

# Imperfect Markets versus Imperfect Regulation in U.S. Electricity Generation

Steve Cicala\*

University of Chicago

[Click here for the latest version.](#)

December 28, 2017

## Abstract

This paper evaluates changes in electricity generation costs caused by the introduction of market mechanisms to allocate production in the United States. I use the staggered transition to markets from 1999-2012 to estimate the causal impact of liberalization using a machine learning-augmented differences-in-differences design on a comprehensive hourly panel of electricity demand and unit-level costs, capacities, and generation. I find that markets reduce production costs by reallocating output: Gains from trade across service areas increase by 20% based on a 10% increase in traded electricity, and costs from using uneconomical units fall 20% from a 10% reduction in their operation.

*JEL Classification:* L50, L94, L98, P18, P51, Q41, Q48

---

\*I am grateful to Gary Becker, Jim Bushnell, Thom Covert, Tatyana Deryugina, Edward Glaeser, Michael Greenstone, David Hémous, Lawrence Katz, Ryan Kellogg, Erin Mansur, Kevin Murphy, Erica Meyers, Morten Olsen, Jim Sallee, Andrei Shleifer, Chad Syverson, Roberton Williams III, Matthew White, and seminar participants at Harvard, MIT, Cornell, Chicago, Yale, the UC Energy Institute, the EEE Session of the 2015 NBER Summer Institute, UIUC, Wharton, Brown, and the 2017 European Econometric Society Meetings for helpful comments and suggestions. Sébastien Phan, Julien Sauvan, Songyuan Ding, Enrique Chazaro-Acosta, Xianying Fan, Mary Vansuch, and Dan Pechi provided excellent research assistance. This paper has been reviewed by the Energy Information Administration to ensure no confidential data have been disclosed. All errors remain my own. e-mail: scicala@uchicago.edu

## 1 Introduction

When regulation brings its own host of distortions and inefficiencies, the mere existence of a market failure is insufficient to ensure government intervention will improve welfare. Instead, by comparing the distortions under potential regulatory regimes, one can identify superior policies as those with relatively fewer imperfections (Kahn (1979); Joskow (2010)). This paper undertakes such an evaluation in the context of U.S. wholesale electricity markets, which have replaced command-and-control-type operations in some areas.

To do so I construct a virtually complete hourly characterization of supply and demand of the U.S. electrical grid from 1999 - 2012. I use data on fuel costs, capacities, heat efficiency, and operations of nearly all generating units at the hourly level to construct power supply curves (known as the “merit order”) for each of 98 “Power Control Areas” (PCAs), as well as observe the units that were chosen to operate to meet demand at any moment in time. These curves allow me to decompose observed costs in to two key welfare measures of allocative efficiency for each PCA-date-hour: “out of merit” losses from dispatching higher marginal cost units relative to installed capacity, and the savings from trading electricity across areas. Market power losses manifest themselves as out of merit production (Borenstein et al. (2002); Mansur (2001)), as do normal grid operations, such as maintenance, refueling, start-up costs, and transmission congestion (Davis and Wolfram (2012); Mansur (2008); Reguant (2014)). In either case, the increased operational costs are observationally equivalent as the distance between the realized cost of operations and cost from utilizing only the lowest-cost installed capacity.

I develop a framework and compile the necessary data to examine both of these welfare measures simultaneously over the history of market transitions since 1999. I use the staggered creation and expansions of wholesale electricity markets over this period to estimate the causal impact of using markets to allocate production on these welfare measures. I employ a differences-in-differences (DD) framework to estimate changes in gains from trade and out of merit losses following the transition to market dispatch against PCAs that have not undergone such operational changes.<sup>1</sup> This approach finds gains from trade increase by upwards of 30% after adopting market dispatch due to a 10% increase in electricity traded. There is also a 10% decrease in

---

<sup>1</sup>The timing of power plant divestitures and expansion of non-utility generators does not line up with the transition to market dispatch: Most divestiture was completed by 2001 when state-level restructuring stalled, while the median transition to market dispatch occurred in 2005. In fact, over half of PCAs adopt market dispatch while retaining their traditional cost of service structure for determining the retail price of electricity. See Appendix B for an event study-type figure.

out of merit operations, reducing these costs by nearly 20%.

The simple DD approach is susceptible to the confounding effects of fuel price fluctuations (over time and across areas) when estimating counterfactual outcomes: One might estimate changes in the gains from trade without any actual changes in production patterns because the value of offset production scales with fuel prices. This issue motivates a policy function approach in which I use historical patterns of unit-level production given demand, unit costs and capacities to estimate each system operators' rules for dispatching units. I estimate these rules for each PCA-year and use them to predict unit-level production in the subsequent year. I show how the treatment effect can be estimated by comparing *changes* in the quality of fit of this rule between areas that switch to market dispatch against areas with no change in regulation.<sup>2</sup>

Estimating dispatch probabilities with out-of-sample validity is a pure prediction problem for which recent developments in the machine learning literature have proven to be particularly effective (Kleinberg et al. (2015)). I use the random forest algorithm of Breiman (2001) to non-parametrically estimate policy functions, then embed the results in a DD framework to estimate causal treatment effects. This part of the paper complements the recent work of Burlig et al. (2017), who also use machine learning methods (Least Absolute Shrinkage and Selection Operator) to predict counterfactual outcomes and contributes to a growing literature that synthesizes program evaluation and machine learning methods (Chernozhukov et al. 2017; Athey, Imbens and Wager, forthcoming XXX).

The policy function approach yields estimates smaller in magnitude than the simple DD estimates for gains from trade, suggesting fuel price confounding. I find that production costs are reduced by about three billion dollars per year due to market-based improvements in allocating output to lower cost units, with these savings split between reduced output from uneconomical units and gains from trade by 2:1.

It should be noted at the outset that my estimates measure changes in how output is allocated *given the installed capacity, costs, and patterns of demand*. It would not be unreasonable to suspect that market dispatch has affected investment incentives, which are likely to be an important source of welfare changes. In addition, my estimates measure the average effect of market dispatch, which itself has been heterogeneous both with respect to pre-existing institutions (i.e. power pools, bilateral

---

<sup>2</sup>This approach avoids issues raised by Lucas (1976) by holding regulatory status fixed in the estimation of counterfactuals: I compare observed outcomes to those predicted if generation units were dispatched according to last year's rules, i.e. the production decisions prior to market dispatch. The Lucas Critique applies to attempting to predict outcomes under a different regulatory structure.

markets, or smoke-filled rooms), and with respect to the rules of the markets implemented (uniform or locational marginal prices, virtual bidding, market monitors, etc.). However, given the even greater differences between market and traditional dispatch methods these estimates should be informative regarding the performance of the relatively new mechanisms that currently determine how over 60% of generating capacity in the United States is utilized.

The paper is organized as follows: in the next section I describe the structure of electricity generation and transmission in the United States, and the institutional details that will facilitate estimation. The third section describes how out of merit costs and gains from trade are measured in electricity generation, and the fourth section describes the data. The fifth section presents an estimation strategy motivated by this setting. The sixth section presents causal estimates of the impact of markets on gains from trade and out of merit costs. The final section concludes. There are two appendices, one provides greater detail on data assembly and the other contains supplementary results.

## **2 Background on Power Control Areas and Dispatch in the United States**

The U.S. electricity grid developed over the 20<sup>th</sup> century based on a mix of Investor Owned Utilities (IOUs), government-owned utilities (municipal, state, and federal), and non-profit cooperatives. All of these organizations tended to be vertically integrated, so they owned the power plants, the transmission system, and the delivery network within their respective, exclusively operated territories. As either government-run or regulated monopolies, I use the term “command-and-control” to refer to the methods of dispatch under this industrial structure. The entity that has historically determined which power plants operate to meet demand is called a “balancing authority.” A single balancing authority controls the transmission system and dispatches power plants within a “Power Control Area,” or PCA. When vertically-integrated, the balancing authority and utility have often been one-in-the-same, as with the service territory and the PCA.<sup>3</sup> These areas operate with relative autonomy over their assets, and transmission lines that connect areas enable flows between them.

The national grid consists of three large Interconnections: East, West, and Texas (with relatively little capacity to transmit power between them). Figure 1 shows the approximate configurations of the U.S. Electricity Grid in 1999 and 2012.<sup>4</sup> The

---

<sup>3</sup>Exceptions include the New York and New England Power Pools, which formed in response to The Great Northeast Blackout of 1965, as well as smaller utilities that do not control dispatch directly. Regional reserve margin coordination was also formalized during this time with the establishment of the National Electric Reliability Council.

<sup>4</sup>The exact geographic boundaries of PCAs often defy straightforward demarcation. This map is

boundaries between Interconnections are denoted in Panel A by the thick black lines separating Texas and the West (unchanged over the period). The red lines denote regions of the North American Electric Reliability Corporation (NERC) that coordinate their operations in order to preserve the stability of the transmission system (when large plants go down for maintenance, for example). The tangle of power control areas reflects the legacy of local monopolies that have been the principal architects of the U.S. electricity grid.

Although the Public Utility Regulatory Policies Act of 1978 (PURPA) opened the door for independent power generation (by requiring IOUs to buy their output at “avoided cost”), the growth of such producers was impeded by discriminatory transmission practices (Joskow (2000)). Because the IOUs owned the transmission system, they could effectively shut independent producers out of wider markets by denying transmission access.<sup>5</sup> This began to change with the Energy Policy Act of 1992, which required the functional separation of transmission system owners and power marketers—they were no longer allowed to use their wires to prevent or extract the surplus from trades across their territory. These changes were codified on April 24, 1996 with FERC orders 888 and 889, which required open-access, non-discriminatory tariffs for wholesale electricity transmission.

Open-access created greater potential for wholesale electricity markets, which were initially conducted through bilateral contracts for power. In this decentralized setting, contracts would typically specify the amount of electricity to be generated by one utility under a set of conditions, transmitted across a particular area, and withdrawn from the system by the purchasing utility. Mansur and White (2012) give examples showing why the nature of congestion in electricity transmission networks renders decentralized markets particularly poorly suited for identifying all of the potential gains from trade. In particular, transmission lines are constrained by net flows of power. When this is the case, there are production externalities that may allow otherwise infeasible bilateral trades to occur by coordinating offsetting transactions to keep net flows below transmission capacity. Identifying such potential trades in this type of decentralized market is a challenge akin to coordinating simultaneous multilateral exchanges (Roth et al. (2004)).

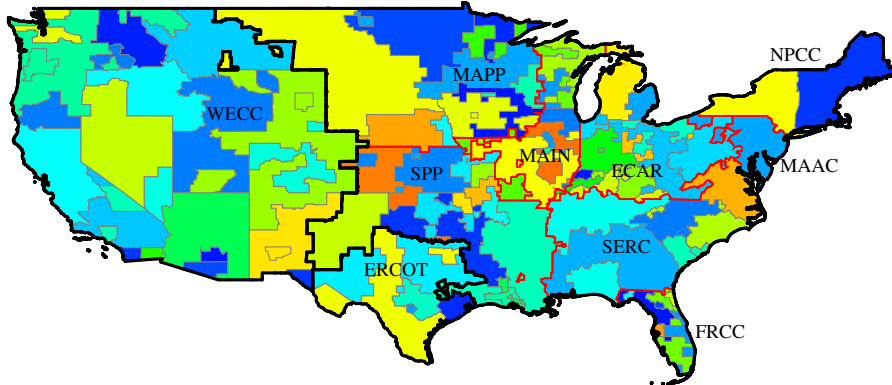
---

based on U.S. counties, with the predominant PCA receiving assignment of the entire county—and is therefore approximate for visualization purposes. In addition, a number of small or hydro-only PCAs are merged with the larger neighboring areas that provide the majority of their (fossil-based) energy.

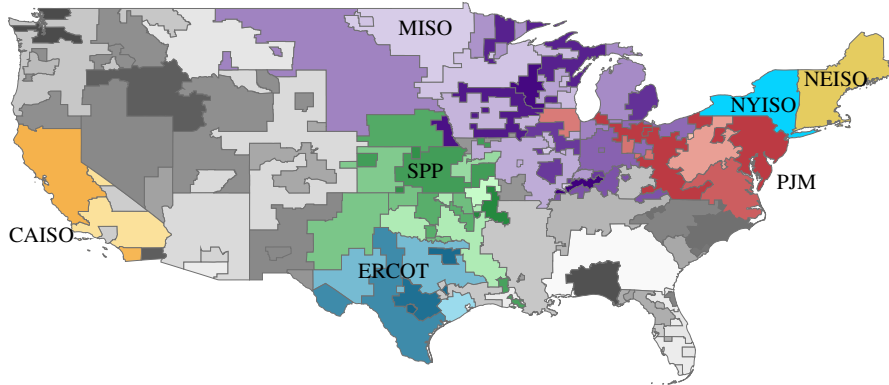
<sup>5</sup>Examples of IOUs exercising market dominance can be found in Appendix C of FERC Order 888.

# Figure 1: U.S. Electrical Grid as Power Control Areas

(a) Approximate PCA Configuration in 1999



(b) PCAs by Market Dispatch in 2012



Note: Thick black lines in Panel A denote Interconnection boundaries, red lines denote NERC Reliability Regions. Boundaries are approximate.

**Figure 2:** Share of Generating Capacity Dispatched by Markets



Note: Vertical red lines indicate dates of transition to market-based dispatch.

Operationally, balancing authorities have relied on engineering estimates of costs to devise dispatch algorithms to determine which plants within the PCA operate, and separately schedule any other operations requested by utilities (for bilateral trades). Centralized wholesale electricity markets (“market dispatch”) integrate dispatch operations in to an auction for electricity. In day-ahead auctions, for example, generators submit bids to produce electricity, and only those below the price needed to meet projected demand are called on to operate. These auctions incorporate feasibility constraints, so calling on higher-priced units to operate due to transmission congestion allows for the direct revelation of the cost of shortcomings in the transmission system.<sup>6</sup> Day-ahead markets establish financial obligations to produce, which are subsequently either met with production in the real time market or unwound by buying back one’s allocated output at the real time price (Wolak (2000); Hortacsu and Puller (2008); Ito and Reguant (2016); Cramton (2003); Jha and Wolak (2013); Borenstein et al. (2008), among others).

As of 2012, 60 of the 98 PCAs operating in 1999 had adopted market dispatch, either during the initial creation of a new market or as part of the expansion of an existing market. Adopting market dispatch is a discrete change in the decision algorithm that allocates output to generating units: the local PCA cedes control of

<sup>6</sup>In particular, auctions using the “Standard Market Design” yield “Locational Marginal Prices” (LMPs) which denote the market-clearing price at each of the points of withdrawal from the system. When LMPs are identical everywhere, the system is said to be uncongested.

their transmission system to an Independent System Operator, who conducts the auctions.

All told, there have been 15 distinct events in which PCAs have transitioned to market dispatch overnight. Figure 2 denotes each of these events with a vertical red line, and shows that over the period of study markets have expanded from covering about 10% of capacity to roughly 60%. The remaining areas have retained their traditional dispatch methods, though a number have continued to explore the possibility of joining existing markets.<sup>7</sup> This variation in market adoption forms the basis of the empirical strategy for causal estimates by allowing the comparison of changes in allocative efficiency following the transition to market dispatch relative to areas that have not undergone such changes over the same period.

The transition from command-and-control to market dispatch is related to, but distinct from the movement toward restructured electricity markets in the United States (Joskow and Schmalensee (1988)). In particular, the changes to dispatch and transmission described thus far were undertaken by the Federal government.<sup>8</sup> The end of cost-of-service regulation of vertically-integrated IOUs was initiated by states. These state-led initiatives halted after the California electricity crisis, while the adoption of market dispatch has continued through the 2000's. It is important to distinguish between these developments, for although all states that adopted restructuring legislation eventually adopted market dispatch, many areas began participating in these markets while preserving their traditional regulatory framework.<sup>9</sup> I therefore focus my attention on the cost of generating electricity, rather than the retail price of power delivered to consumers, whose relationship with their local utility may or may not have changed over this period.

Vulnerability to the exercise of market power has been a primary focus of the research on wholesale electricity markets to date. From the UK (Green and Newberry (1992); Wolfram (1999); Wolak and Patrick (1997)), Spain (Reguant (2014); Ito and Reguant (2016)), New Zealand and Australia (Wolak (2012)) abroad, to California (Borenstein et al. (2002); Borenstein (2002); Joskow and Kahn (2002); Bushnell et al.

---

<sup>7</sup>For example, the East Kentucky Power Cooperative joined PJM on 6/1/2013, there was a major southern expansion of MISO on 12/18/2013, and PacifiCorp has formally begun to explore the possibility of joining CAISO.

<sup>8</sup>The ERCOT system in Texas is the exception because this Interconnection does not cross state lines, and is therefore not subject to FERC jurisdiction on many matters. However, Texas does participate in the North American Electric Reliability Corporation (NERC), which has been designated by FERC as the electricity reliability organization for the United States.

<sup>9</sup>Examples include Indiana, West Virginia, and parts of Kentucky in the Pennsylvania-Jersey-Maryland (PJM) Interconnection, most of the Midwest ISO (MISO), and all of the Southwest Power Pool (SPP).



(2008); Puller (2007)), PJM (Mansur (2001, 2008)), plus New England (Bushnell et al. (2008)) and Texas (Hortacsu and Puller (2008)) in the United States, one could fairly characterize these vulnerabilities as robust. Against these losses, there is sparse evidence of allocative efficiency gains from market dispatch, with the notable exception of Mansur and White (2012) who study one of the 15 market expansion events described above. Instead, liberalization studies have focused on state-led deregulatory events to estimate within-plant changes: reduced maintenance time (Davis and Wolfram (2012), Cropper et al. (2011)), labor and fuel costs (Fabrizio et al. (2007); Cicala (2015)), and capital intensity of pollution abatement equipment (Fowlie (2010); Cicala (2015)). On the other hand, the actual rate at which heat is converted to electricity (heat rate) has proven largely unaffected by the nature of regulatory oversight (Fabrizio et al. (2007); Wolfram (2005); Cropper et al. (2011)).

While market imperfections are certainly cause for concern, evidence of their existence is not proof of their inferiority (Joskow (2010)). The relevant question for policymakers considering what to do about the current regulatory situation is: do markets (including all of their flaws) outperform the alternative methods for deciding which plants should operate in order to satisfy demand for electricity?

### 3 Measuring Welfare in Electricity Generation

The approach I use to measure welfare decomposes observed production costs in to two components of allocative efficiency, within- and across-PCAs. Each PCA has a narrowly defined “merit order” in which the fixed, installed generating capacity is lined up in order of increasing marginal cost (effectively a supply curve for the area). Each generating unit has a nameplate rating that constrains the maximum amount of electricity it is capable of generating at any moment. Its cost per MWh is based on its heat rate, cost of fuel, and emissions fees, making the supply curve a step function.<sup>10</sup> “Economic dispatch” solves this constrained cost minimization problem to meet a given level of demand without damaging plants by exceeding their nameplate capacity.

To fix ideas, let  $C_{pt}(Q_{pt})$  denote the observed cost of producing total quantity of electricity  $Q_{pt}$  in PCA  $p$  at hour  $t$ , which has  $N_{pt}$  MW of capacity installed. Further, define  $C_{pt}^*(Q_{pt})$  as the cost of generation from the  $Q_{pt}$  lowest-cost MW of PCA  $p$  in

---

<sup>10</sup>Labor costs are unobserved, but relatively small compared to fuel costs. Commercial vendors of unit production cost data (such as SNL or Platts) often include a 10-20% markup over fuel costs to account for labor, operations, and maintenance costs.

merit order, indexed by  $i$ :

$$C_{pt}^*(Q_{pt}) = \sum_{i=0}^{Q_{pt}} c_{pt}(i) \quad (1)$$

$$\text{where } Q_{pt} = \sum_{i=0}^{N_{pt}} q_{pt}(i); \quad q_{pt}(i) \in [0, 1] \quad \forall \quad i$$

where  $c_{pt}(i)$  is the cost of dispatching the  $i^{\text{th}}$  lowest cost MW in PCA  $p$  at time  $t$ .<sup>11</sup>

If, instead of producing  $Q_{pt}$ , each area simply produced enough to meet their own demand,  $L_{pt}$ , the merit order cost of autarky production would be  $C_{pt}^*(L_{pt})$ . Total observed costs can therefore be decomposed as

$$\sum_p C_{pt}(Q_{pt}) = \underbrace{\sum_p [C_{pt}(Q_{pt}) - C_{pt}^*(Q_{pt})]}_{\text{Within PCAs}} + \underbrace{\sum_p \{C_{pt}^*(L_{pt}) - [C_{pt}^*(L_{pt}) - C_{pt}^*(Q_{pt})]\}}_{\text{Across PCAs}} \quad (2)$$

The first term holds each PCA's total production fixed, and measures the cost of deviating from the merit order. These are referred to as "out of merit costs." The second term includes the difference in costs from the observed PCA production levels rather than producing in autarky. As shown below, the reduction in costs from reallocating production across PCAs is derived from this second term.

