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.

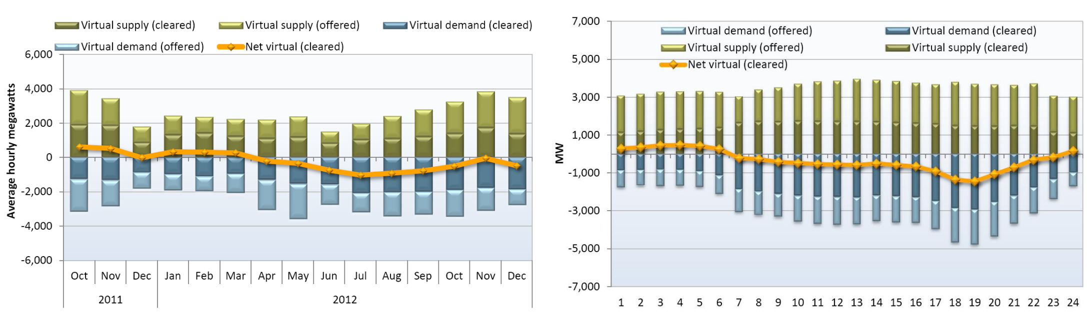
[47]This table is reproduced from CAISO's 2012 Annual Report (CAISO (2012a)).

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

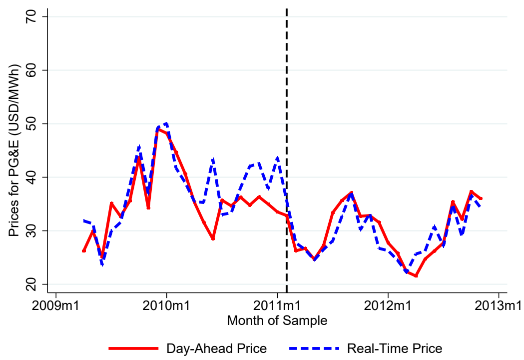
Notes: This is a map of the territories served by each of the three major investor-owned electric utilities in California. These three utilities are Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E). This map is reproduced from FERC (2015).

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

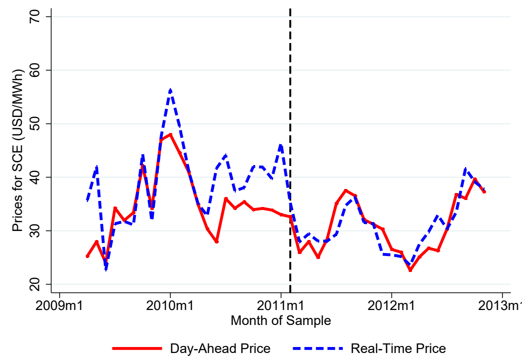
These plots also demonstrate that day-ahead prices are lower than real-time prices on average for the majority of hours of the day for all three utilities prior to FT. This is consistent with the results in Borenstein et al. (2008), which argues that large retailers in California withheld demand from 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.

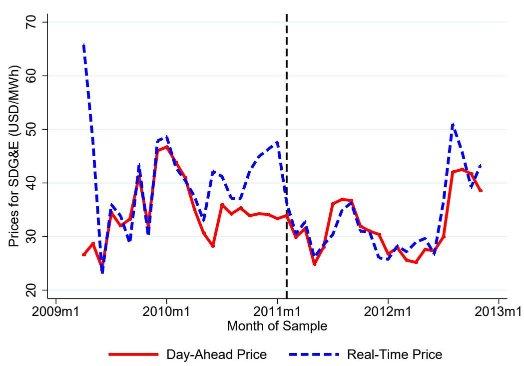
Figure A.3: Monthly Average Day-Ahead and Real-Time Prices By Service Territory



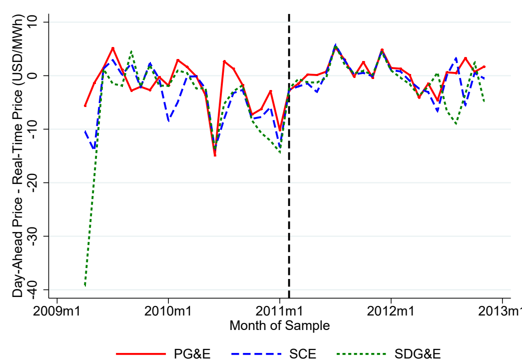
(a) Pacific Gas and Electric



(b) Southern California Edison



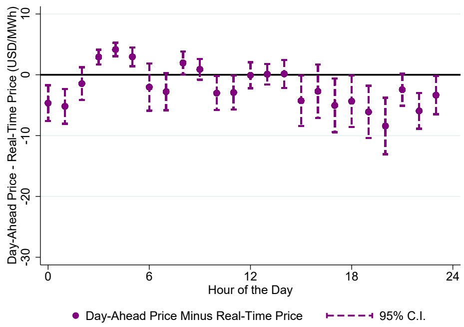
(c) San Diego Gas and Electric



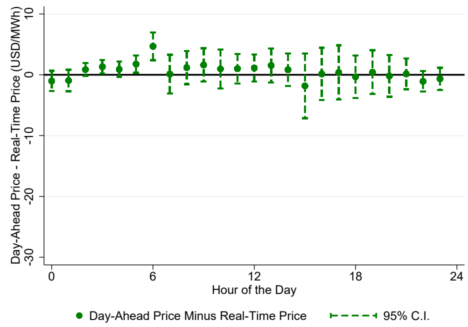
(d) Day-Ahead Price − Real-Time Price

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.

