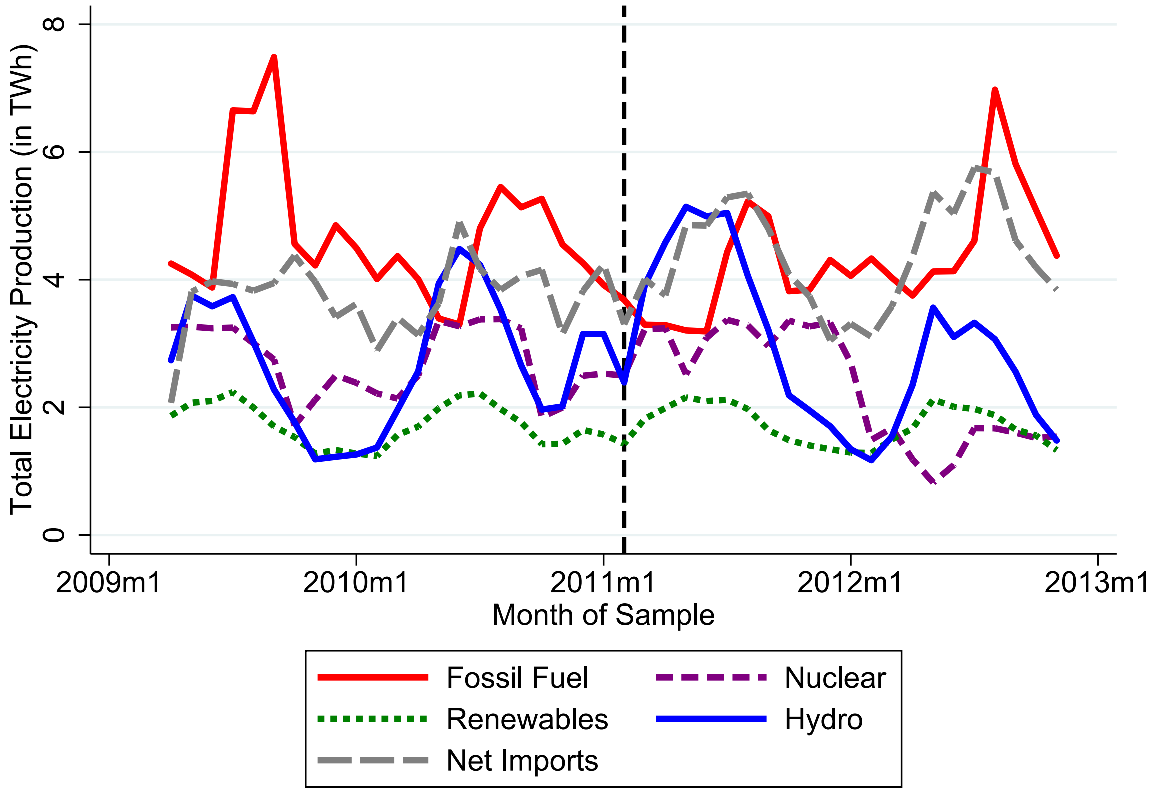
Appendix Figure A.7 plots monthly total electricity production by type: gas-fired, nuclear, renewables, and all hydro.[49] We sum only over sources under the operational control of California's Independent System Operator (CAISO). Appendix Figure A.7 also includes monthly total net electricity imports. Finally, Appendix Figure A.8 plots monthly total electricity demand.[50] A vertical dashed line corresponding to the introduction of FT is included in both figures.

Appendix Figures A.7 and A.8 document that that there are not systematic upward or downward time trends in electricity production by source type, electricity imports, or system-wide total demand over our sample period. In addition, there are not large changes in production from nuclear sources and renewables in the 6-12 months after the introduction of FT, suggesting that production from these sources did not respond to the implementation of this policy. However, we see a reduction in output from gas-fired sources coupled with decreases in electricity demand and increases in production from hydroelectric sources in the roughly 6-7 months around

[49] The classification "renewables" includes wind, solar, and geothermal sources as well as hydro sources with capacity less than 30 MW. Monthly plant-level data on output come from Form EIA-923 (EIA, 2009-2012).

[50] Hourly data on total net electricity imports and demand can be downloaded from the OASIS API (CAISO, 2009-2012).

Figure A.7: Monthly Total Electricity Production By Source



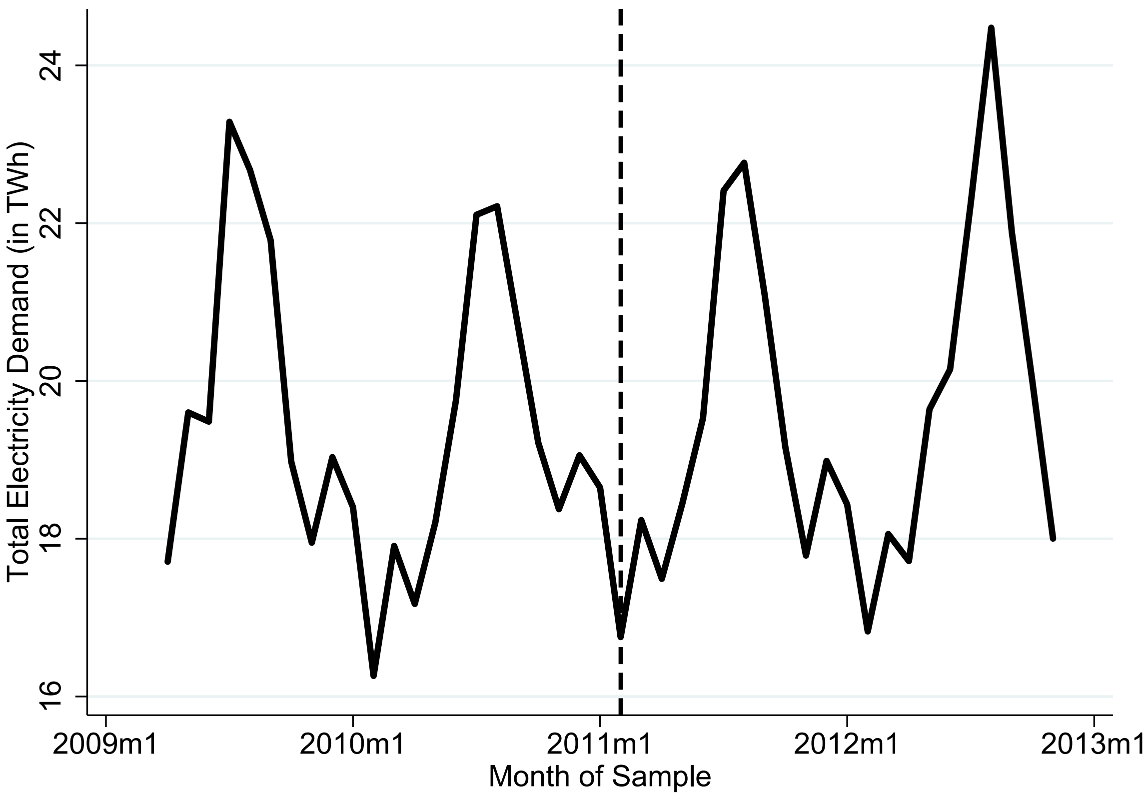
Notes: This figure plots monthly total electricity production by type: fossil-fuel-fired, nuclear, renewables (wind + solar + biomass + biogas + hydro sources less than 30MW), and all hydro. We sum only over sources under the operational control of California's Independent System Operator (CAISO). This figure also plots monthly total net electricity imports. Finally, this figure includes a vertical dashed line denoting the introduction of financial trading.

February 2011. This highlights the importance of flexibly controlling for hydroelectric production and demand in our specifications in Section VI that consider how fuel costs per MWh and input fuel use per MWh change on high complexity days versus low complexity days after FT is introduced.

Appendix Figure A.9 plots the annual total electricity generating capacity in California by source type: fossil-fuel-fired, nuclear, hydro, and wind + solar.[51] The sample period considered in the figure spans the years 2000-2016, with vertical dashed red lines denoting the years 2009 and 2012. We see from this figure that there were no major investments in generating capacity between 2009-2012. That being said, this figure

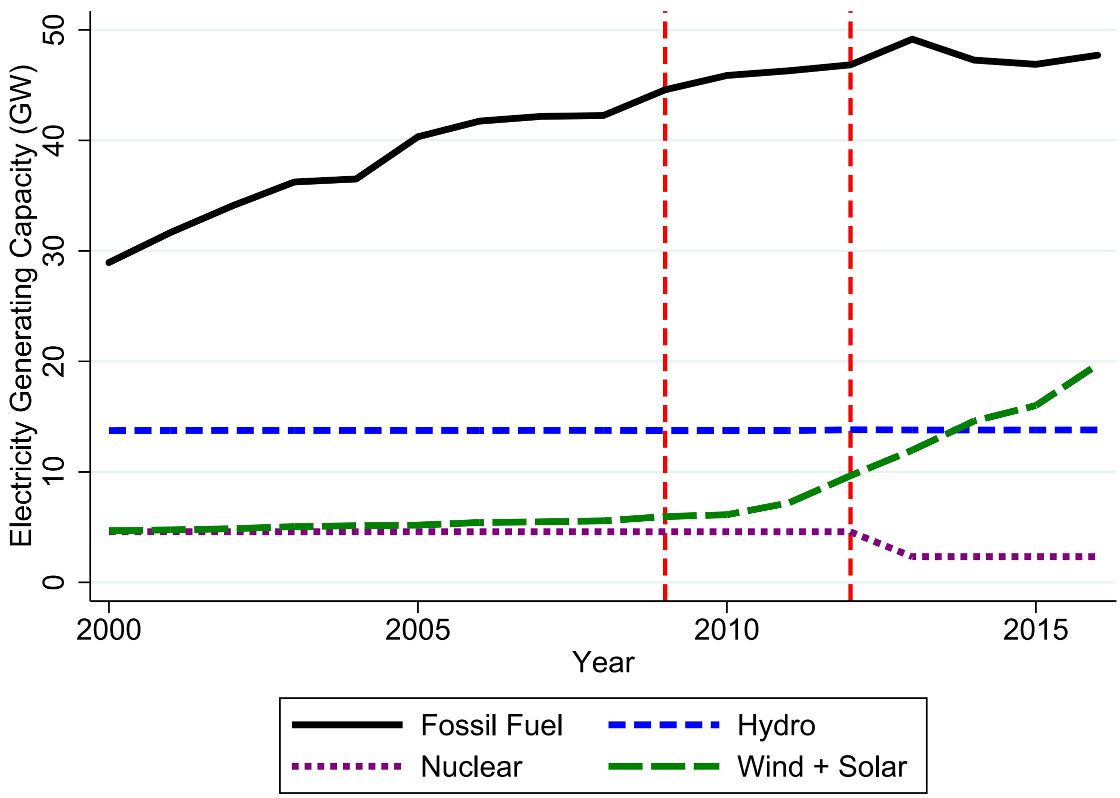
[51]We sum over units of each source type in California using the eGrid database for 2012 provided by the United States Environmental Protection Agency (USEPA, 1996-2012).

Figure A.8: Monthly Total Electricity Demand



Notes: This figure plots monthly total electricity demand. We include a vertical dashed line denoting the introduction of financial trading.

Figure A.9: Annual Total Electricity Generating Capacity By Type



Notes: This figure plots annual total electricity generating capacity by type: fossil-fuel-fired, nuclear, hydro, and wind + solar. We sum over units of each source type in California using the eGrid database for 2012 provided by the United States Environmental Protection Agency (USEPA, 1996-2012). The sample period considered in this figure spans the years 2000-2016, with vertical dashed red lines denoting the years 2009 and 2012.

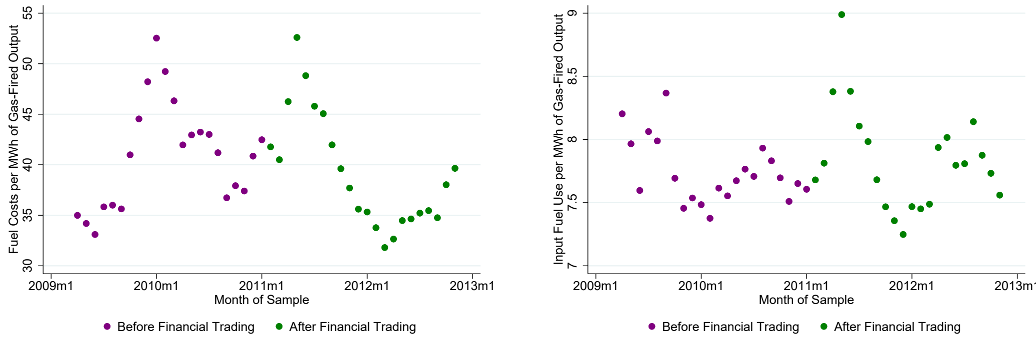
documents a steady increase in the installation of renewable capacity during the latter half of our sample period. In addition, we see a decrease in nuclear generating capacity after 2012 due to the retirement of the San Onofre nuclear power plant (Davis and Hausman, 2016). Based on these trends, all of the specifications considered in Section VI control flexibly for monthly total production from renewables and monthly total production from nuclear plants. We also show that the difference-in-differences estimates in Section VI.C remain similar if we drop all days-of-sample after the shutdown of the San Onofre nuclear power plant.

