

NBER WORKING PAPER SERIES

VERTICAL ARRANGEMENTS, MARKET STRUCTURE, AND COMPETITION:
AN ANALYSIS OF RESTRUCTURED U.S. ELECTRICITY MARKETS

James B. Bushnell
Erin T. Mansur
Celeste Saravia

Working Paper 13507
<http://www.nber.org/papers/w13507>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2007

We thank Soledad Arellano, Severin Borenstein, Joe Bowring, Judy Chevalier, Phil Haile, Stephen Holland, Jun Ishii, Wally Mullin, Steve Puller, Peter Schott, Fiona Scott Morton, Raphael Thomadsen, Larry White, Matt White, Frank Wolak, Catherine Wolfram, two anonymous referees, and seminar participants at Colby College, International Industrial Organization Conference, Massachusetts Institute of Technology, New York University, University of California Energy Institute, University of California at Irvine, University of Toulouse, Wesleyan University, and Yale University for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by James B. Bushnell, Erin T. Mansur, and Celeste Saravia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured U.S.
Electricity Markets

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

NBER Working Paper No. 13507

October 2007

JEL No. L11,L13,L94

ABSTRACT

This paper examines vertical arrangements in electricity markets. Vertically integrated wholesalers, or those with long-term contracts, have less incentive to raise wholesale prices when retail prices are determined beforehand. For three restructured markets, we simulate prices that define bounds on static oligopoly equilibria. Our findings suggest that vertical arrangements dramatically affect estimated market outcomes. Had regulators impeded vertical arrangements (as in California), our simulations imply vastly higher prices than observed and production inefficiencies costing over 45 percent of those production costs with vertical arrangements. We conclude that horizontal market structure accurately predicts market performance only when accounting for vertical structure.

James B. Bushnell
University of California Energy Institute
2547 Channing Way
Berkeley, CA 94720
and NBER
bushnell@haas.berkeley.edu

Celeste Saravia
University of California Energy Institute
2547 Channing Way
Berkeley, CA 94720-5180
csaravia@econ.Berkeley.EDU

Erin T. Mansur
Yale School of Management
135 Prospect Street
P.O. Box 208200
New Haven, CT 06520-8200
and NBER
erin.mansur@yale.edu

While rules concerning horizontal market structure form the basis of antitrust policies in most countries, it is widely recognized that horizontal structure comprises only one piece of the competition puzzle. Vertical integration and other vertical arrangements between wholesalers and retailers will also impact the incentives of firms. In addition, regulators and many economists have focused on the effects that market rules, such as auction design, may have on equilibrium prices. This paper empirically examines the relative importance of horizontal market structure and vertical arrangements in determining prices in imperfectly competitive markets.

We study three U.S. electricity markets: California, New England, and the Pennsylvania, New Jersey, and Maryland (PJM) market. These were the first three U.S. markets to undergo regulatory restructuring. There is substantial evidence that the California market was the least competitive.¹ However, previous studies do not address *why* there were apparent differences in the competitiveness of these markets.

Events like the California electricity crisis have contributed to a perception that electricity markets are fundamentally different from other commodity markets. The complexity of market rules as well as processes for the production and delivery of electricity have added weight to this argument. Our results demonstrate that, in fact, fundamental concepts of oligopoly competition do apply to, and are significantly informative about, the restructured electricity industry. In the markets we study, prices did not diverge greatly,

¹See Severin Borenstein, James B. Bushnell, and Frank A. Wolak (henceforth BBW, 2002), Bushnell and Celeste Saravia (2002), Paul J. Joskow and Edward P. Kahn (2002), Erin T. Mansur (2007), and Steven L. Puller (2007).

especially during the important peak hours, from those predicted by a slightly modified model of Cournot competition, *but only* after we account for the vertical commitments made by producers.

This paper reconstructs market conditions, in detail, for the three markets. First, we calculate cost functions for the important market participants. Then, with data on firms' vertical commitments and hourly demand, we simulate market outcomes under differing assumptions of firm behavior. Specifically, we estimate outcomes under an assumption of perfect competition and under an assumption of Cournot competition. These two counterfactual assumptions bound the space of possible static, noncooperative outcomes. By establishing where actual market outcomes fall within these bounds, we can compare how markets are performing relative to the extremes determined by structural factors alone.²

Our analysis highlights the importance of vertical integration and vertical contracts, although for reasons that have not received much treatment in the literature on vertical structure. Much of the concern about the negative impacts of vertical arrangements has focused on foreclosure, or the ability of the integrated firm to raise rivals' costs.³ However, in the markets we study, third-party independent system operators control the common

²Several papers have applied models of oligopoly competition to electricity markets to forecast possible future market outcomes using hypothetical market conditions (Richard Schmalensee and Bennett W. Golub, 1984; Richard J. Green and David M. Newbery, 1992; Borenstein and Bushnell, 1999; and Benjamin Hobbs, 2001). Unlike those papers, we are applying actual market data to simulate market outcomes within the Cournot framework.