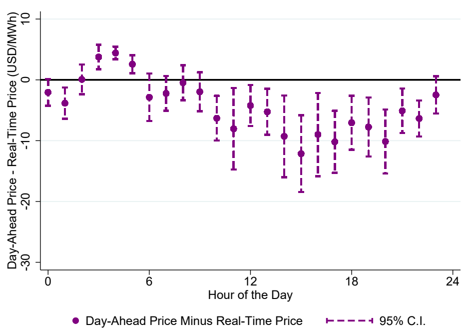
Figure A.4: Hourly Average Day-Ahead/Real-Time Price Spreads: Before and After Financial Trading



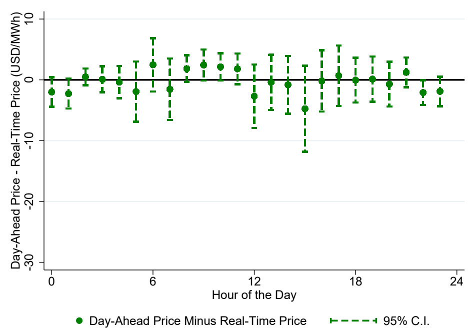
(a) PG&E, Before Financial Trading



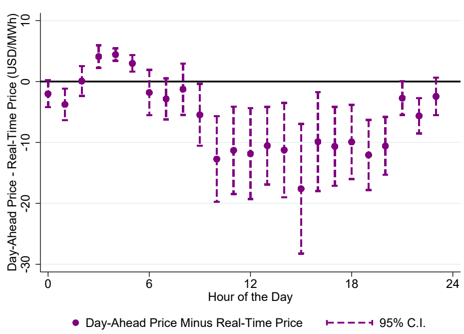
(b) PG&E, After Financial Trading



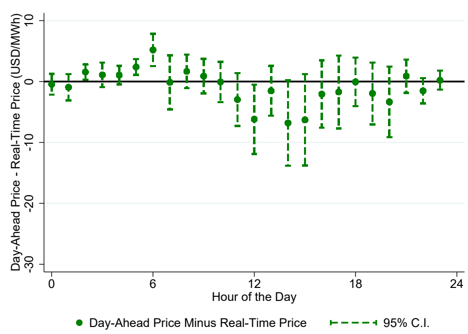
(c) SCE, Before Financial Trading



(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

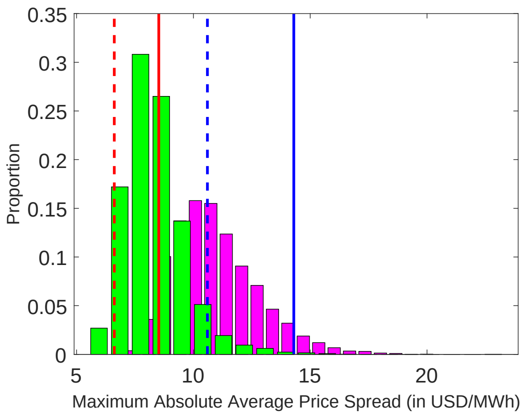
Appendix Table A.2 reports estimates of our two measures of implied trading costs before and after the implementation of FT for the day-ahead/real-time price spreads corresponding to the territories served by PG&E, SCE, and SDG&E. Recall that, as discussed in Section IV.B, c_{lower} is the smallest value 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.

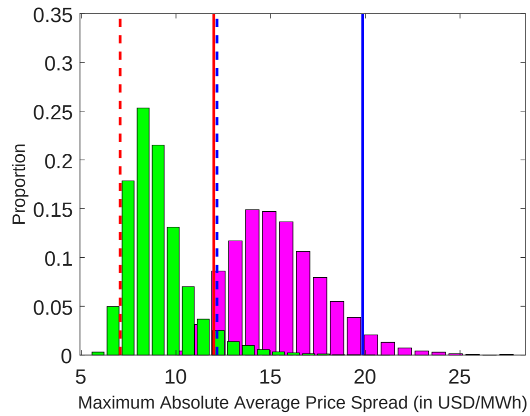
All three panels of Appendix Figure A.5 indicate that both c_{lower} and c_{upper} fell substantially after the introduction of financial trading. That being said, Appendix Figure A.6 presents results from a formal test of the null hypothesis that c_{lower} and c_{upper} remained the same after financial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

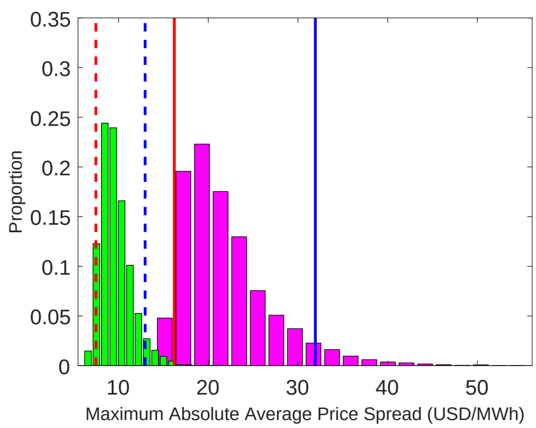
Figure A.5: Bootstrap Distribution of Implied Trading Costs For Each Service Territory: Pre-FT in Purple and Post-FT in Green



(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.

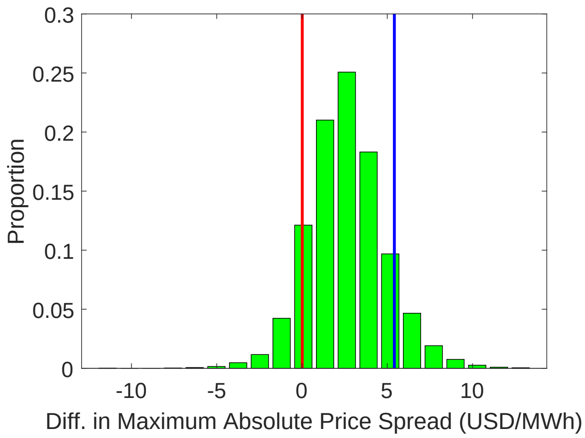
Table A.2: Implied Trading Costs by Territory (in USD/MWh)

	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

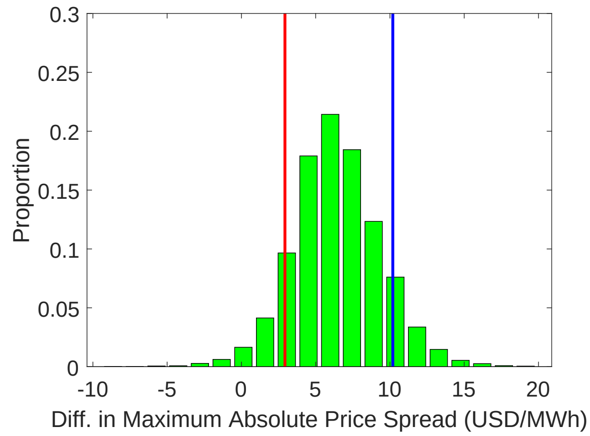
Notes: This table presents the implied trading costs estimated using the modeling framework discussed in Section IV. We estimate implied trading costs separately for each utility service territory for the sample periods before versus after the introduction of FT. The three service territories considered 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.