### Out of Merit Costs

A unit operates "out of the merit order" when it is called on to operate to help meet  $Q_{pt}$  MW of demand although it is not one of the  $Q_{pt}$  cheapest MW of PCA  $p$ 's installed capacity based on its marginal cost. There are a number of reasons to fire up units that are out of merit: Plants must occasionally go off-line for maintenance, or are forced to shutdown unannounced, causing more expensive units to fill the gap. Transmission constraints may make it infeasible for the least-cost units to meet local demand. Large units require time and fuel to substantially change their output (ramping and start-up costs) which may exceed the cost of firing up a more nimble

---

<sup>11</sup>The unit dispatch problem partitions the  $N_{pt}$  MW of capacity into distinct units (with common costs), and chooses how much to generate from each unit subject to nameplate rating constraints. While this is an identical problem, indexing MW according to  $i$  creates a stable metric of the merit order, while indexing units themselves may shuffle as fuel prices vary.

out of merit unit (Reguant (2014); Cullen (2011); Mansur (2008)). Large units may also continue operating when out of merit to prevent having to pay larger start-up costs from a cold start (idling). These are all real physical constraints that make out of merit operation the true cost-minimizing allocation of output. The cost of these constraints can be measured by the incrementally higher cost unit that must be used:  $C_{pt}(Q_{pt}) - C_{pt}^*(Q_{pt})$ .

This can be seen in Panel A of Figure 3, which plots a hypothetical (smooth) supply curve against the perfectly inelastic demand that must be met in a particular moment to avoid a blackout. The welfare costs of dispatching units out of merit is simply the additional cost of output from these units relative to dispatching the lowest cost units installed in the area. It is important to emphasize that these are the *gross* costs, which are often incurred to avoid the even larger costs of following the strict merit order.

This out of merit cost is also the loss borne when market power is exerted. A firm may increase the market clearing price by taking an economical unit “down for maintenance,” forcing an otherwise out of merit unit to operate (presumably to collect rents on co-owned inframarginal units). Because demand is completely inelastic (in real-time operations), the welfare loss is the incremental operating costs caused by taking economical units offline (Borenstein et al. (2002)).

It should be clear that legitimate maintenance, congestion, etc. is observationally equivalent from a welfare perspective to the exertion of market power—they differ by intent only. Mansur (2008) and Reguant (2014) note that failing to account for start-up and ramping costs will lead one to over-attribute the gap between the merit order and observed dispatch to market power when only accounting for normal maintenance and outages. The same is true when failing to account for transmission constraints (Ryan (2013); Borenstein et al. (2000)). Instead of measuring the *levels* of out of merit costs due to market power, as in Borenstein et al. (2002), my focus will be in how a change in regulation brings a PCA closer or farther away from the merit order.<sup>12</sup>

## Gains From Trade

When importing electricity from another area, one saves having to fire up a more expensive unit at the cost of the imports. When exporting, one gains any additional

---

<sup>12</sup>The critique of Mansur (2008) remains if unit operators fail to account for the wear of ramping once facing prices in a wholesale market. In this case short-run operations will move closer to the merit order, but damage to the units will go unaccounted for. In the longer run, operators will have to adjust their bids to cover these additional costs.

revenue beyond that required to generate the power. Panel B of Figure 3 considers the gains from trade between two areas, effectively a fixed-factor Heckscher-Ohlin model. The red line continues to represent demand in the “Local” PCA of Panel A. Superimposed on this is the mirror image supply and demand figure from a “Foreign” PCA. The width of the x-axis is the sum of the demand of the two areas. If the two areas were to operate in autarky, the cost of meeting this demand would be the area under the upper envelope of the supply curves, meeting at the solid demand line. While their joint production costs in the absence of constraints would be minimized at the intersection of their marginal cost curves, sending power between the areas is costly. These constraints prevent them equating their marginal costs, and they instead trade up the dotted line. Extending this logic to a multilateral setting yields a decomposition of reallocation in the across-PCA term of equation (2). Let  $c_{pt}(i = Q_{pt})$  to be the cost of producing the  $Q_{pt}^{th}$  MW in area  $p$ 's merit order and  $c_t(i = Q_t)$  as the marginal cost in the national merit order for total production in hour  $t$ . Noting that total supply equals total demand at any moment in time ( $\sum_p Q_{pt} - L_{pt} = 0$ ), the reduction in costs from autarky becomes:

$$\begin{aligned} \sum_p C_{pt}^*(L_{pt}) - C_t^*(Q_{pt}) &= \sum_p C_{pt}^*(L_{pt}) - C_{pt}^*(Q_{pt}) + c_{pt}(i = Q_{pt}) * [Q_{pt} - L_{pt}] + \dots \\ &\quad - \sum_p [c_{pt}(i = Q_{pt}) - c_t(i = Q_t)] * [Q_{pt} - L_{pt}] \end{aligned} \tag{3}$$

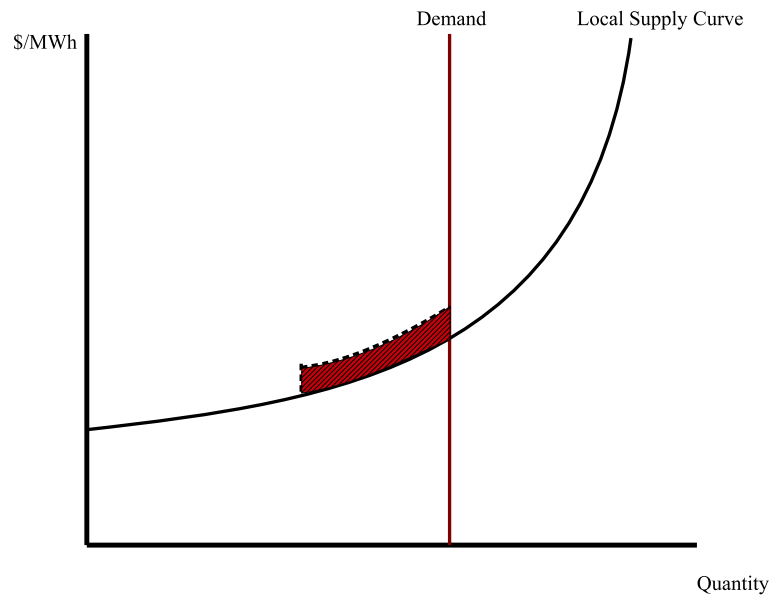
The right side of the first line measures the gains from trade accruing to each area, corresponding to the wedges of Figure 3, Panel B. These are positive, and accrue as cost reductions in the decomposition of equation (2). The second line adds up implied transmission costs: by how much the marginal merit order unit differs from the national merit order (i.e. the intersection of marginal cost curves), times trade volumes. These are barriers to trade, borne as iceberg costs: the sum of transmission bottlenecks and other constraints that prevent a single merit order marginal cost of production from prevailing across all areas. Note that any constant times the sum of trade flows will add to zero, so the national marginal merit order cost simply facilitates interpretation. This term is ultimately a measure of by how much marginal costs in importing areas exceed those of exporting areas, which raises observed production costs in the decomposition of equation (2).

This approach builds upon that of Mansur and White (2012), extending it in a

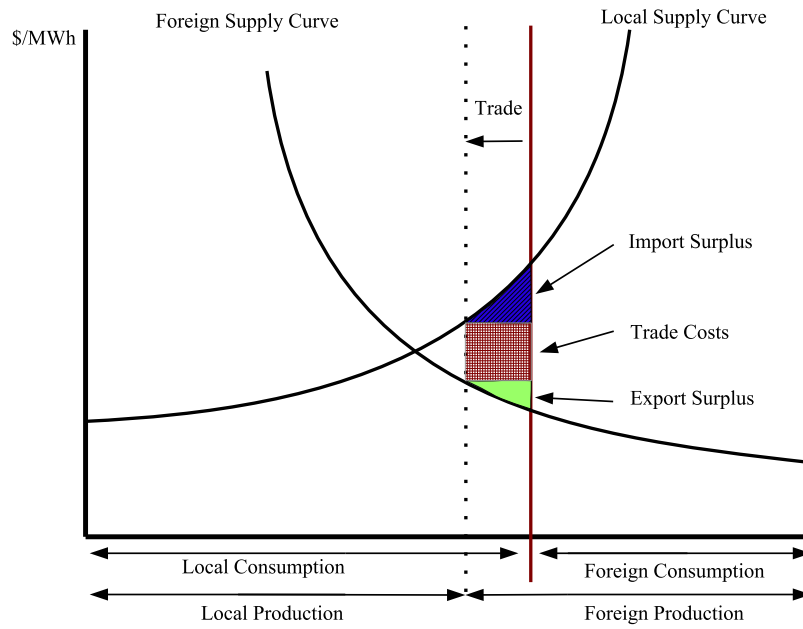
few important ways. First, the gains from trade can be calculated separately for each PCA in a multilateral setting in which it is impossible to determine the origin of electrons. Second, this framework allows for the evaluations of outcomes in the absence of markets, while Mansur and White (2012) use the convergence of market prices—which of course limits analyses to settings where markets of some form are already in use. Finally, I measure gains from trade as wedges accrued PCA-by-PCA rather than complete trapezoids (which would include the cross-hatched area of Figure 3, Panel B). This distinction attributes the remaining barriers to a single marginal cost from prevailing as due to transmission costs. Aside from completing the decomposition of observed production costs, the inclusion of trade costs (broadly defined) provides a natural explanation for marginal cost variance: importers buy power up to the point that the delivery-inclusive price equates their own marginal cost of production, and likewise for exporters. This framework greatly expands the set of questions that can be explored, but comes at the cost of substantially higher data requirements.

**Figure 3:** Welfare Measurement in Electricity Markets

(a) Out of Merit Losses



(b) Gains From Trade



## 4 Data

This study draws from a disparate and incongruous set of data sources to synthesize an essentially complete characterization of U.S. electricity production at the hourly, generating unit level from 1999-2012 (over 530 million unit-hour observations). This section presents an overview of the data, while the details of data assembly can be found in Appendix A.

### Hourly Load Data

The demand side consists of a balanced panel of hourly load (consumption, including line losses) from the 98 major U.S. power control areas (PCAs) that dispatched power plants in 1999 to meet demand. These data have been reported annually to the Federal Energy Regulatory Commission on Form 714, “Annual Electric Balancing Authority and Planning Area Report.” Record-keeping challenges at FERC requires these data to be supplemented with equivalent data from regional authorities and markets (Western Interconnection, ERCOT, PJM, NYISO, NEISO, and NERC). In instances that original administrative data are unavailable (or reporting policies/boundaries change), I employ LASSO to estimate missing demand based on weather, population, and employment. Combined with cross-validation to maximize out-of-sample accuracy, this procedure delivers predictions within 4% of the realized values on average (see the Data Appendix). Small municipal authorities that do not actually conduct dispatch of fossil- or nuclear-powered plants are added to the load of their principal suppliers or customers, yielding 98 total PCAs.

Figure 4 summarizes the electricity load data. The US consumes a bit less than 4,000 TWh (billions of kilowatt-hours) annually. Panel A shows that electricity consumption increased from 1999 until the Great Recession, and was relatively flat through 2012. Panel A also highlights the seasonal nature of electricity usage: summer cooling and winter heating can increase usage by over a third of temperate seasonal usage on a month-to-month basis, with much larger swings during peak usage. Panel B plots hourly usage over the course of the week, averaged over the 14-year study period. Here too there are large swings in usage both over the course of the day and the week. The key fact to remember when interpreting these figures is that production must move exactly in sync with these demand swings, and that utilities must have enough generation capacity to meet demand at the moment of peak usage. Thus every downward swing also represents vast quantities of generating capacity becoming idle.

As a demonstration of real-time patterns of demand, I have animated one year’s

worth of hourly load here. This animation shows the East-to-West flow of electricity demand as usage follows local clocks. It also reflects the daily and seasonal patterns shown in Figure 4, while highlighting the substantial variation around these averages: peak demand can be as much as 2.5 times average annual usage, can be quite persistent during summer months in the South and Southern Plains, and generally varies less in temperate areas of the Pacific Northwest.

### **Hourly Generation Data**

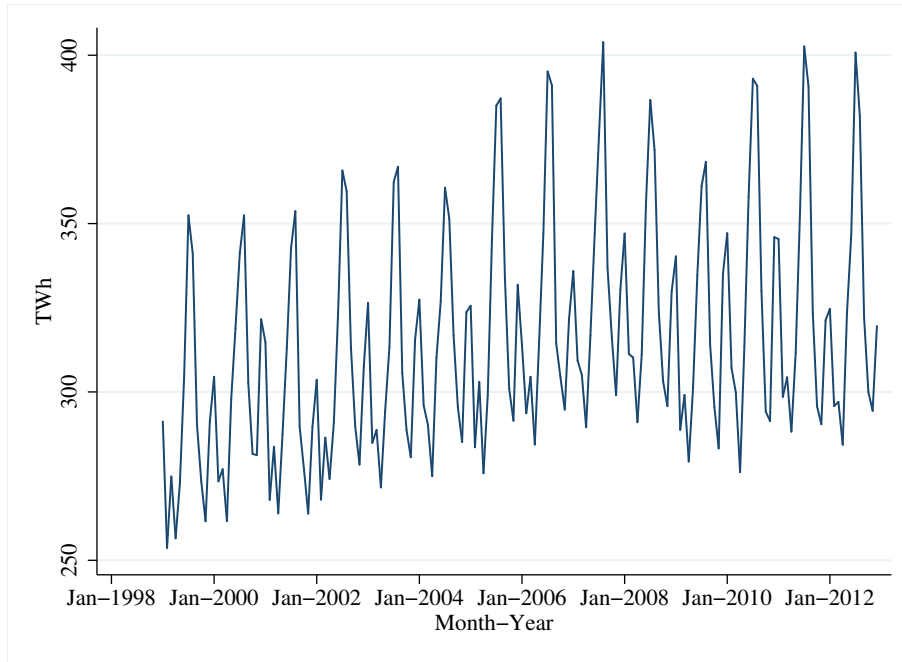
The supply side is based on data from the Energy Information Administration (EIA), merged with hourly gross generation reported to the Environmental Protection Agency (EPA) with Continuous Emissions Monitoring Systems (CEMS), as well as daily production at nuclear-powered units from the Nuclear Regulatory Commission (NRC). Boilers from the EPA are matched to generators' monthly net generation and heat rates via Forms EIA-767 and EIA-923, "Annual Steam-Electric Plant Operation and Design Data / Power Plant Operations Report." Hourly production in the data is the gross generation from CEMS scaled by the ratio of monthly gross-to-net generation from EIA at the unit level. I then merge this data on heat rates and hourly production with coal and oil fuel costs under a non-disclosure agreement with the EIA (from Forms EIA-423, "Monthly Report of Cost and Quality of Fuels for Electric Plants," EIA-923 and Form FERC-423, "Monthly Report of Cost and Quality of Fuels for Electric Plants"). These are shipment-level data, reported monthly by generating facilities with a combined capacity greater than 50MW. I use estimated spot-market coal prices to measure the opportunity cost of coal burned rather than contract prices. Natural gas prices are from 65 trading hubs around the country reported by Platts, Bloomberg, and NGI (not EIA), and are quoted daily. Plants are linked to their nearest trading hub along the pipeline network. Areas with emissions markets for Sulfur and Nitrogen Oxides include the cost of pollution based on measured emissions and monthly market prices from BGC Environmental Brokerage Services.

Generation from hydro-powered units either comes directly from the source (i.e. Tennessee Valley Authority, U.S. Bureau of Reclamation, etc.), or is based on the streamflow of the nearest downstream gage from the U.S. Geological Survey's Stream-gage Network (linked through analysis of the National Hydrography Database). Because the cost of reservoir-based hydropower is the opportunity cost of the water, I price hydropower-based power on the marginal cost of fossil generation in the merit order that is being supplanted. Run-of-river hydro is priced at zero. Hydropower units >10MW were classified as reservoir or run-of-river based on internet searches

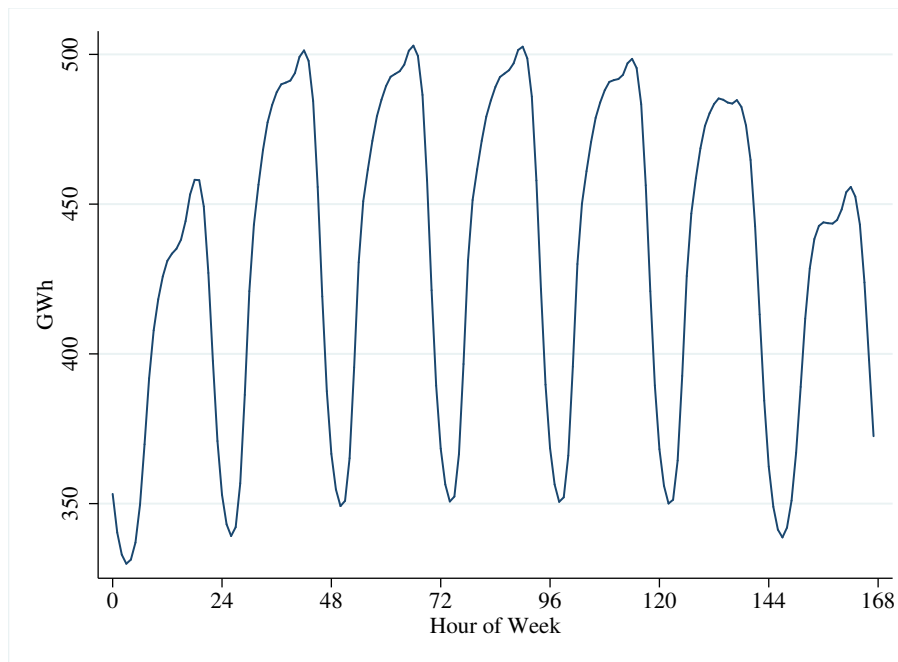


**Figure 4: Electricity Load over Time**

(a) Total Monthly Load



(b) Average Load by Hour of Week



and/or satellite images.<sup>13</sup>

Hourly generation is unavailable from a number of smaller fossil-fired units (whose net generation rarely exceeds 3% by NERC region-year). Power from these units is distributed across the hours of the month in an intuitive manner: having produced  $nW$  MWh in a month, where  $W$  is the unit's nameplate capacity, I assume that the unit produced at maximum capacity during the  $n$  hours of highest demand observed over the course of the month. This replicates the behavior of a dispatcher who employs a threshold rule of when to generate from a unit (assuming no start-up costs or ramping constraints), while allowing observed behavior to dictate what threshold was employed each month.

Figure 5 presents the aggregate annual statistics for electricity generation in the US. Roughly half of the electricity generated from 1999 - 2012 was powered by coal, with a declining share since 2007. From that time, natural gas has grown from rough parity with nuclear (20%) to 30%, almost entirely at the expense of coal-fired generation. Following a nearly three-fold increase from 1999 - 2008, Panel B shows that fossil fuel expenditures fell by approximately 50% from the peak in 2008 from the combined effects of reduced demand overall and the massive reallocation of output to units burning cheaper gas thanks to the advent of hydraulic fracturing (Hausman and Kellogg (2015); Linn et al. (2014); Knittel et al. (2014)). Fossil fuel expenditures averaged about \$72B/year over these 14 years, thus the complete dataset tracks the burning of \$1T of fuel at the plant-generation unit-hour level.

