CSEM WP 126

Market Structure and Competition: A Cross-Market Analysis of U.S. Electricity Deregulation

James Bushnell, Erin T. Mansur, and
Celeste Saravia

March 2004

This paper is part of the Center for the Study of Energy Markets (CSEM) Working Paper Series. CSEM is a program of the University of California Energy Institute, a multi-campus research unit of the University of California located on the Berkeley campus.

2547 Channing Way
Berkeley, California 94720-5180
www.ucei.org

Market Structure and Competition: A Cross-Market Analysis of U.S. Electricity Deregulation

James Bushnell, Erin T. Mansur, and Celeste Saravia*

December, 2003

Abstract

This paper examines the importance of market characteristics in restructured electricity markets. We measure market performance relative to benchmarks that abstract away from market design characteristics but capture important structural elements. Specifically, we estimate market outcomes under an assumption of perfect competition and under an assumption of Cournot competition in three U.S. markets: California, New England, and PJM. These two counter-factual assumptions bound the space of possible static, non-cooperative outcomes. By establishing where actual market outcomes fall within these bounds, we can compare how markets perform relative to the extremes determined by structural factors alone. Our findings suggest that vertical arrangements between suppliers and retailers, dramatically affect estimated market outcomes. When we include vertical arrangements in firms' objective functions, Cournot equilibrium prices in both PJM and New England fall dramatically. California did not have such arrangements. After accounting for vertical arrangements, performance in each market relative to Cournot is similar, particularly during hours of peak demand.

* *Bushnell:* University of California Energy Institute. Email: Bushnell@Haas.Berkeley.edu. *Mansur:* School of Management and School of Forestry and Environmental Studies, Yale University. Email: Erin.Mansur@Yale.edu. *Saravia:* Dept. of Economics, University of California, Berkeley, and University of California Energy Institute. Email: CSaravia@econ.berkeley.edu.

1 Introduction

While rules concerning market structure form the basis of anti-trust policies in most countries, it is widely recognized that market structure comprises only one piece of the competition puzzle. In recent years, regulators and many economists have focused on the effects that market design may have on equilibrium prices. The design of deregulated electricity markets offers economists a unique and challenging opportunity. Wilson (2002) discusses the many physical problems inherent in the design of electricity markets including the necessity of balancing supply and demand instantaneously and the difficulties transmission congestion offers to the problem. Although market design may have a large effect on market outcomes Joskow (2003) notes that the majority of the benefits that result from a well designed electricity market will be concentrated in a few hours when the market structure is less favorable to competition. In particular, he states "All (electricity) wholesale market designs work reasonably well in the short run when demand is low or moderate. The performance challenges of wholesale market institutions arise during the relatively small number of hours when demand is high."

Undoubtedly, the attempts by economists to design well functioning markets that give players the correct incentives can improve production efficiency and limit market power. However, it appears as though some energy regulators have come to the conclusion that market design is more important than market structure. This view that regulating market design may be the best way to proceed is particularly true in deregulated electricity markets. For example, in a November 2000 report, the Federal Energy Regulatory Commission (FERC) concluded that the flawed rules of the California electricity market contributed to the crises in that state.[1] The FERC is currently working on a Standard Market Design (SMD) which it will encourage all deregulated electricity markets to adopt. The FERC has stated that it is willing to trade leniency in its review of market structure for the voluntary submission of firms to its standardized market rules.[2] The focus of electricity policy makers on market design and the lack of attention given to market structure may be due in part to the self-professed, but somewhat misguided, in ability of regulators to do anything to improve structural conditions.

In this paper, we attempt to examine the relative impacts of market design and market structure on equilibrium prices in deregulated electricity markets. In spirit, our paper is closely related to Evans and Green (2003), who address similar questions in the context of a dramatic change in the market rules in England and Wales. Our approach is quite different, however. We compare the market performance of deregulated electricity markets to estimates that abstract away from market design characteristics but capture the important structural elements of the

[1] The FERC found that "electric market structure and market rules for wholesale sales of electricity energy in California were seriously flawed and that these structures and rules in conjunction with an imbalance of supply and demand in California have caused and continue to have potential to cause, unjust and unreasonable rates for short term energy."

[2] See Bushnell (2003a) for a discussion of FERC policy initiatives regarding market structure and market design.

markets. Specifically, we estimate market outcomes under an assumption of perfect competition and under an assumption of Cournot competition. These two counter-factual assumptions bound the space of possible static, non-cooperative outcomes. By establishing where (and if) actual market outcomes fall within these bounds, we can compare how markets are performing relative to the extremes determined by structural factors alone.

We study three markets in the US: California, New England, and PJM. These markets represent opposite ends of the US spectrum in terms of market design and performance. California's market design has been labeled 'dysfunctional' by the Federal Energy Regulatory Commission (FERC), which in turn made reform of that design the centerpiece of its solutions to the California crisis. The PJM market has been widely viewed as the biggest success amongst US markets, and many elements of its design form the core of the FERC's proposals to establish a Standard Market Design in the US. The New England market, which by some measures has performed even better than PJM, has recently abandoned its original market design that was fashioned after the original UK pool in favor of one very similar to PJM.

Generation ownership in the two eastern markets is more concentrated than in California. As we show in this paper, imports into California were also much more substantial in magnitude and more responsive to market prices. End-use demand in all three markets was effectively perfectly inelastic as the vast majority of customers were on flat retail rate structures. Thus the California market had several structural advantages compared to the eastern markets, making the performance of the eastern markets relative to California even more impressive. We demonstrate that both eastern markets were dramatically more competitive than would be predicted by a model of Cournot competition. The California market, by contrast, produced prices somewhat lower, but largely consistent with an assumption of Cournot competition.

Our analysis highlights the importance of the element that differentiated California during its crisis from all other restructured electricity markets, the lack of long-term contracts or other vertical arrangements. Retail providers in both New England and PJM had extensive long-term vertical arrangements with producers. In PJM, most large retail providers had retained their generation resources during our sample period. Although there was much more divestiture of generation plants by incumbent utilities in New England than in PJM, most of these divestitures also coincided with the signing of long-term supply contracts by the new generation owners. Only in California was the bulk of economic activity concentrated in the short-term spot markets.

Although data on many short-term contracts are not publicly available, the details of the long-term retail supply arrangements are for the most part in the public domain. The impact of these vertical arrangements on estimated market outcomes is dramatic. When these retail obligations are included in the objective functions of suppliers, Cournot equilibrium prices in both PJM and New England fall dramatically. After known retail obligations are accounted for, the performance of all three markets relative to Cournot is similar. These results support the

hypothesis that long-term contracts and other vertical arrangements are a major source of the differences in performance of electricity markets.

In Section 2, we provide an general overview of the relevant characteristics of the three markets we study. In Section 3, we present our oligopoly modeling framework and describe our data. Section 4 presents our results and Section 5 our conclusions.

2 Overview of Electricity Markets

The term deregulation has come to mean many things in the U.S. electricity industry. In fact, only some aspects of electricity operations have been deregulated to any extent, and even those aspects are subject to considerable potential regulatory scrutiny. Deregulation efforts have been focused on the pricing of wholesale production (*i.e.* generation) and the retail function.[3] In the markets we study here, most large producers were granted authority to sell power at market-based prices rather than regulatory determined, cost-based rates. The distribution and transmission sectors remain regulated, but have been reorganized to accommodate wholesale markets and retail choice. The markets essentially share the same general organizational structure, but there are significant differences in ownership structure, market rules, and vertical relationships that are described in more detail below.

The PJM and California markets began operating during 1998, while the New England market opened in the spring of 1999. The PJM market was operating under very different circumstances during 1998 than in later years, however, as no firms yet had regulatory authority to sell power at market-based rates. Prices for 1998 were therefore the product of regulated offer prices into the PJM market-clearing process. Figure 1 illustrates the monthly average prices for the major price indices in each of the three markets from 1998 through 2002. As can be seen from this figure, market prices have varied widely across the three markets, with significant price spikes arising in PJM during the summer of 1999 and of course during the California crisis of 2000.

There has been much speculation and debate about the causes of these differences. Relative production costs, fuel prices, and overall demand of course played an important part in market outcomes. For example, the extremely high gas prices of the winter of 2000-01 are reflected in both the California and New England prices, but less so in PJM where coal is often the marginal fuel during the winter. There is also substantial evidence that the California market was less competitive than its eastern counterparts.[4] However, these studies do not address *why* there were

[3]In fact, wholesale electricity markets are not technically deregulated. Under the Federal Power Act, the Federal Energy Regulatory Commission has a mandate to ensure electricity prices remain 'just and reasonable.' In areas the FERC has deemed to be workably competitive, firms are granted permission, through a waiver process, to sell electricity at market-based rates (see Joskow, 2003).

[4]See Borenstein, Bushnell, and Wolak (2002), Bushnell and Saravia (2002), Joskow and Kahn (2002), Mansur (2003), and Puller (2000).

apparent differences in the competitiveness of these markets. Speculation has focused on three possible explanations, market design (or activity rules), horizontal market structure, and vertical arrangements.

Certainly, there is variation in all these factors across the markets. This variation over many attributes of the market makes comparison difficult, but fortunately there are data available that allow us to control for many of these factors for at least a subset of the markets' operating lives. In the following section we provide a brief overview of each of these factors.[5]

2.1 Market Rules

Electricity systems are made up of grids of transmission lines over which electricity is transported from generation plants to end use consumers. In most deregulated electricity markets, utilities have retained ownership of the transmission network, but they have relinquished the day to day control of the network to new institutions, called Independent System Operators (ISOs). ISOs are charged with operating electricity systems and guaranteeing that all market participants have equal access to the network.

In each of the markets studied in this paper, an ISO oversees at least one organized exchange through which firms can trade electricity. The rules governing these exchanges vary quite a bit. During the time period of this study, in California there were two separate markets for electricity: a day-ahead futures market and a real time spot market for electricity. Each day the California Power Exchange (PX), ran a day-ahead market for electricity to be delivered in each hour of the following day. The PX day-ahead market was a double-auction in which both producers and consumers of electricity placed their bid and offer prices. The California ISO held a real-time spot market for electricity. During the time period we study, PJM and New England featured only a single real time spot market for electricity overseen by their respective ISOs.[6] These ISO spot markets, also known as 'balancing' markets, cleared a set of supply offers against an inelastic demand quantity that was based upon the actual system needs for power during that time interval.

There were several variations of auction formats and activity rules across the markets, although each market utilized a uniform-price clearing rule. In the California PX, both supply offers and demand bids took the form of generic portfolio bids, with each firm able to submit an essentially unlimited number of piece-wise segments to their supply or demand function. Because of its close link to physical operating requirements, supply offers into the balancing market were

[5] This review is not meant to be exhaustive. For more details about individual markets, see Borenstein, Bushnell and Wolak (BBW, 2002) [California], Mansur (2003) [PJM], and Bushnell and Saravia (BS, 2003) [New England].

[6] The California PX stopped operating in January 2001. A day-ahead market began in PJM in 2000, while ISO-NE began operating a day-ahead market in early 2003. All these changes happened after the period of our study.

linked to specific generating plants and took the form of step functions. Bids and offers into the California PX and ISO-NE could be adjusted as frequently as hourly, while each supply offer into the PJM market were more 'long-lived,' with the same daily offer being applicable to each of the 24 hourly markets.