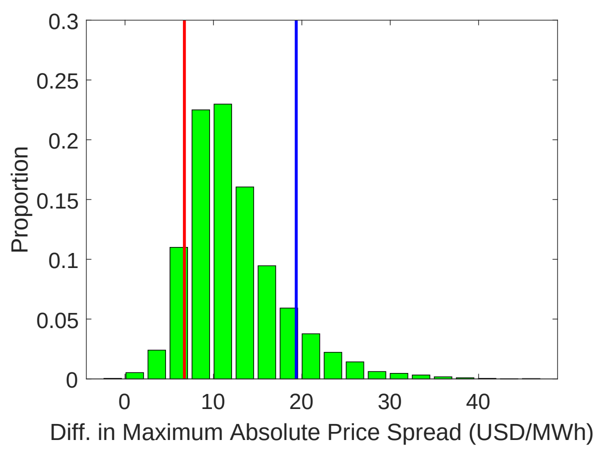
Figure A.6: Bootstrap Distribution of the Difference in Implied Trading Costs



(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

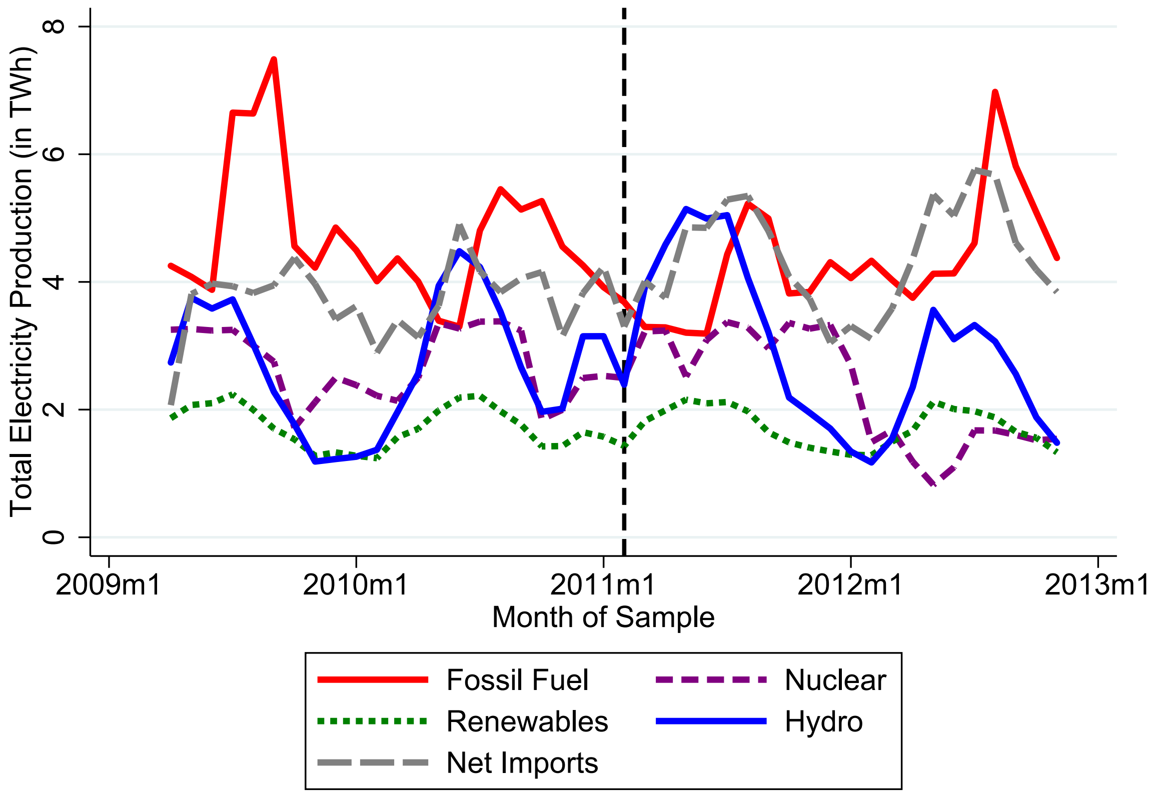
Appendix Figure A.7 plots monthly total electricity production by type: gas-fired, nuclear, renewables, and all hydro.[49] We sum only over sources under the operational control of California's Independent System Operator (CAISO). Appendix Figure A.7 also includes monthly total net electricity imports. 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