A.5 Additional Tables and Figures: Fuel Cost and Fuel Use

The left panel of Appendix Figure A.10 plots the monthly averages of the log of daily total fuel costs incurred by gas-fired plants divided by the daily total output of these plants. The right panel of this figure plots the monthly averages of the log of daily total fuel use by gas-fired plants divided by daily total output from these plants. Appendix Figure A.10 documents that both outcomes exhibit substantial seasonality. The variability induced by this seasonality obfuscates comparisons of the outcomes across the sample periods before versus after FT. For this reason, we include separate sets of month-of-year fixed effects for high complexity days and low complexity days in all specifications. That being said, Appendix Figure A.10 also suggests that neither of the outcome variables are systematically trending up or down over our sample period. This is comforting given that any such trend over time might confound the comparison of outcomes across the pre-FT versus post-FT sample periods.

Appendix Table A.4 presents the asymptotic p-values from two different tests of the null hypothesis that the market outcome considered is nonstationary. The two tests considered are the Augmented Dickey-Fuller unit-root test (Dickey and Fuller (1979); MacKinnon (1994)) and the Phillips-Perron unit-root test (Phillips and Perron, 1988). We can reject the unit root null hypothesis for both outcomes using either of the two statistical tests. This provides formal evidence that market outcomes are

Figure A.10: Monthly Average Outcomes Before vs. After Financial Trading



(a) Fuel Costs Per MWh

(b) Input Fuel Use Per MWh

Notes: The left panel of this figure plots the monthly averages of the daily total fuel costs incurred by gas-fired plants divided by the daily total output from these gas-fired plants. The right panel plots the monthly averages of the daily total fuel use by gas-fired plants divided by daily total output from these plants. Averages corresponding to months before (after) the introduction of financial trading are plotted in purple (green).

Table A.4: P-Values for Tests for Nonstationarity

	Dickey-Fuller	Phillips-Perron
Log Fuel Cost per MWh	0.007	0.063
Log Input Energy per MWh	≈ 0	≈ 0

Notes: This table presents p-values from two tests of the null hypothesis that the daily time series of the relevant market outcome is nonstationary. The two tests considered are the Augmented Dickey-Fuller unit-root test (Dickey and Fuller (1979); MacKinnon (1994)) and the Phillips-Perron unit-root test (Phillips and Perron, 1988). We consider two outcome variables: the log of fuel costs per MWh of gas-fired output and the log of input energy per MWh of gas-fired output.

not trending up or down during our sample period, allowing us to compare outcomes across the pre-FT versus post-FT sample periods without including time trends or first-differencing the outcome.

B Trading Fees for California's Electricity Market

There are three broad types of transaction costs associated with financial trading ("FT") in California's wholesale electricity market: collateral, trading fees and uplift. Purely financial participants must post collateral greater than the total value of the virtual bids they submit each day.[52] This collateral does not earn any rate of return while it is held by California's Independent System Operator (ISO). Moreover, there can be a lag of more than two weeks between when a market participant requests that some or all of its collateral be returned and when this money is actually returned. Consequently, a purely financial participant is foregoing non-trivial financial returns on any collateral posted with the California ISO in order to engage in virtual bidding.[53]

Purely financial participants must pay roughly 0.5 cents for each price and quantity step associated with the virtual bid curve they submit. They must also pay 9 cents per MWh of virtual energy *cleared* in fees associated with "market services". For example, consider a virtual bidder that submits a demand curve with 10 price/quantity steps to the day-ahead market. If 50 MWh of her demand bid clears, she must pay \$4.55 = (\$0.09 $\times$ 50) + (\$0.005 $\times$ 10) in transaction fees. Finally, all financial participants are required to pay a monthly transaction fee of 1,000 dollars regardless of the volume of virtual bids they submit or clear.[54]

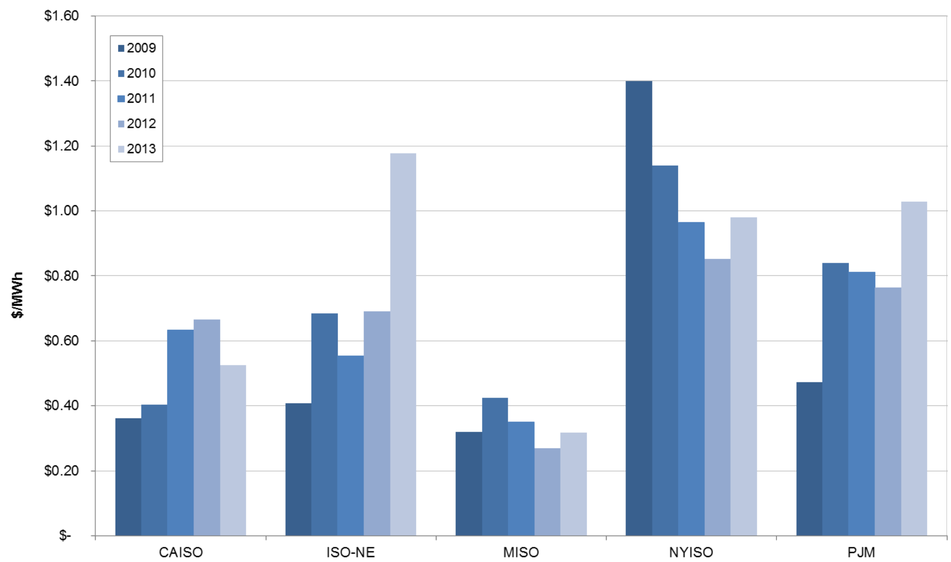
The California ISO clears day-ahead and real-time markets by solving a mixed-integer programming problem. The California ISO is sometimes forced to manually dispatch generation units after the close of the day-ahead market or in real-time to satisfy operational constraints that may not have have been accounted for in the day-ahead or real-time markets. Any generation units forced by the California ISO to change production levels outside of the formal market-clearing mechanism receive

[52]The total value of the virtual bids submitted each day is equal to the sum of the product of the absolute value of megawatt-hours offered times the applicable reference price for a virtual bid at that location. See the California ISO document, "Convergence bidding, participating in markets, credit policy implications," for a description of the process used to compute nodal reference prices.

[53]See the California ISO document, "California ISO Credit Management," for more background.

[54]These transaction fees are listed in Session 7 of the Convergence Bidding tutorial published by California's ISO (CAISO (2015b)).

Figure B.1: Annual Uplift Charges for the Five Major ISOs: 2009-2013



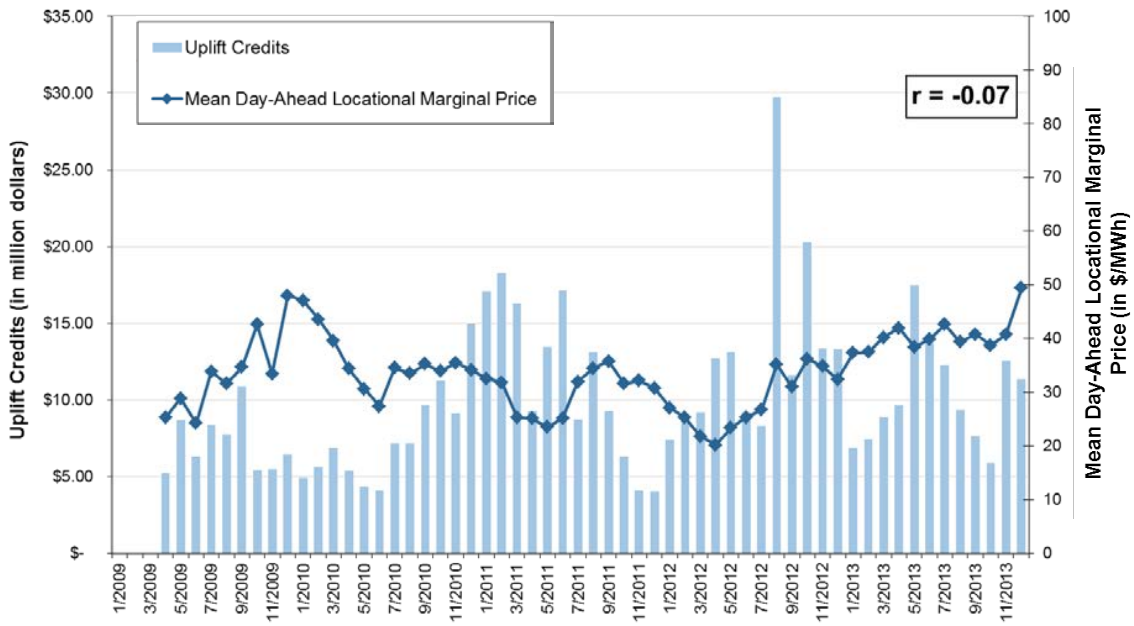
Notes: This figure is taken from FERC (2014). Annual average uplift charges (in dollars per MWh) are calculated for each Independent System Operator (ISO) by dividing total annual uplift charges (in dollars) by total annual electricity demand (in MWh). Total uplift charges and total electricity demand for CAISO for 2009 are based on the nine months of data after April 1st 2009. FERC estimated the total uplift charges and electricity demand for ISO-NE for 2012. Uplift charges for PJM for the years 2012 and 2013 exclude the credits associated with reactive services (these credits amount to approximately 45 million dollars per year).

"uplift" payments. Generation units that are turned on in the day-ahead market and fail to recover their start-up, minimum load and as-offered costs from selling energy and operating reserves also receive a "make-whole payment" to cover this deficit. These make-whole payments are also included in uplift and ensure that any generation unit committed to operate in the day-ahead market will at least recover their as-offered costs.[55]

Uplift charges are paid by the market participants whose bids contributed to the out-of-market dispatch of units. Each participant's contribution is based on a formula subject to fierce policy debate (Kurlinski, 2013). Purely financial participants are required to pay uplift charges to the extent that their trades result in generation unit output levels that deviate from those dictated by the market clearing algorithm. Appendix Figure B.1 shows the annual average uplift charge per MWh of electricity demand for the five major Independent System Operators (ISOs) in the United States

[55] The following link provides more details on uplift charges: http://www.caiso.com/Documents/BriefingISO_MarketPricing-MSCPresentation-May19_2014.pdf.

Figure B.2: Correlation Between Uplift and Day-Ahead Prices



Notes: This figure, taken from FERC (2014), documents the correlation between the monthly total uplift credits paid out by CAISO and the monthly average locational marginal price for the location TH_SP15_GEN-APND from Ventyx. The Pearson correlation coefficient between uplift and day-ahead prices is r = -0.07.

for 2009-2013. This figure indicates that average uplift charges range from roughly 40 to 60 cents per MWh. However, these annual averages conceal significant volatility in daily uplift charges (FERC (2014)).

Appendix Figure B.2 plots monthly total uplift payments in California from April 2009 to December 2013. This figure shows an increase in uplift payments after the introduction of financial trading in February 2011.[56] Kurlinski (2013) argues that much of this increase in uplift payments is due to financial trading at "interties," which are locations where electricity is imported or exported between the California ISO and other balancing authorities. During our sample period, this led to fierce policy debate surrounding both whether trading at interies should be allowed and how uplift payments from trades should be allocated. Consequently, virtual bidding on interties was suspended on November 28, 2011. We leave it as future work to determine how this suspension impacted the market efficiency benefits from introducing FT.

[56]The spike in uplift payments in August 2012 was likely due to an extreme heat wave from August 7th through August 17th (CAISO (2012b)).

Finally, Appendix Figure B.2 also documents that there is little correlation between monthly average day-ahead prices and monthly total uplift charges. Average day-ahead prices are between 30 and 50 dollars per MWh while average uplift charges are between 0.40 to 0.60 dollars per MWh. It is thus unlikely that the increases in uplift charges after financial trading was introduced resulted in substantial increases in the retail electricity prices paid by consumers. Instead, the policy debate has centered on the allocation of uplift charges across financial versus physical market participants.

C Additional Empirical Results: Price Spreads

This Appendix section discusses three additional results pertaining to day-ahead/real-time price differences. The first subsection provides empirical evidence that average day-ahead/real-time price differences are smaller in absolute value after the introduction of FT. These results suggest that day-ahead prices better reflect real-time conditions after purely financial participation was implemented.

The second subsection presents the methodology and results corresponding to the hypothesis test that the distribution of the number of hours of day with positive average price spreads for demand locations first-order stochastically dominates the corresponding distribution for generation locations. We perform this test separately for the sample periods before versus after FT is introduced. Our findings suggest that electricity suppliers are better able to drive real-time prices up at the locations where they own generation units relative to demand locations.

In the final subsection, we test whether the daily 24×1 vector of hourly price spreads is autocorrelated over days-of-sample. The results of this analysis indicate that traders are unlikely to earn significantly more profits by conditioning on day-ahead/real-time price differences from two or more days prior to the current day.

C.1 Absolute Average Price Spreads Before Versus After FT

This subsection describes our statistical test of whether expected day-ahead/real-time price spreads decrease in absolute value after the introduction of financial trading on February 1st 2011. In particular, we formulate a test of the null hypothesis that $|\mu^j_{pre}| > |\mu^j_{post}|$ for $j = 1, 2, ..., 24$, where μ^j_{pre} (μ^j_{post}) is the jth element of the 24×1 vector composed of the expected day-ahead/real-time price differences for each hour of the day for the pre-FT sample period (post-FT sample period). We implement this statistical test separately for each pricing location. In a slight abuse of notation, we represent the above null hypothesis as H_0: $|\mu_{pre}| > |\mu_{post}|$.