In order to accommodate several unique physical characteristics of electricity, the market clearing mechanisms in electricity markets are more complicated than those in other commodity markets. For reliability reasons, supply and demand must always be balanced. This is the reason that every electricity market holds a real time, balancing, market.[7] Electricity markets must take account of transmission network constraints. Absent network congestion, the cost of transporting electricity is relatively low. When congestion exists, however, it can impact the opportunity cost of consuming and producing power in broad regions of the network.

Each of the three markets studied in this paper deals with the issue of transmission congestion in a different way. PJM uses locational marginal pricing (LMP). Under LMP, the price at any point includes the additional congestion costs of injecting or withdrawing power from that point. This pricing scheme means that at any given time there may be thousands of distinct locational prices in the PJM market. Both California and New England aggregation locational prices over larger regions than PJM. California employs 23 price 'zones,' three internal and 20 others at points of interface with neighboring systems. During the period of our study, New England applied only a single pricing zone to its entire system. In both New England and California, generation that did not clear a zonal market, but was required to satisfy intra-zonal network constraints, was paid as bid above the market price.[8] The additional costs of this intra-zonal congestion was shared pro-rata by consumers within the pricing zone.

2.2 Market Structure

Table 1, summarizes the market structure in the three markets we study. By conventional measures, the PJM market, with an HHI of nearly 1400, is much more concentrated then either New England or California, with HHIs of around 850 and 620 respectively. With a peak demand over 45,000 MW and installed capacity of just over 44,000 MW, California is the market that most heavily relies on imports to supply electricity. On average, California imported about 25% percent of the electricity consumed within its system during 1999. New England, with an installed capacity of 27,000 MW and a peak demand of 21,400 MW, is the smallest market we study. New England also imports a substantial amount of power, almost 10% of its consumption, due to the fact that much of its native generation is older, gas & oil fired technology. PJM consists of approximately 57,000 megawatts (MW) of capacity, including coal, oil, natural gas, hydroelectric,

[7] The ISOs also procures ancillary reserve services that require generators to perform under various contingencies.

[8] In some cases, production from some generation needed to be reduced to satisfy network constraints. These generators would then 'buy-back' their production obligation from the ISO at below market prices.

nuclear energy sources. Unlike the other two markets, coal plays a major role in PJM, and is frequently the marginal fuel.

The limitations of conventional structural measures, particularly when applied to electricity markets, has previously been explored.[9] At least as important as concentration in markets for non-storable goods is the relationship of production capacity to overall demand levels. The elasticity of imported supply that could contest the market sales of local producers is also extremely important. These and other aspects of each market are explicitly incorporated into our oligopoly framework described below. One last critical factor that can influence the relative competitiveness of the markets is the extent of long-term contracts and other vertical commitments.

2.3 Retail Policies and Vertical Arrangements

The retail function in electricity differs from most other industries in that the ability of firms to adjust retail prices is severely constrained. This was particularly true for the multiyear 'transition' periods that followed restructuring in most states. In all three markets studied here, as well as most other major US electricity markets, the incumbent utilities were required to freeze retail rates for several years. Although newly entering retail firms were not explicitly bound to these agreements, the freezing of the largest retailers' prices served as an effective cap on all retail rates in a market since customers could always elect to remain with the incumbent. Thus, retail firms were potentially very vulnerable to wholesale price volatility. The strategic response by retailers to the risks imposed by these policies varied substantially across the three markets.

In PJM, most retailers retained their generation assets and thus remained vertically integrated into production. Vertical integration provided a physical hedge against high wholesale prices. It also impacted the incentives of those controlling production. Large producers, such as GPU, also had substantial retail obligations that nearly eliminated the profitability of reducing production to raise market prices. As shown in Table 1, the distribution of retail obligations and production resources was uneven, with some firms frequently in the position of 'net-seller' while others were nearly always 'net-buyers.' Mansur (2003) examines the relative production decisions of these firms using a difference-in-differences approach. Using data from 1998, when bidding was still regulated, and 1999 when firms were first allowed to employ 'market-based' bids, Mansur compares the changes in output quantities of net-sellers with those of net-buyers. While controlling for estimates of how firms in a competitive market would have produced, he finds that the two main net-sellers produced relatively less during 1999 than 1998 as compared to the other, net-buying firms.

The divestiture of generation from vertically integrated utilities was much more widespread in New England, although the process was not completed until after 1999. In order to hedge

[9] See for example, Borenstein, Bushnell, and Knittel (1999).

their price exposure, however, many of the retail utilities signed long-term supply contracts, often with the firms to whom they had divested their generation. The largest producer in New England during our sample period, Northeast Utilities, was in the process of divesting most of its generation during 1999, but these transactions were not finalized until after September. During the summer of 1999, NU is therefore both the largest producer and retailer of electricity. Soon after divesting its generation, NU subsidiary Connecticut Light & Power signed long-term supply arrangements with NRG, Duke Energy, and its own subsidiary, Select Energy. Pacific Gas & Electric's unregulated subsidiary National Energy Group (NEG) also controlled a large generation portfolio, but was obligated to provide power to the non-switching, 'default' retail customers served by NEES, the former owner of the generation. United Illuminating of Connecticut and Boston Edison had also signed supply contracts with the purchasers of their generation, Wisvest and Sithe Co., respectively. The Sithe contract had expired by the summer of 1999, while the Wisvest contract expired the following year. In their study of the New England Electricity market, Bushnell and Saravia (2003) utilize bidding data to compare the bid margins of firms they characterize as obligated to serve substantial retail load with those of firms that were relatively unencumbered by such arrangements. They find that bid margins from both classes of firms increase monotonically with overall market demand, but that the margins of the 'retailing' class of suppliers were often negative, indicating that these firms may have utilized their generation assets to lower overall market prices in hours when they were net-buyers on the market. We revisit the potential for such 'monopsony' production strategies in our results below.

In contrast to New England, where most retailers responded to the risk exposure of rate-freezes by signing long-term supply contracts, the purchases of the utilities in California were notoriously concentrated in the daily spot markets of the California Power Exchange and California ISO. During the summer of 1999, there were almost no meaningful long-term arrangements between merchant generation companies and the incumbent utilities.[10] The largest utilities, PG&E and Southern California Edison (SCE) did retain control of substantial nuclear and hydro generation capacity, as well as regulatory era contracts with many smaller independent power producers. This capacity was nearly always infra-marginal however, so the utilities had limited ability to reduce prices by 'over-producing' from resources that should have been producing anyway. The failure of the utilities to sign long-term contracts has been attributed to regulatory barriers put in place by the California Public Utilities Commission, but the full reasons are more complex and remain a source of disagreement (see Bushnell (2004)). The impact of long-term vertical arrangements has been shown to have significant impact on the performance of markets, but to our knowledge there has been no attempt to assess the degree to which these contracts influenced market outcomes, or how these impacts varied across markets. These are questions that we address below.

[10]The utilities did purchase some power through futures contracts in the PX's block-forward market, we hope to examine the impact of these arrangements in future work.

3 Market Structure and Market Performance

In this section we develop estimates of the impact of market structure on market performance. Our approach is to abstract away from the detail market rules and regulations in each market and examine the range of equilibrium price outcomes that would be predicted from considering market structure alone. We calculate the upper and lower bounds on market prices that could be produced in a static, non-cooperative equilibrium. These bounds are, respectively, represented by the Nash-Cournot and perfectly competitive, or price-taking, equilibria. Several models of oligopoly competition in the electricity industry have employed the supply function equilibrium (SFE) concept developed by Klemperer and Meyer (1989).[11] In many cases there exist multiple SFE, and Klemperer and Meyer show that these equilibria are bounded by the Cournot and competitive equilibria. To the extent that market design influences market outcomes by helping determine which of the many possible equilibria arise, these impacts can be thought of as placing the market price within these bounds.[12] Several papers have applied a model of Cournot competition to electricity markets to forecast possible future market outcomes using hypothetical market conditions.[13] Unlike those papers we are applying actual market data to 'backcast' market outcomes within the Cournot framework.

The consideration of vertical arrangements is critical to our examination of the interaction of market structure and prices. A line of research, beginning with Allaz and Vila (1993), has examined whether the existence of forward markets can increase the competitiveness of a market for a commodity. These considerations have been shown to be relevant to electricity markets.[14] As described in section 4.2, the vertical arrangements in our markets take several different forms, but all have the effect of committing a producer to the supply of an exogenously determined quantity at a pre-determined price. These commitments have a powerful impact on the incentives of various producers, to the extent that, when such arrangements are ignored, market structure appears to provide little information about market outcomes. However, once these arrangements are explicitly taken into consideration, the range of possible non-cooperative equilibria narrows considerably and market outcomes strongly resemble the Nash-Cournot equilibrium.

3.1 Model of Equilibrium Behavior

For each market, we explicitly model a set of strategic producers as well as non-strategic, price-taking, producers. The former are comprised of the firms described in Table 1, the latter are

[11] For example, see Green and Newbery (1992) and Rudkeivich, et. al (1998).

[12] Green and Newbery (1992) show that when capacity constraints apply to producers, the range of possible equilibria narrows as the lower bound becomes less competitive. Even though capacity constraints are sometimes relevant for the producers in the markets we study, we note that the perfectly competitive price still represents a lower bound, albeit a generous one.

[13] See for example, Schmalensee and Golub (1982), Borenstein and Bushnell (1998), and Hobbs (2001).

[14] See Green (1999) and Wolak (2001).

comprised of the aggregation of generation from firms owning less than 800 MW of capacity in any market. For each market, strategic firms are assumed to maximize profit according to the Cournot assumption using production quantities as the decision variable. Furthermore, firms consider forward commitments reached through vertical arrangements and long-term contracts. For each strategic firm $i \in \{1, ..., N\}$ and time period $t \in \{1, ..., T\}$ that are assumed to be independent, firm i maximizes profits:

$$\pi_{i,t}(q_{i,t}, q^c_{i,t}) = p_t(q_{i,t}, q_{-i,t}) \cdot [q_{i,t} - q^c_{i,t}] + p^c \cdot q^c_{i,t} - C(q_{i,t}), \tag{1}$$

where $q_{i,t}$ is the total quantity produced by firm i, $q_{-i,t}$ is the quantity produced by the other $N-1$ strategic firms, and p_t is the market price. Also, q^c_{it} is the forward-contracted quantity for which firm i will receive a predetermined price (p^c), and $C(q_{i,t})$ is the total production cost for firm i. Considering that both the contract quantity and price are sunk at the time production decisions are made, the second term of (1), $p^c \cdot q^c_{i,t}$, drops out of the equilibrium conditions. As described below, marginal production costs are represented using a piece-wise linear convex function. The market price is determined from the strategic firms' residual demand function (Q_t), which equals the market demand ($\overline{Q}_t$) minus fringe supply (q^{fringe}_t). We model Q_t as a linear-log function:

$$Q_t = \overline{Q}_t - q^{fringe}_t = \alpha_t - \beta \ln(p_t). \tag{2}$$

Supply from imports and fringe units–for which there are no cost data–provide price responsiveness ($\beta < 0$). Under these assumptions, we can represent the Cournot equilibrium as the set of quantities that simultaneously satisfy the following first order conditions for each firm i and period t:

$$\frac{\partial \pi_{i,t}(q_{i,t})}{\partial q_{i,t}} = p_t(q_{i,t}, q_{-i,t}) + [q_{i,t} - q^c_{i,t}] \cdot \frac{\partial p_t}{\partial q_{i,t}} - C'_{i,t}(q_{i,t}) \geq 0. \tag{3}$$

The production of non-strategic firms is described by the price-taking condition:

$$p_t(q_{i,t}, q_{-i,t}) - C'_{i,t}(q_{i,t}) = 0, \forall t. \tag{4}$$

The full solution to these equilibrium conditions is represented as a complementarity problem and solved for using the PATH algorithm.[15] The appendix contains a more complete description of the complementarity conditions implied by the equilibrium, given the functional forms of the cost and inverse demand described below. In the following section, we describe the data used to develop the various parameters described in the above equilibrium conditions. To calculate the lower bound on static, non-cooperative outcomes we set the production of all firms according to condition (4). In other words we calculate the equilibrium assuming all firms are acting according to the price-taking condition.