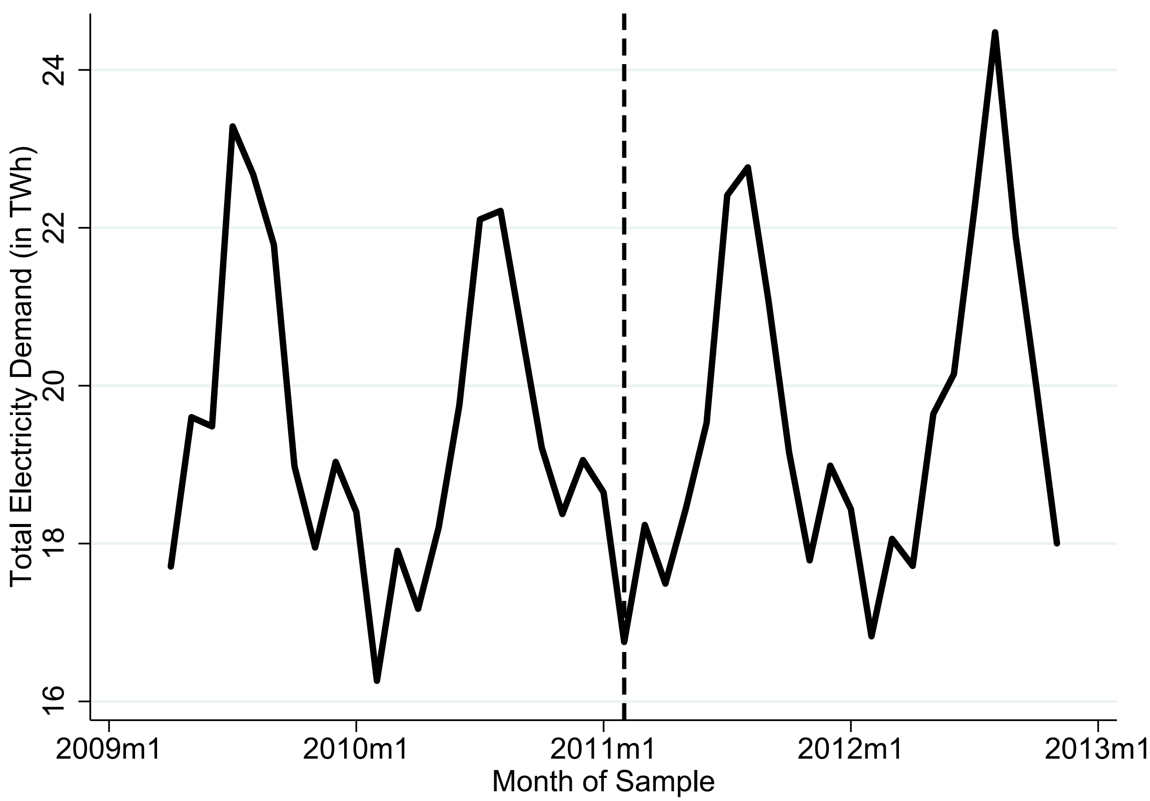
Notes: This figure plots monthly total electricity production by type: fossil-fuel-fired, nuclear, renewables (wind + solar + biomass + biogas + hydro sources less than 30MW), and all hydro. We sum only over sources under the operational control of California's Independent System Operator (CAISO). This 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

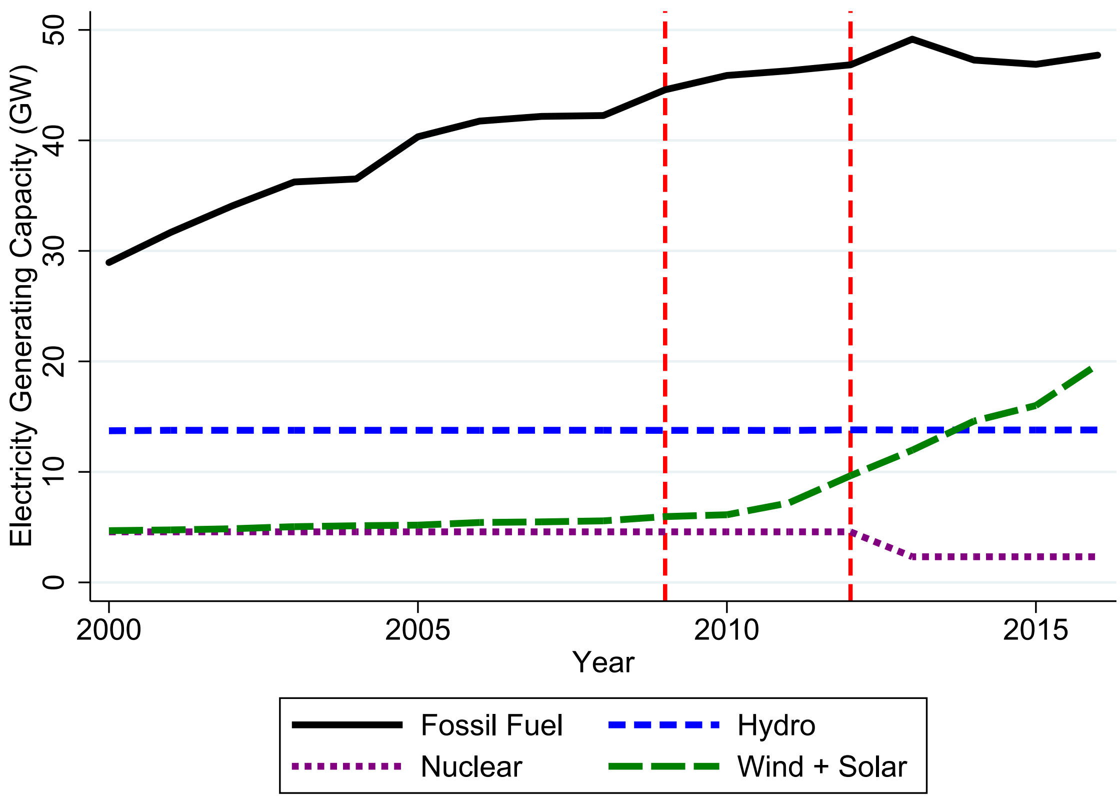
[51]We sum over units of each source type in California using the eGrid database for 2012 provided by the United States Environmental Protection Agency (USEPA, 1996-2012).

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.