Using the methodology derived in Wolak (1989), we compute the following test statistic in order to test the null hypothesis that $|\mu_{pre}| > |\mu_{post}|$:

$$TS = \begin{matrix} min \\ \theta \geq 0 \end{matrix} (|\overline{X}^{pre}| - |\overline{X}^{post}| - \theta)'\hat{V}^{-1}(|\overline{X}^{pre}| - |\overline{X}^{post}| - \theta)$$

where $\overline{X}^{pre}$ ($\overline{X}^{pre}$) is the 24×1 vector of the average day-ahead/real-time price differences for each hour of the day for the pre-FT (post-FT) sample period. We calculate the covariance matrix $\hat{V}$ as follows:

$$\hat{V} = \frac{diag[SIGN(\overline{X}^{pre})]'\hat{\Sigma}^{pre}diag[SIGN(\overline{X}^{pre})]}{N^{pre}} + \frac{diag[SIGN(\overline{X}^{post})]'\hat{\Sigma}^{post}diag[SIGN(\overline{X}^{post})]}{N^{post}}$$

where the $diag[Z]$ operator takes a vector Z and returns a diagonal matrix with the elements of Z on the diagonal. N^{pre} (N^{post}) is the number of days in the sample period before (after) the introduction of financial trading. $\hat{\Sigma}^{pre}$ ($\hat{\Sigma}^{post}$) is an estimate of the asymptotic covariance matrix associated with $\overline{X}^{pre}$ ($\overline{X}^{post}$). We reject the null hypothesis that $|\mu_{pre}| > |\mu_{post}|$ if and only if:

$$\sum_{h=1}^{24} w(24, 24-h, \hat{V}) Pr[\chi^2_{(h)} > TS] < \alpha$$

where $\chi^2_{(h)}$ is a chi-squared random variable with h degrees of freedom, $w(24, 24 -$

Table C.1: Service Territory Level P-values for the Absolute Difference Tests

	H_0: $\lvert\mu_{pre}\rvert > \lvert\mu_{post}\rvert$	H_0: $\lvert\mu_{post}\rvert > \lvert\mu_{pre}\rvert$
PG&E	0.752	0.003
SCE	0.972	0.000
SDG&E	0.832	0.000

Notes: This table reports the p-values associated with the statistical test of the null hypothesis that $|\mu_{pre}| > |\mu_{post}|$ (Column 1) as well as the statistical test of the null hypothesis that $|\mu_{post}| > |\mu_{pre}|$ (Column 2). μ_{pre} (μ_{post}) is a 24×1 vector composed of the expected day-ahead/real-time price spreads for each hour of the day for the sample period before (after) the introduction of financial trading. We perform these statistical tests on the service territory level price spreads faced by each of California's three major electric utilities: PG&E, SCE, and SDG&E.

Table C.2: Proportion of Locations for which we fail to reject the Absolute Difference Test

	H_0: $\lvert\mu_{pre}\rvert > \lvert\mu_{post}\rvert$	H_0: $\lvert\mu_{post}\rvert > \lvert\mu_{pre}\rvert$
Generation Locations	0.999	0.013
Demand Locations	0.987	0.011

Notes: This table reports the proportion of pricing locations for which we fail to reject a size 0.05 test of the null hypothesis that $|\mu_{pre}| > |\mu_{post}|$ (Column 1) and the null hypothesis that $|\mu_{post}| > |\mu_{pre}|$ (Column 2). μ_{pre} (μ_{post}) is a 24×1 vector composed of the expected day-ahead/real-time price spreads for each hour of the day for a given location for the sample period before (after) the introduction of financial trading. There are 653 locations associated with generation units ("Generation Locations") and 3,961 locations not associated with generation units ("Demand Locations") that are present in the sample periods both before and after financial trading.

$h, \hat{V}$) are the weights defined in Wolak (1989), and α is the asymptotic size of the hypothesis test. We consider tests of size $\alpha = 0.05$ in the results presented below. The test statistic and p-value associated with the null hypothesis that $|\mu_{post}| > |\mu_{pre}|$ are computed in a similar manner.

We first perform these statistical tests on the service territory level price spreads faced by each of California's three major electricity distribution utilities: Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E). Appendix Table C.1 presents the p-values associated with these tests. For all three utilities, we fail to reject the null hypothesis that $|\mu_{pre}| > |\mu_{post}|$ but reject the null hypothesis that $|\mu_{post}| > |\mu_{pre}|$.

We also perform our statistical tests separately for each pricing location in California. Specifically, Column 1 of Appendix Table C.2 lists the proportion of locations for which we fail to reject the null hypothesis that $|\mu_{pre}| > |\mu_{post}|$, separately for locations associated with generation units ("Generation Location") versus locations not associated with generation units ("Demand Locations"). We fail to reject this null hypothesis for over 98% of locations for both generation and demand locations. Column 2 of Appendix Table C.2 lists the proportion of locations for which we fail to reject the null hypothesis that $|\mu_{post}| > |\mu_{pre}|$. We fail to reject this null hypothesis for only roughly 1% of locations for both generation and demand locations. Combined, Appendix Table C.2 constitutes strong evidence that absolute average day-ahead/real-time price spreads fell after purely financial participation was allowed.

C.2 Test for First-Order Stochastic Dominance: Generation versus Demand Locations Before versus After Financial Trading

This subsection describes our hypothesis test for whether the distribution across locations of the number of hours of the day with positive average day-ahead/real-time price spreads for locations associated with generation units ("Generation Locations") first-order stochastically dominates the distribution for locations not associated with generation units ("Demand Locations"). These hypothesis tests are implemented using the methodology discussed in Schmid and Trede (1996). First, we calculate the average day-ahead/real-time price spread $X_{n,h,s}$ for each location n in each hour of the day h before versus after the introduction of FT. The subscript $s = 0$ denotes the pre-FT sample period while $s = 1$ denotes the post-FT sample period. We next calculate the number of hours of the day with positive price spreads for each location in each sample period:

$$\text{NUMPOS}_{n,s} = \sum_{h=1}^{24} 1[X_{n,h,s} > 0]$$

Finally, we construct the empirical distribution function (EDF) and empirical probability mass function (PMF) of $\text{NUMPOS}_{n,s}$ separately for generation locations (indexed "G") versus demand locations (indexed "D") before versus after FT. Specifically, note that:

$$\hat{F}_{i,s}(t) = \frac{1}{N_{i,s}} \sum_{n=1}^{N_{i,s}} 1[\text{NUMPOS}_{n,s} \leq t]$$

$$\hat{f}_{i,s}(t) = \frac{1}{N_{i,s}} \sum_{n=1}^{N_{i,s}} 1[\text{NUMPOS}_{n,s} = t]$$

where $N_{i,s}$ is the number of locations of type $i \in \{\text{G}, \text{D}\}$ in sample period s. The argument t for each of these functions can potentially take on the integer values between 0 and 24. For example, $\hat{f}_{i,s}(t)$ measures the probability that the number of hours of the day with positive price spreads at location type i in sample period s is equal to t.

We test the null hypothesis that the EDF for demand locations first-order stochastically dominates the EDF for generation locations. We do so separately for the pre-FT sample ($s = 0$) and the post-FT sample ($s = 1$). Formally, the null hypothesis for a given sample period s is:

$$H_0\text{: } F_{\text{G},s}(t) \geq F_{\text{L},s}(t) \text{ for all } t \in \{0, 1, 2, ..., 24\} \tag{C.1}$$

We also test the reverse hypothesis that the EDF for generation locations first-order stochastic dominates the EDF for demand locations. This null hypothesis is:

$$H_0\text{: } F_{\text{L},s}(t) \geq F_{\text{G},s}(t) \text{ for all } t \in \{0, 1, 2, ..., 24\} \tag{C.2}$$

Schmid and Trede (1996) demonstrate that the test statistic associated with the null hypothesis presented in Appendix Equation (C.1) is:

$$\sqrt{\frac{N_{G,s}N_{L,s}}{N_{G,s} + N_{L,s}}} \sum_{k=1}^{K} (\hat{F}_{\text{G},s}(t_k) - \hat{F}_{\text{L},s}(t))^{+} \hat{f}_{\text{L},s}(t_k)$$

Table C.3: First-Order Stochastic Dominance Tests: Test Statistics

	Before FT	After FT
H_0: $F^{\mathrm{G}}(t) \geq F^{\mathrm{L}}(t)$	0.139	0.297
H_0: $F^{\mathrm{L}}(t) \geq F^{\mathrm{G}}(t)$	0.810	1.429

Notes: We reject the null hypothesis at the 5% level (1% level) if the test statistic is greater than 0.48 (0.68). Schmid and Trede (1996) discusses the derivation of this test statistic.

Additional Notes: This table presents the test statistics associated with null hypotheses pertaining to the first-order stochastic dominance of the distribution function of the number of hours of the day that average day-ahead/real-time price spreads are positive. Specifically, the top row focuses on the null hypothesis that the distribution for locations associated with generation units ("Generation Locations") is first-order stochastically dominated by the distribution for locations not associated with generation units ("Demand Locations") for all points where the probability mass function for demand locations is positive. The bottom row focuses on the null hypothesis that the distribution for demand locations is first-order stochastically dominated by the distribution for generation locations for all points where the probability mass function for generation locations is positive. The first row presents test statistics calculated for the sample period before the introduction of financial trading ("FT") while the second row presents test statistics calculated for the sample period after FT.

where $(y)^+ = max(0, y)$ and we evaluate the EDFs and PMF at all points $t_k \in \{t_1, t_2, ..., t_K\}$ such that $\hat{f}_{\mathrm{L},s}(t_k) > 0$. We reject the null hypothesis at the 5% level (1% level) if the test statistic is greater than 0.48 (0.68). The test statistic for the null hypothesis presented in Appendix Equation (C.2) is similar in form. Simply reverse the "G" and "L" subscripts in the computation of the test statistic.

Appendix Table C.3 presents the test statistics associated with testing the null hypotheses listed in Appendix Equations (C.1) and (C.2). These results indicate that we fail to reject the null hypothesis that the distribution for demand locations first-order stochastically dominates the distribution for generation locations for both the pre-FT and post-FT sample periods. They also support rejection of the null hypothesis that the distribution for generation locations first-order stochastically dominates the distribution for demand locations for both sample periods.

Combined, the results from Appendix Table C.3 suggest that more elements of the vector of average day-ahead/real-time price differences are positive for demand locations relative to generation locations. This result is consistent with two features of California's wholesale electricity market. First, retailers must submit territory-level bid curves to the day-ahead market, which greatly limits their ability to exercise market power at specific nodes. Second, except for a very small quantity of flexible loads, only electricity suppliers are able to influence real-time prices by submitting price-elastic, location-specific offer curves into the real-time market. Our results thus suggest that suppliers have a greater ability to raise real-time prices relative to day-ahead prices throughout the day at the locations where they own generation units relative to demand locations during both the pre-FT and post-FT sample periods.

C.3 Testing for Autocorrelation in Price Spreads

The methodology for measuring implied trading costs discussed in Section IV considers trading strategies that vary only by hour of the day. Specifically, we do not allow our hypothetical trader to update her strategy based on information from past days. We justify this restriction on trading strategies in this subsection.

Traders submit virtual bids to buy (sell) one MWh of electricity in the day-ahead market at a given location for a given hour with the obligation to sell (buy) this electricity back in the real-time market at the same location for the same hour. Traders simultaneously submit virtual bids for all 24 hours of the following day. Therefore, trading strategies can potentially be a function of lagged values of the 24×1 vector of realized day-ahead/real-time price spreads for each hour of the day.

However, trading strategies for day d cannot be a function of information from the values of the 24×1 vector of day-ahead/real-time price differences for day $d-1$. This is because the vector of real-time prices for day $d-1$ is not known before virtual bids are submitted to the day-ahead market for day d. Therefore, traders cannot use correlation between X_d and X_{d-1} in their strategies. However, if X_d and X_{d-h} are correlated for $h > 1$, then conditioning on X_{d-h} can improve a trader's forecast of the mean of X_d. Therefore, restricting consideration to trading strategies that do not condition on past values of price differences is only reasonable if all of the autocorrelation matrices associated with the time series process governing the daily vector of price spreads are zero except for the autocorrelation matrix associated with the first lag.

We denote the τ^{th} autocovariance matrix associated with the 24×1 vector of price spreads $\Gamma(\tau) = E[(X_t - \mu)(X_{t-\tau} - \mu)']$. Consistent with our above discussion, we expect $\Gamma(1)$ to be non-zero but test whether $\Gamma(\tau) = 0$ for all $\tau > 1$. We thus formulate a statistical test of the following null hypothesis:

$$\mathrm{H}_0 \colon \Gamma(2) = 0, \Gamma(3) = 0, ..., \Gamma(R) = 0$$

for a fixed value of R. Empirically, we set $R = 10$.

To implement this hypothesis test, we first define:

$$\xi \equiv [vec(\Gamma(2))', vec(\Gamma(3))', ..., vec(\Gamma(R))']'$$

where the $vec(.)$ operator takes each 24×24 autocovariance matrix and stacks it columnwise to create a 576×1 vector. Therefore, ξ has 5,184 ($= 576 \times 9$) elements, all of which must equal zero under the null hypothesis. We use the moving block bootstrap discussed in Section III.C to estimate the 5,184 $\times$ 5,184 covariance matrix associated with $\hat{\xi}$. Our Wald statistic $TS = \hat{\xi}'\hat{\Sigma}^{-1}_{\xi,boot}\hat{\xi}$ is asymptotically chi-squared distributed with $576 \times (R-1)$ degrees of freedom under the null hypothesis, where we use a moving block bootstrap procedure in order to estimate the covariance matrix $\hat{\Sigma}_{\xi,boot}$.

We first conduct this statistical test separately for the sample periods before and after the introduction of financial trading ("FT") using the day-ahead/real-time price spreads faced by each of California's three major investor-owned utilities. Appendix Table C.4 reports the resulting test statistics; the upper $\alpha = 0.05$ critical value for these test statistics is $\chi^2_{(5{,}184)} = 5{,}352.6$. We fail to reject the null hypothesis that the second through tenth autocovariance matrices are zero for all three utilities both before and after the introduction of FT.