³For example, Thomas G. Krattenmaker and Steven C. Salop (1986), Richard Gilbert and Justine Hastings (2005), and Patrick Rey and Jean Tirole (forthcoming) describe various strategies for exclusion and raising rivals' costs.

distribution networks. This, combined with the fact that electricity is a homogenous commodity, makes it somewhat more difficult for suppliers to foreclose competitors or discriminate in favor of their own retail affiliates.

Instead, it is the rigidity of retail prices that creates a strong relationship between vertical structure and competitive performance. Regulators constrain retailers in their ability to adjust electricity prices. The rate-making process restricts the frequency with which retail prices can be adjusted, usually no more frequently than annually. This is particularly true during our sample period, where transition arrangements put in place at the time of restructuring strongly restricted retailers' ability to adjust prices. Even in completely deregulated retail markets, however, price commitments more than a year in length are not unusual. The key attribute is that integrated firms are making retail price commitments *before* committing production to the wholesale market.

A restriction on retail price adjustment means that a producer is effectively making a long-term forward commitment when it integrates with downstream retailers. Vertical relationships take the form of long-term price commitments to retail customers. The integrated firm has committed to supplying a portion of its output at fixed prices to its retail customers and therefore has an effectively smaller position on the wholesale market and less incentive to raise wholesale prices. The impact on the incentives of wholesale producers is analogous to that provided by a futures contract or other hedging instrument, which is generally thought to be procompetitive.⁴ Firms effectively undercut each other in the

⁴In particular, Blaise Allaz and Jean-Luc Vila (1993) note that an oligopoly equilibrium will be more competitive when there are more opportunities for firms to contract ahead of the time of delivery.

forward market in an attempt to gain a Stackelberg leader position. These considerations have been shown to be relevant to electricity markets.⁵ Many other industries exhibit similar vertical relationships whereby the presence of long-term, fixed-price contracts are likely to influence the spot market.⁶

In our study, the impact of vertical arrangements on estimated market outcomes is striking. When vertical arrangements are not taken into account, 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. We use publicly available data on long-term retail supply arrangements, which were present in New England and PJM but not in California. When these retail obligations are included in the objective functions of suppliers, Cournot equilibrium prices in both New England and PJM fall substantially.

In each market, actual prices are similar to our simulated prices that assume Cournot

⁵Green (1999) discusses the theoretical implications of how hedge contracts impact the England and Wales electricity market. In the context of the Australian electricity market, Wolak (2000) examines firm bidding behavior for supplying electricity given long-term contracts. He finds that financial hedging mitigates market power. Natalia Fabra and Juan Toro (2005) find that the retail commitments arising from a regulatory transition mechanism in Spain not only strongly influence producer behavior but provide the foundation for tacit collusion between those producers. Puller and Ali Hortaçsu (forthcoming) incorporate estimates of producer contract positions into their estimates of the optimality of the bidding of Texas energy producers.

⁶Other energy markets—such as railroad coal deliveries and natural gas—are structured similarly where some of the supply is procured through long-term, fixed-price contracts and some through a spot market (or short-term contracts). By contrast, in the gasoline industry retailers change prices with impunity and great frequency. Vertical contracts between refiners and retailers guarantee the supply of physical product, but almost never set an advanced fixed price for that product. To some degree, many imperfectly competitive industries—including concrete, construction, telecommunications, and pharmaceuticals—feature both wholesale forward price commitments and spot markets that foster competition.

behavior and that account for vertical arrangements. This is not to imply that firms followed strict Cournot strategies, but rather that the more complex supply function strategies they did adopt produced prices similar to Cournot equilibrium ones. With the vertical arrangements, the upper bounds on these static, noncooperative oligopoly prices are greatly reduced. In this sense, horizontal structure does explain performance but only when coupled with these arrangements. After accounting for these structural factors, there is relatively little variation between markets left to be explained by market rules, local regulation, and other factors. 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. Furthermore, had regulators in PJM and New England impeded vertical arrangements as in California, our simulations suggest that not only would prices have been much higher than they were, but also there would have been substantial welfare loss: production costs would have increased more than 45 percent relative to those costs with vertical arrangements.

This paper proceeds as follows. In Section I, we provide a general overview of the relevant characteristics of the three markets we study. Section II describes our model. We present our results in Section III and state our conclusions in Section IV.

I Electricity Markets' Structure and Design

The markets we study essentially share the same general organizational structure. Electricity deregulation has been quite limited in scope, focusing on wholesale pricing and re-

tailing.⁷ Most large producers were granted authority to sell power at deregulated prices. Distribution and transmission sectors remain regulated, but have been reorganized to accommodate wholesale markets and retail choice. Most utilities have retained ownership of transmission lines, but have relinquished the day-to-day control of the network to Independent System Operators (ISOs). ISOs operate electricity systems and provide market participants with equal access to the network. ISOs oversee at least one organized exchange through which firms can trade electricity. These ISO spot, or “balancing,” markets clear a set of supply offers against a perfectly inelastic demand quantity based upon the actual system needs for power during that time interval.