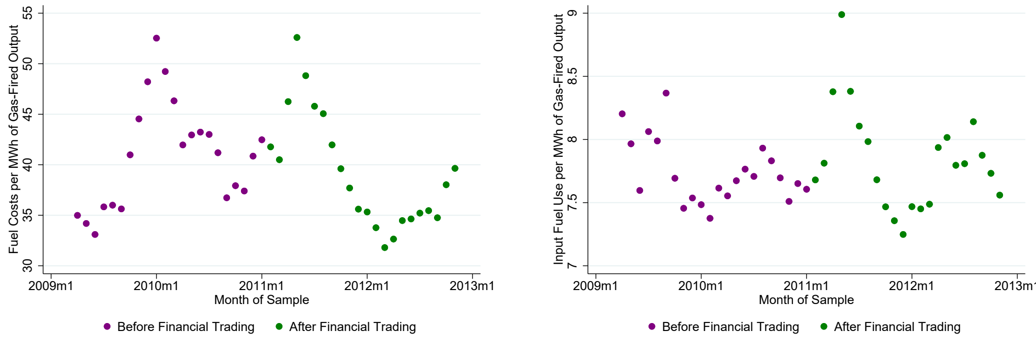
documents a steady increase in the installation of renewable capacity during the latter half of our sample period. In addition, we see a decrease in nuclear generating capacity after 2012 due to the retirement of the San Onofre nuclear power plant (Davis and Hausman, 2016). Based on these trends, all 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]

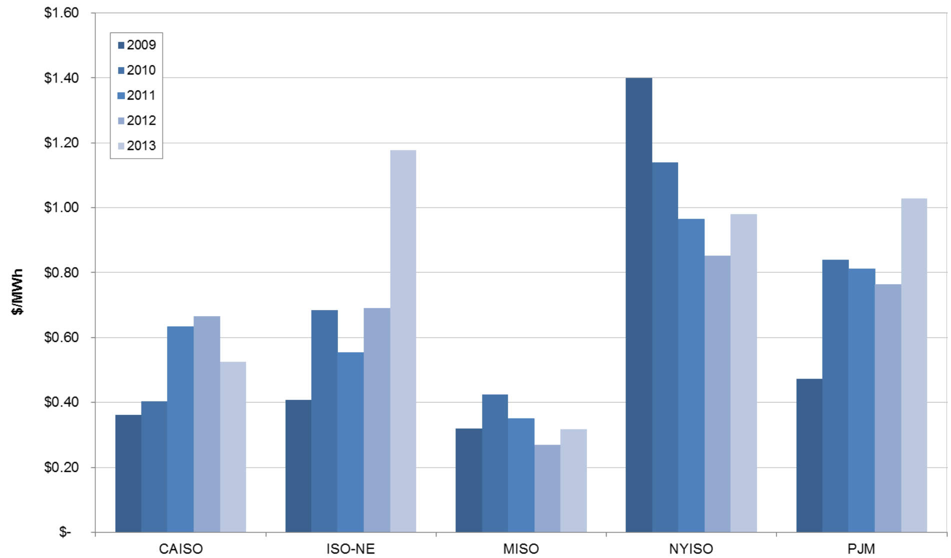
The California ISO clears day-ahead and real-time markets by solving a mixed-integer programming problem. The California ISO is sometimes forced to manually dispatch generation units after the close of the day-ahead market or in real-time to satisfy operational constraints that may not have have been 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



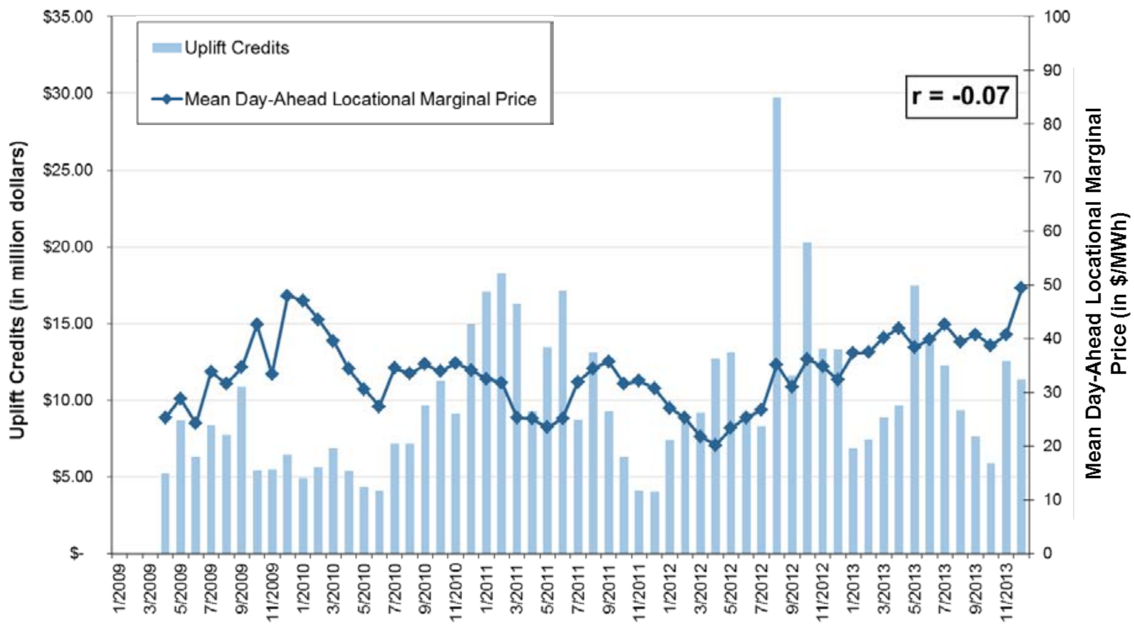
Notes: This figure is taken from FERC (2014). Annual average uplift charges (in dollars per MWh) are calculated for each Independent System Operator (ISO) by dividing total annual uplift charges (in dollars) by total annual electricity demand (in MWh). Total uplift charges and total electricity demand 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 $|\mu_{pre}| > |\mu_{post}|$:

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

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.

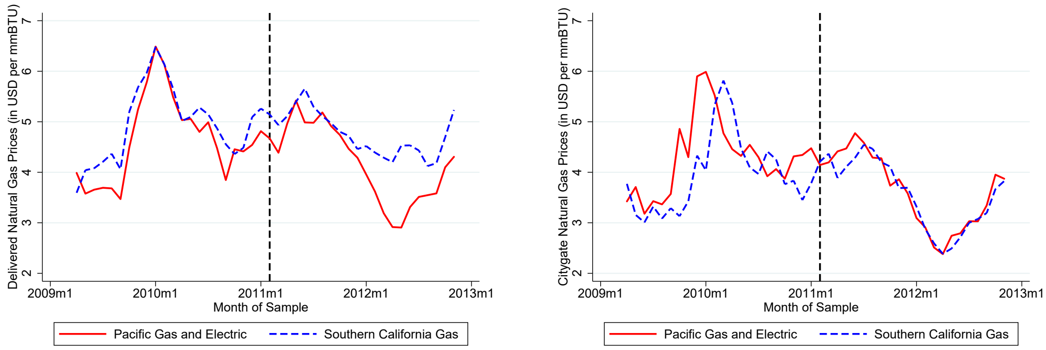
D.2 Data Construction: Natural Gas Prices

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.