[15]See Dirkse and Ferris (1995).

4 Data Description

In this section, we describe both the sources and application of the data used in our equilibrium calculations. In most instances, the sources of the data are identical to those used in the Borenstein, Bushnell, and Wolak (BBW), Mansur, and Bushnell and Saravia (BS), studies of California, PJM, and New England, respectively. There are several substantial differences in the application of those data to the model used in this paper, however, and we focus our discussion on those issues.

4.1 Cost Functions

In general there are two classes of generation units in our study: those for which we are able to explicitly model their marginal cost and those for which it is impractical to do so due to either data limitations or the generation technology. Fortunately, the vast majority of electricity is provided by units that fall into the first category. Most of the units that fall into the second category, which includes nuclear and small thermal and hydro-electric plants, are generally thought to be low-cost, infra-marginal technologies. Therefore, as we explain below, we apply the available capacity from units in this second category to the bottom of their owner's cost function.

Fossil-Fired Generation Costs

We explicitly model the major fossil fired thermal units in each electric system. Because of the legacy of cost-of-service regulation, relatively reliable data on the production costs of thermal generation units are available. The cost of fuel comprises the major component of the marginal cost of thermal generation. The marginal cost of a modeled generation unit is estimated to be the sum of its direct fuel, environmental and variable operation and maintenance (O&M) costs. Fuel costs can be calculated by multiplying a unit's 'heat rate,' a measure of its fuel-efficiency by an index of the price of fuel, which is updated as frequently as daily. Many units are subject to environmental regulation that require them to obtain nitrogen oxides (NO_x) and sulfur dioxide (SO_2) tradable pollution permits. Thus, for units that must hold permits, the marginal cost of polluting is estimated to be the emission rate (lbs/mmbtu) multiplied by the price of permits and the unit's heat rate.

The capacity of generation units is reduced to reflect the probability of forced outage of each unit. The available capacity of generation unit i, is taken to be $(1-fof_i)*cap_i$. where cap_i is the summer rated capacity of the unit and fof_i is the forced outage factor reflecting the probability of the unit being completely down at any given time.[16] By ordering all the generation units

[16]This approach to modeling unit availability is a departure from the methods used BBW, Mansur, and BS. In those studies, unit availability was modeled using Monte Carlo simulation methods. Because of the additional computational burden of calculating Cournot equilibria, we have simplified the approach to modeling outages. As we discuss below, the impact of this simplification on estimates of competitive prices is minimal.

owned by firm i, one can construct a step-wise function for the production cost from that firm's portfolio. For computational reasons, we approximate this step function with a piece-wise linear function with five segments. For each firm, we calculated from the step-wise cost function the available operating capacity (*i.e.,* capacity with costs less than or equal to price) for each quintile of marginal costs in that firm's portfolio. These quantities became the available capacity in each of the five segments of the piece-wise linear cost function for that firm. The marginal costs of these linear segments connected the five quintiles of marginal costs in a firm's portfolio.

We do not explicitly represent scheduled maintenance activities. This is in part due to the fact that maintenance scheduling can be a manifestation of the exercise of market power and also because these data are not available for PJM and California. The omission of maintenance schedules is unlikely to significantly impact our results for the summer month, high demand periods, when few units traditionally perform scheduled maintenance. This is one reason why we limit our comparisons to summer months.

Nuclear, Cogeneration, and Energy Limited Resources

There are several categories of generation for which it is impractical to explicitly model marginal production costs. Much of this energy is produced by conventional generation sources, but there is also a substantial amount of production from energy-limited (primarily hydro-electric) resources. Most of this generation is produced by firms considered to be non-strategic. Because the production decisions for firms controlling energy limited resources are quite different from those controlling conventional resources, we treat the two categories differently.

With one exception, the vast majority of production from conventional non-modeled sources is controlled by firms considered to be non-strategic. Because of this, we include the production from such capacity in our estimates of the residual demand elasticity faced by the strategic firms described below. The exception applies to the substantial nuclear capacity retained as part of large portfolios in PJM.[17] While nuclear production is an extreme infra-marginal resource, and unlikely to be strategically withheld from the market for both economic and technical reasons, the substantial amount of infra-marginal production could likely have a significant impact on amounts nuclear firms may choose to produce from the other plants in their portfolios. We therefore take the hourly production from nuclear resources as given and apply that production quantity as a zero-cost resource at the bottom of its owners cost-function.

Energy-limited units (*i.e.,* hydroelectric units) present a different challenge than other units in the non-modeled category since the concern is not over a *change* in output relative to observed

[17] It should also be noted that a large amount of production in California from smaller generation sources providing power under contract to the three utilities. In one sense, this generation can be thought of as 'controlled' by the utilities as they have purchased it under contracts left over from the 1980's and early 1990's. However these contracts are essentially 'take-or-pay' contracts, and the utilities have extremely limited influence over the quantity of such production. Because of this, we include production from all 'must-take' resources, as they are called in California, in our estimates of residual demand for the California market.

levels but rather a *reallocation* over time of the limited energy that is available. The production cost of hydroelectric units do not reflect a fuel cost but rather a cost associated with the lost opportunity of using the hydroelectric energy at some later time. In the case of a hydroelectric firm that is exercising market power, this opportunity cost would also include a component reflecting that firm's ability to impact prices in different hours (Bushnell, 2003b). Because the overall energy available is fixed, we do not consider supply from these resources to be price-elastic in the conventional sense and did not include fringe-hydro production in our residual demand estimates. Rather, we take the amount of hydro produced as given for each hour and apply that production to the cost function of each firm.[18]

Thus a firm's estimated marginal cost function consists of a piece-wise linear function of fossil-fuel production costs, where each segment of the piece-wise linear function represents a quintile of the firm's portfolio marginal cost, beginning at the marginal cost of its least-expensive unit and ending at the marginal cost of its most expensive unit. This piece-wise linear function is shifted rightward by an amount equal to the quantity of electricity produced by that firm from hydro-electric and nuclear resources. The aggregate production capacity of a firm can therefore change from hour to hour if that firm has volatile hydro-electric production.

4.2 Market Data

Market Clearing Quantities and Prices

Since the physical component of all electricity transactions is overseen by the system operators, it is relatively straightforward to measure market volume. We measure energy demand as the metered output of every generation unit within the respective system plus the net imports into the system for a given hour. Because of transmission losses, this measure of demand is somewhat higher than the metered load in the system. To this quantity we add an adjustment for an operating reserve service called automated generation control, or AGC. Units providing this service are required to be able to respond instantaneously to dispatch orders from the system operator. These units are therefore 'held-out' from the production process and the need for this service effectively increases the demand for generation services. This reserve capacity typically adds about three percent to overall demand.

In California, the market price is the day-ahead unconstrained price from the California Power Exchange. About 85% of California's volume traded in this market between 1998 and 2000. There were no day-ahead markets in New England or PJM during 1999. We use the ISO-NE's Energy Clearing Price (ECP) for the New England market price and the PJM market's real-time Locational Marginal Prices (LMP). Neither the California PX-UCP or the New

[18]Due to the lack of available data on PJM, Mansur (2003) develops an approximation of these hourly output levels. In PJM units in this category constituted an relatively modest level of production.

England ECP reflect any geographic variation in response to transmission constraints. The PJM market, however, reports no single 'generic' market-wide price, and instead provides up to several thousand different geographic prices that implicitly reflect the costs of transmitting electricity within that system. As in Mansur (2003), we therefore utilize a demand weighted average of these locational PJM prices.[19]

Vertical Arrangements and Long-term Contracts

Data on the contractual arrangements reached by producers is more restricted than data on spot market transactions. We focus on the large, long-term vertical arrangements between generation firms and retail companies responsible for serving end-use demand. These arrangements have for the most part been reached with regulatory participation and have been made public knowledge. For PJM, where all major producers remained vertically integrated, we calculate the retail obligation by estimating the utilities' hourly distribution load and multiplying it by the fraction of retail demand that remained with that incumbent utility. Monthly retail migration data are available for Pennsylvania, but were relatively stable during the summer of 1999 so a single firm-level summer average was used to calculate the percentage of customers retained. A utilities hourly demand was calculated by taking the peak demand of each utility and dividing it by overall PJM demand. This ratio was applied to all hours. We therefore assume that the relative demand of utilities in the system is constant.[20]

In New England, we apply the same methodology for the vertically integrated NU. Wisvest had assumed responsibility for the retail demand of United Illuminating during 1999, so they are treated as effectively integrated with each other. NEG was responsible for the remaining retail demand of NEES, so their obligation is estimated as the hourly demand in the NEES system multiplied by its percentage of retained customers. These estimates of retail obligations as a fraction of system load are given in Table 1.

Estimating Residual Demand

For most power plants in each market, detailed information enables us to directly predict performance given assumptions over firm conduct. For other plants, either we lack information on costs and outside opportunities–such as imports into and exports out of a market–or the plants' owners have complex incentives, such as "must-take" contracts. For these fringe plants, we estimate a supply function, which we then use to determine the residual demand for the remaining plants. Recall that the derived demand in wholesale electricity markets is completely inelastic; therefore, the residual demand curve slope will equal market demand minus the slope of the supply of net imports (imports minus exports) and other fringe plants not modeled. In

[19] All market quantities and prices are available from the respective ISO websites: www.caiso.com, www.pjm.com, and www.iso-ne.com.

[20] Hourly utility level demand data are available for some, but not all utilities in our study. A comparison of our estimation method to the actual hourly demand of those utilities for which we do have data shows that the estimation is reasonably accurate.

California, this supply includes net imports and must-take plants.[21] In New England, net imports from New York and production from small firm generation comprise this supply.[22] We estimate only net import supply in PJM. For all markets, the sample period is the summer of 1999 (June to September).