## Matching Supply and Demand

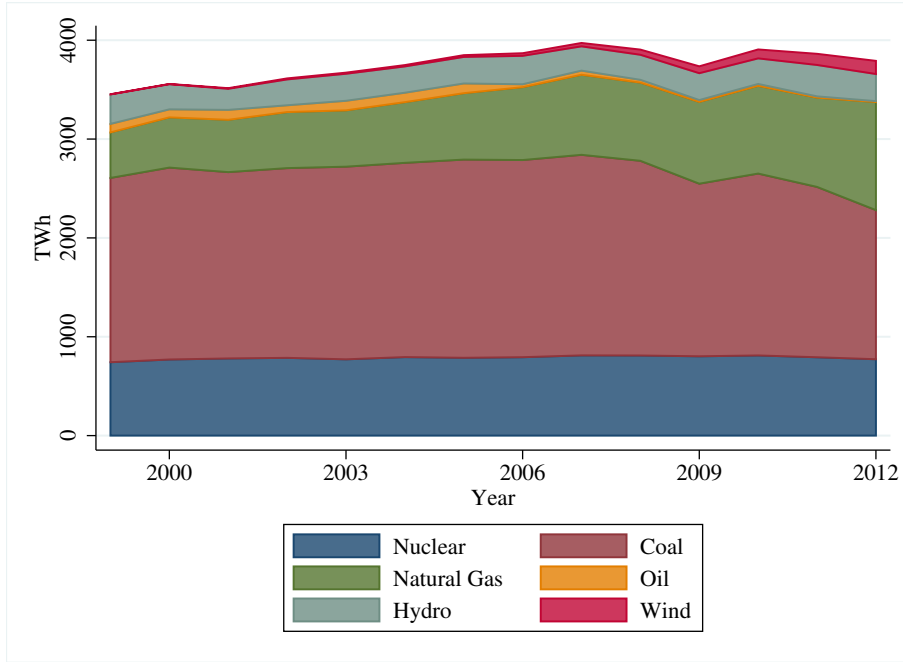
Because the supply data are built up from microdata independently from the demand side, it is important to ensure congruence between the data sources—there is nothing institutional about their reporting to ensure they agree. Beginning with the 1999 configuration of the electrical grid, I match plants to their initial PCA from the EPA eGRID database. New capacity since that time is matched to PCA either directly or based on historical utility service territory in the case that the PCA territory has changed. These associations are then checked based on power plant names reported by PCAs in FERC 714. I then compare the implied monthly totals from the supply side of the data against those reported by the PCAs to FERC. In total, about 99%

---

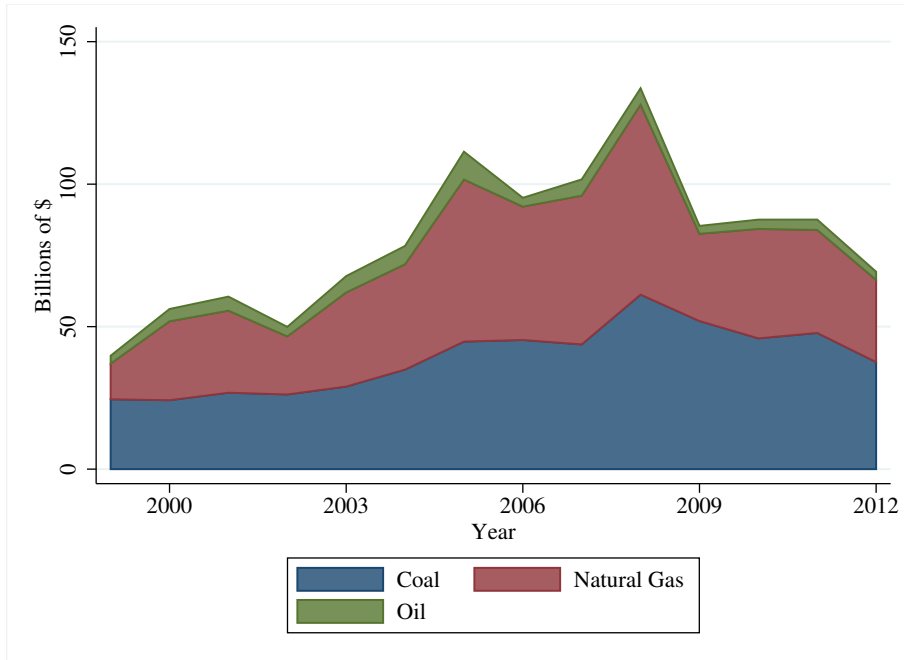
<sup>13</sup>Pricing hydropower at the merit order marginal cost assumes that it is never operated out of merit, so all dynamics of hydro production are orthogonal to out of merit results: the difference between observed and merit order production costs will net out hydro at any price. Pricing reservoir hydropower in this manner infers that the value of storing 1MWh worth of power (either for production later, or flood management, irrigation, etc.) is equal to the marginal cost of fossil power required to offset hydro production. See Archsmith (2017) for a recent application.

**Figure 5:** Annual Net Generation and Fuel Cost by Source

(a) Net Generation



(b) Fuel Cost



of reported generation from FERC 714 can be accounted for in the supply-side data. About 3% of net generation does not fit neatly in to a single power control area because multiple PCAs report a share of output from large plants as their own. In these cases, the plant is assigned to the PCA with greatest dispatch authority.

Figure 6 breaks down generation by data source, and shows the quality of the match between supply and demand. The top black line in Panel A is identical to the total monthly load shown in Panel A of Figure 4. After totaling the generation observed (or calculated) based on high-frequency data, the remaining numbers reported at the monthly level result in totals that almost exactly match the demand side of the data. Panel B gives a closer view of what is missing by calculating the gap (as imports or exports) every hour across PCAs, then adding them separately up to the monthly level, measuring the volume of trade across areas. The first striking statistic is that roughly 90% of generation is effectively consumed in its local PCA—while PCAs are interconnected, they continue to largely produce energy for their own consumption. To my knowledge, these statistics are new: regulatory bodies typically report the *net* flow of electricity between areas, which fails to reflect the real-time interdependence among PCAs (or lack thereof).

The remaining gap between imports and exports as I observe them is due to imports from outside of the US (which have grown over this period to about 1% of supply (Energy Information Administration (2012) Table 2.13). Based on the framework presented in Section 3, not observing this generation effectively values it as an import from outside of each PCA, which is valued as displaced local generation. The production costs and exporter surpluses (mostly from Canada) are outside of the data.

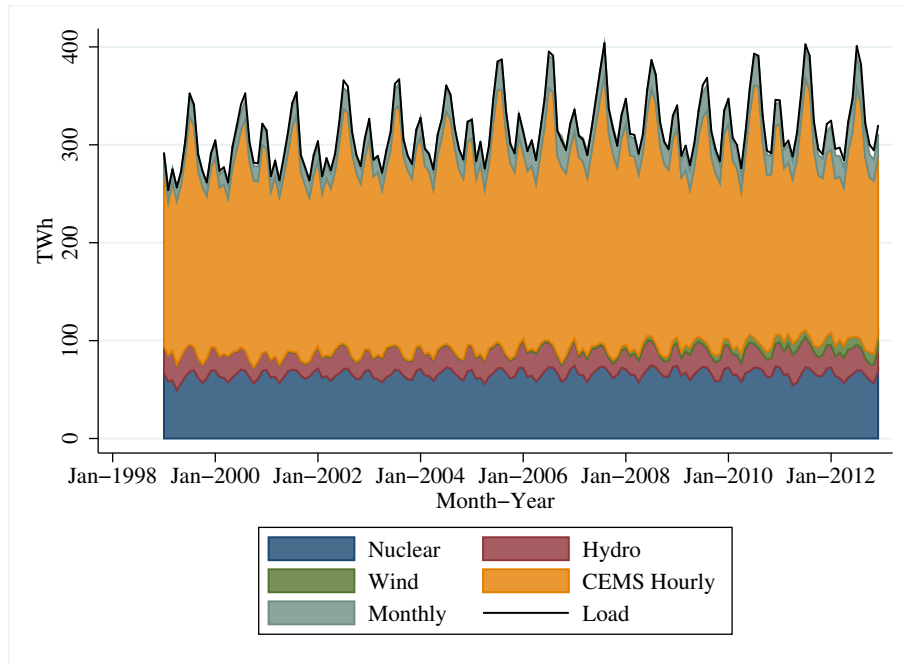
## 5 Estimating Counterfactual Operations

I use the staggered timing of market creation and expansion to arrive at an estimate of the causal impact of the transition to market-based electricity dispatch. These events are defined as the PCAs formally ceding control of their transmission system to an Independent System Operator, who conducts auctions to allocate output to generating units. As demonstrated in Figure 2, these are discrete events—typically demarcated prominently in the history of each market. These events suggest a DD approach, using areas without regulatory change to estimate counterfactual outcomes after one has adjusted for common shocks and time-invariant differences:

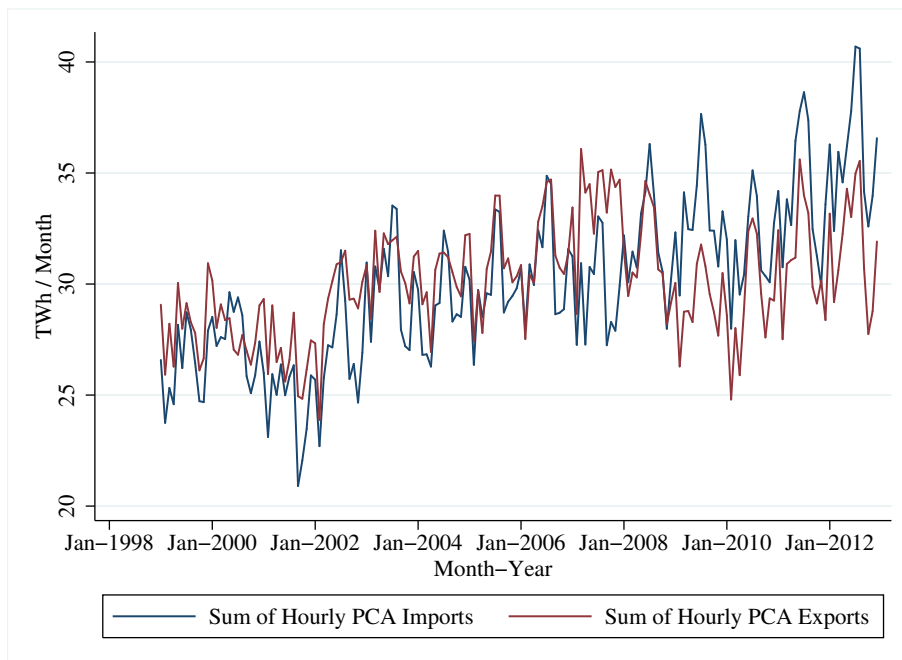
$$y_{pt} = \tau D_{pt} + \gamma_p + \delta_{tr} + \varepsilon_{pt} \tag{4}$$

**Figure 6:** Monthly Net Generation by Data Source and Trade Across PCAs

(a) Monthly Net Generation by Data Source



(b) Monthly Trade Across PCAs



where  $y_{pt}$  is the logged value of the outcome variable for PCA  $p$  in date-hour  $t$ , and  $D_{pt}$  is an indicator of market dispatch. This approach may also allow PCA-specific coefficients on flexible measures of demand (taken to be exogenous since consumers do not face the real-time cost of electricity).<sup>14</sup> The time fixed effects  $\delta_{tr}$  are included at the date-hour-region level to account for spatial and time-varying unobservables, particularly with respect to fuel prices (Cicala (2015)).  $\tau$  measures the average effect of market dispatch, and should be interpreted as an Average Treatment on the Treated (ATT)—it measures the effect in the areas that have adopted market dispatch. Interpreting this as an Average Treatment Effect (ATE) requires the stronger assumption that PCAs in the South and West have the same potential benefits from market integration—rather than the continued business-as-usual assumption required for the validity of the ATT. One should keep in mind that markets themselves are heterogeneous, and their rules change over time. Thus a single “treatment effect” of markets as conceived here takes the average of these various institutional changes, compared to the various institutions that preceded the transition to market dispatch.

### A Policy Function Approach to Counterfactuals

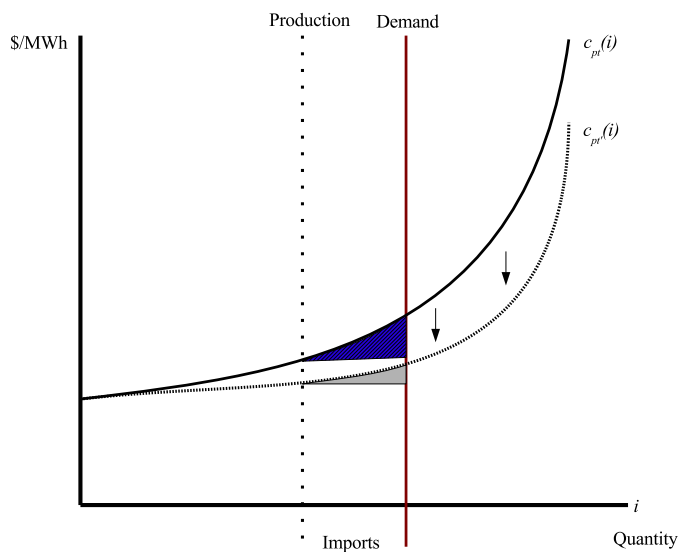
The causal effect of markets on gains from trade or out of merit dispatch is the difference between an observed outcome in a market area and what that outcome *would have* been if not for the market—holding production capacity, fuel costs, and demand fixed. Although DD forms a natural starting point for the analysis, it is insufficient to simply estimate the change in outcomes following market introduction, even relative to areas without any regulatory change: Within a PCA, outcomes (holding demand fixed) are confounded by varying fuel prices over time, which change the cost of operating a given unit out of merit, or the value of offset production through trade. Contemporaneous differences across areas are confounded by the fact that PCAs differ in their installed capacity, and are therefore differentially affected by common time-varying shocks.

To illustrate this problem, Figure 7 presents the import gains of the “local” supply curve from Figure 3. The lighter grey addition represents the import gains realized in this same area, but under a different set of fuel prices, represented by the dashed curve. Though there may be no differences in production between these two curves, the difference in prices yields different gains from trade. If one curve were realized

---

<sup>14</sup>This would account for heterogeneous time-invariant relationships between the outcome variable and demand—the fact that some areas are more prone to congestion in times of high demand, for example. In addition, controlling for demand also accounts for the possibility that differences in outcomes across areas might be driven by regional trends (such as population), instead of how dispatchers allocate production taking demand as given.

**Figure 7:** Cost Changes Unrelated to Deregulation Confound Counterfactual Estimates



Note: This figure shows how measured gains from trade change with the price of fuel, holding demand and traded quantities constant. Using gains from trade in period  $t'$  as the counterfactual for what would have happened in  $t$  in the absence of treatment would yield a predicted change in outcome in spite of no behavioral change.

during regulation, and the other after the introduction of markets, a simple difference would indicate that gains from trade had changed, although there had been no change in behavior. Because each PCA differs in the composition of units, common fuel price shocks affect areas differently, comparing changes in neighboring areas will fail to correct for this confounding.

I propose a policy function approach that builds upon the “generation regressions” of Davis and Hausman (2016) and emissions regressions of Mansur (2007); Holland and Mansur (2008); Graff Zivin et al. (2014); Holland et al. (2016) to overcome this issue: I use historical patterns of unit-level production given load, unit capacity, and position in the merit order to estimate predicted allocations of production.<sup>15</sup> I apply these predicted quantities to observed unit costs and demand to estimate what production costs would have been if not for treatment.

Let a policy function for PCA  $p$  in year  $y$  be the probability that the PCA orders generation from the  $i^{th}$  MW of capacity of the merit order in hour  $t$ , conditional upon covariates  $X_{ipt}$  (such as load, month of year, hour of day, and nameplate capacity of

<sup>15</sup>Here, policy refers to a rule that maps states in to actions, without any reference to the optimality of that rule, as is typically implied in the use of this term in the control theory literature.

the unit producing the  $i^{\text{th}}$  MW) and treatment,  $D_{pt}$

$$\psi_{py}(i, X_{ipt}, D_{pt}) = Pr [q_{pt}(i) = 1 | X_{ipt}, D_{pt}] \quad (5)$$

With this notation, a PCA can be expected to produce the total amount

$$\mathbb{E}(Q_{pt} | X_{ipt}, D_{pt}) = \sum_{i=0}^{N_{pt}} \psi_{py}(i, X_{ipt}, D_{pt})$$

Expected costs of production are based on the inner product of costs and the policy function in vector form,  $\mathbf{c}_{pt} / \psi_{pt}(\mathbf{X}_{pt}, D_{pt})$ .

To operationalize these policy functions for causal inference, some assumptions are required. To economize on notation I adopt the ‘Potential Outcomes’ framework popularized by Rubin (1974), in which a generic outcome can be thought of taking on value  $Y_{pt}^0$  in the absence of treatment, and  $Y_{pt}^1$  if treated. Thus estimating the causal impact when  $Y_{pt}^1$  is observed requires estimating  $Y_{pt}^0$ , which is not. Here the outcomes being evaluated are functions of production allocations and costs, capacities, and demand:  $Y^D = F(\mathbf{q}_{pt}^D, \mathbf{X}_{pt}^D)$ , such as gains from trade in the first line of equation (3).

**Assumption 1.** *Demand, unit production costs and capacities are invariant to treatment:*

$$X_{ipt}^0 = X_{ipt}^1 = X_{ipt}$$

This assumption narrows the set of potential outcomes to focus on the question: how does market dispatch affect the allocative efficiency of meeting demand? Real-time pricing for retail customers is nearly nonexistent during the sample period, so that consumers’ behavior is invariant to hourly production costs. Although I have shown elsewhere (Cicala (2015)) that prices paid for coal (but not gas) depend on plant-level regulations, this study focuses on allocative efficiency changes—how production moves across power plants holding costs fixed. The brief time horizon evaluated after the introduction of market dispatch is intended to hold the capital stock fixed so that the observed supply function for each area is invariant to treatment.



**Assumption 2.** *Parallel trends in unobservables and evolution of the policy function:*

$$\begin{aligned}
Y_{pt}^0 &= F(\mathbf{q}_{pt}^0, \mathbf{X}_{pt}) \\
&= F(\psi_{\mathbf{p},y-1}^0, \mathbf{X}_{pt}) + \underbrace{F(\mathbf{q}_{pt}^0, \mathbf{X}_{pt}) - F(\psi_{\mathbf{p},y}^0, \mathbf{X}_{pt})}_{\text{Contemporaneous Error}} + \dots \\
&\quad + \underbrace{F(\psi_{\mathbf{p},y}^0, \mathbf{X}_{pt}) - F(\psi_{\mathbf{p},y-1}^0, \mathbf{X}_{pt})}_{\text{Change in Policy Function}} \\
&= F(\psi_{\mathbf{p},y-1}^0, \mathbf{X}_{pt}) + \delta_{tr} + \gamma_p + v_{pt}
\end{aligned}$$

where  $E(v_{pt}) = 0$

This assumption forms the basis of the estimation strategy, using the allocation of production based on operations in year  $y - 1$  to predict operations *in the absence of markets* in year  $y$ . There are two forms of error with this approach: the difference between the true outcome and the value based on the contemporaneous policy function, and the difference induced by the evolution of policy functions from year-to-year. Assumption 2 decomposes these errors into a PCA-specific, time-invariant component, a regional contemporaneous shock, and noise. This allows, for example, for out-of-sample predictions based on last year's operations to persistently be off by an amount that varies by PCA, while also accounting for contemporaneous regional shocks to fuel prices.