We also conduct these autocorrelation tests at each pricing location, reporting the

Table C.4: Test Statistics for Autocorrelation ($1 < L \leq 10$) in Daily Price Spreads

	Before FT	After FT
PG&E	4,863.4	3,531.3
SCE	7,541.0	3,635.9
SDG&E	12,003.1	3,334.0

Notes: This table presents chi-squared test statistics corresponding to the null hypothesis that the second through tenth autocovariance matrices associated with the 24×1 vector of day-ahead/real-time price spreads for each hour of the day are zero. Formally, we are testing the null hypothesis that $\Gamma(2) = \Gamma(3) = ... = \Gamma(10) = 0$. We perform this test separately for the sample periods before versus after the introduction of financial trading using the day-ahead/real-time price spreads faced by each of California's three major investor-owned utilities. The upper $\alpha = 0.05$ critical value for these test statistics is $\chi^2_{(5,184)} = 5{,}352.6$.

Table C.5: Proportion of Locations for which we fail to reject the Autocorrelation Test

	Before FT	After FT
Demand Locations	0.562	0.981
Generation Locations	0.586	0.943

Notes: This table presents the proportion of locations for which we fail to reject a size $\alpha = 0.05$ test of the null hypothesis that the second through tenth autocovariance matrices of the 24×1 vector of day-ahead/real-time price spreads for each hour of the day are zero. Formally, we are testing the null hypothesis that $\Gamma(2) = \Gamma(3) = ... = \Gamma(10) = 0$.

results in Appendix Table C.5. Prior to FT, we fail to reject the null hypothesis of no second through tenth degree autocorrelation at 58.6 percent and 56.2 percent of generation and demand locations respectively. After FT, we fail to reject the null hypothesis of no second through tenth degree autocorrelation at 94.3 percent and 98.1 percent of generation and demand locations respectively. This is consistent with the logic that financial traders quickly take advantage of any systematic autocorrelation in price spreads after financial trading is introduced. The results from this subsection provide evidence that traders cannot earn significantly greater profits by conditioning on previous realizations of price spreads. This helps to justify our focus in Section IV on trading strategies that do not condition on past lags of daily price spreads.

D Data Appendix: Event Study and Difference-in-Differences

This Appendix section discusses how we construct the daily total fuel cost, fuel use, output and number of start-ups across all gas-fired units located in the territory served by California's ISO. The first subsection focuses on the Continuous Emissions Monitoring Systems (CEMS) database used in the analyses in Section VI while the second subsection discusses how we construct the monthly average natural gas price paid by each power plant.

D.1 Data Construction

We estimate the event study and difference-in-differences specifications discussed in Section VI using the Continuous Emissions Monitoring Systems (CEMS) database administered by the United States Environmental Protection Agency (USEPA, 2009-2012). These data are publicly available from the USEPA's website. CEMS provides us with the hourly output in MWh produced by each fossil-fired unit with capacity greater than 25MW in each hour-of-sample. CEMS also lists the input heat energy used by each unit in each hour, including the input energy used to start up or operate the unit at its minimum safe operating level. For this analysis, we only consider electricity generation units located in California.

We impose additional sample restrictions using plant-level characteristics from 2009, 2010, and 2012 from the eGRID database provided by the USEPA (USEPA, 1996-2012). We construct two variables from these data: (1) an indicator that is equal to one if and only if the plant lists natural gas as its primary fuel in 2009, 2010, or 2012, and (2) an indicator that's equal to one if and only if the plant lists the California ISO as its balancing authority in 2009, 2010, or 2012. We merge primary fuel type and balancing authority from eGrid into the CEMS database using the plant code (i.e., "orispl code"). Only plants listing natural gas as their primary fuel in at least one of the three years are kept for the analysis. We also drop plants that do not list California ISO as their balancing authority in 2009, 2010 or 2012.

Finally, we construct monthly average prices for natural gas supplied by PG&E and Southern California Gas (SCG) as discussed below in Appendix Section D.2. A plant in the CEMS data is assigned the natural gas price time series for PG&E if the eGrid data lists PG&E as either the utility service territory associated with the plant or the plant's operator in 2009, 2010, or 2012. Similarly, the plant is assigned the natural gas price series for SCG if either the utility service territory associated with the plant or the plant's operator is listed as either SCE or SDG&E in 2009, 2010, or 2012. All remaining plants are assigned the overall monthly gas price averaged over all transactions listing either PG&E or SCG as the supplier.

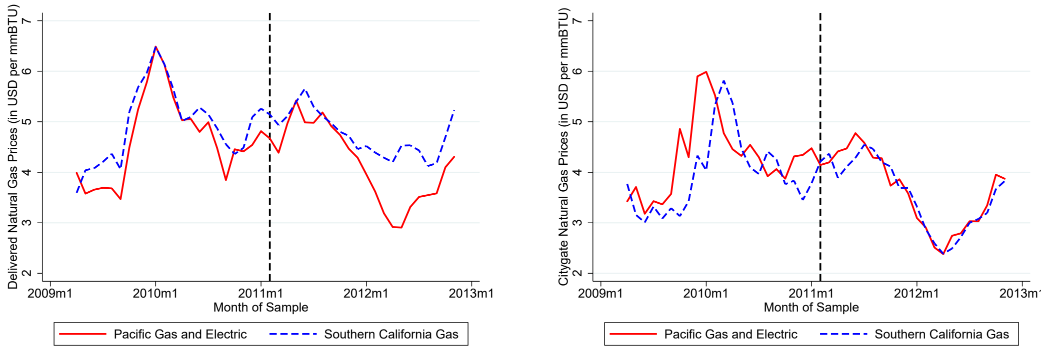
D.2 Data Construction: Natural Gas Prices

We calculate the monthly average natural gas price paid by power plants in California using transaction-level data from the Energy Information Administration (EIA, 2009-2012). Among other variables, the data contain the month-of-transaction, supplier, fuel price, and quantity sold. The natural gas prices paid by power plants owned by independent power producers not subject to output price regulation are not made publicly available. Fortunately, Cicala (2015) demonstrates that the average natural gas prices paid by price-regulated plants are similar to those paid by market-based plants.

From these transaction-level data, we construct monthly average natural gas prices for each of two suppliers: Pacific Gas and Electric (PG&E) and Southern California Gas (SCG). The resulting monthly average gas prices are plotted in the left panel of Appendix Figure D.1. We see from this figure that the two time series track each other fairly well.

Moreover, natural gas prices do not seem to respond to the introduction of finan-

Figure D.1: Monthly Average Natural Gas Prices By Supplier



Notes: The left panel of this figure plots the monthly average natural gas prices paid by plants supplied by Pacific Gas and Electric (PG&E) versus Southern Calfornia Gas (SCG). Monthly average natural gas prices for each supplier are constructed using transaction-level data for U.S. power plants from Form EIA-923 administered by the Energy Information Administration (EIA, 2009-2012). The right panel plots the monthly average gas prices paid at the PG&E and SCG citygates; we collect daily data on the spot gas prices paid at the PG&E and SCG citygates from S&P Global Platts (S&P Global Platts, 2009-2012). The vertical black dashed line denotes the introduction of financial trading in February 2011.

cial trading on February 1st 2011. This is not surprising because natural gas is a homogeneous product used for many purposes other than electricity generation; it is thus unlikely that shocks to local electricity demand transmit to natural gas prices. Finally, the gas price series constructed from the EIA data exhibit very similar trends over time to the monthly average gas prices paid at the PG&E versus SCG citygates.[57]

[57] We obtain daily data on the spot gas prices paid at the PG&E and SCG citygates from S&P Global Platts (S&P Global Platts, 2009-2012).

E Robustness Checks: Event Study and Difference-in-Differences

This Appendix section describes robustness checks pertaining to the event study and difference-in-differences results presented in Section VI.

E.1 Results For Ancillary Services Costs

This subsection explores how ancillary service costs change after the introduction of FT. The California ISO incurs ancillary service costs in order to ensure that electricity supply equals electricity demand at every instant even in the face of unanticipated changes in physical conditions such as generation unit outages or transmission outages as discussed in Wolak (2019) and Buchsbaum et al. (2020). For example, the market operator may pay a supplier to keep capacity available from a generation unit that is currently operating or can turn on quickly in order to balance supply and demand if a currently operating generation unit fails. We collect data on the costs associated with ancillary services from the Open Access Same-time Information System (OASIS) API administered by the California ISO (CAISO, 2009-2012).[58]

We first assess how ancillary service costs per MWh of gas-fired output change after FT was introduced for high complexity days versus low complexity days. To do so, we estimate the following regression specification:

$$Y_t = \alpha_{m,\text{HIGH}} + \theta_w + \gamma_{y,m} + \sum_{s=1}^{S} \sum_{k=1}^{K} \left[(X_{k,t} - \overline{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[X_{k,t} \in \text{BIN}_{k,b}] \right] + u_t \quad \text{(E.1)}$$

where Y_t is the logarithm of ancillary services cost per MWh of natural gas-fired generation for hour t. We define HIGH_t to be an indicator that is equal to one if and only if the relevant measure of complexity on day-of-sample t is above the 75th

[58] During our sample period, the California ISO operated short-term ancillary services markets for Frequency Regulation Up (RegUp), Frequency Regulation Down (RegDn), Spinning Reserve, and Non-Spinning Reserve.

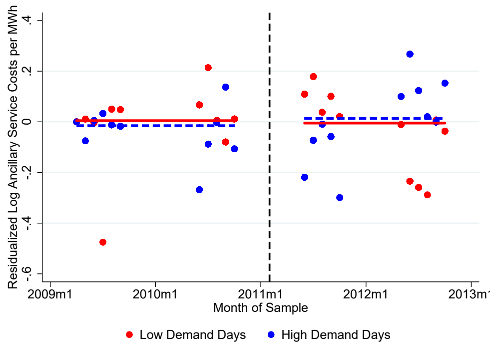
percentile of the distribution of this measure across our sample period. As discussed in Section VI.B, we estimate separate specifications based on three different measures of complexity: daily total demand, the daily standard deviation across locations and hours of real-time prices, and daily total starts.

Appendix Equation (E.1) controls for separate sets of calendar month fixed effects for high complexity days and low complexity days ($\alpha_{m,\text{HIGH}}$), an indicator for weekend-versus-weekday (θ_w), month-of-sample fixed effects ($\gamma_{y,m}$) and the variables in X_t: the log of total electricity demand, the log of net electricity imports, the log of the monthly average natural gas price, as well as separate controls for the logs of monthly total production from: (1) renewables, (2) nuclear sources, and (3) hydro sources. Specifically, we center each control variable in X_t; for each centered variable x in X_t, we include x, x^2, x^3, x^4 and ten separate indicators defined using the deciles of the distribution of x.

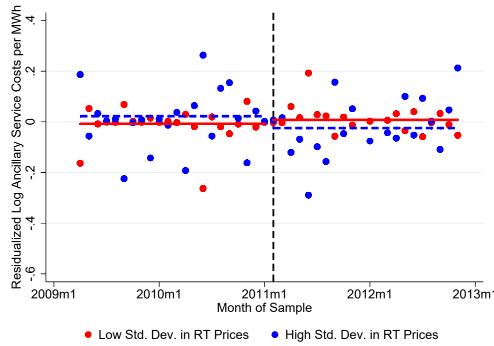
Appendix Figure E.1 plots the monthly average residuals from estimating Appendix Equation (E.1). In the top left and top right panels, we define "high complexity" using daily total demand and the daily standard deviation across locations and hours in real-time prices respectively. The bottom panel is based on defining complexity using the daily total number of starts by gas-fired units. The vertical black dashed line denotes the introduction of financial trading on February 1st 2011. The solid red horizontal lines plot the overall averages of residuals for low complexity days taken separately over the pre-FT and post-FT sample periods. Similarly, the dashed blue horizontal lines plot overall averages for high complexity days in the pre-FT versus post-FT sample periods.

The top left and bottom panels of Appendix Figure E.1 suggest that there is not much difference in residualized ancillary service costs per MWh before versus after the introduction in FT on either high or low complexity days when complexity is measured using either daily total demand or daily total starts. In contrast, the top right panel of Appendix Figure E.1 indicates that ancillary service costs per MWh fell on average

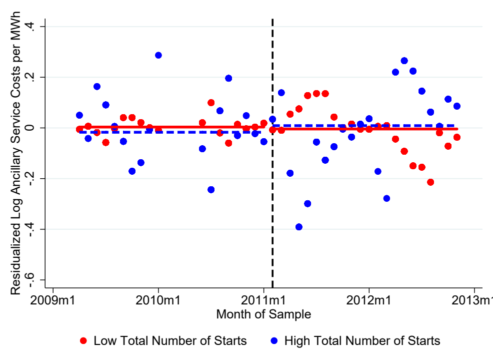
Figure E.1: Monthly Average Residualized Ancillary Service Costs per MWh



(a) Measure of Complexity: Daily Total Demand



(b) Measure of Complexity: Std. Dev. Across Locations and Hours of RT Price



(c) Measure of Complexity: Daily Total Starts

Notes: This figure plots the monthly averages of the residualized logarithm of ancillary service costs per MWh of gas-fired output for high complexity days versus low complexity days. We plot only months with both high and low complexity days. The top left, top right, and bottom panels of this figure define complexity using daily total demand, the daily standard deviation across locations and hours of real-time prices, and daily total number of starts by gas-fired units respectively. For each measure, day t is classified as "high complexity" if the value of the measure on day t is larger than the 75th percentile of the distribution of this measure across the sample period. Log ancillary service costs per MWh are residualized using the daily-level regression shown in Appendix Equation (E.1). The vertical black dashed line denotes the introduction of financial trading ("FT"). The solid red horizontal lines plot the overall averages of residuals for low complexity days taken separately over the pre-FT and post-FT sample periods. Similarly, the dashed blue horizontal lines plot overall averages for high complexity days in the pre-FT and post-FT sample periods.

after FT on days with a high standard deviation in real-time prices but not on days with a low standard deviation in real-time prices. Combined, the evidence suggests that, at the very least, ancillary service costs per MWh did not increase substantially after purely financial participation was allowed.[59]

E.2 Event Study: Additional Tables and Figures

Appendix Figure E.2 presents the monthly average residualized outcome for high complexity days minus the monthly average residualized outcome for low complexity days. We consider two measures of complexity: daily total demand and the daily standard deviation across locations and hours in real-time prices. For a given measure, a day is considered to have "high complexity" if the value of the measure on the day exceeds the 75th percentile of the distribution of this measure. We only plot average differences for months-of-sample with both high and low complexity days.