Firms importing and exporting decisions depend on relative prices. If firms located within the modeled market increase prices above competitive levels, then actual fringe supply will also exceed competitive levels. With less fringe supply and completely inelastic demand, more expensive units in the market will operate. We assume that firms exporting energy into the restructured markets behave as price takers because they are numerous and face regulatory restrictions in their regions. When transmission constraints do not bind, the interconnection is essentially one market. However, the multitude of prices and "loop flow" concerns make assuming perfect information implausible. The corresponding transaction costs make fringe supply dependent on both the sign and magnitude of price differences.

For each hour t, we proxy regional prices using daily temperature in bordering states ($Temp_{st}$),[23] and fixed effects for hour h of the day ($Hour_{ht}$) and day j of week (Day_{jt}). For each market and year, we estimate fringe supply (q_t^{fringe}) as a function of the natural log of actual market price ($\ln(p_t)$),[24] proxies for cost shocks (fixed effects for month i of the summer ($Month_{it}$)), proxies for neighboring prices ($Temp_{st}, Day_{jt}, Hour_{ht}$), and an idiosyncratic shock (ε_t):

$$\begin{aligned} q_t^{fringe} &= \sum_{i=6}^{9} \alpha_i Month_{it} + \beta \ln(p_t) + \sum_{s=1}^{S} \gamma_s Temp_{st} \\ &\quad + \sum_{j=2}^{7} \delta_j Day_{jt} + \sum_{h=2}^{24} \phi_h Hour_{ht} + \varepsilon_t. \end{aligned} \tag{5}$$

As price is endogenous, we estimate (5) using two stage least squares (2SLS) and instrument using hourly quantity demanded. The instrument is the natural log of hourly quantity demanded inside each respective ISO system. Typically quantity demanded is considered endogenous to price, however, since the derived demand for wholesale electricity is completely inelastic, this unusual instrument choice is valid in this case. We exclude demand from the second stage as it only indirectly affects net imports through prices.

[21] Borenstein, Bushnell, and Wolak (2002) discuss must take plants and why they are not modeled directly in measuring firm behavior. These plants include nuclear and independent power producers.

[22] Canadian imports are constant as cheap Canadian power almost always flows up to the available transmission capacity into New England. Small generation includes those generators not owned by the major firms. These include small independent power producers and municipalities. See Bushnell and Saravia (2002) for further discussion.

[23] For California, this includes Arizona, Oregon, and Nevada. New York is the only state bordering New England, while in PJM, bordering states include New York, Ohio, Virginia, and West Virginia. The temperature variables for bordering states are modeled as quadratic functions for cooling degree days (degrees daily mean below 65° F) and heating degree days (degrees daily mean above 65° F). As such $Temp_{st}$ has four variables for each bordering state. These data are state averages from the NOAA web site daily temperature data.

[24] Unlike a linear model, this functional form is smooth, defined for all net imports, and accounts for the inelastic nature of imports nearing capacity. A log-log model, with constant elasticity, would drop observations with negative net imports, a substantial share of the data in some markets.

For each market, Table 2 reports the 2SLS coefficient and standard error estimates that account for serial correlation and heteroskedasticity.[25] Panel A shows the coefficients on the instruments in the first stage, which suggest strong load instruments, while panel B displays the β coefficients for each year. In California, has the most elastic import and fringe supply, with a $\beta = 5041$ (with a standard error of 686). In New England, β is 1246 (s.e. 155) in 1999, and is 595 (s.e. 139) in 2000. Finally, in PJM during 1999, we estimate β as 856 (s.e. 117).

These coefficient estimates are then used to determine the N strategic firms' residual demand (Q_t). In equilibrium, $Q_t = \sum_{i=1}^{N} q_{i,t}$ so we define α_t as the vertical intercept:

$$\alpha_t = \sum_{i=1}^{N} q_{i,t}^{actual} + \beta \ln(p_t^{actual}), \tag{6}$$

where p_t^{actual} and $q_{i,t}^{actual}$ are the actual price and quantities produced. Therefore, for each hour, we model the inverse residual demand:

$$p_t = \exp(\frac{\alpha_t - \sum_{i=1}^{N} q_{i,t}}{\beta}). \tag{7}$$

5 Results

Our sample period is the summer of 1999, which featured extreme weather in the mid-Atlantic states, but relatively mild weather in California, although the 1999 summer peak in California was actually higher than the peak demand during 2000. Figure 2 illustrates, the distributions of demand in each market. These distributions are normalized by dividing by the maximum observed demand in each market. While our sample period includes a substantial range of market conditions in each market, some notable periods, such as the summer of 2000 in California, are not represented. However, the analyses in BBW and BS, which examine nearly three years worth of market operations each, indicate that the overall competitiveness of the market is consistent across the years when one controls for overall load levels. We first set $q_{i,t}^c$ to the approximate levels we have been able to determine from public data sources. To assess the impact of horizontal market structure alone, we also examine the range of possible equilibria when the incentive effects of vertical arrangements and long-term contracts are ignored. In other words, we set the contract quantity, $q_{i,t}^c$ in equation (3) equal to zero for all firms. In both cases the 'lower' bound on the equilibrium prices is the same, as the contract quantity is not relevant to the price-taking conditions.

While we have used the phrase 'lower' bound to refer to the competitive equilibrium and 'upper' bound to refer to the Cournot equilibrium, it is important to recognize that the use of

[25] We test the error structure for autocorrelation (Breusch-Godfrey LM statistic) and heteroscedasticity (Cook-Weisberg test). First we estimate the 2SLS coefficients assuming *i.i.d.* errors in order to calculate an unbiased estimate of ρ, the first-degree autocorrelation parameter. After quasi-differencing the data, we re-estimate the 2SLS coefficients while using the White technique to address heteroscedasticity.

these terms should be qualified. As we describe below, there are observations where the Cournot outcome yields *lower* prices than the perfectly competitive outcome, and observations where both the Cournot and competitive outcomes are above the actual market price, as well as observations when the actual price was greater than both the Cournot and competitive estimates.

These phenomenon are influenced by several factors. First, it should be noted that each observation of actual prices reflects a single realization of the actual import elasticity and outage states that are estimated with error. So the structure of the markets in any given hour will be somewhat different than our aggregate estimates, and therefore may result in individual prices outside of our estimated bounds.

Second, the oligopoly and competitive outcomes are functions of our estimates of marginal costs, which are also subject to measurement error. To the extent we overstate the marginal cost of production, observed market prices during very competitive hours, which will be close to marginal cost, will be lower than our estimated prices. Our treatment of production cost as independent of the hour-of-day will likely bias our estimates of costs upward during off-peak hours, and downward during peak hours. This is because power plants in fact have non-convex costs and inter-temporal operating constraints, such as additional fuel costs incurred at the start-up of a generation unit and limits on the rates in which the output of a unit can change from hour to hour. These constraints are most severe in the PJM market, as is reflected in the fact that estimates of competitive prices in PJM are above actual prices in a large number of off-peak hours. The inclusion of NOx emissions costs in our cost estimates may also be biasing our costs upward. The markets for NOx permits were illiquid and there is reason to believe that the available NOx price indices may overstate those costs.

Lastly, even without any measurement error the Cournot equilibrium can produce prices lower than perfectly competitive ones when vertical arrangements are considered. To the extent that large producers also have even larger retail obligations, they may find it profitable to over-produce in order to drive down their wholesale cost of power purchased for retail service. In terms of equation 3, when $q^c_{i,t} > q_{i,t}$ marginal revenue is greater than price, and therefore it is profit maximizing to produce at levels where marginal cost is greater than price. Thus, when the load obligations exceed the production levels of key producers, the Cournot price in fact becomes the 'lower' bound, and the competitive price the 'upper' bound.

Table 3 summarizes the prices for the Nash-Cournot equilibrium with and without vertical arrangements, as well as the price-taking equilibrium and the actual market prices. Note that the California market effectively had no long-term vertical arrangements between utility retailers and suppliers during 1999. There was considerable generation retained by the two largest, still-partially vertically integrated, utilities. However, the overwhelming majority of this capacity was either nuclear or other 'must-take' resources such as regulatory era contracts with small producers, or hydro production. Functionally, this means that there is no meaningful difference

between a 'no vertical arrangements' and 'with vertical arrangements' case in California.[26]

Errors in our cost estimates will have a much larger proportional impact on our estimates of competitive prices and Cournot prices during very competitive hours, where prices closely track marginal cost, than on hours where there is substantial potential market power. At low levels of demand even strategic firms are not able to exercise a great deal of market power, and thus, the Cournot prices are very close to the competitive prices. When firms are able to exercise a great deal of market power, the quantity they produce will be more sensitive to the slope of the residual demand curve than to their own marginal costs. This implies that if our cost estimates are biased, the bias will have a differential impact on the fit of the two models at different demand levels. In particular, for low levels of demand both models very closely track marginal costs and therefore they will both have similar degrees of bias. At high levels of demand, the competitive prices still track marginal costs and thus they will still have the same degree of bias, while at high levels, the Cournot estimates are more sensitive to residual demand than to marginal costs and thus a cost bias will have less of an effect. We therefore separate our results into peak and off-peak hours to better reflect this differential impact of any bias in cost measurement, where peak hours are defined as falling within 10 AM and 8 PM on weekdays. In all three markets, actual prices appear to be consistent with Cournot prices in comparison to competitive prices during the peak hours of the day. Our off-peak competitive price estimates exceed actual prices the most in the PJM market, where inflexible coal units with high NOx emissions are frequently on the margin during off-peak hours. The low prices do not appear to be caused by monopsony behavior, as Cournot prices converge to competitive prices from above at the lowest demand levels.

By contrast, the negative price-cost margins during off-peak hours in New England are in fact consistent with strategic behavior. The Cournot equilibrium price at lower demand levels in New England is significantly below the competitive price. The September Cournot price in New England, for example, averaged \$28.77/MWh compared with an estimated competitive price of \$32.09/MWh. The actual price in September averaged \$28.37/MWh. The average Cournot price in August was also below the competitive price. The New England market is the only market where we see this phenomenon, as it is the only market where the dominant producers also have large retail obligations and sufficient extra-marginal resources with which to over-produce and drive down prices.

[26] As we have argued above, firms have no ability to impact equilibrium prices with must-take resources since they would be producing in the market under all possible market outcomes. A firm could allocate production from its energy-limited hydro resources with the goal of driving-down prices (as opposed to raising them as an oligopolist, or allocating to the highest price hours as would a price-taker. Any attempts to do so by PG&E, the large hydro producer in California, would be reflected in the actual production numbers, and therefore already incorporated into the residual demand of the oligopoly producers. A fully accurate 'no contracts' case in California would consider the ability of a hypothetical 'pure-seller' PG&E to allocate water in a way that maximizes generation revenues. However the strategic optimization of hydro resources is beyond the scope of this paper. See Bushnell (2003b) for an examination of the potential impacts of strategic hydro production in the western U.S.