**Assumption 3.** *Conditional Independence of Treatment for Control Outcomes and Policy Function Measurement Error*

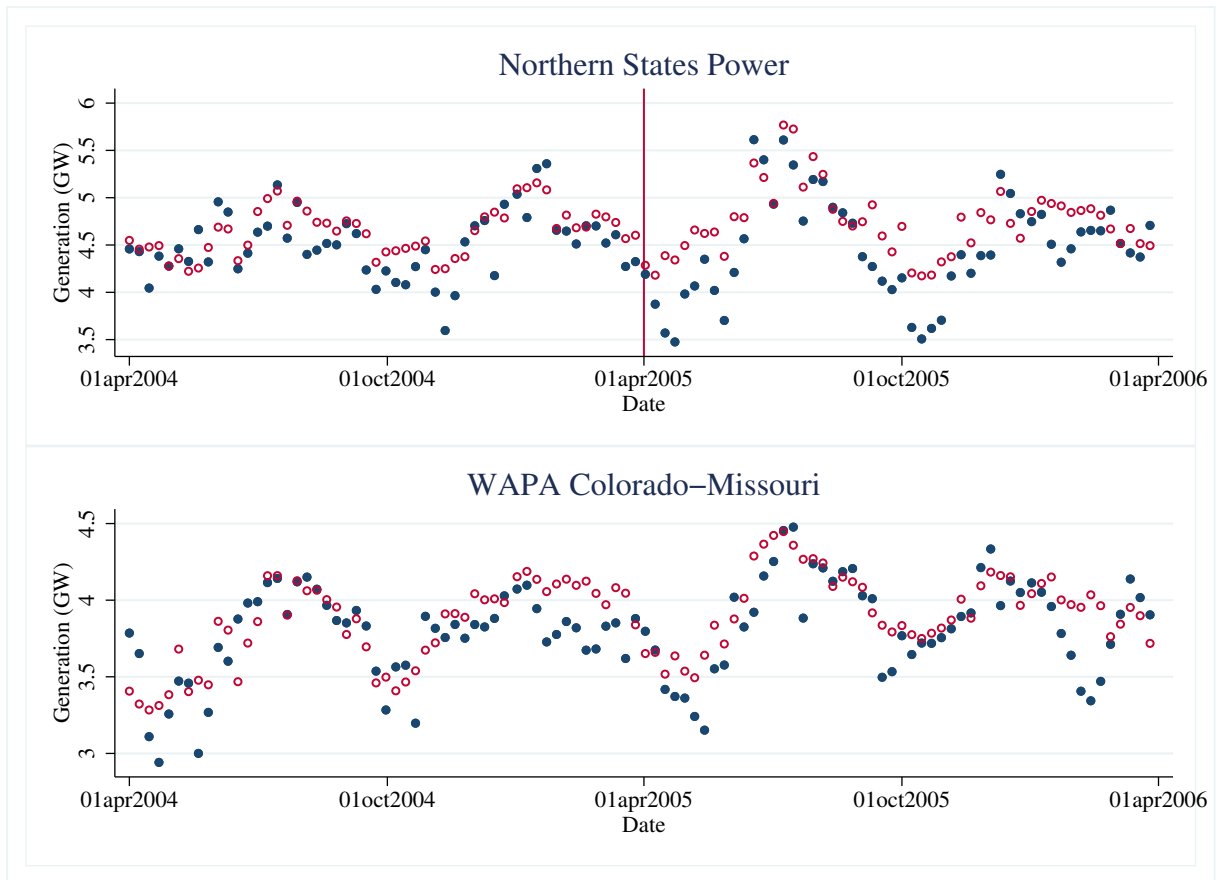
$$Y_{pt}^0, \left( F(\psi_{\mathbf{p},y-1}^0, \mathbf{X}_{pt}) - F(\hat{\psi}_{\mathbf{p},y-1}^0, \mathbf{X}_{pt}) \right) \perp\!\!\!\perp D_{pt} | X_{pt}$$

That treatment is conditionally independent of control outcomes allows the identification of an average treatment on the treated (ATT). The second part of this assumption ensures that using estimated values of counterfactual outcomes will not bias estimates of the treatment effect. Rather than including these estimates as a generated regressor, this assumption allows a modified DD-type estimating equation in which the dependent variable is the departure from the outcome predicted by the estimated policy function:

$$Y_{pt} - F(\hat{\psi}_{\mathbf{p},y-1}^0, \mathbf{X}_{pt}) = \tau D_{pt} + \delta_{tr} + \gamma_p + u_{pt} \quad (6)$$

To develop intuition for how estimates are constructed, Figure 8 plots weekly gen-

**Figure 8:** Comparing Observed and Predicted Generation After Market Dispatch



Notes: Unit-level generation is aggregated by PCA-hour, then averaged by week. Observed data are denoted with solid symbols. Out of sample predictions based on the prior year's data are hollow symbols. The vertical red line indicates the beginning of market dispatch for Northern States Power. WAPA Colorado-Missouri did not adopt market dispatch during this period.

eration for two PCAs, Northern States Power and WAPA Colorado-Missouri. Starting with unit-level generation, the solid symbols aggregate each PCA’s hourly production, and depicts weekly averages. Northern States Power joined the Midwest on April 1, 2005, so the observed production after the red line in the top panel is an example of  $Y_{pt}^1$  in the nomenclature above, and the rest of observed production is  $Y_{pt}^0$ . The hollow symbols are predicted values based on the prior year’s estimated policy function,  $(\hat{\psi}_{p,y-1}^0)$  and contemporaneous covariates,  $(X_{pt})$ . These are combined to produce predicted outcomes  $F(\hat{\psi}_{p,y-1}^0, \mathbf{X}_{pt})$ , that estimate what the allocation would have been if units were run as they had been in the prior year.<sup>16</sup> The dependent variable in equation (6) is the difference between these observed and predicted values—the vertical distance between solid and hollow symbols in Figure 8. The treatment effect is estimated by comparing the change in how well predictions match outcomes before and after the red line, between the upper and lower panels. The figure shows that predictions were matching outcomes well for both areas in the year before market dispatch, but that following market dispatch, Northern States Power began generating less electricity than predicted, while predictions continue to match observed operations in WAPA Colorado-Missouri.

There are a number of potential threats to the validity of this research design. First and foremost, the stable unit treatment value assumption (SUTVA) requires that the treatment status of markets that become PCAs does not affect the outcomes of other areas. This will be violated, for example, if the expansion of markets in Ohio facilitates the delivery of electricity from the Tennessee Valley Authority (TVA), which is not dispatched by markets. Using TVA as a control PCA will understate the true effect of market dispatch when their exports change due to the policy change. This estimation framework also assumes that outcomes change immediately with the change in treatment status. However, sudden massive changes tend not to be conducive to keeping the lights on. The pre-period may be contaminated if PCAs began to change their dispatch policies in preparation for the transition to markets. On the other hand, the treatment effect may take time to fully manifest itself as PCAs learn how to use the market to improve their operations (or exert market power).

As is standard in DD research designs, unrelated, differential trends between treatment and control also threaten the validity of estimates. The policy function approach mitigates this issue by transforming the dependent variable in to the residual of behavior predicted by the prior year’s policy. This kicks the threat of differential trends

---

<sup>16</sup>In this example,  $F()$  is the sum of  $\hat{\psi}_{p,y-1}^0$  over the merit order, yielding predicted PCA generation.

up one level (for quantities, not prices), requiring an unrelated trend in how well last year’s policy matches that of this year. Such problems become evident with event study-style estimates leading up to the time of treatment.

On interpretation, the supply curves I construct are based on fuel and emissions prices, while variable labor, operations, and maintenance costs are ignored. Although these other costs are small relative to total variable cost, they create distance between my measured merit order and the true marginal cost of power. The treatment effect on the costs I observe may be well-measured, but it will be a biased estimate of the overall change in allocative efficiency if something about the transition to market dispatch changes these errors—such as reduced labor costs in markets as in Fabrizio et al. (2007). The small share of non-fuel costs multiplied by the modest impact of restructuring renders the potential magnitude of this bias quite small. There is likely to be greater measurement error concerning exact fuel prices and unit capacities. I reduce these errors to the extent possible by using daily gas prices at geographically disperse hubs (to account for pipeline congestion), and by using the implied capacity based on observed operations from CEMS (maximum hourly net generation by season) rather than the round figures reported to EIA. Again, these errors bias my causal estimates only to the extent that they are non-stationary and correlated with market dispatch.

Regarding inference, estimates using this approach are presented with standard errors calculated by block-bootstrapping PCA-months, with regular DD estimates clustered at the PCA-month. This reflects the thought experiment that the observed data (a complete census of operations) are drawn from a super-population of operations to allow for the inference of potential outcomes—and that each months’ fluctuations in demand allow for an independent observation for each PCA. If one believes that there are really only 98 (PCA) independent observations, the reported standard errors roughly double. Conversely, the standard errors become infinitesimal if one follows the existing literature, having studied one area at a time with independence (or serially correlation) assumed across fine time units.

### **Machine Learning Estimation of the Policy Functions**

The policy function approach removes the role of fuel price variation in the estimation of counterfactual outcomes for a given allocation of output: Instead it is the quantities themselves that are predicted, then applied to the observed production costs and demand to calculate counterfactual behavior.

Estimating the policy functions requires balancing flexibility and risk of over-

fitting. On one hand, the probability of running a unit is a complex, unknown function of the variables system operators use to make decisions—simple approximations are unlikely to deliver high-quality predictions of behavior. On the other hand, overly-flexible specifications may provide the illusion of superior fit, but perform poorly out-of-sample. Since the estimated treatment effect comes from changes in the quality of fit between predicted and observed behavior, it is particularly important to prevent overfitting from showing up as illusory treatment effects.

This is a pure prediction problem, for which recent tools from machine learning are well-suited. I use the “random forest” algorithm of Breiman (2001) as implemented by Wright and Ziegler (2016). This nonparametric estimation algorithm draws bootstrap samples of the data and calculates means of the outcome variable for random partitions of the explanatory variables. It then aggregates these weak predictions across the bootstrap samples to form robust estimates without functional form assumptions.

More formally, for PCA  $p$  and year  $y - 1$  with sample size  $N_{p,y-1}$ , random forest draws  $N_{p,y-1}$  pairs  $(q_{ipt}, X_{ipt})$  with replacement from that PCA-year. It then “grows” a regression tree as follows: starting from an origin node, it randomly selects a set of variables from  $\mathbf{X}_{p,y-1} \subseteq \mathbb{R}^p$  where  $p$  is the dimension of  $\mathbf{X}_{p,y-1}$ . It then splits the data along these dimensions at cut-points that make the subsequent nodes as homogeneous as possible with respect to the outcome (Breiman et al. (1984)), forming two nodes. Each of these nodes are subsequently split using the same method until a pre-specified (and here, cross-validated) number of observations remain at each final node, referred to as leaves (or perfect uniformity is achieved). Using  $\theta_m$  to denote the random vector used to draw the bootstrap sample and determine which explanatory variables are used to split at each node of tree  $m$ , the tree produces a set of leaves  $l = 1, \dots, L$  that partition the space of explanatory variables ( $\mathbb{R}^p$ ) in to rectangular subspaces,  $\mathbb{R}_l$ . The prediction of the tree given a particular  $x$  is obtained by averaging over the outcomes of the observations in the leaf to which  $x$  belongs,  $l(x, \theta_m)$ . Following Meinshausen (2006), the prediction for a vector of covariates  $x$  can be thought of as a weighted mean of the entire sample of the original data, depending upon each observation’s inclusion in the bootstrapped sample and terminal leaf position

$$\hat{\psi}_m(x) = \sum_{i=1}^{N_{p,y-1}} w_i(x, \theta_m) q_i$$

where

$$w_i(x, \theta_m) = \frac{1 \{X_i \in \mathbb{R}_{l(x, \theta_m)}\}}{\sum_{j=1}^{N_{p,y-1}} 1 \{X_j \in \mathbb{R}_{l(x, \theta_m)}\}}$$

with  $1 \{ \cdot \}$  denoting an indicator function that is one when the statement in the braces is true, zero otherwise. The prediction from a single tree provides a poor prediction—it does not use all of the underlying data and over-fits the data it does use—Breiman (2001) shows that as the number of trees grown in this way increases, the quality of out-of-sample predictions stabilizes.<sup>17</sup> Continuing with the weighted-average interpretation, one draws a number of iid  $\theta_m$  vectors to grow  $M$  trees, then calculates the final weights each observation receives in the final prediction as

$$w_i(x) = \frac{1}{M} \sum_{m=1}^M w_i(x, \theta_m)$$

Predictions for policy functions are made out-of-sample for year  $y$  for data with explanatory variables  $\mathbf{X}_{\mathbf{p},y}$  by calculating

$$\hat{\psi}_{\mathbf{p},y-1}(\mathbf{X}_{\mathbf{p},y}) = \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^{N_{p,y-1}} w_i(\mathbf{X}_{\mathbf{p},y}, \theta_m) q_i$$

where each  $i$  indexes observations of the production data of PCA  $p$  in year  $y - 1$ .

The core motivation for methods such as random forest from the machine learning literature has been its performance in out-of-sample prediction. This remains true in this setting as well, as demonstrated in Figure 9. The metric of performance here is the out-of-sample residual sum of squares, divided by that of a simple OLS regression of unit operations on the covariates used in the random forest estimation (separately by PCA, including month and hour of day as dummies). The x-axis separates units by their position in the merit order, as a percentile of costs of installed capacity for each PCA-hour to create a common scale. The figure is constructed using data from areas without market dispatch to show quality of fit in the control group. The dashed line shows the performance of a more flexible OLS specification: a second-order polynomial of all terms, estimated separately by PCA for each month and hour of day. While this specification fits the data better than the simpler one, random forest far outperforms throughout the merit order. It delivers a superior fit to observed operations uniformly

---

<sup>17</sup>Scornet et al. (2015) establish the consistency of random forests grown in this way as estimators of the conditional expectation function in the presence of an additive error. Wager and Athey (2016) establish consistency and asymptotic normality results more broadly in the context of causal inference using the unconfoundedness assumption for estimating treatment effects conditional upon terminal leaf partitions, and review related results.

across the merit order, and is particularly good at predicting baseload and peaking operations.

It is important to note that this estimation framework has been designed so that *all* predicted values of the policy function come from out-of-sample estimates. The treatment effect is based on *changes* in how well last year’s operations predict this year’s operations. Using the change in how well observed behavior in year  $y - 1$  fits predictions estimated during year  $y - 1$  against using year  $y - 1$ ’s predictions for year  $y$  risks baking-in an in-sample/out-of-sample change in fit.<sup>18</sup> It also requires iterative estimation of placebo treatment dates among areas that never receive treatment, an exceptionally high computational burden in this setting (or the assumption of time invariance of overfitting issues).

## 6 Results

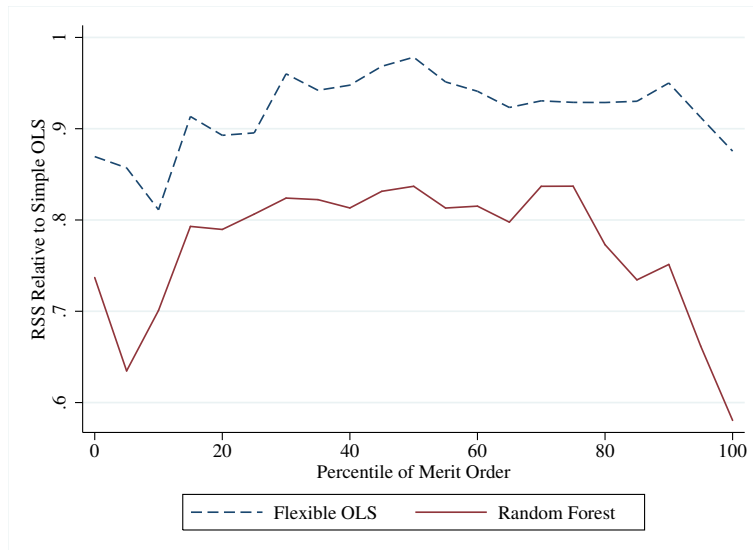
Tables 1 through 4 present the main results as Average Treatment on the Treated estimates to measure the impact of market dispatch on allocative efficiency in electricity production. The first columns are based on straight DD estimates that include date-hour-region and PCA fixed effects. The second column flexibly controls for the effect of load on the outcome variables (allowing a different slope for each quartile of each PCA’s load distribution). This permits each area to have persistent idiosyncratic relationships between demand and how it goes about meeting that demand with out of merit generation and trade. The third column adds PCA-specific time trends. The final column transforms the outcome variable to be the difference between the observed outcome and that predicted by the policy function, as described in Section 5.

All specifications also include separate dummies for greater than 24 months prior, and greater than 24 months after the transition to markets. This serves two functions: For the first three specifications, this prevents long-term responses to market dispatch (and potential confounders) from loading on to the short-term DD estimates. For the policy function estimates, “treatment” only occurs when predicting behavior for a period with a different status of market dispatch. I predict from the year before dispatch out two years afterwards (and year-on-year otherwise). Subsequent predic-

---

<sup>18</sup>Burlig et al. (2017) deal with this issue by randomly selecting a placebo date to separate in-sample/out-of-sample data in the control group, then including an indicator of out-of-sample prediction. This makes the estimated treatment effect the relative deterioration of fit going from in-sample to out-of-sample (non-contemporaneously). The approach taken here avoids making assumptions on the in-sample/out-of-sample transitions, instead evaluating the quality of out-of-sample predictions made from contemporaneous training periods. The cost of this approach is that a year of baseline outcomes lacks out-of-sample predictions.

**Figure 9:** Relative Prediction Quality of Random Forest in Control Group



Notes: This figure plots the relative residual sum of squares (RSS) based on out-of-sample predictions of random forest and a flexible OLS specification where the numeraire is the RSS of an OLS specification with linear terms and indicators for month and hour of day. Explanatory variables include: position in the merit order, nameplate capacity, and load. All predictions are estimated separately by PCA.



tions are based on behavior after markets have begun, making treatment effectively an impulse during this initial window. A post-24-month indicator allows this new period to have a different mean than pre-treatment. Changes in observation counts between these tables indicates the extent to which PCAs operate exactly according to my measure of the merit order: zeroes are dropped in the logarithmic specifications when the merit order is followed so that no generation is out of merit. The drop in observations between the DD and policy function specifications in Tables 1 through 4 is because the baseline period is held out to ensure all observations for the policy function estimates are from out-of-sample calculations.<sup>19</sup>

Beginning with quantities, Table 1 indicates a roughly 10% increase in traded volumes following the adoption of market dispatch. Because the data is structured PCA-by-PCA, an increase in exports in one area will be complemented with increases in imports in other areas. In a DD framework, this yields an underestimate of the true treatment effect if that power is being sent to a control PCA in a way that increases its trade volumes (netting out the increased trade between treatment and control). Increases in trade between market PCAs are not double-counted in coefficient estimates, as their changes are being compared to changes in control PCAs. These estimates are relatively stable across specifications, and do not change in a statistically significant manner when using policy functions to predict counterfactual operations. This is also true for out of merit generation, with the exception of a drop when including PCA-specific trends. However, estimates return to their original levels in the final specification, suggesting the coarser linear trend projects over changes that are more subtly accounted for with policy functions.

One striking measure of the predictive power of the policy functions is to compare the  $R^2$  of the models across specifications. Removing the outcomes predicted by the machine learning algorithm leaves substantially less variation in the dependent variable, and the control variables have far less power in explaining the variation that remains.

To ensure these results are not the artifacts of pre-existing time trends, Figure 10 estimates the model of column (2) to control for load, including separate dummies for each month measuring the time until (or since) market dispatch adoption. Note that this specification only measures the effect for the initial transition to market dispatch: performance changes among incumbents (with whom the area is trading)

---

<sup>19</sup>To avoid losing the first year completely (which includes the New York and New England transitions), the held-out data are from every-other week for the first year. The impacts in these markets were relatively large, but dropping them does not change the overall estimates substantially.

following market expansion are not included. While not as clean as one might like for event study-style figures, they make clear that the overall estimates are not due to long-term trends. There is an overall level shift in Panel (a) corresponding with the onset of treatment, while it appears the slide to a new, lower level of out of merit generation occurs over a few months.

Tables 3 and 4 estimate the welfare impacts of this reallocation of output based on changes in production costs. These results indicate substantial impacts of market dispatch: over 30 log points for gains from trade, and 20% reductions in out of merit costs. One should note the substantial reduction in observations between the two tables. This is because the specifications in Table 3 condition upon positive gains from trade: in roughly 25% of PCA-hours, there is sufficiently little trade that both supply and demand land on the same generator, which yields zero surplus (as described in Section 3).

That the value of this trade exceeds the change in volume implies that there is a substantial gap between the cost of electricity whose production increases versus that being displaced—equating the marginal cost of power across areas would yield zero net benefits of an additional MWh traded. Similarly for out of merit costs, these results imply that it is the relatively expensive out of merit units whose production is reduced by market dispatch.