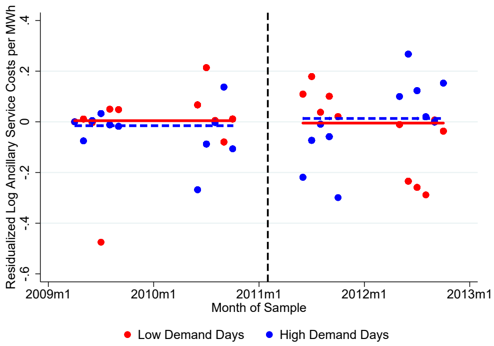
percentile of the distribution of this measure across our sample period. As discussed in Section VI.B, we estimate separate specifications based on three different measures of complexity: daily total demand, the daily standard deviation across locations and hours of real-time prices, and daily total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

Appendix Equation (E.1) controls for separate sets of calendar month fixed effects for high complexity days and low complexity days ($\alpha_{m,\mathrm{HIGH}}$), 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.

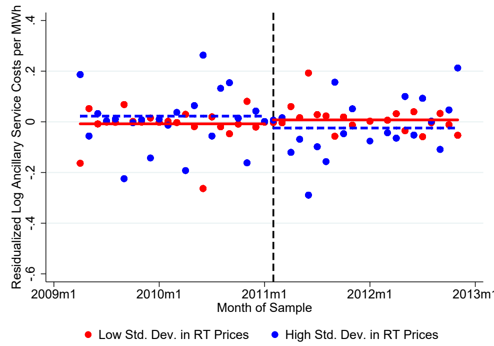
Appendix Figure E.1 plots the monthly average residuals from estimating Appendix Equation (E.1). In the top left and top right panels, we define "high complexity" using daily total demand and the daily standard deviation across locations and hours in real-time prices respectively. The bottom panel is 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

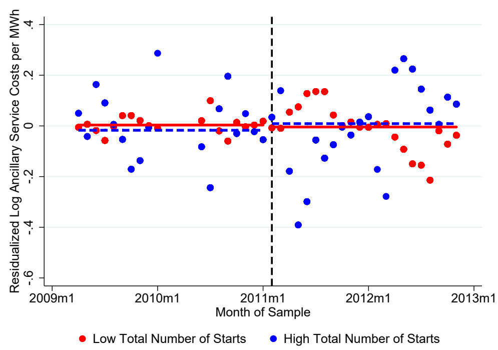
Figure E.1: Monthly Average Residualized Ancillary Service Costs per MWh



(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:

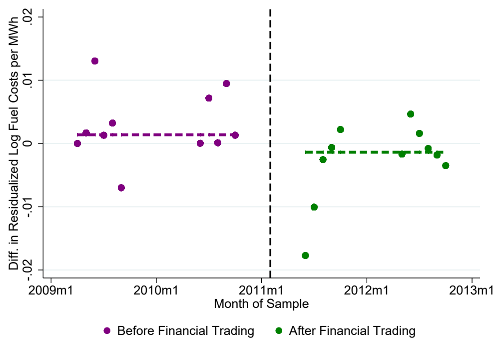
$$Y_t = \alpha_{m,\text{HIGH}} + \theta_w + \gamma_{y,m} + \sum_{s=1}^{S} \sum_{k=1}^{K} [(X_{k,t} - \overline{X}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[X_{k,t} \in \text{BIN}_{k,b}]] + u_t \quad \text{(E.2)}$$

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.

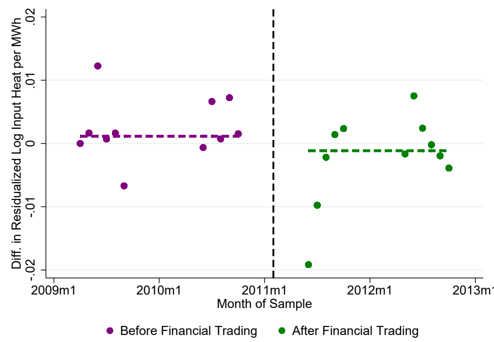
The two left panels of Appendix Figure E.2 focus on differences in the log of fuel costs per MWh of gas-fired output while the two right panels focus on differences in the log of input heat use per MWh of gas-fired output. This figure includes a vertical

[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.

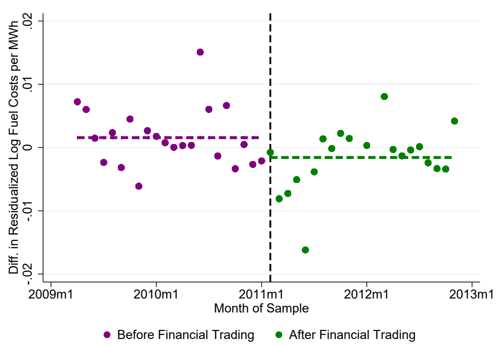
Figure E.2: Monthly Average Differences in Residualized Outcomes



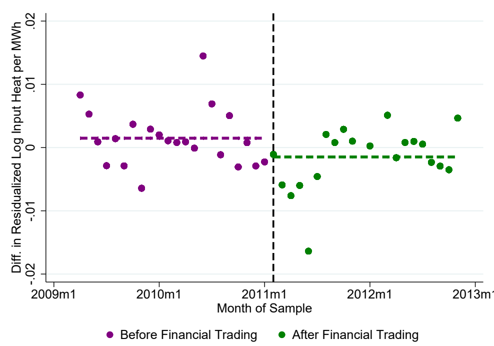
(a) Log Fuel Cost Per MWh
Measure of Complexity: Demand



(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.