Figure 3 plots actual hourly prices in California from June 1 to September 30, 1999. We estimate a non-parametric kernel regression of the relationship between the actual hourly prices and the share of installed capacity.[27] In Figure 3, this is shown with a black line, In addition, we estimate the kernel regression for our estimates of prices from each hour's Cournot equilibrium (gray line) and the prices that we estimate would arise under competitive behavior (dotted line). In the case of California, the actual prices and the Cournot estimates are similar except at low demand levels, where both competitive and Cournot prices exceed actual prices, likely for the reasons described above. Figure 4 presents the same type of analysis for the New England market. As with California, the Cournot prices are similar to the actual prices at high demand levels. At lower demand levels, note that the Cournot prices lie below the competitive prices. This is consistent with the monopsony over-production strategy. Figure 5 illustrates the same analysis for PJM. Again, Cournot prices are quite close to actual prices at higher demand levels and actual exceed Cournot prices at the very highest levels of demand.

We can examine the relative 'fit' of the two estimated price series, competitive and Cournot, to actual prices, although one must be very cautious in interpreting the results as both of our price series are estimated with error. We first calculate the relative R^2 of the two price estimates from the difference between our two estimates and the actual price for each hour. The R^2 in California is 0.88 for the Cournot estimates and 0.77 for the competitive prices. In New England, the R^2 is 0.85 for Cournot and 0.80 for competitive. In PJM the values are 0.76 and 0.46 for the Cournot and competitive prices, respectively. In all three markets, our point estimates of Cournot prices are therefore more highly correlated with actual prices than are our estimates of competitive prices.[28]

In order to determine the sensitivity of these findings to the uncertainty from estimating the residual demand function in Table 2, Figures 6, 7, and 8 display a 90 percent confidence interval on our estimates. This is done by perturbing the coefficient estimate of fringe supply in Table 2 by plus and minus two standard errors. As expected, the variation in elasticity produce more substantial differences in Cournot prices during very high demand hours, but the range of prices is

[27] We use the 100 nearest neighbor estimator, namely the Stata command 'knnreg.'

[28] A more formal test can examine whether these values are in fact meaningfully different. The empirical model is actual price either equals the competitive price or the Cournot price, but is not a function of both. Since there does not exist a mapping of one pricing model to the other, a non-nested test is required. We follow the methodology of an encompassing test, as described in Davidson and MacKinnon (1993, pages 386-387), which is done by testing one hypothesis and including the variables from the second hypothesis that are not already in the model. In our case, this is just regressing actual prices on the competitive and Cournot prices. We estimate this equation using generalized least squares (GLS) estimates that account for serial correlation and heteroskedasticity (the Prais-Winsten AR(1) regression with robust standard errors). We find that the coefficient on the Cournot prices is positive and significant (1.31 with a standard error of 0.08), while the coefficient on the competitive prices is negative and insignificant (-0.16, with s.e. 0.11). The encompassing test for New England results in similar findings as in California. We find that the coefficient on the Cournot prices is positive and significant (2.28 with a standard error of 0.43), while the coefficient on the competitive prices is negative and insignificant (-1.01, with s.e. 0.63). However, in PJM the encompassing test fails to reject the competitive model in PJM: the coefficient on competitive prices is 4.32 with a standard error of 0.90. The model also fails to reject the Cournot model: the coefficient on Cournot prices is 0.88 and is more tightly estimated with a standard error of 0.082.

still relatively narrow compared to the effects of eliminating the vertical arrangements. Figure 9 presents the Cournot prices with vertical arrangements as in Figure 4 and also the Cournot prices without the vertical arrangements. In both markets, the prices without vertical commitments far exceed actual prices for even moderate demand levels, for example, once 60 percent of the installed capacity is in use to meet residual demand. The results for Cournot pricing without contracts in PJM are most startling. In Figure 10, we show that for any level of residual demand above 55 percent of installed capacity, the Cournot price would have been at the price cap of $1000/MWh had firms divested as in California.

These results reinforce the perception that the horizontal market structure in the eastern markets, particularly in PJM, is quite uncompetitive, and that vertical arrangements are playing a critical role in mitigating the exercise of market power in the spot market. Several important caveats about our analysis should also be noted. First, our data about long-term contracts is incomplete. Although we observe what we believe are the major long-term arrangements between suppliers and retailers, details of other arrangements, particularly more short-term trades, have not been made public. However, we do know that the contracts signed by retailers in California were minimal, so that any arrangements we have missed will be in the eastern markets. Additional purchase arrangements by retailers in the East would make those markets look less competitive relative to a Cournot calculation, and therefore reinforce our general observation that California's market design should not bear the blame for the crisis there.

Second, a discussion about the potential endogeneity of market design and market structure is appropriate. We have shown that, after controlling for known vertical arrangements, the market structure in each market produces outcomes relatively similar to Cournot equilibria. However, one could argue that the market design influenced both the horizontal structure and the decision to enter into vertical arrangements. We feel that it is implausible that the market design, at least when defined as meaning market rules, significantly impacted the market structure during the period that we study. The horizontal structure was determined from sales that were for the most part initiated, if not completed, before these markets began operating. The vertical arrangements for the period that we study were determined at roughly the same time as the asset sales. This is one of the reasons why we focus on a time period relatively early in the life of these markets: the vertical arrangements are more well-understood and can reasonably be considered to be exogenous. Going forward, given the apparent importance of vertical arrangements, it will be an important line of research to better understand what kinds of market environments produce various vertical arrangements.

Last, our analysis has focused on only one, albeit central, aspect of restructured markets: the average wholesale price of electric energy. Many other attributes of electricity markets, such as the costs of reserve capacity, the ability to disseminate accurate prices over space and time, and the efficiency of power plant operations and investment, should be considered before rendering

judgment over which market has produced the 'best' performance.

6 Conclusions

Within the U.S., experiences with electricity restructuring have varied dramatically. While the consequences of such initiatives has proved disastrous in California and Montana, regulators and policy-makers are for the most part satisfied with the performance of restructured markets in New York, Texas, New England and PJM. There has been much speculation and dispute over the reasons why the deregulation experiment has produced such dramatically different results to date in the various regions of the country. Some have argued that high costs and low production capacity is to be blamed in the markets that have been deemed failures. Others point to difference in the design of the wholesale market rules as the source of the variation. The FERC, the regulatory institution responsible for oversight of wholesale electricity markets, has continued to focus on market design both in its proposed remedies for California, and in its broader initiatives to apply a standard market design to all the markets under its jurisdiction.

Each market appears to have experienced at least some hours with competitiveness problems, yet previous studies have supported the conventional wisdom that the eastern markets were much more competitive than was California. However, no previous study has systematically addressed the question as to why the eastern markets were more competitive. Three aspects of the markets appear to have contributed, the horizontal structure, the market design, and the vertical relationships between producers and retailers, most of whom faced serious constraints on their ability to raise retail prices. While it is extremely difficult to examine the impacts of myriad differences in market rules in a cross-market comparison, we can estimate the effect of horizontal and vertical structure. The differences that remain likely reflect the impacts of market rules.

We examine the impact of market structure by abstracting away from specific market rules and estimating the market prices that would result from Cournot competition in each of these markets. We also estimate the prices that would result from all firms adopting a price-taking position. While other non-cooperative equilibrium concepts, notably the supply function equilibrium, could be applied these other forms of oligopoly competition are bounded by the two sets of equilibria we do model. We estimate market outcomes under two vertical structures, one in which suppliers have no long-term retail obligations, and one in which the retail obligations of producers that are currently public information are included in the objective function of those producers. We apply this approach to three of the oldest and largest markets operating in the U.S., California, New England, and PJM and exam the summer of 1999.

We find that the vertical relationships between producers and retailers play a key role in determining the competitiveness of the spot markets in the markets that we study. The concen-

tration of ownership and low elasticity of import supply combine to give PJM by far the least competitive horizontal structure. If one ignores the vertical arrangements a Cournot equilibrium reaches the price-cap in PJM in a majority of hours. Yet the PJM market was in fact fairly competitive except during very high demand hours. Although not as severe, we find a similar dramatic contrast between a Cournot equilibrium with no vertical arrangements and actual market prices in New England. Once the known vertical arrangements are explicitly modeled as part of the Cournot equilibrium, the Cournot prices are dramatically reduced and are reasonably similar to actual prices.

These results carry important implications for both electricity restructuring and anti-trust policies. The horizontal structure of the markets is important, but similar horizontal structures can produce dramatically different outcomes under different vertical arrangements. While much energy has been spent on establishing a standardized market design and on regulating the behavior of individual producers in the electricity industry, far less attention has been paid to the retail side of these markets. Retail competition, at least for residential customers, is nearly moribund in many parts of the country including the 'successful' eastern markets. Yet the transition arrangements that produced many of the vertical arrangements in the East will be expiring over the next several years. Whether and how those arrangements are continued or replaced will likely play a key role in the future success of these markets.

References

[1] Allaz, Blaise and Jean-Luc Vila. 1993. "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, 53(1), February, pp. 1-16.

[2] Borenstein, S., J.B. Bushnell, and C. Knittel (1999). "Market Power in Electricity Markets: Beyond Concentration Measures." *Energy Journal*, 20(4): 65-88.

[3] Borenstein, S., J.B. Bushnell, and F. Wolak (2002). "Measuring Market Inefficiencies in California's Deregulated Wholesale Electricity Market." *American Economic Review*, 92(5): 1376-1405.

[4] Bushnell, J.B. (2003a). "Looking for Trouble: Competition Policy in the U.S. Electricity Industry."*Electricity Restructuring: Choices and Challenges.* J. Griffen and S. Puller, eds. University of Chicago Press.

[5] Bushnell, J.B. (2003b). "A Mixed Complementarity Model of Hydro-Thermal Electricity Competition in the Western U.S." *Operations Research.* 51(1): 80-93.

[6] Bushnell, J.B. (2004). "California's Electricity Crisis: A Market Apart?." *Energy Policy.* 32(9):1045-1052. June.

[7] Bushnell, J.B., and C. Saravia (2002). "An Empirical Analysis of the Competitiveness of the New England Electricity Market." CSEM Working Paper WP-101, University of California Energy Institute. May. Available at www.ucei.org.

[8] Davidson, Russell and James G. MacKinnon (1993) *Estimation and Inference in Econometrics*, Oxford University Press, 875 pages.

[9] Dirkse, S.P., and M.C. Ferris (1995). "The PATH solver: Anon-monotone stabilization scheme for mixed complementarity problems.*Optimization Methods and Software,* 5:123-156.

[10] Evans, J., and R. Green (2003). "Why did British electricity prices fall after 1998?" Working Paper. Center for Economic Policy. University of Hull. Available at http://www.hull.ac.uk/hubs/staff/profiles/staff/evans_green_neta_prices.pdf.

[11] Green, R.J. (1999). "The Electricity Contract Market in England and Wales" *Journal of Industrial Economics*, Vol XLVII, No 1, pp.107-124.

[12] Green, R.J., and D.M. Newbery (1992). "Competition in the British Electricity Spot Market," *Journal of Political Economy,* October, 100(5): 929-53.

[13] Hobbs, B. (2001). "Linear Complimentarity Models of Nash-Cournot Competition in Bilateral and POOLCO Power Markets." *IEEE Transactions on Power Systems.* 16(2): 194-202, May.

[14] Joskow, P.J., and E.P. Kahn (2002). "A Quantitative Analysis of Pricing Behavior In California's Wholesale Electricity Market During Summer 2000," *Energy Journal,* 23(4): 1-35.