We residualize each outcome Y_t in day-of-sample t by estimating the following equation:

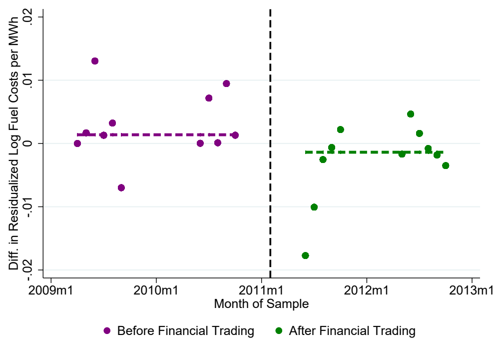
$$Y_t = \alpha_{m,\text{HIGH}} + \theta_w + \gamma_{y,m} + \sum_{s=1}^{S} \sum_{k=1}^{K} [(X_{k,t} - \overline{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[X_{k,t} \in \text{BIN}_{k,b}]] + u_t \quad \text{(E.2)}$$

where we include separate sets of calendar month fixed effects for high versus low complexity days ($\alpha_{m,\text{HIGH}}$), an indicator for whether the day-of-sample is weekday versus weekend (θ_w), and month-of-sample fixed effects ($\gamma_{y,m}$). We also control for the variables in X_t as discussed in Section VI.B.

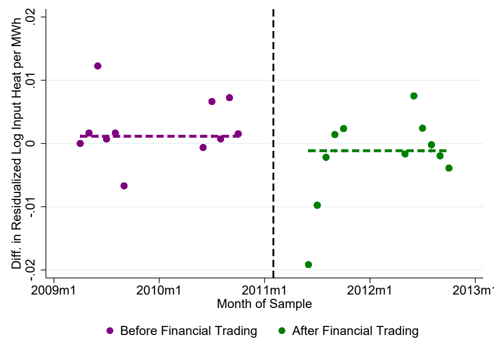
The two left panels of Appendix Figure E.2 focus on differences in the log of fuel costs per MWh of gas-fired output while the two right panels focus on differences in the log of input heat use per MWh of gas-fired output. This figure includes a vertical

[59] This is borne out by estimating the difference-in-differences regression specified in Equation (6) considering the log of ancillary service costs per MWh as the dependent variable. Specifically, we do not find a statistically significant increase in ancillary service costs per MWh for high complexity days relative to low complexity days after FT is introduced regardless of the measure of complexity considered, sets of controls included, or whether the outcome is trimmed or not. These results are available upon request.

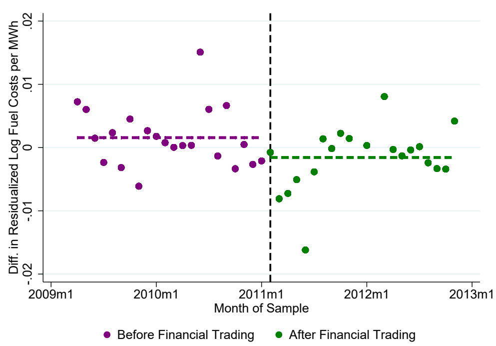
Figure E.2: Monthly Average Differences in Residualized Outcomes



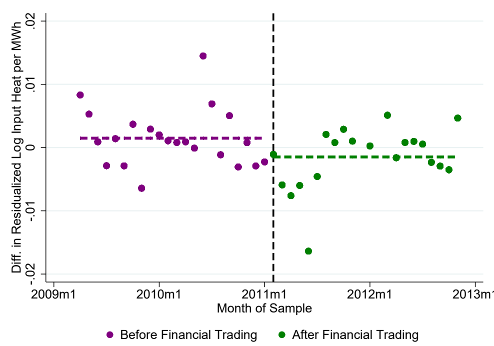
(a) Log Fuel Cost Per MWh
Measure of Complexity: Demand



(b) Log Input Energy Per MWh
Measure of Complexity: Demand



(c) Log Fuel Cost Per MWh
Measure of Complexity: SD[RT Price]



(d) Log Input Energy Per MWh
Measure of Complexity: SD[RT Price]

Notes: This figure plots the monthly average residualized outcome for high complexity days minus the monthly average residualized outcome for low complexity days. We plot only months with both high and low complexity days. For the top two panels, day t is classified as "highly complex" if daily total demand on the day is larger than the 75th percentile of the distribution of daily total demand. For the bottom two panels, day t is classified as "highly complex" if the daily standard deviation across locations and hours in the day is above the 75th percentile of the distribution of this measure. The relevant outcome is residualized using the daily-level regression shown in Appendix Equation (E.2). We consider the log of fuel costs per MWh of gas-fired output in the two left panels and the log of input heat per MWh of gas-fired output in the two right panels. The horizontal solid purple line (dashed green line) presents the overall average of the difference in residualized outcome across high versus low complexity days for the sample period before (after) the introduction of financial trading. Finally, the vertical dashed line denotes the introduction of financial trading.

dashed line denoting the introduction of financial trading ("FT"). The horizontal solid purple line (dashed green line) in the figure presents the overall average of the difference in residualized outcomes across high versus low complexity days for the sample period before (after) the introduction of FT. Appendix Figure E.2 documents that there are not substantial differences in the trends of monthly average residualized outcomes for high versus low complexity days prior to FT being introduced. Moreover, we see that residualized outcomes fell on average for high complexity days relative to low complexity days after the introduction of FT.

One might be concerned that the reduction in average residualized outcomes on high complexity days is driven by the six months before and after the introduction of FT. To assuage this concern, we plot the monthly averages of residualized outcomes for high complexity days and low complexity days excluding the six months before and after February 1, 2011. The overall averages for high complexity days and low complexity days, denoted using red and blue horizontal lines respectively, are also calculated excluding the six months before and after February 1st 2011. We see that average residualized outcomes fall after FT on high complexity days but not low complexity days even after excluding the six month window around February 1st 2011.

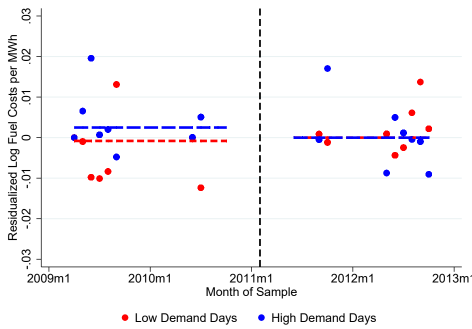
One might also be concerned that the base specification in Equation (5) "over-controls" for the economic factors in X_t. To assuage this concern, we consider specifications that control only linearly for the variables in X_t. Specifically, for Appendix Figure E.4, we residualize each outcome Y_t in day-of-sample t by estimating the following equation:

$$Y_t = \alpha_{m,\text{HIGH}} + \theta_w + \gamma_{y,m} + X_t\phi + u_t \tag{E.3}$$

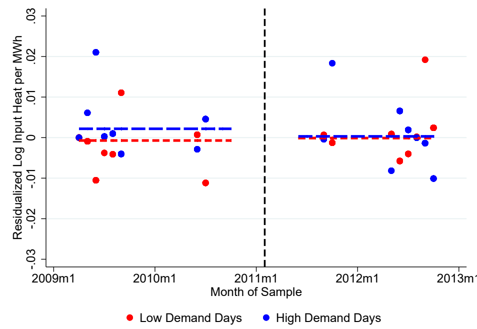
As before, the set of control variables included in X_t is the log of total electricity demand, the log of net electricity imports, the log of the monthly average natural gas price, as well as the logs of monthly total production from: (1) renewables, (2) nuclear sources, and (3) hydro sources.

Appendix Figure E.4 documents that the trends in monthly residualized outcomes

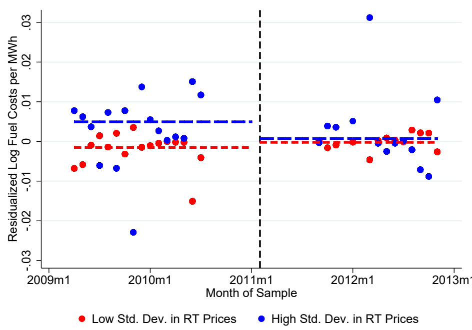
Figure E.3: Monthly Average Residualized Outcomes Before versus After Financial Trading Dropping the 6 Months Before and After FT



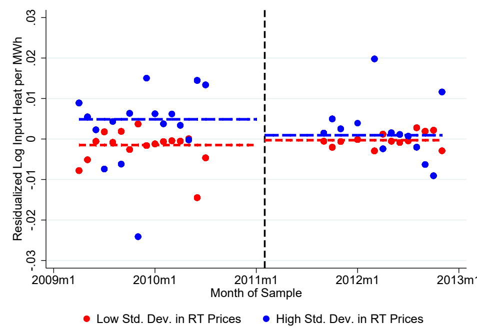
(a) Log Fuel Cost Per MWh
Measure of Complexity: Demand



(b) Log Input Energy Per MWh
Measure of Complexity: Demand



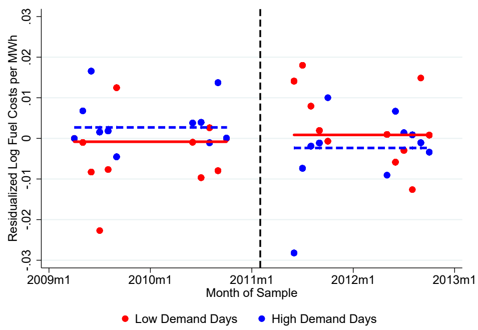
(c) Log Fuel Cost Per MWh
Measure of Complexity: SD[RT Price]



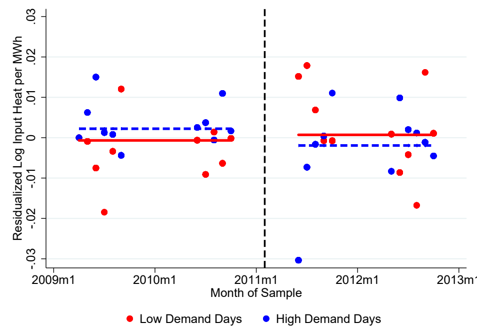
(d) Log Input Energy Per MWh
Measure of Complexity: SD[RT Price]

Notes: This figure plots the monthly averages of the residualized outcome for high complexity days versus low complexity days. We plot only months with both high complexity days and low complexity days. Moreover, we do not plot the six months before and after the introduction of financial trading ("FT") on February 1st 2011. Complexity is measured using daily total demand for the top two panels and the daily standard deviation over locations and hours of real-time prices for the bottom two figures. For a given measure of complexity, day t is defined as being "highly complex" if the value of the measure on the day is above the 75th percentile of the distribution of this measure across the sample period. Outcomes are residualized using the daily-level regression shown in Equation (5). We consider the log of fuel costs per MWh of gas-fired output in the two left panels and the log of input heat use per MWh of gas-fired output in two right panels. The vertical black dashed line denotes the introduction of FT. The solid red horizontal lines plot the overall averages of residuals for low complexity days taken separately over the pre-FT and post-FT sample periods. The dashed blue horizontal lines plot overall averages for high complexity days in the pre-FT versus post-FT sample periods. The six months before and after February 1, 2011 are not included when calculating the four overall averages denoted by the blue and red horizontal lines.

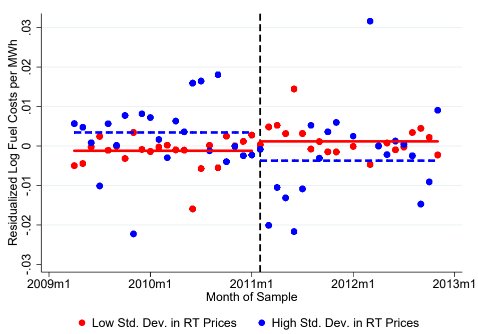
Figure E.4: Monthly Average Residualized Outcomes Before versus After Financial Trading: No Nonlinear Controls



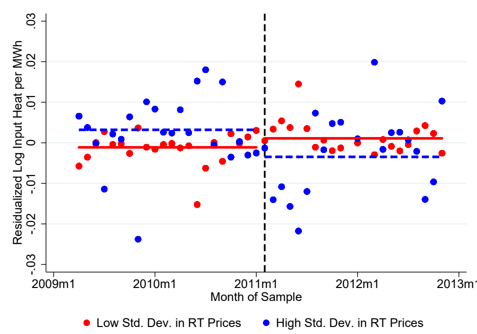
(a) Log Fuel Cost Per MWh
Measure of Complexity: Demand



(b) Log Input Energy Per MWh
Measure of Complexity: Demand



(c) Log Fuel Cost Per MWh
Measure of Complexity: SD[RT Price]



(d) Log Input Energy Per MWh
Measure of Complexity: SD[RT Price]

Notes: This figure plots the monthly averages of the residualized outcome for high complexity days versus low complexity days. We plot only months with both high complexity days and low complexity days. Complexity is measured using daily total demand for the top two panels and the daily standard deviation over locations and hours of real-time prices for the bottom two figures. For a given measure of complexity, a day is defined as being "high complexity" if the value of the measure on the day is above the 75th percentile of the distribution of this measure. In contrast to Equation (5), residuals are calculated using the daily-level regression specified in Appendix Equation (E.3) which does not include nonlinear functions of the control variables in X_t. The vertical black dashed line denotes the introduction of financial trading ("FT"). The solid red horizontal lines plot the overall averages of residuals for low complexity days taken separately over the pre-FT and post-FT sample periods; the dashed blue horizontal lines plot overall averages for high complexity days in the pre-FT and post-FT sample periods.

for both high complexity days and low complexity days remain similar even if we only control linearly for the variables in X_t. Indeed, the trends are quite similar to those from our primary specification presented in Figure 5. Specifically, we see that the overall average of each residualized outcome falls after the introduction of FT on high complexity days but not low complexity days, which is consistent with the mechanism described in Section II.