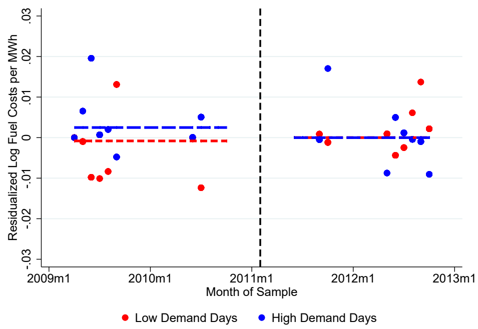
One might also be concerned that the base specification in Equation (5) "over-controls" for the economic factors in X_t. To assuage this concern, we consider specifications that control only linearly for the variables in X_t. Specifically, for Appendix Figure E.4, we residualize each outcome Y_t in day-of-sample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

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

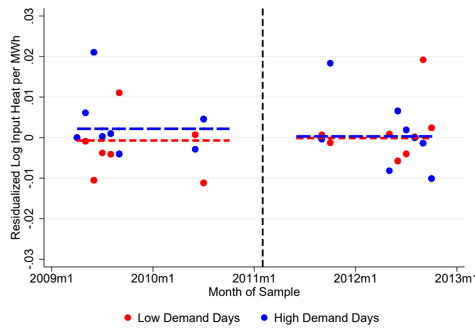
As before, the set of control variables included in X_t is the log of total electricity demand, the log of net electricity imports, the log of the monthly average natural gas price, as well as the logs of monthly total production from: (1) renewables, (2) nuclear sources, and (3) hydro sources.

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

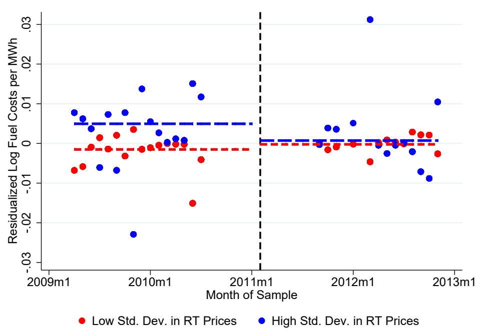
Figure E.3: Monthly Average Residualized Outcomes Before versus After Financial Trading Dropping the 6 Months Before and After FT



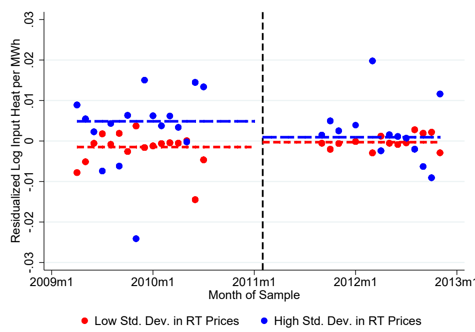
(a) Log Fuel Cost Per MWh
Measure of Complexity: Demand



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



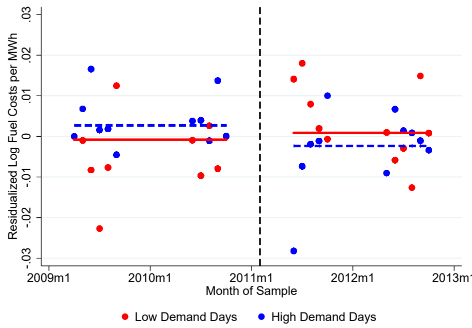
(c) Log Fuel Cost Per MWh
Measure of Complexity: SD[RT Price]



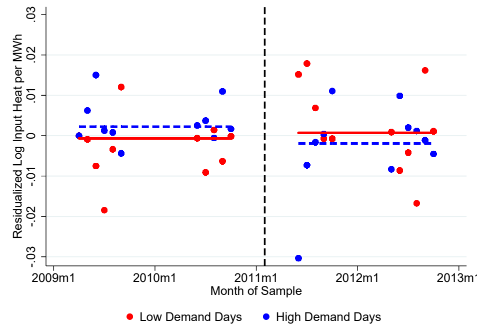
(d) Log Input Energy Per MWh
Measure of Complexity: SD[RT Price]

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.

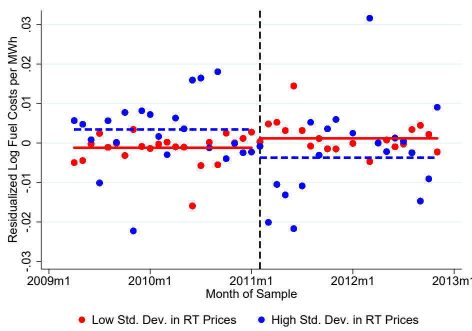
Figure E.4: Monthly Average Residualized Outcomes Before versus After Financial Trading: No Nonlinear Controls



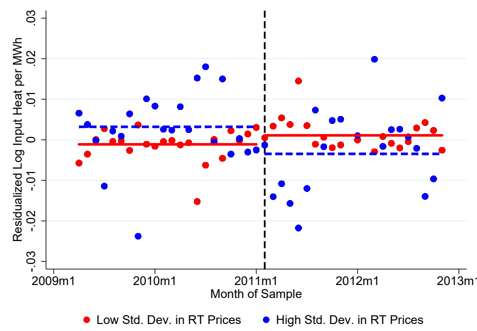
(a) Log Fuel Cost Per MWh
Measure of Complexity: Demand



(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)
$HIGH_t \times POSTFT_t$	-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)
$HIGH_t \times POSTFT_t$	-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").