Figure 11 presents the main results on cost reductions relative to the onset of treatment. The additional volatility of coefficients in these figures relative to the quantity estimates highlights the dependence of the welfare estimates on fuel prices mentioned above: production costs scale with input prices in this Leontief setting, so volatility in fuel prices is directly translated in to volatility of the welfare impacts of a given change in behavior.

While there is a nice jump in gains from trade in Panel (a), longer-term trends play a more prominent role in the overall shape of the plots than in Figure 10. That differential trends in fuel prices might confound estimates as presented in Figure 11 motivates the policy function estimates of Figure 12. This figure plots the analogous event-time coefficients, but with the dependent variable transformed in to the residual between observed outcomes and those predicted by the policy function. Although this approach adds volatility relative to the straight DD estimates, it makes clear that the estimated treatment effects are not due to differential trends: there are unambiguous breaks in the series for both outcomes and an absence of pre-trends. The timing of these breaks also correspond with the transition to markets, though reductions in out of merit costs begin the month prior to market dispatch.

There are two possible contributing factors to why the results are lower for the policy function approach for gains from trade. First, estimating a slightly smaller shift in output naturally yields smaller cost reductions (less output is offset). Second, as highlighted in Figure 7, the simple DD estimates are potentially confounded by fuel price changes. Higher fuel prices steepen the supply curve, yielding greater gains from trade compared to periods with lower prices. Straight DD estimates make such comparisons, while the basis of the policy function approach is to compare a given supply curve to itself.

With over \$4.2B in trade surplus in market PCAs annually, applying the gains from trade treatment uniformly over the treated territories is worth nearly \$1B/year. For out of merit costs, these estimates applied to the \$10B accrued in market PCAs raises the overall impact of this institutional change on cost reductions to about \$3B per year.

**Table 1:** Market Dispatch on  $\text{Log}(\text{Trade Volume})$ 

|                           | (1)                 | (2)                 | (3)                 | (4)                 |
|---------------------------|---------------------|---------------------|---------------------|---------------------|
| Market Dispatch           | 0.125***<br>(0.021) | 0.118***<br>(0.020) | 0.095***<br>(0.021) | 0.098***<br>(0.029) |
| $\text{Log}(\text{Load})$ |                     | Yes                 | Yes                 | Yes                 |
| PCA Trend                 |                     |                     | Yes                 | Yes                 |
| Policy Function           |                     |                     |                     | Yes                 |
| Clusters                  | 16464               | 16464               | 16464               | 15910               |
| PCAs                      | 98                  | 98                  | 98                  | 98                  |
| $R^2$                     | 0.557               | 0.595               | 0.613               | 0.099               |
| Obs.                      | 12001882            | 12001882            | 12001882            | 11352820            |

Note: All specifications include PCA and Region-Date-Hour Fixed Effects. Demand controls are PCA-specific. Standard errors clustered by PCA-Month in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

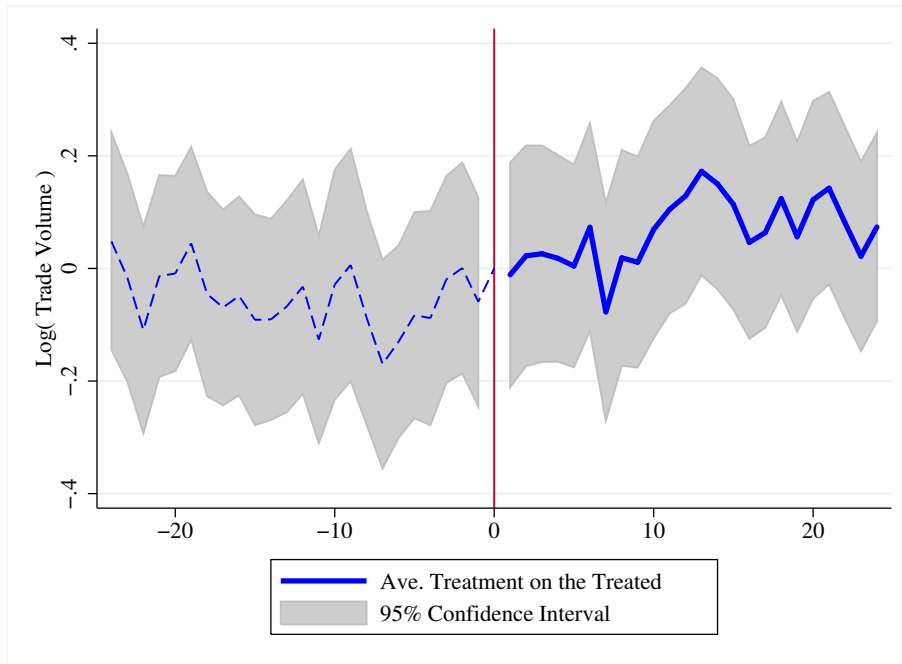
**Table 2:** Market Dispatch on  $\text{Log}(\text{MWh Out of Merit})$ 

|                           | (1)                  | (2)                  | (3)                  | (4)                  |
|---------------------------|----------------------|----------------------|----------------------|----------------------|
| Market Dispatch           | -0.111***<br>(0.018) | -0.113***<br>(0.017) | -0.061***<br>(0.017) | -0.117***<br>(0.019) |
| $\text{Log}(\text{Load})$ |                      | Yes                  | Yes                  | Yes                  |
| PCA Trend                 |                      |                      | Yes                  | Yes                  |
| Policy Function           |                      |                      |                      | Yes                  |
| Clusters                  | 16448                | 16448                | 16448                | 16430                |
| PCAs                      | 98                   | 98                   | 98                   | 98                   |
| $R^2$                     | 0.841                | 0.852                | 0.863                | 0.157                |
| Obs.                      | 11648909             | 11648909             | 11648909             | 11428353             |

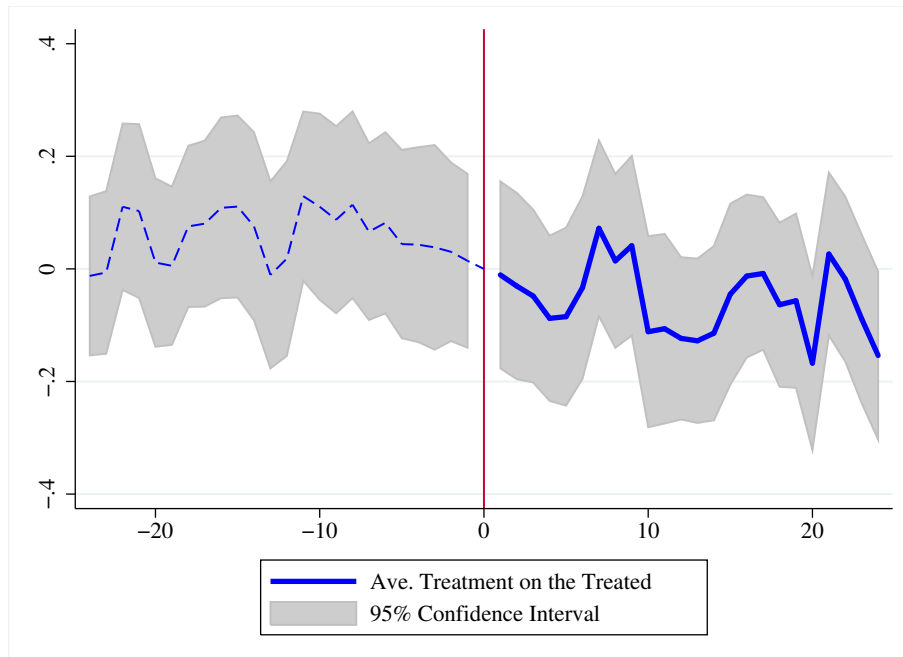
Note: All specifications include PCA and Region-Date-Hour Fixed Effects. Demand controls are PCA-specific. Standard errors clustered by PCA-Month in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

**Figure 10:** Treatment Effects by Months to Market: Quantities

(a)  $\text{Log}(\text{Trade Volume})$



(b)  $\text{Log}(\text{MWh Out of Merit})$



Note: These figures are based on regressing logged outcomes on a set of indicator variables for each month until (after) the transition to market dispatch, PCA-specific controls for load, date-hour-region and PCA fixed effects. The month prior to treatment is normalized to zero. Confidence intervals are based on clustering at the PCA-month level.

**Table 3:** Market Dispatch on  $\text{Log}(\text{Gains from Trade})$ 

|                           | (1)                 | (2)                 | (3)                 | (4)                 |
|---------------------------|---------------------|---------------------|---------------------|---------------------|
| Market Dispatch           | 0.312***<br>(0.045) | 0.350***<br>(0.042) | 0.258***<br>(0.041) | 0.191***<br>(0.048) |
| $\text{Log}(\text{Load})$ |                     | Yes                 | Yes                 | Yes                 |
| PCA Trend                 |                     |                     | Yes                 | Yes                 |
| Policy Function           |                     |                     |                     | Yes                 |
| Clusters                  | 16424               | 16424               | 16424               | 15814               |
| PCAs                      | 98                  | 98                  | 98                  | 98                  |
| $R^2$                     | 0.494               | 0.579               | 0.617               | 0.125               |
| Obs.                      | 8671235             | 8671235             | 8671235             | 8098935             |

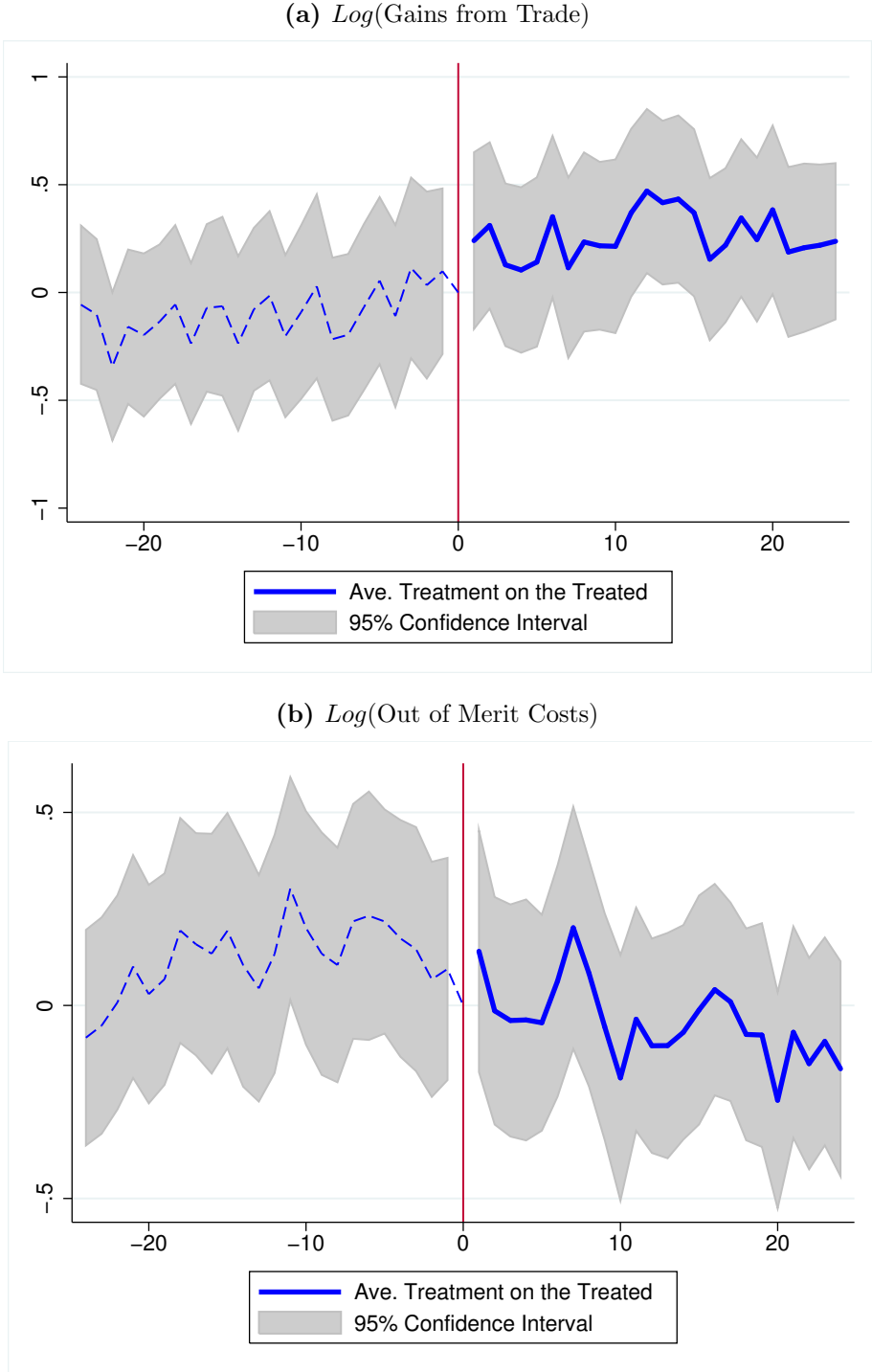
Note: All specifications include PCA and Region-Date-Hour Fixed Effects. Demand controls are PCA-specific. Standard errors clustered by PCA-Month in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

**Table 4:** Market Dispatch on  $\text{Log}(\text{Out of Merit Costs})$ 

|                           | (1)                  | (2)                  | (3)               | (4)                  |
|---------------------------|----------------------|----------------------|-------------------|----------------------|
| Market Dispatch           | -0.186***<br>(0.034) | -0.164***<br>(0.033) | -0.009<br>(0.033) | -0.179***<br>(0.033) |
| $\text{Log}(\text{Load})$ |                      | Yes                  | Yes               | Yes                  |
| PCA Trend                 |                      |                      | Yes               | Yes                  |
| Policy Function           |                      |                      |                   | Yes                  |
| Clusters                  | 16450                | 16450                | 16450             | 16444                |
| PCAs                      | 98                   | 98                   | 98                | 98                   |
| $R^2$                     | 0.775                | 0.793                | 0.812             | 0.302                |
| Obs.                      | 11648731             | 11648731             | 11648731          | 11427266             |

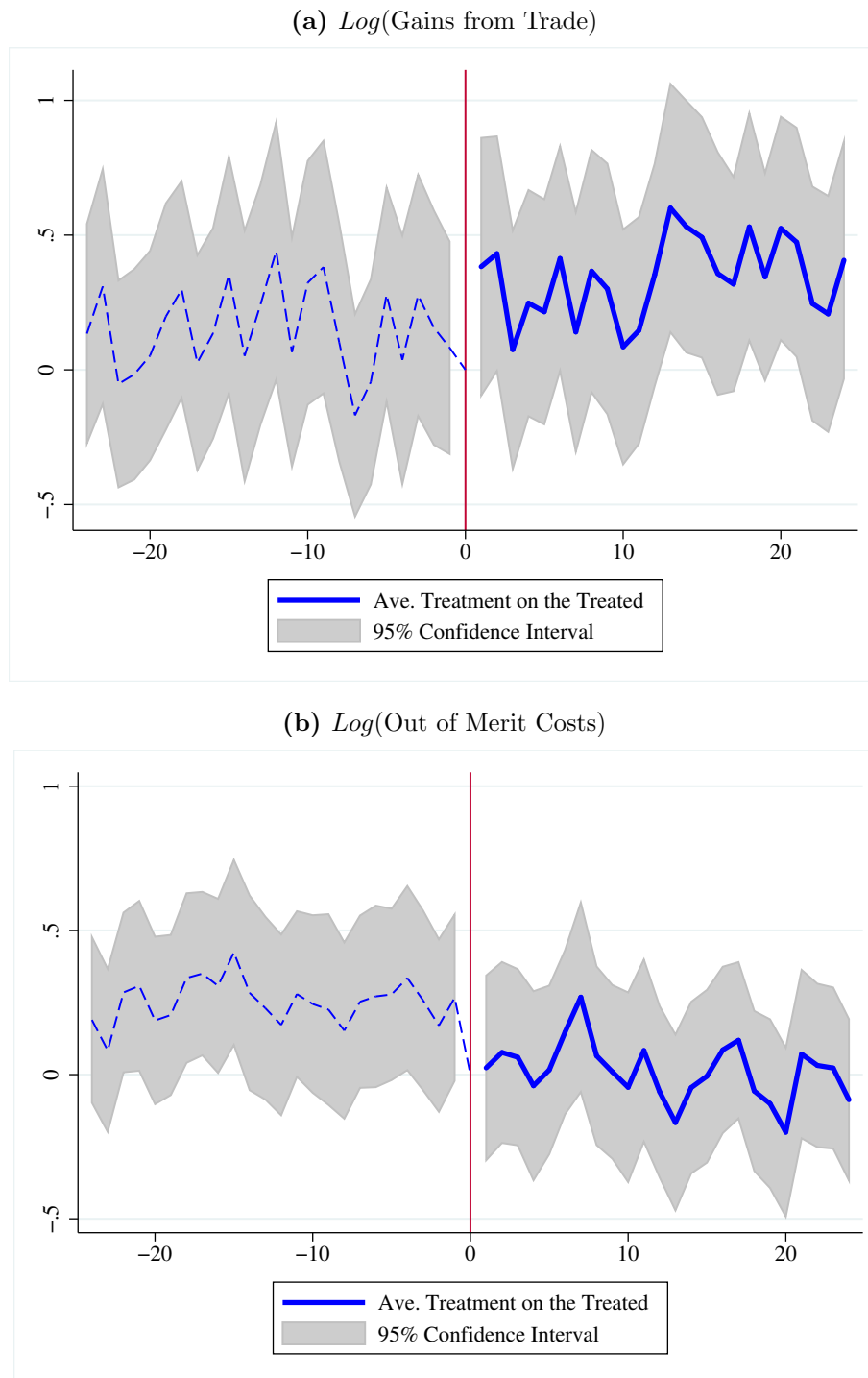
Note: All specifications include PCA and Region-Date-Hour Fixed Effects. Demand controls are PCA-specific. Standard errors clustered by PCA-Month in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

**Figure 11:** Treatment Effects by Months to Market: Welfare



Note: These figures are based on regressing logged outcomes on a set of indicator variables for each month until (after) the transition to market dispatch, PCA-specific controls for load, date-hour-region and PCA fixed effects. The month prior to treatment is normalized to zero. Confidence intervals are based on clustering at the PCA-month level. The presence of trends in these figures motivates the use of policy functions below.

**Figure 12:** Policy Function-Based Treatment Effects by Months to Market



Note: These figures are based on regressing the difference between logged outcomes and those predicted by the policy function on a set of indicator variables for each month until (after) the transition to market dispatch, PCA-specific controls for load, date-hour-region and PCA fixed effects. The month prior to treatment is normalized to zero. Confidence intervals are based on clustering at the PCA-month level.



## Heterogeneity over the Year

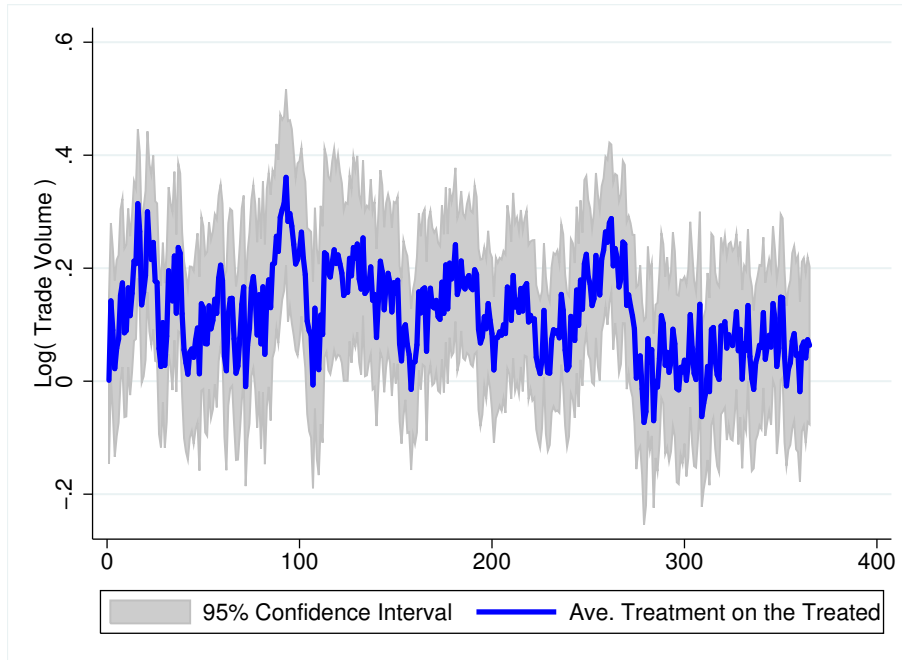
The richness of the data allow for the examination of heterogeneous treatment effects in order to better understand the forces driving the overall point estimate. In that spirit, Figures 13 and 14 interact the treatment variable with the day of the year, and plot the corresponding coefficients. For quantities, Figure 13 shows the strong complementarity between the two measures: The biggest reductions in out of merit generation occur during the low demand periods when utilities use to perform maintenance on their large units (and refuel nuclear-powered units). How do they manage to reduce these outage costs? Panel (a) shows that trade volumes increase during these periods in market areas. This indicates that markets keep utilities from favoring their own higher-cost units during maintenance, and instead coordinate supply of lower-cost power across PCAs. These results complement the prior findings of Davis and Wolfram (2012), who show that merchant nuclear units reduce their down-time. The timing of those divestitures largely precede the transition to market dispatch, so the results presented here should be interpreted as mostly in addition to those found previously. Furthermore, the decompositions of equations (2) and (3) make clear that there is no *a priori* reason that these effects would go hand-in-hand—a PCA that reduces its out of merit generation may find itself less reliant on outside generation sources.