E.3 Statistical Test of Common Trends Using First-Differences

The definition of "common pre-existing trends" is that the slope over time in outcomes is the same for high versus low complexity days. The "slope over time" is simply the first difference in outcomes: $\Delta Y_t = Y_t - Y_{t-1}$. Thus, to formally test the "common pre-existing trends" assumption, we estimate the following regression model using only data from before the introduction of FT:

$$\Delta Y_t = (\Delta \overrightarrow{M_t})\phi + \beta \text{HIGH}_t + \epsilon_t \tag{E.4}$$

For Columns 1 and 3 of Appendix Table E.1, HIGH_t is an indicator variable that is equal to one if and only if daily total demand on day-of-sample t is larger than the 75th percentile of the distribution of daily total demand across our sample period. For Columns 2 and 4 of this table, HIGH_t is equal to one if the standard deviation in real-time prices across locations and hours on day t is larger than the 75th percentile of the distribution of daily standard deviations. For ease of exposition, we refer to days with $\text{HIGH}_t = 1$ as high complexity days, recognizing that this indicator is defined based on demand in some specifications and the standard deviation in real-time prices in other specifications.

All specifications control for the first differences of the variables in $\overrightarrow{M_t}$. The variables included in $\overrightarrow{M_t}$ are indicators corresponding to separate sets of fixed effects for

Table E.1: Checking For Common Pre-Existing Trends Using First-Differences

Dep. Var.	First Diff. of Log Fuel Cost per MWh		First Diff. of Log Input Energy per MWh	
	(1)	(2)	(3)	(4)
POSTFT_t	0.006	0.005	0.005	0.005
	(0.007)	(0.005)	(0.007)	(0.005)
R^2	0.972	0.972	0.653	0.660
Mean of Dep. Var.	0.095	0.095	-0.003	-0.003
Measure: Total Demand	Y	N	Y	N
Measure: SD RT Price	N	Y	N	Y
Number of Obs.	670	670	670	670

Notes: This table presents evidence that pre-existing differential trends in outcomes across high versus low complexity days are not driving the difference-in-differences results presented in Table 4. The unit of observation for these regressions is day-of-sample; the regressions are estimated using only days before the introduction of financial trading. For Columns 1 and 3 of this table, the indicator variable HIGH_t is equal to one if and only if daily total demand on day t is greater than the 75th percentile of the distribution of daily total demand across the sample period. For Columns 2 and 4, HIGH_t is equal to one if the daily standard deviation across locations and hours of real-time prices on day t is greater than the 75th percentile of the distribution of daily standard deviations. The dependent variable considered in the first two columns of this table is the first difference of the log of fuel costs per MWh; the dependent variable considered in Columns 3 and 4 of this table is the first difference of the log of input energy use per MWh. The row titled "Mean of Dep. Var." reports the mean of the relevant dependent variable. All of the regressions listed in this table control for the first differences of the fixed effects and control variables described for Equation (6) in Section VI.C; see Appendix Equation (E.4) for more details. Standard errors are clustered by week-of-sample and are reported in parentheses.

high versus low complexity days, month-of-sample fixed effects, and weekend versus weekday fixed effects as well as the linear and nonlinear functions of X_t specified in Equation (6). Standard errors are clustered by week-of-sample.

Appendix Table E.1 presents the results from estimating Appendix Equation (E.4). These results indicate that, for both outcome variables and both indicators of complexity, we cannot reject the null hypothesis that the first difference of the outcome is the same in high versus low complexity days prior to February 1st 2011. This provides statistical evidence that the findings from our difference-in-differences framework are not driven by pre-existing differences in the time trend of our outcomes in high versus low complexity days.

E.4 Robustness to Percentage Cut-Off for Complexity

In this subsection, we estimate the difference-in-differences regression specified in Equation (6) defining days with a "high" complexity based on different cut-offs. Specifically, in Appendix Table E.2, we define day t as having "high complexity" if daily total demand on day t is higher than the Xth percentile of the distribution of daily total demand; X is equal to 50, 60, 70, 80, and 90 for Columns 1, 2, 3, 4, and 5 of Appendix Table E.2 respectively. In Appendix Table E.3, we define complexity using the standard deviation across locations and hours of real-time prices. As with Appendix Table E.2, Columns 1, 2, 3, 4, and 5 consider the 50th, 60th, 70th, 80th, and 90th percentiles of the distribution of daily standard deviations respectively.

The top panel of Appendix Table E.2 shows that the estimated reduction in average fuel costs per MWh after financial trading on relatively high demand days remains statistically significant whether "high demand" is defined as days-of-sample above the 50th, 60th, 70th, 80th, or 90th percentiles of daily total demand. The corresponding reductions in input energy use per MWh also remain statistically significant regardless of the cut-off used to define high demand days. This demonstrates that our results are not an artifact of choosing the 75th percentile of the distribution of daily total demand as the cut-off in our primary specifications. Moreover, the results remain similar when defining high complexity days using different percentiles of the daily standard deviation across locations and hours in real-time price rather than daily total demand (see Appendix Table E.3).

Focusing on the top panel of Appendix Table E.2, the estimated effects using the 50th, 60th, or 70th percentiles imply similar fuel cost savings. Specifically, these estimates suggest that fuel costs fell by roughly 24-38 million dollars on high demand days after financial trading was introduced. The first three columns of the bottom panel indicate that the corresponding reductions in input energy resulted in a decrease in CO_2 emissions of roughly 258-428 thousand tons on high demand days. However, the estimates of the aggregate fuel cost savings and carbon emissions reductions are

Table E.2: Diff-in-Diff Robustness Check: By Percentage of Demand

Dependent Variable: Log of Average Fuel Cost Per MWh					
	(1)	(2)	(3)	(4)	(5)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.019	-0.032	-0.026	-0.022	-0.028
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
Cut-Off Percentage	50	60	70	80	90
Fuel Cost Savings (Million USD)	27.525	37.757	24.650	15.353	11.049
R^2	0.964	0.965	0.964	0.960	0.960
Mean of Dep. Var.	3.680	3.680	3.680	3.680	3.680
Number of Obs.	1,340	1,340	1,340	1,340	1,340
Dependent Variable: Log of Average Input Heat Use Per MWh					
	(1)	(2)	(3)	(4)	(5)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.017	-0.030	-0.022	-0.020	-0.025
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
Cut-Off Percentage	50	60	70	80	90
CO_2 Reductions (Tons)	295,929	427,562	258,475	168,086	120,846
R^2	0.735	0.741	0.735	0.714	0.712
Mean of Dep. Var.	2.051	2.051	2.051	2.051	2.051
Number of Obs.	1,340	1,340	1,340	1,340	1,340

Notes: This table presents the difference-in-differences estimates of the change in fuel costs per MWh and input heat energy per MWh after financial trading ("FT") is introduced on high demand days relative to low demand days. The unit of observation for these regressions is day-of-sample. The "Post FT" indicator is equal to one if and only if the day-of-sample is on or after February 1st 2011. The indicator variable HIGH_t is equal to one if and only if daily total demand in day t is greater than the Xth percentile of the distribution of daily total demand across our sample period; X is equal to the 50th, 60th, 70th, 80th, or 90th percentile depending on whether we're considering the specification estimated in Columns 1, 2, 3, 4, or 5 respectively. All of the regressions listed in this table include the sets of fixed effects and control variables specified in Equation (6) in Section VI.C. Standard errors are clustered by week-of-sample and are reported in parentheses.

Table E.3: Diff-in-Diff Robustness Check: By Percentage of SD[Real-Time Prices]

Dependent Variable: Log of Average Fuel Cost Per MWh					
	(1)	(2)	(3)	(4)	(5)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.006	-0.006	-0.011	-0.014	-0.013
	(0.003)	(0.004)	(0.004)	(0.005)	(0.006)
Cut-Off Percentage	50	60	70	80	90
Fuel Cost Savings (Million USD)	8.008	6.203	8.810	7.289	3.395
R^2	0.960	0.960	0.960	0.960	0.960
Mean of Dep. Var.	3.680	3.680	3.680	3.680	3.680
Number of Obs.	1,340	1,340	1,340	1,340	1,340
Dependent Variable: Log of Average Input Heat Use Per MWh					
	(1)	(2)	(3)	(4)	(5)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.006	-0.005	-0.010	-0.014	-0.014
	(0.003)	(0.004)	(0.004)	(0.005)	(0.006)
Cut-Off Percentage	50	60	70	80	90
CO_2 Reductions (Tons)	89,056	64,765	95,512	84,647	45,576
R^2	0.715	0.714	0.715	0.715	0.712
Mean of Dep. Var.	2.051	2.051	2.051	2.051	2.051
Number of Obs.	1,340	1,340	1,340	1,340	1,340

Notes: This table presents the difference-in-differences estimates of the change in fuel costs per MWh and input heat energy per MWh after financial trading ("FT") is introduced on days with a relatively high daily standard deviation in real-time prices. The unit of observation for these regressions is day-of-sample. The "Post FT" indicator is equal to one if and only if the day-of-sample is on or after February 1st 2011. The indicator variable HIGH_t is equal to one if and only if the standard deviation across locations and hours in real-time prices for day t is greater than the Xth percentile of the distribution of daily standard deviations across our sample period; X is equal to the 50th, 60th, 70th, 80th, or 90th percentile depending on whether we're considering the specification estimated in Columns 1, 2, 3, 4, or 5 respectively. All of the regressions listed in this table include the sets of fixed effects and control variables specified in Equation (6) in Section VI.C. Standard errors are clustered by week-of-sample and are reported in parentheses.

far smaller if we instead consider the 80th or 90th percentiles of the distribution of daily total demand. This is because we are applying a similarly sized effect to far fewer days when considering the 80th or 90th percentiles of daily total demand as the cut-off. For this reason, we consider the 75th percentile of daily total demand as the cut-off for our primary specifications.

E.5 Excluding Months After the San Onofre Nuclear Plant Shutdown

Davis and Hausman (2016) studies the shut down of the San Onofre nuclear power plant in February 2012. One may be concerned that this shut down impacts our estimates of the reductions in fuel cost per MWh after financial trading on high complexity days relative to low complexity days. To assuage this concern, we note that our primary specifications control for a host of economic factors, including monthly total output from nuclear plants in California, as follows. First, we center each control variable; for each centered control variable x, our specification includes x, x^2, x^3, x^4 and ten separate indicators defined using the deciles of the distribution of x.

To further assuage this concern, we estimate the difference-in-differences regression specified in Equation (6) excluding the months after the San Onofre plant shut down. Namely, we estimate Equation (6) considering only the sample period 4/1/2009-1/31/2012

Appendix Table E.4 presents the results from this estimation. Columns 1, 2, and 3 define high complexity days based on the 75th percentile of the distribution of daily total demand, daily standard deviation in real-time prices, and daily total starts by gas-fired units respectively. The top panel considers the log of fuel costs per MWh while the bottom panel focuses on the log of input fuel use per MWh. Regardless of the measure of complexity considered, the reductions in fuel cost per MWh and input fuel use per MWh on high complexity days after FT remain precisely estimated and similar

Table E.4: Diff-in-Diff Specification Excluding Days After January 31, 2012

Dep. Var.: Log Fuel Cost per MWh			
	(1)	(2)	(3)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.028	-0.018	-0.018
	(0.007)	(0.006)	(0.005)
R^2	0.958	0.955	0.957
Mean of Dep. Var.	3.717	3.717	3.717
Number of Obs.	1,036	1,036	1,036
Measure: Total Demand	Y	N	N
Measure: SD RT Price	N	Y	N
Measure: Total Starts	N	N	Y

Dep. Var.: Log Input Energy Use per MWh			
	(1)	(2)	(3)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.025	-0.018	-0.018
	(0.007)	(0.006)	(0.005)
R^2	0.746	0.735	0.746
Mean of Dep. Var.	2.051	2.051	2.051
Number of Obs.	1,036	1,036	1,036
Measure: Total Demand	Y	N	N
Measure: SD RT Price	N	Y	N
Measure: Total Starts	N	N	Y

Notes: This table presents the difference-in-differences estimates of the change in outcome after the introduction of financial trading ("FT") on high complexity days relative to low complexity days. The unit of observation for these regressions is day-of-sample. The dependent variable considered in the top (bottom) panel of this table is the log of fuel costs per MWh (the log of input energy per MWh). Columns 1, 2, and 3 of each panel of the table measure complexity using daily total demand, daily standard deviation in real-time prices, and daily total starts respectively. For a given measure of complexity, the indicator variable HIGH_t is equal to one if and only if the value of the measure on day t is higher than the 75th percentile of the distribution of this measure across the 4/1/2009-1/31/2012 sample period used for this table. The "Post FT" indicator is equal to one if and only if the day-of-sample is on or after February 1st 2011. All of the regressions listed in this table include the sets of fixed effects and control variables specified in Equation (6) in Section VI.C. Standard errors are clustered by week-of-sample and are reported in parentheses.

in magnitude when estimated on the 4/1/2009-1/31/2012 sample period rather than the full 4/1/2009-11/30/2012 sample period. This suggests that our primary estimates do not stem from the shut down of the San Onofre nuclear plant.