[15] Joskow, P.J.(2003). "The Difficult Transition to Competitive Electricity Markets in the US,"*Electricity Restructuring: Choices and Challenges.* J. Griffen and S. Puller, eds. University of Chicago Press.

[16] Klemperer, P.D., and M.A. Meyer (1989). "Supply Function Equilibria in Oligopoly Under Uncertainty," *Econometrica.* November. 57(6): 1243-77.

[17] Mansur, E.T. (2003). "Vertical Integration in Restructured Electricity Markets: Measuring Market Efficiency and Firm Conduct," *CSEM Working Paper PWP-117.* October. Available at www.ucei.org.

[18] Puller, Steven L. 2000. "Pricing and Firm Conduct in California's Deregulated Electricity Market," *POWER Working Paper PWP-080.*

[19] Rudkevich, A., M. Duckworth, and R. Rosen (1998), "Modeling Electricity Pricing in a Deregulated Generation Industry: The Potential for Oligopoly Pricing in a Poolco," *Energy Journal*, 19(3): 19-48.

[20] Schmalansee, R. and Golub, B. (1984), "Estimating Effective Concentration in Deregulated Wholesale Electricity Markets," *RAND Journal of Economics*, 15, pp. 12-26.

[21] Wilson, R.B. (2002). "Architecture of Power Markets," *Econometrica* 70(4): 1299-1340.

[22] Wolak, Frank A. 2000. "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market," *International Economic Journal*, 14(2): 1-39.

Appendix: Complementarity Formulation of Cournot-Nash Equilibrium

We assume that the market demand $Q_t = a_t - b\ln(p_t)$, or $p_t = e^{\frac{a_t - Q_t}{b}}$. While the marginal cost curves of most electricity companies are not strictly linear, they can be very closely approximated with a piecewise linear function. Let $q_i^{Th,j}$ represent the thermal production of type j from firm i with associated marginal cost $c(q_i^{Th,j}) = K_i^j + c_i^j q_i^{Th,j}$where each thermal production type represents a different segment along a piecewise linear marginal cost curve. The production capacity of each segment, $q_{i,\max}^{Th,j}$is such that $K_i^j + c_i^j q_{i,\max}^{Th,j} \leq K_{ij+1}$, thereby producing a non-decreasing marginal cost curve.

The thermal capacity of fringe firms is aggregated into a single, price-taking fringe firm, with piecewise linear marginal production cost, where each segment j, of thermal production has a corresponding marginal cost of $c(q_f^{Th,j}) = K_f^j + c_f^j q_f^{Th,j}$.

Equilibrium Conditions

Under the assumptions of piecewise linear marginal costs and linear demand, the first order conditions presented in section 3 reduce to the following set of mixed linear complementarity conditions.

$$\text{For } q_{it}^{Th,j}, \forall i \neq f, j, t: \qquad \text{(CO1)}$$

$$0 \geq \left(1 - \frac{\left(q_{it} - q_{it}^C\right)}{b_t}\right) e^{\left(\frac{a_t - \sum_l q_{lt}}{b_t}\right)} - K_i^j - c_i^j q_{it}^{Th,j} - \psi_{it}^j \perp q_{it}^{Th} \geq 0 \; ;$$

$$\text{For } q_{ft}^{Th,j}, \forall j, t \; : \qquad \text{(FR1)}$$

$$0 \geq e^{\left(\frac{a_t - \sum_l q_{lt}}{b_t}\right)} - K_f^j - c_f^j q_{ft}^{Th,j} - \psi_{ft}^j \perp q_{ft}^{Th,j} \geq 0 \; ;$$

$$\text{For } \psi_{it}^j, \forall i, j, t: \quad 0 \leq \psi_{it}^j \perp q_{it}^{Th,j} \leq q_{it,\max}^{Th,j}; \; (g1)$$

where the symbol $\perp$ indicates complementarity. Simultaneously solving for the dual and primal variables $\{q_{it}^{Th,j}, \psi_{it}^j\}$ for all i, t produces an equilibrium of the multi-period game. For n producers (including the fringe), J segments to the thermal marginal cost curve, and T time periods, the above conditions produce $2nJT$complementarity conditions or equality constraints for the same number of variables. The system of equations is therefore a 'square' complementarity problem with a solution. Although the profit function is not strictly concave, it can be shown that profits are quasi-concave and strictly concave at the point where the first-order condition (CO1) is satisfied. The solution to this system of equations therefore constitutes a Nash-Cournot equilibrium where each firm has set output at a globally profit-maximizing level, given the output of the other firms.

Figures and Tables

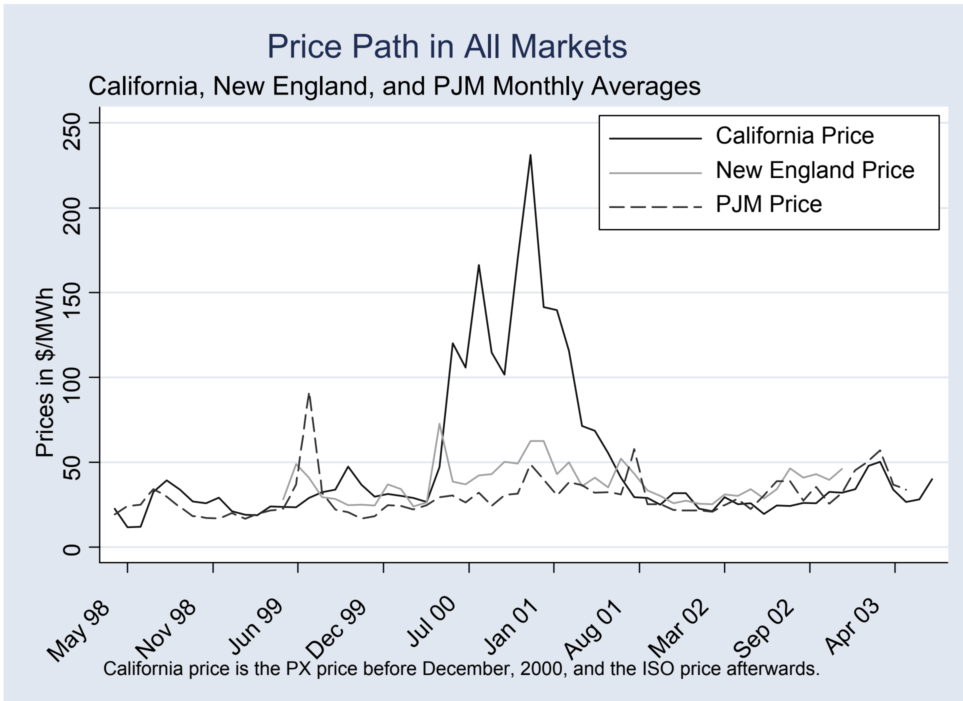


Figure 1:

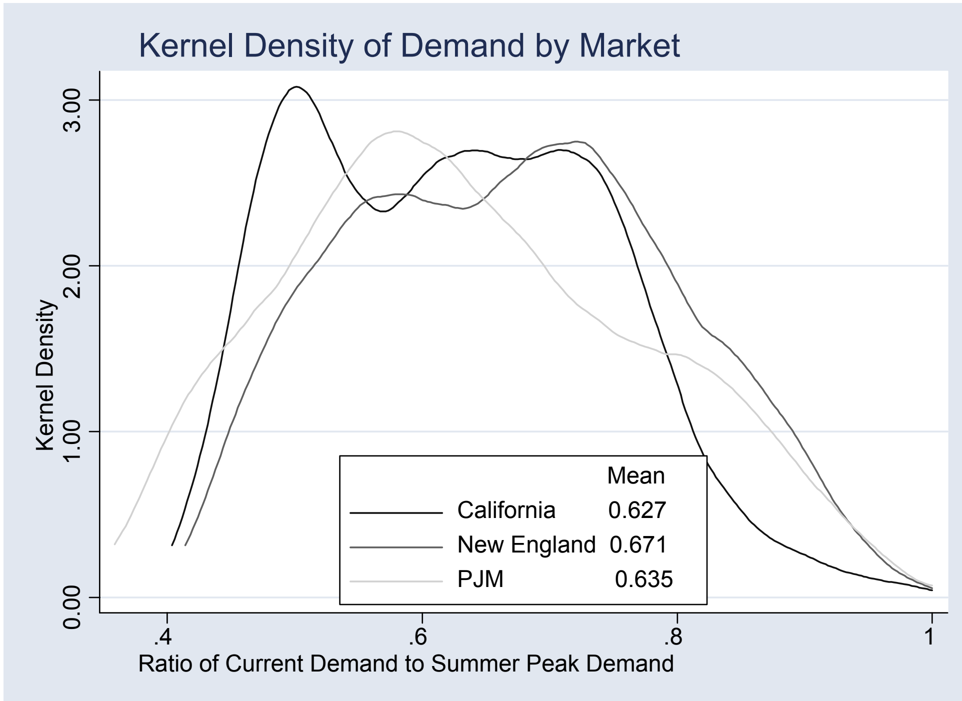


Figure 2:

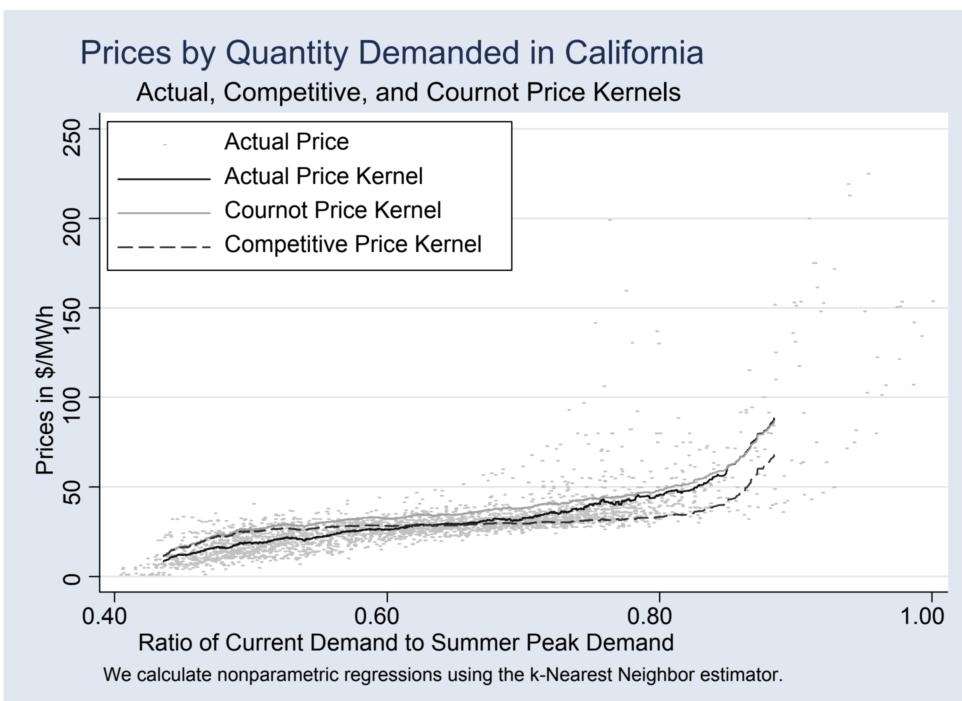


Figure 3:

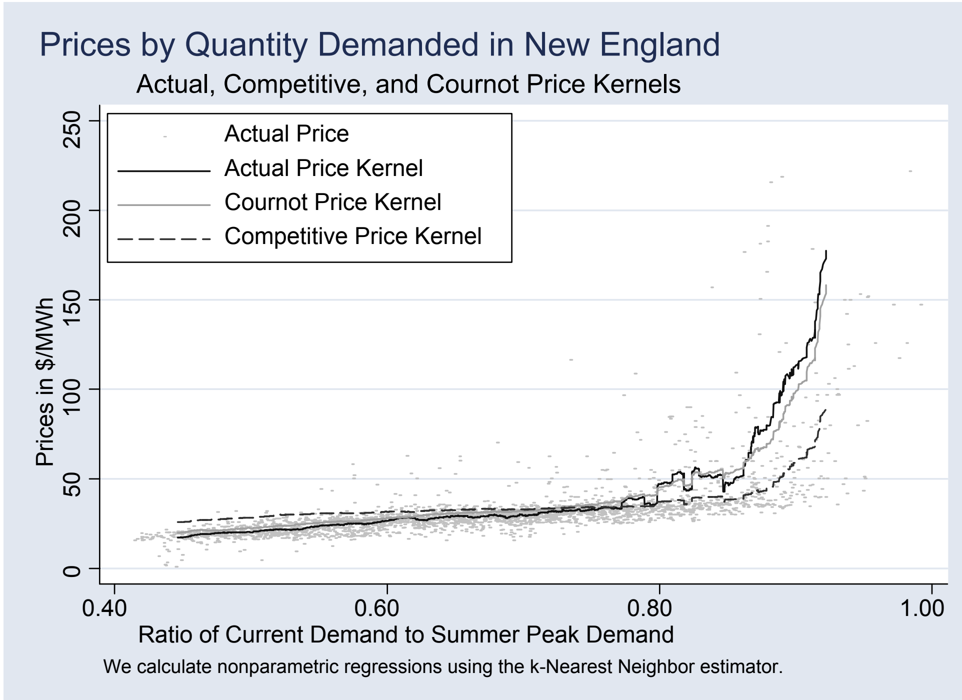


Figure 4:

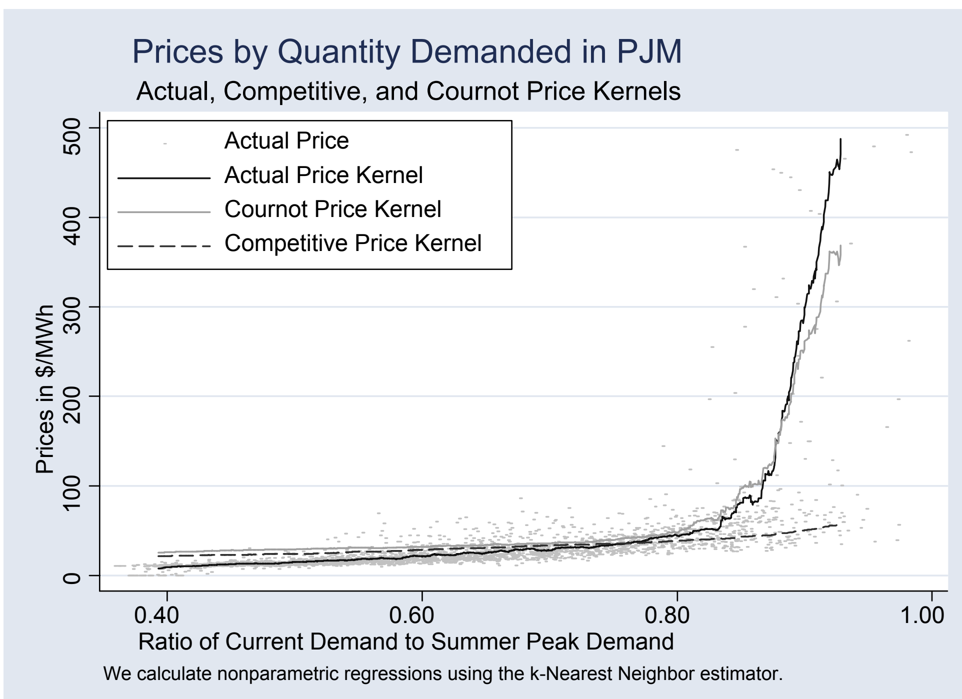


Figure 5:

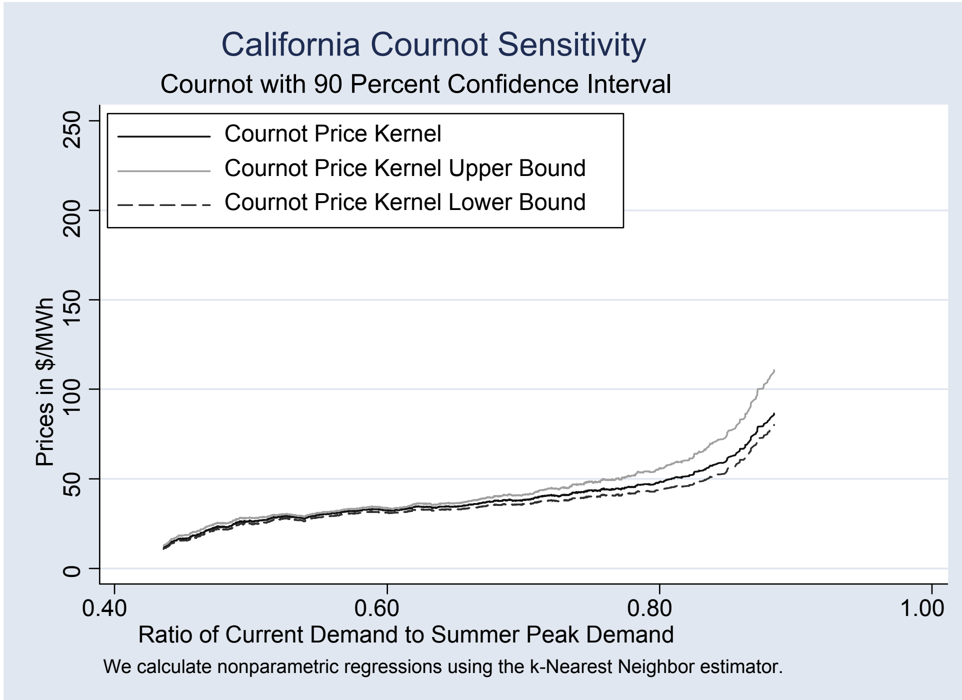


Figure 6:

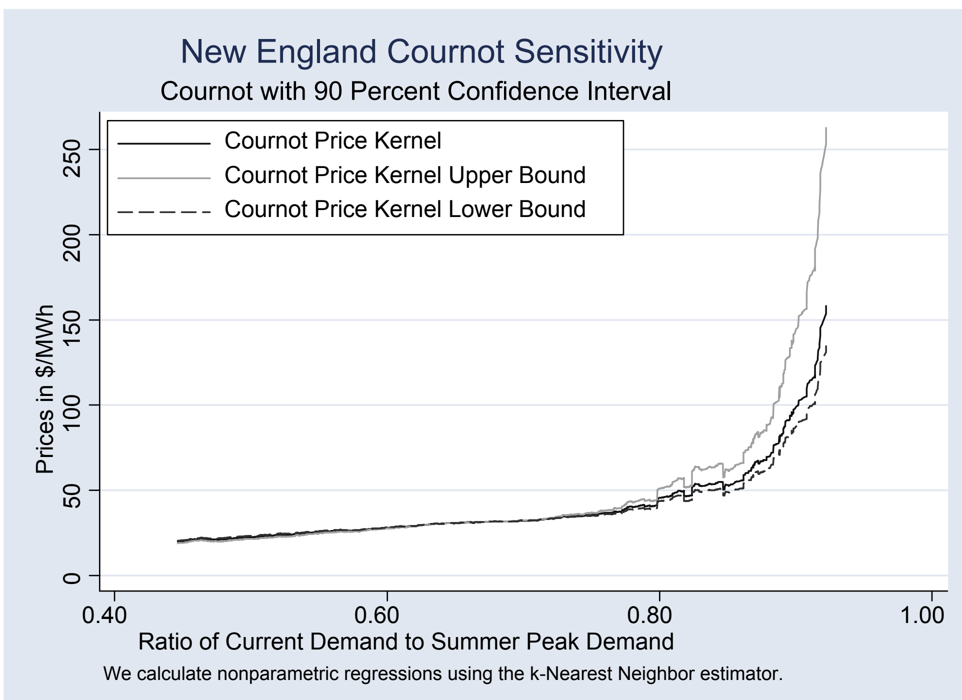


Figure 7:

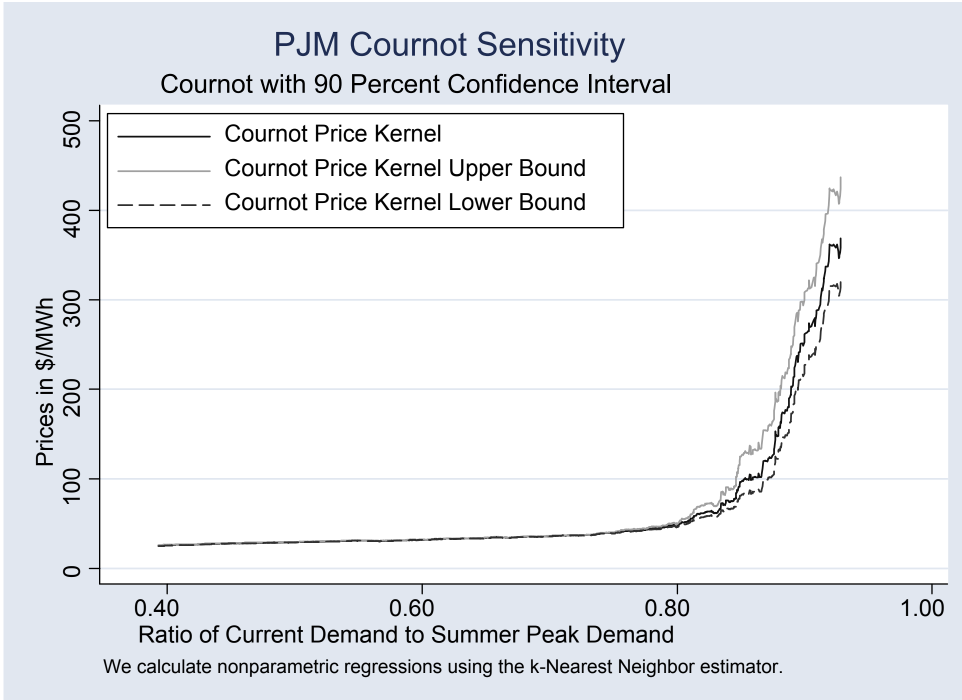


Figure 8:

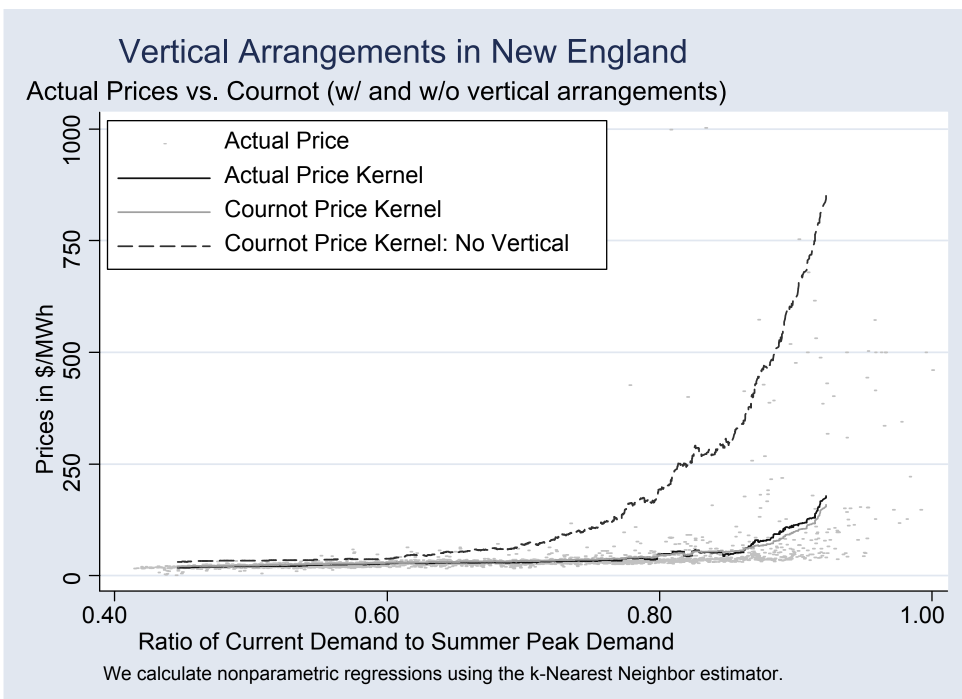
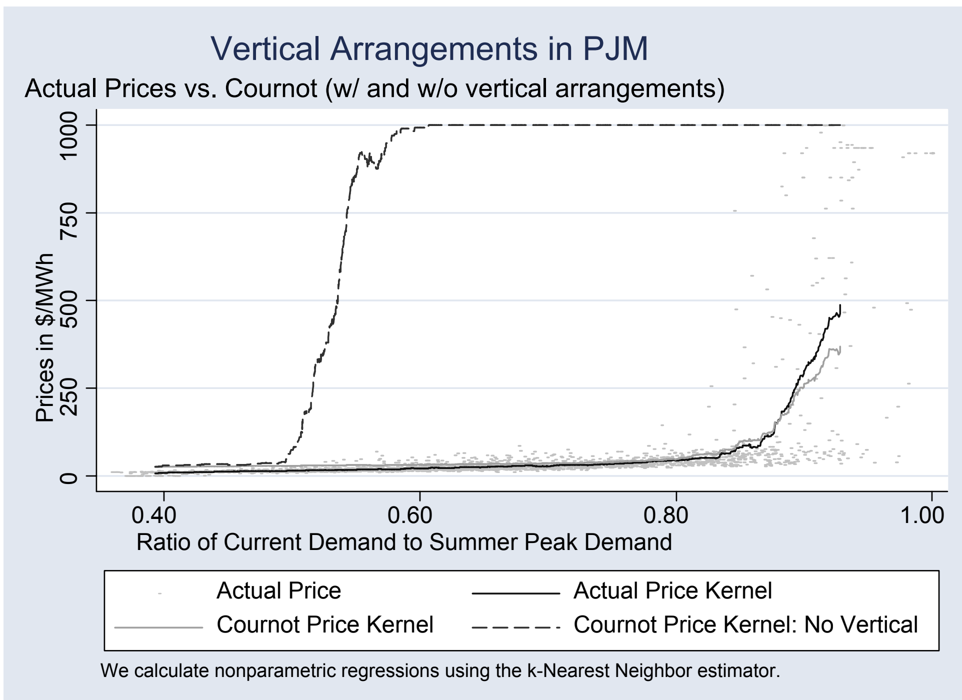


Figure 9:



Vertical Arrangements in PJM
Actual Prices vs. Cournot (w/ and w/o vertical arrangements)
Prices in $/MWh
0
250
500
750
1000
0.40
0.60
0.80
1.00
Ratio of Current Demand to Summer Peak Demand
Actual Price
Actual Price Kernel
Cournot Price Kernel
Cournot Price Kernel: No Vertical
We calculate nonparametric regressions using the k-Nearest Neighbor estimator.

Figure 10:

Table 1: Firm Characteristics for Each Market: Summer 1999

Panel A: California Firm Characteristics.

Firm	Fossil	Water	Nuclear	Other	Output Max	Output Share	Load Max	Load Share
PG&E	570	3,878	2,160	793	7,400	17%	17,676	39%
AES/Williams	3,921	-	-	-	3,921	9%	-	-
Reliant	3,698	-	-	-	3,698	8%	-	-
Duke	3,343	-	-	-	3,343	8%	-	-
SCE	-	1,164	2,150	-	3,314	8%	19,122	42%
Mirant	3,130	-	-	-	3,130	7%	-	-
Dynegy/NRG	2,871	-	-	-	2,871	6%	-	-
Other	6,617	5,620	-	4,267	16,504	37%	9,059	20%
Total	24,150	10,662	4,310	5,060	44,181		45,857	

Panel B: New England Firm Characteristics.

Firm	Fossil	Water	Nuclear	Other	Output Max	Output Share	Load Max	Load Share
Northeast Util.	3,884	1,406	-	175	5,465	21%	7,560	34%
PG&E N.E.G.	3,264	1,152	-	165	4,581	18%	4,440	20%
Mirant	1,310	-	-	16	1,943	8%	-	-
Sithe	1,904	-	-	-	1,904	7%	-	-
FP&L Energy	970	365	-	-	1,335	5%	-	-
Wisvest	980	-	-	-	980	4%	1,200	5%
Duke Energy	474	-	-	13	487	2%	-	-
Other	2,291	1,663	4,359	819	9,132	35%	9,281	41%
Total	17,276	3,180	4,359	1,013	25,828		22,481	

Panel C: PJM Firm Characteristics.

Firm	Fossil	Water	Nuclear	Output Max	Output Share	Load Max	Load Share
Public Service Elec.	6,760	-	3,510	10,270	18%	8,947	17%
PECO	3,682	1,274	4,534	9,490	17%	4,551	9%
GPU, Inc.	7,478	454	1,513	9,445	17%	7,602	15%
PP&L Inc.	6,102	148	2,304	8,554	15%	5,120	10%
Potomac Electric	6,507	-	-	6,507	11%	5,378	10%
Baltimore G & E	3,945	-	1,829	5,774	10%	5,792	11%
Delmarva P & L	2,458	-	-	2,458	4%	3,103	6%
Edison	2,012	-	-	2,012	4%	0	0%
Atlantic City Electric	1,309	-	-	1,309	2%	2,224	4%
Other	428	439	-	867	2%	8,998	17%
Total	40,681	2,316	13,690	56,685		51,714	

Table 2: Two Stage Least Squares Estimation of Fringe Supply from June to September, 1999

Panel A: First-stage dependent variable is log of hourly prices by market.

	California	New England	PJM
ln(Load)	2.46*	1.97*	3.50*
	(0.22)	(0.17)	(0.25)
R-squared	0.41	0.19	0.26
AR(1) coef (ρ)	0.59	0.83	0.87
Sample size	2,922	2,927	2,890

Panel B: Second-stage dependent variable is hourly fringe supply by market.

	California	New England	PJM
ln(Price)	5040.8*	1246.1*	855.8*
	(686.5)	(155.7)	(117.2)
R-squared	.	0.23	.

Notes: Table presents 2SLS coefficients. First we estimate 2SLS and use the errors to correct for serial correlation by estimating an AR(1) coefficient (ρ). Then we quasi-difference the data by calculating $\Delta x = x_t - \rho x_{t-1}$ for all data. We re-estimate the 2SLS results using these quasi-differenced data. Robust standard errors are given in parentheses. Significance is marked with (*) at the 5% level and ($^{\#}$) at the 10% level. Regression includes fixed effects for month of year, day of week, and hour of day. Also weather variables for bordering states are included and modeled as quadratic functions for cooling degree days (degrees daily mean below 65° F) and heating degree days (degrees daily mean above 65° F). In the first stage, we regress ln(price) on the exogenous variables and instruments of log of hourly load (MWh) in each market.

Table 3: Actual Prices and Estimates of Competitive and Cournot Prices

Prices by Market and Time of Day (Peak and Off-Peak) during the Summer of 1999.

Panel A: All Hours.

Variable	Mean	Std. Dev.	Min	Max
California Actual	29.69	19.10	1.00	224.99
Competitive	28.65	12.90	1.18	233.00
Cournot	34.03	15.36	1.19	233.34
New England Actual	36.96	56.66	1.00	1003.21
Competitive	34.66	19.67	10.27	357.43
Cournot	38.69	48.17	7.09	712.80
Cournot n.v.a.	150.85	223.33	0.00	999.99
PJM Actual	45.92	122.76	0.00	999.00
Competitive	31.13	8.10	18.63	72.79
Cournot	60.46	131.73	21.30	1000.00
Cournot n.v.a.	797.45	394.07	18.51	999.99

Panel B: Peak Hours (11AM to 8PM Weekdays).

Variable	Mean	Std. Dev.	Min	Max
California Actual	43.15	26.99	17.16	224.99
Competitive	35.06	19.92	24.70	233.00
Cournot	44.50	20.73	24.97	233.34
New England Actual	55.05	82.86	17.67	753.17
Competitive	41.11	32.19	29.75	357.43
Cournot	59.05	76.97	25.55	712.80
Cournot n.v.a.	303.32	312.36	30.26	999.99
PJM Actual	97.31	210.17	11.22	999.00
Competitive	37.41	9.16	22.77	72.79
Cournot	115.14	225.90	24.61	1000.00
Cournot n.v.a.	998.89	32.65	31.29	999.99

Panel C: Off-Peak Hours.

Variable	Mean	Std. Dev.	Min	Max
California Actual	23.90	9.85	1.00	96.90
Competitive	25.90	6.50	1.18	50.90
Cournot	29.54	9.24	1.19	68.58
New England Actual	29.18	37.95	1.00	1003.21
Competitive	31.88	9.10	10.27	298.37
Cournot	29.95	22.79	7.09	611.73
Cournot n.v.a.	85.33	123.01	0.00	999.99
PJM Actual	23.83	30.90	0.00	677.50
Competitive	28.43	5.80	18.63	56.17
Cournot	36.96	32.62	21.30	771.20
Cournot n.v.a.	710.90	443.46	18.51	999.99

* Note: There are 2,928 hourly observations: 880 peak and 2,048 off-peak. Cournot n.v.a. means no vertical arrangements.