For gains from trade, Figure 13 shows how fuel prices are not simply confounders, but also drivers of treatment effect heterogeneity. There are substantial increases in gains from trade during peak summer months even with smaller trade volumes because more expensive units' production is supplanted with traded generation. The overall treatment effects measured earlier are in fact weighted averages of quite large treatment effects during shoulder seasons and summer, but much smaller effects during the winter months.

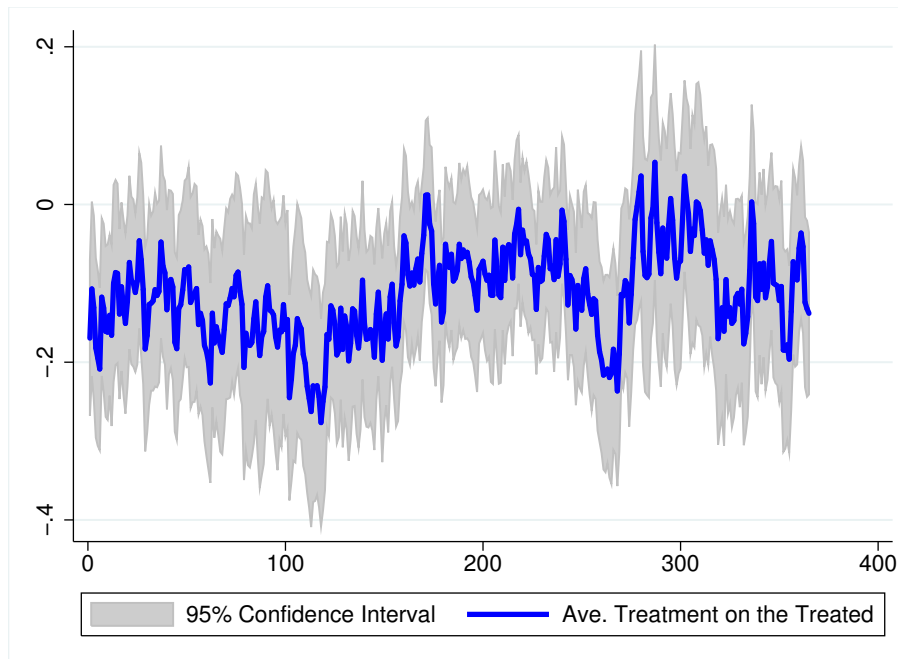
Panel (b) highlights the key opposing forces at play when switching to market dispatch: On one hand, generators have an increased incentive to ensure their low-cost generators are available for production. On the other hand, peak periods of demand create the potential to profitably exercise market power by taking an economical unit offline. Thus even though there is a reduction in the quantity of out of merit generation in 13.b, the reduction in out of merit costs is low relative to the value of offset generation (high during these peak periods). On net, diligent market monitoring has made these strategies of withholding production more difficult, and the effect overall is reduced out of merit costs throughout the year.

**Figure 13:** Treatment Effects by Day of Year: Quantities

(a)  $\text{Log}(\text{Trade Volume})$

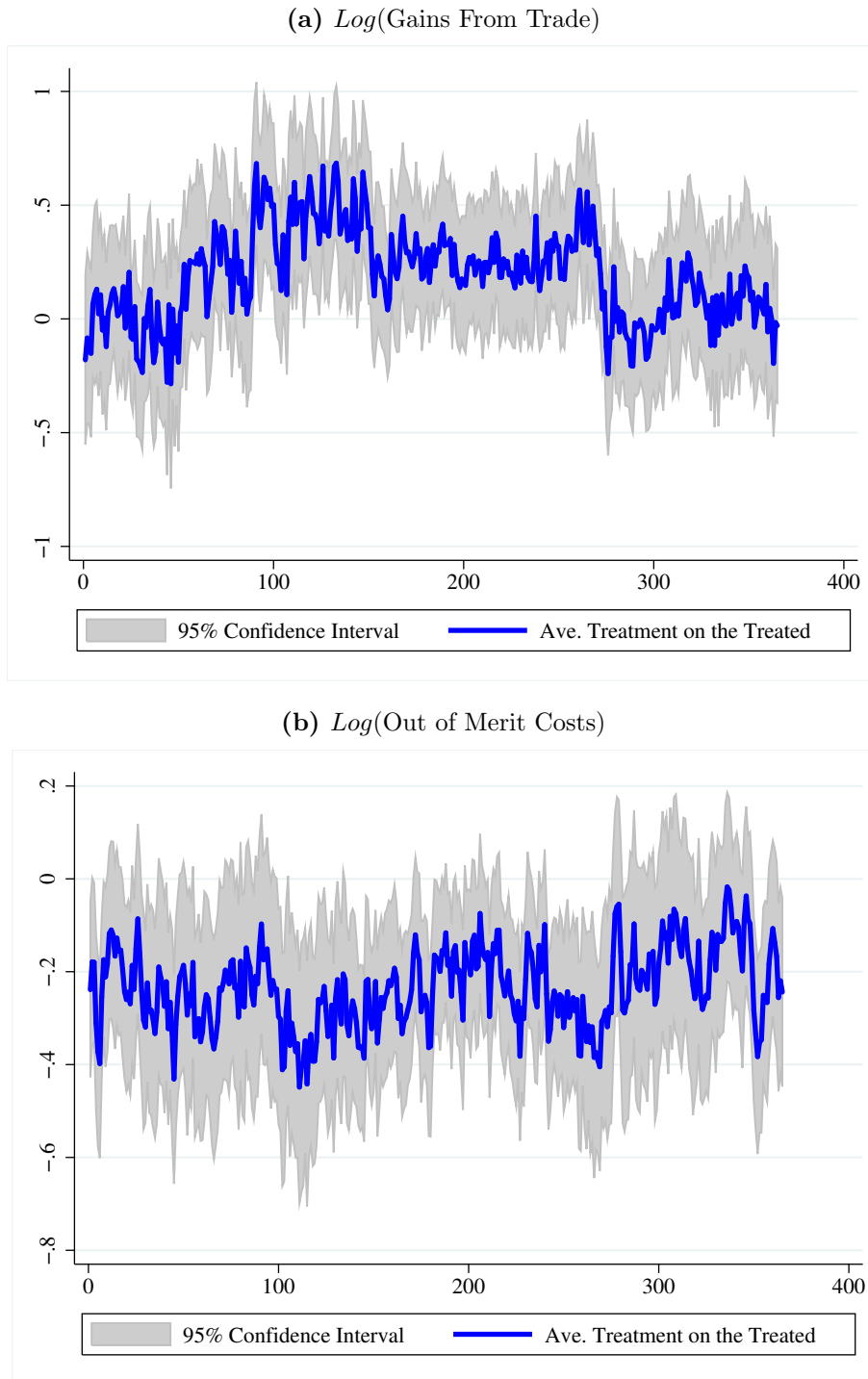


(b)  $\text{Log}(\text{MWh Out of Merit})$



Note: These figures are based on regressing logged outcomes on a set of indicator variables for each day of the year interacted with market dispatch, along with date-hour-region and PCA fixed effects. Confidence intervals are based on clustering at the PCA-month level.

**Figure 14:** Treatment Effects by Day of Year: Policy Function Estimates



Note: These figures are based on regressing the difference between logged outcomes and those predicted by the policy function on a set of indicator variables for each month until (after) the transition to market dispatch, PCA-specific controls for load, date-hour-region and PCA fixed effects. The month prior to treatment is normalized to zero. Confidence intervals are based on clustering at the PCA-month level.

## 7 Conclusion

In this paper I use the recent introduction of wholesale electricity markets in some areas as a natural experiment to evaluate the performance of markets relative to the policy-relevant counterfactual: centralized dispatch by a regulated private or government local monopolist. In constructing a fourteen year panel of hourly operations, I am able to infer gains from trade at any moment of time based on the amount of electricity being produced and consumed in an area, and the installed generating capacity that might have been used to equate local supply and demand. Observing hourly production at the generating unit level allows me to calculate the difference between actual production costs, and those that would have been realized if only the most economical (based on marginal fuel cost) units were utilized. I show how production costs can be decomposed in to these two measures, and estimate how the introduction of wholesale markets affected them. These effects are interpreted as the net impact of market power problems and improved coordination on production costs. I find that market-based dispatch has caused a roughly 20% increase in the gains from trade due to reallocated production across power control areas, while also reducing out of merit costs by 20%—a reduction in production costs of about \$3B per year.

While the estimated allocative efficiency improvements caused by market dispatch are substantial, they are likely part of a much bigger story. These short-run estimates are based on responses to institutional changes imposed on a grid that was built for reliability rather than massive trans-regional exchange. This inherently imposes an upper bound on the potential gains that might be observed with this estimation strategy, but is a constraint that may be relaxed over time as locational marginal prices reveal profitable transmission investments.

About 40% of electricity in the United States continues to be generated by plants called upon to operate based on the decision-making of the local balancing authority. Policymakers are therefore faced with the question of whether markets should be expanded or scaled-back. This is difficult to answer for a particular PCA, but aggregating over experiences thus far provides evidence on the balance between market failures and regulatory shortcomings. While market power is certainly a concern for market monitors (Wolak (2012) shows their work is critical), my results suggest the benefits realized by more efficient allocation of output though market-based dispatch have far outweighed such losses thus far.

## References

- Archsmith, James**, “Dam Spillovers: Direct costs and spillovers from environmental constraints on hydroelectric generation,” 2017.
- Borenstein, Severin**, “The Trouble with Electricity Markets: Understanding California’s Restructuring Disaster,” *The Journal of Economic Perspectives*, 2002, *16* (1), 191–211.
- , **James B. Bushnell**, and **Frank A. Wolak**, “Measuring Market Inefficiencies in California’s Restructured Wholesale Electricity Market,” *The American Economic Review*, 2002, *92* (5), 1376–1405.
- , – , and **Steven Stoft**, “The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry,” *The RAND Journal of Economics*, 2000, *31* (2), 294.
- , – , **Christopher R. Knittel**, and **Catherine D. Wolfram**, “Inefficiencies and Market Power in Financial Arbitrage: A Study of California’s Electricity Markets,” *The Journal of Industrial Economics*, 2008, *56* (2), 347–378.
- Breiman, Leo**, “Random forests,” *Machine Learning*, 2001, *45* (1), 5–32.
- , **Jerome Friedman**, **Charles J Stone**, and **Richard A Olshen**, *Classification and Regression Trees*, Chapman and Hall / CRC Press, 1984.
- Burlig, Fiona**, **Christopher Knittel**, **David Rapson**, **Mar Reguant**, and **Catherine Wolfram**, “Machine Learning from Schools about Energy Efficiency,” 2017.
- Bushnell, James B.**, **Erin T. Mansur**, and **Celeste Savaria**, “Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured US Electricity Markets,” *The American Economic Review*, 2008, *98* (1), 237–266.
- Cicala, Steve**, “When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation,” *American Economic Review*, 2015, *105* (1), 411–444.
- Cramton, Peter**, “Electricity Market Design: The Good, the bad, and the Ugly,” *Proceedings of the 36th Hawaii International Conference on System Sciences*, 2003.

- Cropper, Maureen L., Alexander Limonov, Malik Kabir, and Singh Anoop**, “Estimating the Impact of Restructuring on Electricity Generation Efficiency: The Case of the Indian Thermal Power Sector,” *NBER Working Paper*, 2011, 17383.
- Cullen, Joseph A.**, “Dynamic Response to Environmental Regulation in the Electricity Industry,” 2011.
- **and Erin T. Mansur**, “Inferring Carbon Abatement Costs in Electricity Markets: A Revealed Preference Approach using the Shale Revolution,” *American Economic Journal: Economic Policy*, (Forthcoming).
- Davis, Lucas W. and Catherine D. Wolfram**, “Deregulation, Consolidation, and Efficiency: Evidence from U.S. Nuclear Power,” *American Economic Journal: Applied Economics*, 2012, 4 (4), 194–225.
- **and Catherine Hausman**, “Market impacts of a nuclear power plant closure,” *American Economic Journal: Applied Economics*, 2016, 8 (2), 92–122.
- Energy Information Administration**, “Annual Energy Review,” Technical Report 2012.
- Fabrizio, Kira R., Nancy Rose, and Catherine D. Wolfram**, “Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency,” *The American Economic Review*, 2007, 97 (4), 1250–1277.
- Fowlie, Meredith L.**, “Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement,” *The American Economic Review*, 2010, 100 (3), 837–869.
- Graff Zivin, Joshua S., Matthew J. Kotchen, and Erin T. Mansur**, “Spatial and temporal heterogeneity of marginal emissions: Implications for electric cars and other electricity-shifting policies,” *Journal of Economic Behavior and Organization*, 2014, 107 (PA), 248–268.
- Green, Richard J and David M. Newberry**, “Competition in the British Electricity Spot Market,” *Journal of Political Economy*, 1992, 100 (5), 929–953.
- Hausman, Catherine and Ryan Kellogg**, “Welfare and Distributional Implications of Shale Gas,” *Brookings Papers on Economic Activity*, 2015.

- Holland, S.P. and E.T. Mansur**, “Is Real-Time Pricing Green?: the Environmental Impacts of Electricity Demand Variance,” *The Review of Economics and Statistics*, 2008, *90* (3), 550–561.
- Holland, Stephen P, Erin T Mansur, and Andrew J Yates**, “Are There Environmental Benefits from Driving Electric Vehicles ? The Importance of Local Factors,” *American Economic Review*, 2016, *106* (12), 3700–3729.
- Hortacsu, A and Steven L. Puller**, “Understanding Strategic Bidding in Multi-Unit Auctions: A Case Study of the Texas Electricity Spot Market,” *The RAND Journal of Economics*, 2008, *39* (1), 86–114.
- Ito, Koichiro and Mar Reguant**, “Blowin’ in the Wind: Sequential Markets, Market Power and Arbitrage,” *The American Economic Review*, 2016, *106* (7), 1921–57.
- Jha, Akshaya and Frank A. Wolak**, “Testing for Market Efficiency with Transactions Costs: An Application to Convergence Bidding in Wholesale Electricity Markets,” 2013.
- Joskow, Paul L.**, “Deregulation and Regulatory Reform in the U.S. Electric Power Sector,” in Clifford Winston, ed., *Deregulation of Network Industries: What’s Next?*, Washington, DC: AEI-Brookings Joint Center for Regulatory Studies, 2000, pp. 113–54.
- , “Market Imperfections versus Regulatory Imperfections,” *CESifo DICE Report*, 2010, *8* (3), 3–7.
- **and Alfred Edward Kahn**, “A Quantitative Analysis of Pricing Behavior in California’s Wholesale Electricity Market During Summer 2000,” *The Energy Journal*, 2002, *23* (4), 1–36.
- **and Richard Schmalensee**, “Markets for Power: An Analysis of Electrical Utility Deregulation,” *MIT Press Books*, 1988, *1*.
- Kahn, Alfred E.**, “Applications of Economics to an Imperfect World,” *American Economic Review*, 1979, *69* (2), 1–13.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer**, “Policy Prediction Problems,” *American Economic Review*, 2015, *105* (5), 491–495.

- Knittel, Christopher R., Konstantinos Metaxoglou, and Andre Trindade**, “Dash for Gas: The Sequel,” 2014.
- Linn, Joshua, Lucija Muehlenbachs, and Yushuang Wang**, “How Do Natural Gas Prices Affect Electricity Consumers and the Environment?,” 2014, pp. 1–42.
- Lucas, Robert E.**, “Economic Policy Evaluation: A Critique,” *Carnegie-Rochester Conference Series on Public Policy*, 1976, 1, 19–46.
- Mansur, Erin T.**, “Pricing Behavior in the Initial Summer of the Restructured PJM Wholesale Electricity Market,” 2001.
- , “Do Oligopolists Pollute Less? Evidence from a Restructured Electricity Market,” *The Journal of Industrial Economics*, 2007, 55 (4), 661–689.
- , “Measuring Welfare in Restructured Electricity Markets,” *The Review of Economics and Statistics*, 2008, 90 (2), 369–386.
- **and Matthew W. White**, “Market Organization and Efficiency in Electricity Markets,” 2012.
- Meinshausen, Nicolai**, “Quantile Regression Forests,” *Journal of Machine Learning Research*, 2006, 7, 983–999.
- PRISM Climate Group**, “PRISM Gridded Climate Data,” 2004.
- Puller, Steven L.**, “Pricing and Firm Conduct in California’s Deregulated Electricity Market,” *The Review of Economics and Statistics*, 2007, 89 (1), 75–87.
- Reguant, Mar**, “Complementary Bidding Mechanisms: An Application to Electricity Markets,” *The Review of Economic Studies*, 2014, 81 (4), 1708–1742.
- Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver**, “Kidney Exchange,” *The Quarterly Journal of Economics*, 2004, 119 (2), 457–488.
- Rubin, Donald B.**, “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology*, 1974, 66 (5), 688–701.
- Ryan, Nicholas**, “The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market,” 2013, pp. 1–59.
- Scornet, Erwan, Gérard Biau, and Jean-Philippe Vert**, “Consistency of random forests,” *The Annals of Statistics*, 2015, 43 (4), 1716–1741.



- Wager, Stefan and Susan Athey**, “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests,” 2016.
- Wolak, Frank A.**, “An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market,” *International Economic Journal*, 2000, 14 (2), 1–39.
- , “Regulating Competition in Wholesale Electricity Supply,” in Nancy Rose, ed., *Economic Regulation and Its Reform: What Have We Learned?*, NBER, 2012, chapter 5.
- **and Robert H Patrick**, “The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market,” 1997.
- Wolfram, Catherine D.**, “Measuring Duopoly Power in the British Electricity Spot Market,” *American Economic Review*, 1999, 89 (4), 805–826.
- , “The Efficiency of Electricity Generation in the U.S. After Restructuring,” in James M. Griffin and Steven L. Puller, eds., *Electricity Deregulation: Choices and Challenges.*, Chicago: University of Chicago Press, 2005.
- Wright, Marvin N. and Andreas Ziegler**, “ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R,” *Journal of Statistical Software*, 2016, *in Press*.

## A Data Appendix: For Online Publication

### A.1 Power Control Area Definitions

The definition of a Power Control Area (PCA), or Balancing Authority (BA) is somewhat flexible, varying across regions, regulatory agencies, and over time. For the purposes of this paper, I am interested in identifying decisionmaking units responsible for allocating production to generating units to keep supply and demand for electricity in balance at all moments in time. As a practical matter, the method I use to measure the value of reservoir hydropower (discussed below) relies on the opportunity cost of water based on offsetting fossil power. I therefore classify PCAs as those recognized as such in load reporting by FERC in 1999, also reporting control of fossil-fired units based on a combination of reporting in FERC Form 714 (Part II, Schedule 1), and the 1999 configuration of the grid based on EPA’s eGRID database.<sup>20</sup> This results in the consolidation of a number of “planning areas” that report their own load, though they do not dispatch plants, as well as a few hydropower-only PCAs in the Pacific Northwest (see Tables A.1, A.2, and A.3). I use county-level approximations of these 1999 PCA configurations when using demographic, meteorological, and economic variables to predict load (see Section A.2). To construct these maps, I begin with the 1999 EIA Form 861, “Annual Electric Power Industry Report”, which connects local utilities to PCAs, and reports the counties in which respondent utilities have generation equipment. I then use individual service territory maps (via internet search) to refine these boundaries.