F Potential Mechanism Underlying Efficiency Gains from Financial Trading

This section is split into four parts. In the first subsection, we present descriptive evidence that increases in our three measures of complexity are associated with increases in systemwide fuel costs per MWh. Our three measures of complexity are daily total demand, the daily standard deviation across pricing locations and hours of real-time prices, and daily total number of unit start-ups. The second subsection presents suggestive evidence that the aggregate marginal cost curve becomes steeper as the residual demand to be served by the gas-fired fleet increases. The third subsection discusses results from difference-in-differences specifications defining "high complexity" days using daily total number of starts. The final subsection explores differences in the start-up behavior of units with larger versus smaller fuel costs per MWh before versus after financial trading on high versus low complexity days (i.e., a "triple-differences approach").

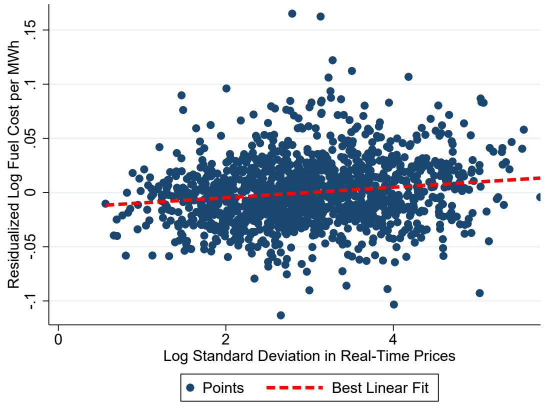
F.1 Measures of Complexity and Fuel Costs

Appendix Figure F.1 plots the relationship between our three measures of system complexity and residualized log fuel costs per MWh. We residualize the log of fuel costs per MWh of gas-fired output using the following equation:

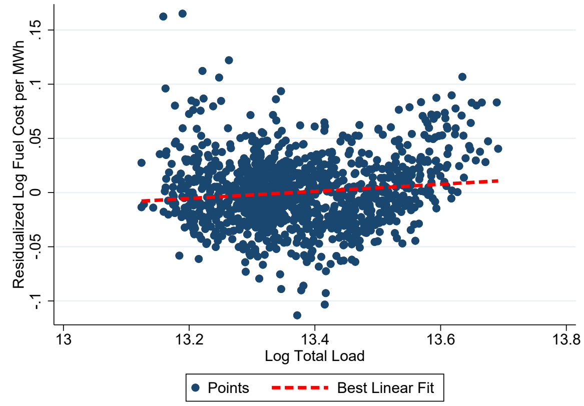
$$Y_t = \theta_w + \gamma_{y,m} + \sum_{s=1}^{S}\sum_{k=1}^{K}[(Z_{k,t} - \overline{Z}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[Z_{k,t} \in \text{BIN}_{k,b}]] + u_t \qquad \text{(F.1)}$$

for day-of-sample t in calendar month m and year-of-sample y. This specification includes month-of-sample fixed effects ($\alpha_{m,y}$) and an indicator for whether the day-of-sample is a weekday versus weekend (θ_w). We also control for the variables in Z_t: the log of daily net electricity imports, the log of the monthly average natural gas price paid by power plants in CAISO, as well as logs of monthly total production from: (1)

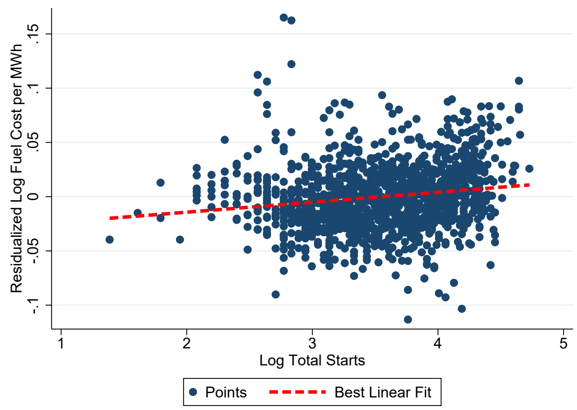
Figure F.1: Residualized Fuel Cost Per MWh and Measures of Complexity



(a) Log Std. Dev. of Real-Time Prices



(b) Log Daily Total Load



(c) Log Daily Total Starts

Notes: This figure documents the relationship between residualized daily total fuel costs per MWh and each of our three measures of complexity. We residualize log fuel costs per MWh using the regression specified in Appendix Equation (F.1). The x-axis plots the relevant measure of complexity: (1) the log of the daily standard deviation in real-time prices across locations and hours of the day in the top left panel, (2) the log of daily total demand in the top right panel and (3) the log of daily total number of starts by gas-fired units in the bottom middle panel.

renewables, (2) nuclear sources, and (3) hydro sources. Specifically, we center each variable in Z_t; for each centered control variable z, our specification includes z, z^2, z^3, z^4 and ten separate indicators defined using the deciles of the distribution of z. In contrast to Equations (5) and (6), we do not control for the log of daily total demand because daily total demand is one of our three measures of complexity.

All three panels of Appendix Figure F.1 document substantial variation in residualized log fuel costs per MWh that is not explained by the relevant measure of complexity. Nevertheless, the best linear fit between residualized log fuel costs per

MWh and each measure of complexity has a positive slope. The estimated slopes are 0.005, 0.033, and 0.009 for the log of the daily standard deviation in real-time prices, the log of daily total demand, and the log of daily total number of starts respectively. The correlation between residualized log fuel cost per MWh and the relevant measure of complexity is 0.150, 0.120, and 0.162 for the log of daily standard deviation in real-time prices, log demand, and log number of starts respectively. Combined, this evidence indicates that increases in each of our three measures of complexity are associated with increases in fuel costs per MWh.

F.2 Marginal Fuel Cost Curves

In this subsection, we present crude estimates of the aggregate marginal fuel cost curve in California's wholesale electricity market. The goal of this subsection is only to provide suggestive evidence that the marginal fuel cost of the marginal unit increases at an increasing rate as the residual demand to be served by the gas-fired fleet increases. We fully acknowledge that we ignore several important factors that enter marginal costs, such as variable operating and maintenance costs and the allowance costs associated with nitrogen oxide emissions.

We calculate each unit's marginal fuel cost quite simply: each unit's marginal fuel cost is its aggregate fuel costs over the sample period divided by its output over the sample period. Appendix Figure F.2 plots the resulting marginal cost curve as a function of the cumulative output of the gas-fired fleet. The x-axis for the two left panels is hourly cumulative output while the x-axis for the two right panels is daily cumulative output. For the top two panels of Appendix Figure F.2, we assume each unit is producing at capacity, as measured by its maximum hourly output over the sample period. For the bottom left (right) panel, we choose an example hour (day) where the total output produced by the gas-fired fleet is especially high; we then simply use the unit's observed output in the hour (day).[60] Finally, we plot the 50th,

[60]The example day chosen is August 13, 2012. We use the 1pm-2pm interval on this day for the

75th, 90th, and 95th percentiles of the distribution of hourly (daily) total gas-fired output as vertical dashed lines in the two left (right) panels of Appendix Figure F.2.

It is clear from Appendix Figure F.2 that the marginal cost curve becomes significantly steeper as the residual demand to be served by the gas-fired fleet increases. The marginal cost curve is especially steep at the very highest levels of residual demand. That being said, even the 95th percentile of residual demand falls well short of the steepest portion of the marginal cost curve. Combined, these figures provide suggestive evidence that there are larger potential gains from reallocation of output across units at higher levels of residual demand to be served by the gas-fired fleet.

F.3 Specifications Based on Number of Starts

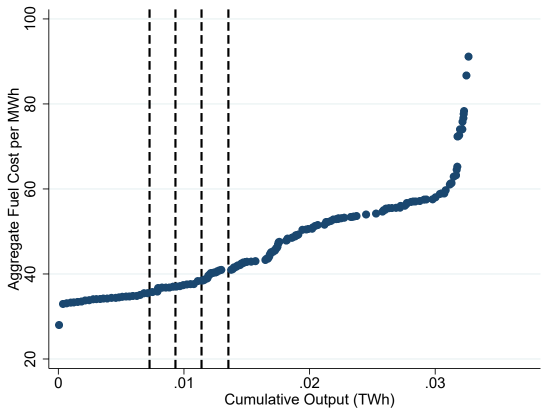
This subsection compares market outcomes before versus after the introduction of financial trading on days with more versus less starts by gas-fired units. We estimate the following specification in order to quantify how our two outcome variables change after financial trading on days with a relatively high number of starts:

$$\begin{gathered} Y_t = \alpha_{m,\mathrm{HIGH}} + \theta_w + \gamma_{y,m} + \delta_{DD}(\mathrm{HIGH}_t \times \mathrm{POSTFT}_t) + \\ \sum_{s=1}^{4}\sum_{k=1}^{K}\left[(X_{k,t} - \overline{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10}\theta_{k,b}1[X_{k,t} \in \mathrm{BIN}_{k,b}]\right] + u_t \end{gathered} \tag{F.2}$$

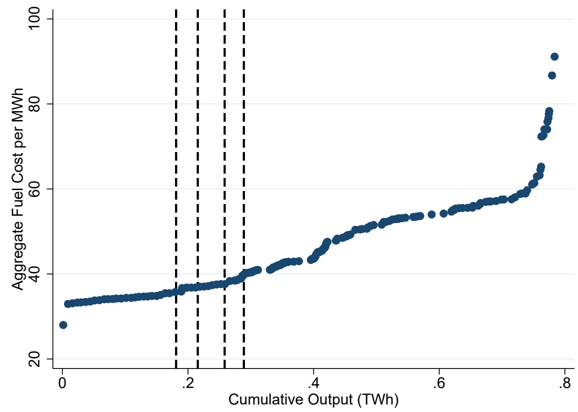
where we define HIGH_t to be an indicator that is equal to one if and only if daily total number of starts on day-of-sample t is above the k^{th} percentile of the distribution of daily total starts across our 4/1/2009-11/30/2012 sample period; we consider specifications based on the 50th, 60th, 70th, 80th and 90th percentiles of the distribution of starts. All regressions include separate sets of calendar month fixed effects for days with a high versus low number of starts ($\alpha_{m,\mathrm{HIGH}}$), weekend versus weekday fixed effects (θ_w), and month-of-sample fixed effects ($\gamma_{m,y}$). In addition, we control for the variables in $\overrightarrow{X_t}$ in the same way as discussed in Section VI.C. Finally, standard errors

hourly figure.

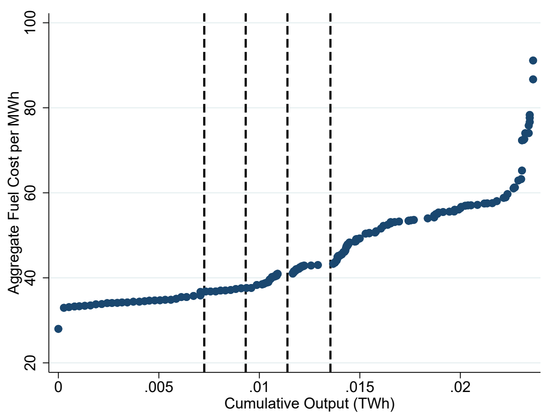
Figure F.2: Hourly and Daily Marginal Cost Curves



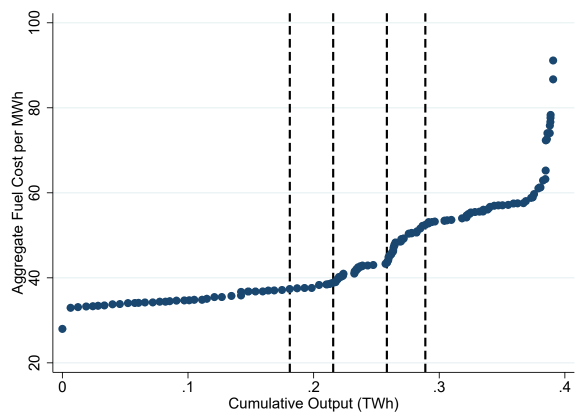
(a) Hourly, Capacity-Based



(b) Daily, Capacity-Based



(c) Hourly, Observed Output



(d) Daily, Observed Output

Notes: This figure presents aggregate marginal cost curves constructed by stacking units based on their aggregate fuel costs per MWh. We calculate each unit's marginal cost as the unit's fuel cost over the sample period divided by the unit's total output over the sample period. The x-axis of each figure is the cumulative output of gas-fired units with marginal cost less than the value listed: the left panels plot hourly cumulative output while the right panels plot daily cumulative outputs. We assume that each unit produces at its capacity for the top two panels; each unit's capacity is defined to be its maximum hourly output across the sample period. The bottom right panel uses each unit's observed output from August 13 2012; the bottom left panel uses each unit's output from the 1pm-2pm interval on August 13 2012. Finally, the left (right) panels also include four vertical dashed lines with the 50th, 75th, 90th, and 95th percentiles of the distribution of hourly (daily) total observed output from gas-fired units.