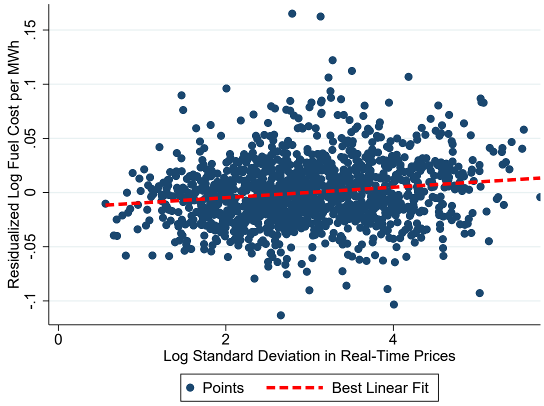
F.1 Measures of Complexity and Fuel Costs

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:

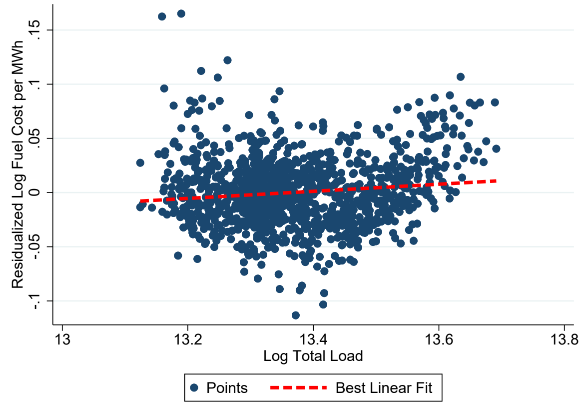
$$Y_t = \theta_w + \gamma_{y,m} + \sum_{s=1}^{S}\sum_{k=1}^{K}[(Z_{k,t} - \overline{Z}_k)^s \phi_{s,k} + \sum_{b=1}^{10} \theta_{k,b} 1[Z_{k,t} \in \mathrm{BIN}_{k,b}]] + u_t \tag{F.1}$$

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)

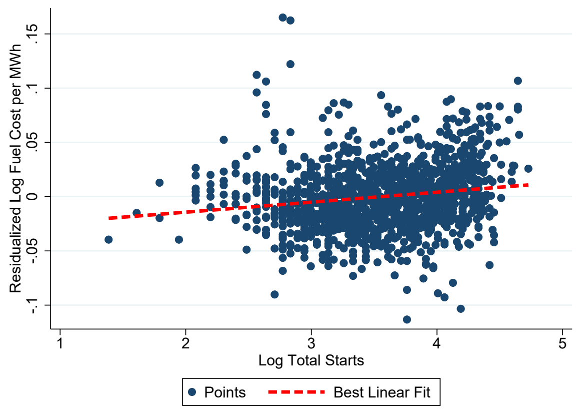
Figure F.1: Residualized Fuel Cost Per MWh and Measures of Complexity



(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

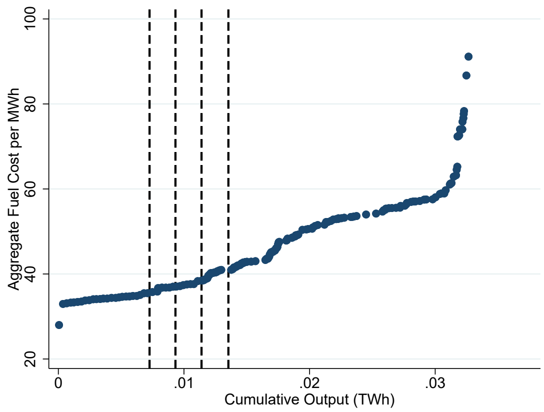
This subsection compares market outcomes before versus after the introduction of financial trading on days with more versus less starts by gas-fired units. We estimate the following specification in order to quantify how our two outcome variables change after financial trading on days with a relatively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

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

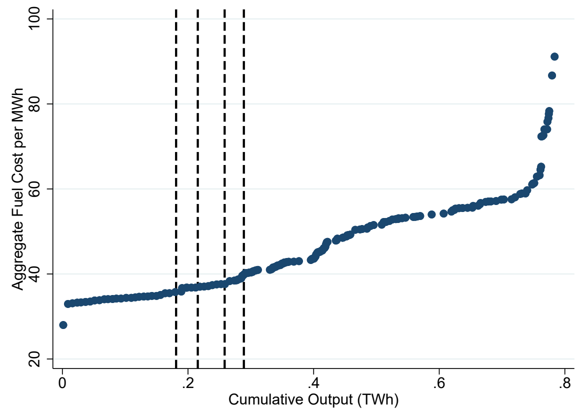
where we define HIGH_t to be an indicator that is equal to one if and only if daily total number of starts on day-of-sample t is above the k^{th} percentile of the distribution of daily total starts across our 4/1/2009-11/30/2012 sample period; we consider specifications based on the 50th, 60th, 70th, 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.

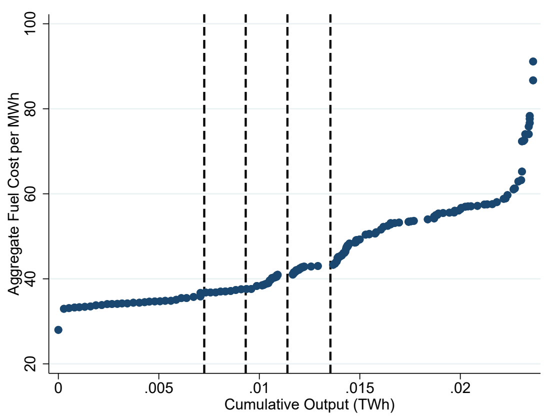
Figure F.2: Hourly and Daily Marginal Cost Curves



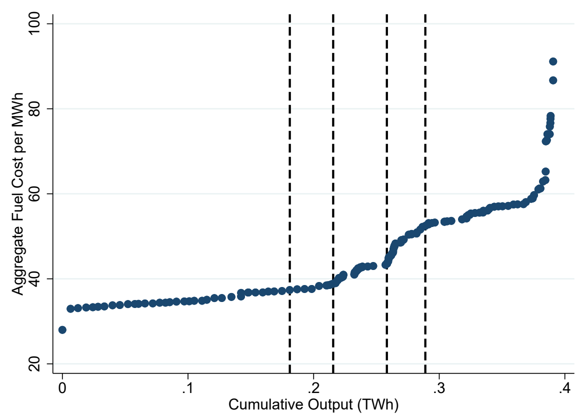
(a) Hourly, Capacity-Based



(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

Log of Fuel Cost Per MWh	(1)	(2)	(3)	(4)	(5)
$HIGH_t \times POSTFT_t$	-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)
$HIGH_t \times POSTFT_t$	-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:

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

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 $\text{HIGH}_t \times \text{POSTFT}_t$, 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:

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

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.