Much like the Neighborhood Change Database, the goal is to create a time-invariant characterization of the grid, which has indeed changed over time. New generation units, for example, are associated with their contemporary ISO rather than historical PCA. To determine the 1999 PCA in which new generation *would have* been located (ignoring differential investment issues), I first use local utility association: many of these utilities are unchanged in spite of changes to the bulk electricity system. If that utility belonged to a 1999 PCA, the plant inherits that association. If no other information is available, the 1999 PCA maps are used to associate new generation with historical areas.

---

<sup>20</sup>eGRID is used as the starting point, then corrections are made by hand based on FERC reporting because these forms are only available as (occasionally handwritten) pdfs of plant names, rather than EIA facility codes.

## A.2 Load Data

Hourly data on electricity usage (load) are compiled from a combination of the Federal Energy Regulatory Commission (FERC) Form 714, local system operators, and the North American Electricity Reliability Corporation (NERC), depending on data availability. While this data, in theory, are publicly-available from a straightforward download from the FERC site, this is emphatically not the case in practice. Until 2006, there was no required submission format for hourly load data, so that each PCA's data might be submitted in anything from an Excel file to free-form text, often without a codebook. In addition, there is no standard procedure for accounting for daylight savings time: some areas ignore it completely, others report zero at the start and double-report the final hour, etc. Annual reports are missing altogether for some PCA-years, or are reported as part of the load of an adjacent area (again, often without documentation). A number of smaller areas that do not own generation (and are therefore *planning* areas, rather than *control* areas) are combined with the neighboring PCA that conducts dispatch. To avoid PCAs composed entirely of estimated hydro generation, a handful of areas in the Pacific Northwest are combined as well. Areas that join an ISO often have their load included in the ISO total, and may not be available as a single PCA.

When missing, hourly load data are estimated separately for each PCA using LASSO to uncover the best functional form in a disciplined manner. One benefit of consumers' insulation from electricity market conditions is that electricity load can be estimated extremely well as a function of time (of year, week and day), population, weather, and economic conditions. The day of week/year variables used in prediction are a set of trigonometric functions with varying periodicity over the course of the week and year to account for regular calendar fluctuations. Temperature variables measure heating and cooling degrees (degrees above or below  $65^{\circ}F$ ) on the daily minimum, maximum, and average temperatures, as well as relative humidity and precipitation. This data come from PRISM Climate Group (2004), and collapses county-level data with population weights for PCA-wide measures. Economic data include unemployment rates as well as electricity-intensive employment in manufacturing and mining sectors aggregated from the county level to the approximate footprint of the PCA in 1999 as with the meteorological data.

These variables are used in a LASSO estimation procedure to avoid over-fitting by including a regularization term in the standard OLS procedure that sets less important predictors to exactly zero rather than fit on noise. When estimated using data for even years, it produces estimates that have a mean absolute deviation of less than

5% when validated against odd years.<sup>21</sup>

### **Western Interconnection**

Load data for the Western Interconnection come from both FERC and the Western Electricity Coordinating Council (WECC), depending upon availability. PCAs as of 1999, and constituent load-reporting areas are reported in Table A.1. The abundance of hydropower requires the consolidation of a number of Public Utility Districts in Washington and Irrigation Districts in California to arrive at a level of aggregation such that reservoir power is offsetting fossil power. In addition, there is relatively inconsistent reporting of load in the former Southern California Edison territory, though total load from the California Independent System Operator (CAISO) is well-reported, as is load from the other territories in the CAISO footprint. Southern California Edison load is therefore calculated from the remaining CAISO load after subtracting off load from Pacific Gas & Electric, San Diego Gas & Electric, and their respective constituent load areas.

### **Texas Interconnection**

The Electric Reliability Council of Texas (ERCOT) is a separate interconnection that consolidated ten PCAs in to a single market on 31 July, 2001. After a period of only reporting total ERCOT load in 2001 and 2002, the ISO began reporting load by eight “weather zones” that do not cleanly overlap with the original PCAs. The ERCOT total is consistently reported throughout the sample period. I therefore run LASSO using the ERCOT total and the population-weighted characteristics for the entire ISO, then use the ISO-derived coefficients projected upon the PCA-level characteristics to predict PCA-level load. Final estimates are scaled by the ratio of observed ERCOT load to the sum of predicted PCA loads to ensure that the totals match those observed in the data. This method delivers estimates of load in 1999 and 2000 for the original PCAs that have an absolute mean deviation from the true loads of about 6%, in line with the out of sample estimates delivered by estimating fixed footprints to years without load data.

### **Eastern Interconnection**

Load data for the Eastern Interconnection varies in the consistency of reporting. The Northeastern ISOs in New York and New England did not consolidate multiple PCAs upon transition to markets, but simply changed the method for allocating output over

---

<sup>21</sup>Using only PCA-specific means yields an error of about 20%, which is reduced to 15% by using PCA-hour means, and no other explanatory variables.

a fixed territory-load reporting is consistent throughout. PCAs in the Pennsylvania-Jersey-Maryland (PJM) market have delegated load reporting to the ISO, but PJM has helpfully preserved the original footprints as the basis for more detailed reporting available through their website. The Southwest Power Pool (SPP) has also made hourly load data available by original PCAs in spite of aggregate reporting to FERC. This has, unfortunately, not been the case for the Midwest ISO (MISO), which declines to release the disaggregated data which was previously publicly available before the ISO took over load reporting in 2009. Instead, the most disaggregated load available from MISO is broken down by three large regions spanning many former PCAs each. Fortunately MISO began market dispatch three years before taking over load reporting, so demand is largely observed through the transition to markets. For the areas without directly-reported load data from 2009-2012, it is estimated via LASSO. Adding up the predicted loads and comparing to the regional totals to which they roughly correspond lines up reasonably well, including predicting the drop in electricity demand in 2009 due to the Great Recession based on pre-recession data.

**Table A.1: Power Control Areas in the Western Interconnection**

| 1999 PCA                          | Constituent Load                      | Unreported Periods | ISO   | Market Date  |
|-----------------------------------|---------------------------------------|--------------------|-------|--------------|
| Avista                            | Avista Corp                           | 1999               |       |              |
| Arizona Public Service            | Arizona Public Service Co             |                    |       |              |
|                                   | WAPA Lower Colorado River             | 2001-2007          |       |              |
| Bonneville Power Authority        | Bonneville Power Authority            |                    |       |              |
|                                   | PUD 1 of Chelan County                | 2011-2012          |       |              |
|                                   | PUD 1 of Douglas County               |                    |       |              |
|                                   | PUD 2 of Grant County                 |                    |       |              |
| El Paso Electric                  | El Paso Electric Company              |                    |       |              |
| Imperial Irrigation District      | Imperial Irrigation District          |                    |       |              |
| Idaho Power                       | Idaho Power Company                   |                    |       |              |
| Los Angeles Dept Of Water & Power | Los Angeles, City Of                  |                    |       |              |
| Montana Power                     | Montana Power Company                 |                    |       |              |
| Nevada Power                      | Nevada Power Company                  |                    |       |              |
| Pacificorp                        | Pacificorp                            |                    |       |              |
| Pacific Gas & Electric            | Pacific Gas & Electric                | 2011-2012          | CAISO | 1 April 1998 |
|                                   | Modesto Irrigation District           |                    | CAISO | 1 April 1998 |
|                                   | WAPA Sierra Nevada Region             |                    | CAISO | 1 April 1998 |
|                                   | City of Redding                       |                    | CAISO | 1 April 1998 |
|                                   | Sacramento Municipal Utility District |                    | CAISO | 1 April 1998 |
| Portland General Electric         | Portland General Electric Co          |                    |       |              |
| Public Service Co Of New Mexico   | Public Service Co Of New Mexico       |                    |       |              |
| Public Service Co Of Colorado     | Public Service Co Of Colorado         |                    |       |              |
| Puget Sound Energy                | Puget Sound Energy                    |                    |       |              |
|                                   | Seattle Department of Lighting        |                    |       |              |
|                                   | Tacoma Power                          |                    |       |              |
| Southern California Edison        | Southern California Edison            |                    | CAISO | 1 April 1998 |
|                                   | City of Vernon                        |                    | CAISO | 1 April 1998 |
|                                   | CA Dept of Water Resources            |                    | CAISO | 1 April 1998 |
|                                   | City of Anaheim                       |                    | CAISO | 1 April 1998 |
|                                   | City of Santa Clara                   |                    | CAISO | 1 April 1998 |
|                                   | City of Riverside                     |                    | CAISO | 1 April 1998 |
|                                   | City of Pasadena                      |                    | CAISO | 1 April 1998 |
| San Diego Gas & Electric          | San Diego Gas & Electric              |                    | CAISO | 1 April 1998 |
| Sierra Pacific Power              | Sierra Pacific Power Co               |                    |       |              |
| Salt River Project                | Salt River Project                    |                    |       |              |
| Tucson Electric Power             | Tucson Electric Power Co              |                    |       |              |
| WAPA Colorado-Missouri            | WAPA Colorado-Missouri                | 1999               |       |              |

**Table A.2: Power Control Areas in the Texas Interconnection**

| 1999 PCA                         | Constituent Load                      | ISO   | Market Date  |
|----------------------------------|---------------------------------------|-------|--------------|
| Central And South West Services  | Central And South West Services (AEP) | ERCOT | 31 July 2001 |
|                                  | South Texas Electric Cooperative      | ERCOT | 31 July 2001 |
|                                  | Brownsville Public Utilities Board    | ERCOT | 31 July 2001 |
| Lower Colorado River Authority   | Lower Colorado River Authority        | ERCOT | 31 July 2001 |
|                                  | Austin Energy                         | ERCOT | 31 July 2001 |
| Reliant Energy                   | Reliant Energy                        | ERCOT | 31 July 2001 |
| San Antonio Public Service Board | San Antonio Public Service Board      | ERCOT | 31 July 2001 |
| Texas Municipal Power Pool       | Texas Municipal Power Pool            | ERCOT | 31 July 2001 |
| TXU Energy                       | Texas Utilities                       | ERCOT | 31 July 2001 |
|                                  | Texas-New Mexico Power Company        | ERCOT | 31 July 2001 |

### Table A.3: Power Control Areas in the Eastern Interconnection

| NERC Region | 1999 PCA                              | Constituent Load                        | Unreported Periods                | ISO   | Market Date      |
|-------------|---------------------------------------|---|-----------------------------------|-------|------------------|
| ECAR        | American Electric Power               | American Electric Power                 |                                   | PJM   | 1 October 2004   |
| ECAR        |                                       | Buckeye Power                           |                                   | PJM   | 1 October 2004   |
| ECAR        |                                       | American Municipal Power - Ohio         |                                   | PJM   | 1 October 2004   |
| ECAR        | Allegheny Power Service               | Allegheny Power Service                 |                                   | PJM   | 1 April 2002     |
| ECAR        | Big Rivers Electric                   | Big Rivers Electric Corp                | 2011-2012                         | MISO  | 1 December 2010  |
| ECAR        | Cinergy                               | Cinergy                                 | 2009-2012                         | MISO  | 1 April 2005     |
| ECAR        | Consumers Energy                      | Consumers Energy                        | 2009-2012                         | MISO  | 1 April 2005     |
| ECAR        | Detroit Edison                        | Detroit Edison Co                       | 2009-2012                         | MISO  | 1 April 2005     |
| ECAR        | Duquesne Light                        | Duquesne Light Company                  |                                   | PJM   | 1 January 2005   |
| ECAR        | Dayton Power & Light Co               | Dayton Power & Light Co                 |                                   | PJM   | 1 October 2004   |
| ECAR        | East Kentucky Power Coop              | East Kentucky Power Coop                |                                   |       |                  |
| ECAR        | Firstenergy                           | Firstenergy                             | 2009-2012                         | MISO  | 1 April 2005     |
| ECAR        | Hoosier Energy                        | Hoosier Energy                          | 2006-2012                         | MISO  | 1 April 2005     |
| ECAR        | Indianapolis Power & Light            | Indianapolis Power & Light              | 2009-2012                         | MISO  | 1 April 2005     |
| ECAR        | Louisville Gas & Electric             | Louisville Gas & Electric               |                                   | MISO  | 1 April 2005     |
| ECAR        | Northern Indiana Public Service       | Northern Indiana Public Service         | 2009-2012                         | MISO  | 1 April 2005     |
| ECAR        | Southern Indiana G & E                | Southern Indiana G & E                  | 2006-2012                         | MISO  | 1 April 2005     |
| FRCC        | Florida Municipal Power Agency        | Florida Municipal Power Agency          |                                   |       |                  |
| FRCC        |                                       | Orlando Utilities                       | 2003, 2004                        |       |                  |
| FRCC        |                                       | City of Lakeland                        | 2004                              |       |                  |
| FRCC        | Florida Power                         | Florida Power Corporation               |                                   |       |                  |
| FRCC        | Florida Power & Light                 | Florida Power & Light                   |                                   |       |                  |
| FRCC        | Gainesville Regional Utilities        | City Of Gainesville                     |                                   |       |                  |
| FRCC        | Jacksonville Electric Authority       | Jacksonville Electric Authority         |                                   |       |                  |
| FRCC        | Seminole Electric Coop                | Seminole Electric Coop                  | 2003                              |       |                  |
| FRCC        | City Of Tallahassee                   | City Of Tallahassee                     |                                   |       |                  |
| FRCC        | Tampa Electric                        | Tampa Electric Company                  | 2004                              |       |                  |
| MAAC        | Pennsylvania-Jersey-Maryland          | Pennsylvania-Jersey-Maryland            |                                   | PJM   | 1 April 1997     |
| MAIN        | Alliant East                          | Alliant East                            | 2009-2012                         | MISO  | 1 April 2005     |
| MAIN        | Ameren                                | Ameren                                  | 2009-2012                         | MISO  | 1 April 2005     |
| MAIN        |                                       | Columbia Water & Light                  | 2004,2009-2012                    | MISO  | 1 April 2005     |
| MAIN        | Commonwealth Edison                   | Commonwealth Edison Co                  |                                   | PJM   | 1 May 2004       |
| MAIN        | Central Illinois Light                | Central Illinois Light Co               | 2007-2012                         | MISO  | 1 April 2005     |
| MAIN        | Illinois Power                        | Illinois Power                          | 2009-2012                         | MISO  | 1 April 2005     |
| MAIN        | Southern Illinois Power Coop          | Southern Illinois Power Coop            | 2009-2012                         | MISO  | 1 April 2005     |
| MAIN        | Springfield (IL) Water Light & Power  | City of Springfield, IL                 | 2009-2012                         | MISO  | 1 April 2005     |
| MAIN        | Wisconsin Energy                      | Wisconsin Electric Power                | 2009-2012                         | MISO  | 1 April 2005     |
| MAIN        | Wisconsin Public Service              | Wisconsin Public Service                | 2004, 2009-2012                   | MISO  | 1 April 2005     |
| MAIN        |                                       | Madison Gas & Electric Co               | 2004, 2009-2012                   | MISO  | 1 April 2005     |
| MAPP        | Alliant West                          | Alliant West                            | 2009-2012                         | MISO  | 1 April 2005     |
| MAPP        | WAPA Upper Missouri                   | WAPA Upper Missouri East Basin          | 2000, 2001, 2004, 2005            | MISO  | 1 April 2005     |
| MAPP        |                                       | Basin Electric Power Cooperative        | 2000, 2001, 2004, 2005            | MISO  | 1 April 2005     |
| MAPP        |                                       | WAPA Upper Missouri West Basin          | 2000, 2001, 2004, 2005            | MISO  | 1 April 2005     |
| MAPP        | Dairyland Power Coop                  | Dairyland Power Coop                    | 2000, 2001, 2004, 2011, 2012      | MISO  | 1 June 2010      |
| MAPP        | Great River Energy                    | Great River Energy                      | 2000-2005, 2009-2012              | MISO  | 1 April 2005     |
| MAPP        | Midamerican Energy                    | Midamerican Energy                      | 2010-2012                         | MISO  | 1 September 2009 |
| MAPP        |                                       | Muscatine Power & Water                 | 2000, 2001, 2004, 2006-2012       | MISO  | 1 September 2009 |
| MAPP        | Minnesota Power & Light               | Minnesota Power & Light                 | 2000-2004, 2009-2012              | MISO  | 1 April 2005     |
| MAPP        | Nebraska Public Power District        | Nebraska Public Power Dist              | 2004                              | SPP   | 1 April 2009     |
| MAPP        |                                       | Lincoln Electric System                 | 2004                              | SPP   | 1 April 2009     |
| MAPP        | Northern States Power                 | Northern States Power Co                | 2001, 2004, 2009-2012             | MISO  | 1 April 2005     |
| MAPP        |                                       | Southern MN Municipal Power             | 1999, 2000, 2004, 2006, 2009-2012 | MISO  | 1 April 2005     |
| MAPP        | Omaha Public Power District           | Omaha Public Power District             | 2000, 2001, 2004                  | MISO  | 1 April 2005     |
| MAPP        | Otter Tail Power                      | Otter Tail Power                        | 2001, 2004, 2009-2012             | MISO  | 1 April 2005     |
| MAPP        |                                       | Minnkota Power Cooperative              | 1999-2001, 2004                   | MISO  | 1 April 2005     |
| NPCC        | New England Power Pool                | New England Power Pool                  |                                   | NEISO | 1 May 1999       |
| NPCC        | New York Power Pool                   | New York Power Pool                     |                                   | NYISO | 18 November 1999 |
| SERC        | Alabama Electric Cooperative          | Alabama Electric Cooperative            |                                   |       |                  |
| SERC        | Associated Electric Cooperative       | Associated Electric Cooperative         |                                   |       |                  |
| SERC        | Carolina Power & Light                | Carolina Power & Light                  |                                   |       |                  |
| SERC        | Duke Energy                           | Duke Energy                             |                                   |       |                  |
| SERC        | South Carolina Electric & Gas         | South Carolina Electric & Gas           |                                   |       |                  |
| SERC        | South Carolina Pub Serv Auth          | South Carolina Public Service Authority |                                   |       |                  |
| SERC        | South Mississippi Electric Power      | South Mississippi Electric Power        | 2001                              |       |                  |
| SERC        | Southern                              | Southern Co                             |                                   |       |                  |
| SERC        |                                       | Oglethorpe Power                        |                                   |       |                  |
| SERC        | Tennessee Valley Authority            | Tennessee Valley Authority              |                                   |       |                  |
| SERC        | Dominion Virginia Power               | Dominion Virginia Power                 |                                   | PJM   | 1 May 2005       |
| SPP         | CLECO                                 | Central Louisiana Electric Co           | 2002                              | SPP   | 1 February 2007  |
| SPP         |                                       | Lafayette Utility System                |                                   | SPP   | 1 February 2007  |
| SPP         | Empire District Electric              | Empire District Electric                |                                   | SPP   | 1 February 2007  |
| SPP         | Entergy                               | Entergy                                 |                                   |       |                  |
| SPP         | Grand River Dam Authority             | Grand River Dam Authority               | 2002                              | SPP   | 1 February 2007  |
| SPP         | Kansas City Power & Light             | Kansas City Power & Light               | 2002, 2006                        | SPP   | 1 February 2007  |
| SPP         |                                       | City of Independence                    | 2002, 2006                        | SPP   | 1 February 2007  |
| SPP         | Kansas City Board of Public Utilities | Kansas City Board of Public Utilities   | 2002                              | SPP   | 1 February 2007  |
| SPP         | Louisiana Energy & Power Authority    | Louisiana Energy & Power Auth           |                                   |       |                  |
| SPP         |                                       | Louisiana Generating                    | 2000, 2007-2012                   |       |                  |
| SPP         | Aquila Networks - MPS                 | Missouri Public Service Co              |                                   | SPP   | 1 January 2010   |
| SPP         | Oklahoma Gas & Electric Co            | Oklahoma Gas & Electric Co              | 2002                              | SPP   | 1 February 2007  |
| SPP         | Sunflower Electric Cooperative        | Sunflower Electric Cooperative          |                                   | SPP   | 1 February 2007  |
| SPP         | Public Service of OK (SWEPCO)         | Public Service of OK (SWEPCO)           | 2002, 2006                        | SPP   | 1 February 2007  |
| SPP         | Southwestern Power Admin              | Southwestern Power Admin                |                                   | SPP   | 1 February 2007  |
| SPP         |                                       | Southwestern Public Service             | 2002                              | SPP   | 1 February 2007  |
| SPP         |                                       | Golden Spread Electric Cooperative      | 2003                              | SPP   | 1 February 2007  |
| SPP         | Western Farmers Elec Coop             | Western Electric Farmers Coop           | 2002                              | SPP   | 1 February 2007  |
| SPP         |                                       | Oklahoma Municipal Power Authority      | 2000, 2001, 2003-2005             | SPP   | 1 February 2007  |
| SPP         | Aquila Networks - WPK                 | Aquila - WestPlains                     | 2007-2012                         | SPP   | 1 February 2007  |
| SPP         | Western Resources                     | Western Resources                       | 2002                              | SPP   | 1 February 2007  |