Table F.1: Diff-in-Diff Robustness Check: By Percentage of Daily Starts

	(1)	(2)	(3)	(4)	(5)
Log of Fuel Cost Per MWh					
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.012	-0.012	-0.010	-0.007	-0.009
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Cut-Off Percentage	50	60	70	80	90
Fuel Cost Savings (Million USD)	16.185	13.701	8.836	4.871	3.164
R^2	0.962	0.962	0.961	0.961	0.961
Mean of Dep. Var.	3.680	3.680	3.680	3.680	3.680
Number of Obs.	1,340	1,340	1,340	1,340	1,340
Log of Input Heat Per MWh	(1)	(2)	(3)	(4)	(5)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.010	-0.011	-0.009	-0.007	-0.009
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Cut-Off Percentage	50	60	70	80	90
CO_2 Reductions (Tons)	168,160	153,552	97,466	51,248	38,281
R^2	0.726	0.727	0.723	0.721	0.718
Mean of Dep. Var.	2.051	2.051	2.051	2.051	2.051
Number of Obs.	1,340	1,340	1,340	1,340	1,340

Notes: This table presents the difference-in-differences estimates of the change in fuel costs per MWh and input heat energy per MWh after the introduction of financial trading ("FT") on days with a high versus low number of times that gas-fired units started up. The unit of observation for these regressions is day-of-sample. The "Post FT" indicator is equal to one if and only if the day-of-sample is on or after February 1st 2011. The indicator variable HIGH_t is equal to one if the daily total number of starts in day t is greater than the Xth percentile of the distribution of daily total starts across our sample period, where X is equal to the 50th, 60th, 70th, 80th, or 90th percentile depending on whether we're considering the specification estimated in Columns 1, 2, 3, 4, or 5 respectively. All of the regressions listed in this table include the sets of fixed effects and control variables specified in Equation (6) in Section VI.C. Standard errors are clustered by week-of-sample and are reported in parentheses.

are clustered by week-of-sample.

Appendix Table F.1 demonstrates that our estimates are negative and precisely estimated regardless of whether we consider days-of-sample with total number of starts above the 50th, 60th, 70th, 80th, or 90th percentiles of the distribution of daily total starts. The estimated reductions in fuel costs per MWh after financial trading on days with a relatively large number of starts are roughly 1% across specifications. These estimates are similar in magnitude to the corresponding estimates for high demand days and high standard deviation days from Appendix Tables E.2 and E.3

respectively. This suggests that one mechanism by which purely financial participation lowers production costs is changes in the type of units that start up on days requiring a larger number of unit start-ups. We explore this hypothesis in the next subsection.

F.4 Starts on High Complexity Days Before versus after Financial Trading

This subsection presents estimates of the differences in the number of starts by gas-fired units before versus after the introduction of financial trading on high complexity days versus low complexity days. We first employ the same difference-in-differences specification as in Section VI.C:

$$
\begin{aligned}
Y_t &= \alpha_{m,\mathrm{HIGH}} + \theta_w + \gamma_{y,m} + \delta_{DD}(\mathrm{HIGH}_t \times \mathrm{POSTFT}_t) \\
&+ \sum_{s=1}^{S}\sum_{k=1}^{K}[(X_{k,t} - \overline{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[X_{k,t} \in \mathrm{BIN}_{k,b}]] + u_t
\end{aligned} \tag{F.3}
$$

where t indexes day-of-sample in calendar month m in year y. The outcome variable Y_t is the log of the total number of starts by gas-fired units on day t divided by the total output from gas-fired units on day t. The indicator variable POSTFT_t is equal to one if day-of-sample t is on or after the introduction of FT.

As before, we consider two different indicators of the complexity of the optimization problems to be solved to clear real-time markets: total daily demand and the daily standard deviation of real-time prices. For the first three columns of Appendix Table F.2, the indicator variable HIGH_t is equal to one if and only if daily total demand on day t is higher than the 75th percentile of the distribution of daily total demand across our sample period. For the last three columns of this table, HIGH_t is equal to one if the standard deviation across locations and hours of real-time prices on day t is larger than the 75th percentile of the distribution of these daily standard deviations.

The independent variable of interest is $\mathrm{HIGH}_t \times \mathrm{POSTFT}_t$, which captures the difference in starts per MWh on high complexity days relative to low complexity days

Table F.2: Change in Starts After FT on Relatively High Complexity Days

	Log Total Number of Starts per MWh					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{HIGH}_t \times \text{POSTFT}_t$	-0.117	-0.115	-0.132	-0.078	-0.088	-0.079
	(0.068)	(0.067)	(0.067)	(0.046)	(0.044)	(0.045)
R^2	0.610	0.598	0.598	0.656	0.647	0.643
Mean of Dep. Var.	-8.456	-8.456	-8.456	-8.456	-8.456	-8.456
Trimmed Dep. Var.?	N	Y	N	N	Y	N
No Nonlinear Controls	N	N	Y	N	N	Y
Measure: Total Demand	Y	Y	Y	N	N	N
Measure: SD RT Price	N	N	N	Y	Y	Y
Number of Obs.	1,340	1,314	1,340	1,340	1,314	1,340

Notes: This table presents the difference-in-differences estimates of the change in outcome after the introduction of financial trading ("FT") on relatively high complexity days. The unit of observation for these regressions is day-of-sample. For the first three columns of each panel, the indicator variable HIGH_t is equal to one if and only if daily total demand on day t is higher than the 75th percentile of the distribution of daily total demand across the sample period. For the last three columns, HIGH_t is equal to one if the daily standard deviation over locations and hours of real-time prices in day t is greater than the 75th percentile of the distribution of daily standard deviations. The dependent variable considered is the log of the total number of starts by gas-fired units divided by daily total output by gas-fired units. The "Post FT" indicator is equal to one if and only if the day-of-sample is on or after February 1st 2011. The regressions underlying the estimates presented in Columns 1, 2, 4 and 5 include the sets of fixed effects and control variables described in Appendix Equation (F.3). The set of controls X_t is included only linearly for the regressions underlying Columns 3 and 6. In Columns 2 and 5, we trim the top and bottom 1% of the outcome before estimating the regression. Standard errors are clustered by week-of-sample and are reported in parentheses.

after relative to before FT. As before, our primary specifications control for X_t: the log of total electricity demand, the log of net electricity imports, the log of monthly average natural gas prices, as well as separate controls for the log of monthly total production from: (1) renewables, (2) nuclear sources, and (3) hydro sources. Specifically, we center each variable in X_t; for each centered variable x in X_t, the specification includes x, x^2, x^3, x^4 and ten separate indicators defined using the deciles of the distribution of x. Finally, standard errors are clustered by week-of-sample.

The results are presented in Appendix Table F.2. The estimated reductions in starts per MWh are precisely estimated regardless of which of the two indicators of complexity are used. Moreover, Columns 2 and 5 demonstrate that the results remain similar if we trim the top 1% and bottom 1% of the distribution of the dependent variable prior to estimating the regressions. Finally, in Columns 3 and 6, we show that the estimates remain similar if we control for the set of variables in X_t only linearly rather than include the nonlinear terms specified in Appendix Equation (F.3). All told, the estimates in Appendix Table F.2 indicate that the number of gas-fired units that start up to produce a given level of gas-fired output falls after the introduction of financial trading on relatively high complexity days.

To explore which types of units are less likely to start up after FT, we categorize a unit as "baseload" if the unit's aggregate fuel costs per MWh are in the bottom half of the distribution across units of this magnitude; units in the top half of the distribution of aggregate fuel costs per MWh are categorized as "peakers". With this categorization in hand, we estimate the following regression:

$$\begin{aligned} Y_{i,t} = \alpha_{i,m,\text{HIGH}} + \gamma_{i,m,y} + \theta_w + \delta_{DDD}(\text{PEAKER}_i \times \text{HIGH}_t \times \text{POSTFT}_t) \\ + \sum_{s=1}^{S}\sum_{k=1}^{K}[(X_{k,t} - \overline{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[X_{k,t} \in \text{BIN}_{k,b}]] + u_{i,t} \end{aligned} \tag{F.4}$$

where i indexes type of unit (either baseload or peaker) and t indexes day-of-sample in calendar month m in year y. For the first two columns of Appendix Table F.3, the outcome variable $Y_{i,t}$ is the log of total starts. We drop observations with zero total starts from this regression. As a robustness check, we also consider the inverse hyperbolic sine of total starts as the dependent variable (see Columns 3 and 4). Finally, we estimate the model using a Poisson regression in Columns 5 and 6 of Appendix Table F.3. Both these models allow us to include observations with zero total starts.

As before, the indicator variable POSTFT_t is equal to one if and only if day-of-sample t is on or after the introduction of FT. For Columns 1, 3, and 5 of Appendix Table F.3, the indicator variable HIGH_t is equal to one if and only if daily total demand on day t is larger than the 75th percentile of the distribution of daily total demand across our sample period. For Columns 2, 4, and 6, HIGH_t is equal to one if the standard deviation across locations and hours of real-time prices on day t is greater than the 75th percentile of the distribution of daily standard deviations.

All specifications include separate sets of type of unit by calendar month fixed effects for days with $\text{HIGH}_t = 1$ versus $\text{HIGH}_t = 1$ (i.e.: $\alpha_{i,m,\text{HIGH}}$), type by month-of-sample fixed effects (i.e.: $\gamma_{i,m,y}$), and an indicator for weekday versus weekend (i.e.: θ_w). We control for the same variables X_t in the same way as discussed above for Appendix Equation (F.2). Finally, standard errors are clustered by week-of-sample.

The independent variable of interest is $\text{PEAKER}_i \times \text{HIGH}_t \times \text{POSTFT}_t$, which captures the difference in starts for peakers relative to baseload units on high complexity days relative to low complexity days after relative to before the introduction of financial trading. Of course, we also include each of the three "main effects" as well as the three two-way interactions defined by these three variables. Note that some of the main effects and interactions are absorbed by the fixed effects considered in the specification.

The estimated reductions in starts for peaker units relative to baseload units after financial trading on relatively complex days remains precisely estimated regardless of: (1) whether complexity is measured using daily total demand or the daily standard deviation in real-time prices (Columns 1, 3, and 5 versus Columns 2, 4, and 6), (2) whether we take the log or the inverse hyperbolic sine before estimating the linear regression (Columns 1 and 2 versus Columns 3 and 4), and (3) whether we estimate the model using linear regression or Poisson regression (Columns 1-4 versus Columns 5 and 6).

In the previous subsection, we documented that fuel costs per MWh fell after financial trading was introduced on days with a relatively high number of starts. We hypothesized that this reduction in fuel costs came from a switch in the type of units that were started up to meet demand during times when solving the optimization problems required to clear the real-time market were complex. Appendix Table F.3 provides evidence consistent with this hypothesis. Namely, focusing on Column 1, our estimates indicate that peakers start up roughly 35% less times than baseload units on relatively high demand days after financial trading was introduced. This concords

Table F.3: Changes in Starts By Plant Type After Financial Trading on Relatively High Complexity Days

	(1)	(2)	(3)	(4)	(5)	(6)
$\text{PEAKER}_i \times \text{HIGH}_t \times \text{POSTFT}_t$	-0.359	-0.269	-0.463	-0.262	-0.294	-0.180
	(0.136)	(0.075)	(0.149)	(0.081)	(0.113)	(0.056)
Measure of Complexity: Total Demand	Y	N	Y	N	Y	N
Measure of Complexity: SD RT Price	N	Y	N	Y	N	Y
Dep. Var. in Logs	Y	Y	N	N	N	N
Dep. Var. in Asinh	N	N	Y	Y	N	N
Poisson Spec.	N	N	N	N	Y	Y
Peaker/Month/High Day FE	Y	Y	Y	Y	Y	Y
Peaker/Month-of-Sample FE	Y	Y	Y	Y	Y	Y
Weekday versus Weekend FE	Y	Y	Y	Y	Y	Y
R^2	0.531	0.579	0.532	0.576		
Mean of Dep. Var.	2.767	2.767	3.450	3.450	19.313	19.313
Number of Obs.	2,669	2,669	2,680	2,680	2,680	2,680

Notes: This table presents the estimated difference in start-ups by baseload versus peaker gas-fired units before versus after the introduction of financial trading ("FT") on high versus low complexity days. We categorize a unit as "baseload" if the unit's aggregate fuel costs per MWh are in the bottom half of the distribution across units of this magnitude; units in the top half of the distribution of aggregate fuel costs per MWh are categorized as "peakers". The unit of observation considered for these regressions is type-of-unit/day-of-sample. For Columns 1, 3, and 5, the indicator variable HIGH_t is equal to one for days-of-sample with daily total demand greater than the 75th percentile of the distribution of daily total demand across the sample period. For Columns 2, 4, and 6, HIGH_t is equal to one if the standard deviation across locations and hours in real-time prices on day t is higher than the 75th percentile of the distribution of daily standard deviations in real-time prices. The "Post FT" indicator is equal to one if and only if the day-of-sample is after FT is introduced on February 1st 2011. The row titled "Mean of Dep. Var." reports the mean of the relevant dependent variable: the log of total number of starts by gas-fired units of the type in the day for Columns 1 and 2, the inverse hyperbolic sine of starts for Columns 3 and 4, and number of starts in levels for Columns 5 and 6. We estimate the model using linear regression for Columns 1-4 but Poisson regression for Columns 5 and 6. All of the regressions listed in this table include the sets of fixed effects and control variables specified in Appendix Equation (F.4). Standard errors are clustered by week-of-sample and are reported in parentheses.

with the intuition that the locational bids and offers submitted by purely financial participants in the day-ahead market resulted in the use of lower cost baseload units rather than higher cost peaker units to satisfy demand during high complexity days.