### A.3 Generation

#### EPA Continuous Emissions Monitor System (CEMS) Data

Roughly 90% of fossil-powered generation in the United States from 1999-2012 is reported at the boiler-hour level in the EPA CEMS data. While reporting for the largest units begins in 1996, comprehensive reporting does not begin until 1999. This data system was designed to monitor emissions for compliance with  $NO_x$  and  $SO_2$  programs of the 1990 Amendments to the Clean Air Act. I adjust for net-to-gross ratios, as CEMS reports gross generation for unit  $i$  in hour  $t$ , but power station usage (typically about 6%) must be subtracted to measure how much power is being sent to the grid. The adjustment is a unit-level version of that used by Cullen and Mansur (n.d.) to measure hourly net generation, except at the interconnection-fuel level, rather than generating unit:

$$\text{Net Generation}_{it} = \frac{\text{EIA Net Generation}_{im}}{\sum_{t \in m} \text{CEMS Gross Generation}_{it}} * \text{CEMS Gross Generation}_{it}$$

The hourly gross load data from CEMS are merged to the monthly EIA data on net generation and heat rate (at the unit level from EIA-767 when possible, otherwise at the plant-prime mover level from EIA-906), and scaled by the ratio of monthly net generation to monthly gross load. This accomplishes two tasks: First, it ensures output represents net generation at the hourly level, smoothing start-up and ramping costs over the month. Second, a number of units (especially Combined Cycle units) only include the steam portion of the unit in CEMS, leaving the generation from the second cycle unreported. This scaling treats the (hourly) unreported generation as dispatched at the same rate as the main unit. Figure A.1 plots the kernel density estimates of the net-to-gross ratios of CEMS units. While a density of the scaling is overwhelmingly concentrated at about 94% (the gross-to-net scale), combined cycle units feature a bimodal distribution with a second (much smaller) peak around 1.4 (reflecting the contribution of the unmatched cycles). The top and bottom percentile of scales is trimmed for outliers (outside of 0.2 and 2), and estimated scales based on a regression of observable unit characteristics are used instead.



For heat rate: top and bottom percent of heat rates are trimmed for outliers (outside of 6 and 100), as well as monthly heat rates that are more than three standard deviations from the unit’s mean heat rate (2% of unit-months). Estimated heat rates are used for these outliers, as well as in months with zero net generation. Estimation allows for unit-level fixed effects (which captures most of the variance), as well as flexible unit-specific trends to account for decaying heat rates with unit age.<sup>22</sup>

### Nuclear Generation

Daily output from nuclear-powered units is reported to the Nuclear Regulatory Commission (NRC) as a share of potential output for each generating unit. The exact potential output is calculated by taking the ratio of the monthly total net generation reported on EIA Form 906 to the sum of these daily potential outputs:<sup>23</sup>

$$\text{Net Generation}_{it} = \frac{\text{EIA Net Generation}_{im}}{\sum_{t \in m} \text{NRC Share of Capacity}_{it}} * \text{NRC Share of Capacity}_{it}$$

Output levels are determined by multiplying the calculated potential output by daily share of output generated, and distributing the generation over the hours of the day. There is minimal potential for error in this last step because nuclear units are typically running at maximum capacity, down for maintenance, or transitioning between the two over the course of days.

### Hydro Generation

Monthly hydro generation is reported on Form EIA-906. I use discharge and/or streamflow data to distribute this aggregated generation across the hours of the month. Inquiries at individual hydropower administrators yielded exact hourly turbine discharge numbers. These sources include the Tennessee Valley Authority, U.S. Army Corps of Engineers, and U.S. Bureau of Reclamation. Over one third of hy-

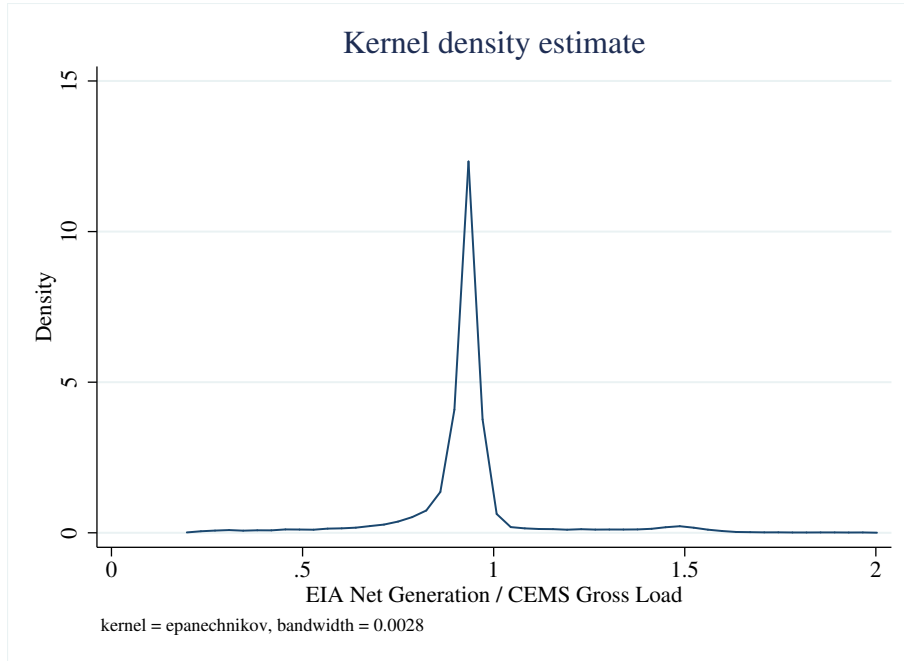
---

<sup>22</sup>Such effects are small: these regressions show heat rates tend to rise about 0.2% with each year of operation, particularly for smaller units.

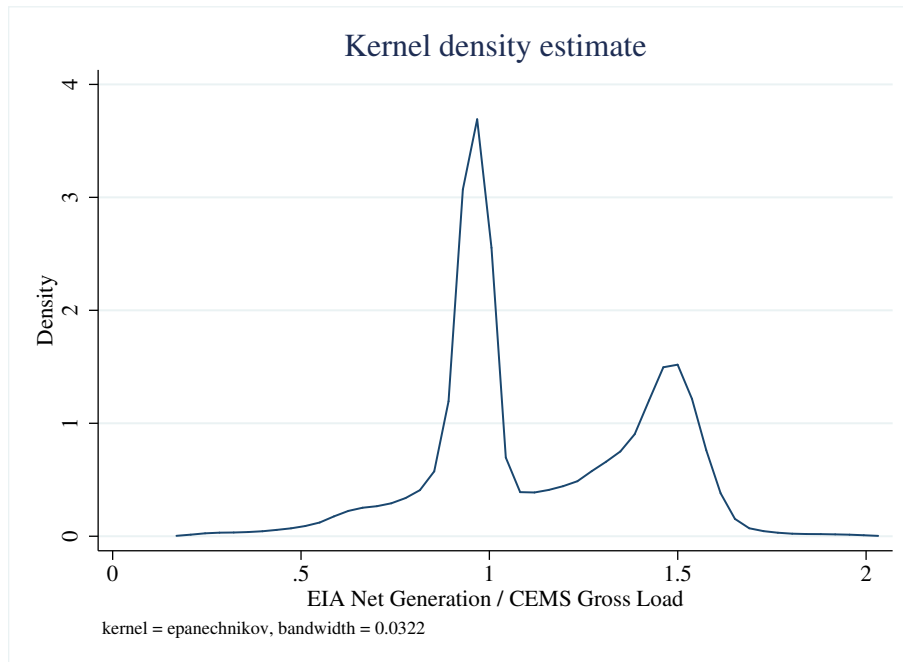
<sup>23</sup>Note that the ratio of monthly net generation to the sum of capacity shares is equal to the unit’s capacity:  $\text{Capacity}_{im} = \frac{\text{EIA Net Generation}_{im}}{\sum_{t \in m} \text{NRC Share of Capacity}_{it}}$

**Figure A.1:** Net-Gross Scaling of CEMS Units with Monthly Data

(a) All CEMS units



(b) Combined Cycle Units in CEMS



dropower generation is collected from these sources.

When exact numbers were not available, I infer hourly generation from streamflow data collected by the U.S. Geological Survey's streamgage network. To do so, I use plant coordinates from EIA to locate generators along the National Hydrography Database's stream network. I then use network analysis software to identify the nearest downstream gage that collects streamflow data (in many cases the dam's discharge monitor *is* the USGS streamflow recorder). I then allocate the month's generation to hours based on the share of monthly streamflow released in that hour in the same spirit as the generation allocation methods described above.

### **Fossil Generation Reported Monthly**

About 10% of generation from fossil-powered units does not appear in the CEMS data. For these units, I use monthly data at the unit level from EIA-767 (mostly small steam-powered boilers), or plant-prime mover level from EIA-906 (mostly small gas turbines and internal combustion generators). To allocate this production to hours of the month, monthly generation is allocated over hours by ranking the hours of the month by load, and producing at maximum capacity in the highest load hours up to the total reported monthly generation (or analogously for annual generation if that is the level of EIA-906 reporting).

For example, if a turbine with a capacity of 10MW reports on Form 906 that it produced 10MWh in a month, it is assumed that it only produced in the hour of maximum load that month. If it produced 50MW, it is assumed to have produced 10MWh in the five highest demand hours that month, and was idle otherwise. This approach is motivated by demand sweeping through the merit order, so that generating units are only briefly marginal, and therefore typically producing at maximum capacity, if at all. In this case, generation is proportional to the number of hours in which demand is sufficiently high for the unit to be "in" the merit order.

This algorithm is applied consistently throughout the sample period, and is invariant to the institutions used to allocate production. The share of production allocated in this manner is also stable over time.

## Other Generation

Monthly wind generation by farm is reported in EIA-906. I merge this data with hourly wind speed data from nearby weather stations, as reported in National Oceanic and Atmospheric Administration's (NOAA) Integrated Surface Database. I estimate the potential output of the farm based on a cut-in wind speed of 3 m/s, a cut-out speed of 20 m/s, a rated wind speed of 10 m/s (at which point the farm produces at maximum capacity), and an increasing cubic between the cut-in and rated wind speeds. I then use predicted power to distribute the observed monthly total over the month:

$$\text{Net Generation}_{it} = \frac{\text{EIA Net Generation}_{im}}{\sum_{t \in m} \text{Predicted Wind Generation}_{it}} * \text{Predicted Wind Generation}_{it}$$

For areas that report hourly wind generation in the ISO footprint (ERCOT since 2007, SPP and MISO since 2008), I use the hourly analog for wind farms in the ISO  $n$ 's footprint:

$$\text{Net Generation}_{it} = \frac{\text{ISO Wind Generation}_{nt}}{\sum_{i \in n} \text{Predicted Wind Generation}_{it}} * \text{Predicted Wind Generation}_{it}$$

Finally, geothermal generation is also reported in EIA-906 at the plant-month level. Geothermal plants are used as baseload, run at maximum possible capacity at zero marginal cost. The monthly generation is evenly distributed over the hours of the month.

### A.4 Heat Rates and Capacities

With a Leontief production function, the ratio of output to heat inputs measures the productivity of generation unit. A substantial literature has developed in industrial organization to consistently measure (Hicks-neutral) productivity, which is typically unobserved and time-invariant. It is possible to measure supply curves in the electricity setting because unit productivity is (more or less) time invariant and capacities are known whether the unit is operating or not. This simplifies matters

quite a bit.

Heat rates when operating are observed at the unit-month level in EIA-767, and EIA-906 at the plant-prime mover-month. When not operating, I use estimated heat rates based on regressions including unit-specific trends.

Heat rates for cogeneration units are a bit trickier: these are units that also provide useful steam energy, making it economical to run even if would not be operating otherwise. These units tend to have higher heat rates, which unaccounted for, would show up as out-of-merit generation. This then becomes a question of how much of the cost of fuel should be attributed to electricity versus steam production. The approach I use is to estimate what the unit's heat rate would be if not for cogeneration based on its vintage, capacity, etc. Fuel usage in excess of this estimated heat rate is attributed to steam generation, and not counted as a cost of running the unit for electricity production.

Unit capacities are reported in EIA-860 (as well as EIA-767), but to ensure out of merit costs are always positive, I use the maximum net generation observed from CEMS units in a year to measure the capacity of the unit.<sup>24</sup> For nuclear units, I calculate capacity as described above: the ratio of monthly output to the share of capacity utilized, as reported by the NRC. Because energy inputs are a fundamental constraint on production from renewables units, I assume that wind and hydro units are never withheld, and are therefore their observed production is their maximum capacity at that moment in time. Differential capacity factors of wind farms, for example, do not contribute to out of merit calculations. For hydro reservoirs this means that dry years, for example, similarly do not show up as economical units sitting idle.

---

<sup>24</sup>This effectively defines the merit order frontier as the lowest possible cost of production, inclusive of the ability to produce slightly higher than the rated limit when necessary, or less than the rated limit due to constraints not considered in nameplate rating reporting. Results are substantially unchanged when using EIA-860 nameplate capacity instead of observed capacities for CEMS units.

## A.5 Fuel Prices

### Coal Prices

As in Cicala (2015), this paper detailed data on coal deliveries to power plants from the Energy Information Administration (Forms EIA-423, “Monthly Report of Cost and Quality of Fuels for Electric Plants,” and EIA-923, “Power Plant Operations Report”) and Federal Energy Regulatory Commission (Form FERC-423, “Monthly Report of Cost and Quality of Fuels for Electric Plants”). These are shipment-level data, reported monthly for nearly all of the coal burned for the production of electricity in the United States (all facilities with a combined capacity greater than 50MW are required to report). The reader is referred to the Online Appendices of that paper for more details.

For this paper, the extensive use of bilateral contracts for coal procurement is potentially problematic: the merit order is determined by spot prices, not contract prices. This is because it is the opportunity cost of coal that determines its value when allocating production to plants. If coal were very expensive, one would want to dispatch those plants judiciously, regardless of whether a particular plant received its coal free of charge. This is a conceptually important distinction, though in practice the main results of the paper are largely invariant to using the observed contract prices instead of estimated spot prices.

The approach I use to estimate spot prices is to separate the delivered price of coal delivered to plant  $i$  in region  $d$  and month  $m$  from mine county origin  $o$  in to minemouth and shipping costs using hedonic regressions that include the characteristics of the coal and a third order polynomial in distance shipped:

$$c_{oim} = X_{im}\beta_{om} + f_{rm}(\text{distance}_{oim}) + \varepsilon_{iom}$$

I then estimate the minemouth spot prices by removing the shipping cost component from the hedonic estimates for the deliveries procured from the spot market only. To this I add the shipping portion of the hedonic regressions from all deliveries.<sup>25</sup>

---

<sup>25</sup>Although all coal regions deliver some coal to the spot market in all periods, they do not deliver them to all areas, making the origin-destination pairs sparse for estimating spot shipping separately.

$$\hat{c}_{oim}^{spot} = X_{im} \hat{\beta}_{om}^{spot} + \hat{f}_{rm}(distance_{oim})$$

## Oil Prices

Oil-burning units also report deliveries in forms EIA-423/923. I estimate spot prices at the state-month level separately by fuel type (diesel, residual fuel oil, etc.), or wider geographic if state-month deliveries are not reported.

## Natural Gas Prices

Spot prices for natural gas deliveries are based on daily prices from 65 major trading hubs across the country with consistent prices series from 1999-2012. These data come from the Platts, Bloomberg, and NGI. I use plant coordinates and natural gas pipeline network shapefiles from EIA to locate plants along the pipeline network. I then use network analysis software to connect each plant to the nearest pricing hub. These hubs are not uniformly distributed across the US, as illustrated in Figure A.2, which displays the connections between power plants and hubs. The southeast generally lacks pricing points in particular. Missing daily prices for weekends and holidays are based on carrying the most recent trading day forward.

**Figure A.2:** Gas Hub - Power Plant Links

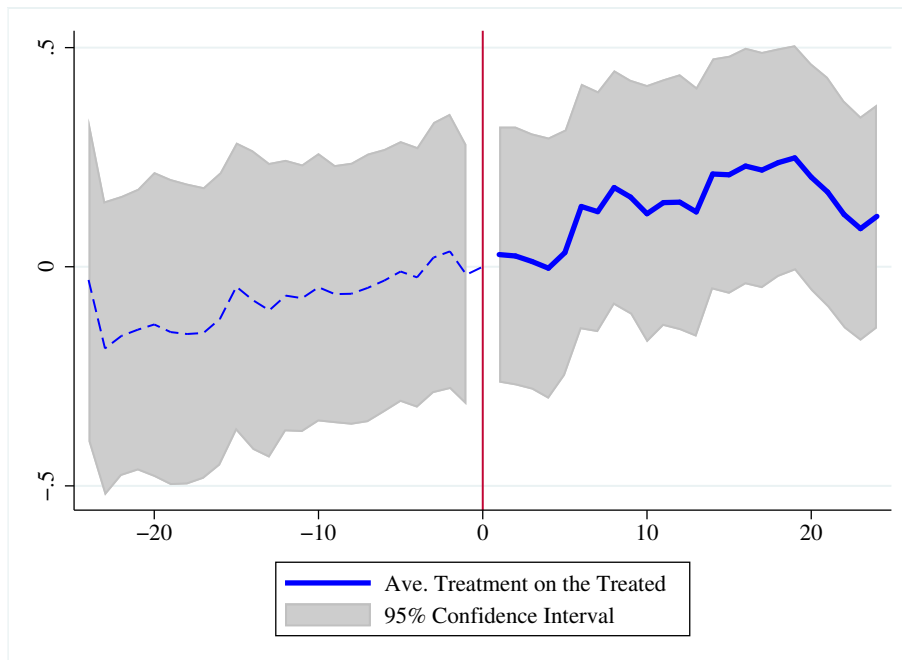




## **B Supplementary Results: For Online Publication**

As described in the paper, the adoption of market mechanisms to determine production was a separate development from state-led restructuring that allowed generators to become the residual claimants of operational revenues. A natural concern is that these non-utility generators have a greater incentive to operate when economical, and that their expansion was correlated with the adoption of market dispatch. To gauge the extent to which this other policy change might bias estimates of market dispatch's impact, Figure A.3 presents an event study-type figure analogous to the main results, but with the logarithm of non-utility generation capacity as the dependent variable. The figure shows that there has indeed been differential growth of non-utility capacity in areas that adopted market dispatch. However, the trend through the onset of treatment is smooth and continues at the same pace throughout the four-year period relative to market adoption. The divestiture and/or installation of new non-utility generation does not line up with the onset of treatment, but these results underscore the importance of the policy function's role in accounting for differential trends.

**Figure A.3: Non-Utility Generating Capacity**



Note: This figure is based on regressing logged non-utility capacity on a set of indicator variables for each month until (after) the transition to market dispatch, PCA-specific controls for load, date-hour-region and PCA fixed effects. The month prior to treatment is normalized to zero. Confidence intervals are based on clustering at the PCA-month level.