Despite commonalities in their general organization, market performance has varied dramatically across the California, New England, and PJM markets. Figure 1 illustrates the monthly average prices for the major price indices in each of the three markets from 1998 through the spring of 2003. California began operating in April, 1998. New England opened in May, 1999. The PJM market opened in the spring of 1998, but firms did not receive permission to sell at market-based rates until April, 1999. Prices for 1998 were therefore the product of regulated offer prices into the PJM market-clearing process. 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.

⁷In fact, wholesale electricity markets are not technically deregulated. Under the Federal Power Act, the Federal Energy Regulatory Commission (FERC) 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 rather than regulatory determined, cost-based rates (see Joskow, 2003).

There has been much speculation and debate about the causes of these price differences. Relative production costs, fuel prices, and overall demand play an important part in market outcomes.⁸ In addition, we examine the extent to which the price variation among the three markets resulted from substantial variation in horizontal structure and vertical relationships.

Sample Period

For several reasons, we restrict our sample period to the summer of 1999. Our study period (June 1 to September 30) was the initial high-demand period after all three markets were restructured. As discussed below, the vertical arrangements of firms are arguably *exogenously* determined in this period. Furthermore, market rules were relatively *stable* in our sample period. For example, the number of organized exchanges changed in 2000.⁹ As this study aims to parse out the effects of horizontal structure, vertical structure, and market rules, we chose a period when these features are both stable and potentially exogenously assigned.

As with any study, one should be cautious in making out-of-sample predictions based on our sample period.¹⁰ Nevertheless, the conditions in 1999 were fairly typical for this

⁸For example, the extremely high gas prices of the winter of 2000-2001 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.

⁹During the time period of this study, there were two separate markets for electricity in California: a day-ahead futures market (the Power Exchange, or PX) and a real-time spot market for electricity. Early in their operations, the PJM and New England markets featured only a single real-time spot market for electricity overseen by their respective ISOs. After the period of our study, several changes occurred. The California PX stopped operating in January 2001. A day-ahead market began in PJM in June 2000, while ISO-NE began operating a day-ahead market in early 2003.

¹⁰For example, even though we find that prices are consistent with Cournot competition (accounting for vertical arrangements), it is possible that extreme changes in market rules would have substantial

industry. This period featured some hot spells in the mid-Atlantic states but relatively mild weather in California (although California's summer peak demand was actually greater in 1999 than during 2000). Our sample period includes a substantial range of market conditions in each market.¹¹

A Horizontal Structure

For each market, Table 1 summarizes the market structure. We report the megawatts (MW) and market shares of generation capacity (*Output Max*) and retail demand during the highest peak hour (*Load Max*).¹² The firms described in the table compose the set of strategic producers. The non-strategic, competitive fringe equals the aggregation of generation from firms owning less than 800 MW of capacity in any market.¹³ By conventional measures, the PJM market, with a Herfindahl-Hirschman Index (HHI) of nearly 1400, is much more concentrated than either New England (850) or California (620).

With a peak demand of 45,000 MW and similar installed capacity, California relies heavily on imports. In 1999, California imported about 25 percent of the electricity it consumed. New England, with an installed capacity of about 26,000 MW, is the smallest market we study. It imports energy to meet approximately 10 percent of demand because

implications on market performance.

¹¹However, some notable periods, such as the summer of 2000 in California, are not represented. BBW and Bushnell and Saravia (2002) examine nearly three years' worth of market operations each and find that the overall competitiveness of the market is consistent across the years when one controls for overall demand levels.

¹²Appendix A discusses the data sources for generation output and retail obligations, including all long-term contracts.

¹³Thus, any individual fringe firm owns less than 1.8 percent of total capacity in California, 3.1 percent in New England, and 1.4 percent in PJM. This is less than half the size of the smallest modeled strategic player.

many of its power plants use older, gas- and oil-fired technology. PJM has about 57,000 MW capacity using primarily coal, nuclear, and natural gas energy sources. PJM is largely self-contained and imports relatively little power.

Borenstein, Bushnell, and Christopher R. Knittel (1999), among others, note the limitations of conventional structural measures when applied to electricity markets. At least as important as market concentration for non-storable goods are the relationship between production capacity and demand levels, and the elasticity of imported supply. These and other aspects of each market are explicitly incorporated into our oligopoly framework. One last critical factor that can influence the relative competitiveness of markets is the presence of long-term contracts and other vertical commitments. Kenneth Hendricks and R. Preston McAfee (2006) explore the applicability of horizontal measures in the context of markets where vertical considerations are important.¹⁴ We now examine the vertical structure of our three markets.

B Retail Policies and Vertical Arrangements

The retail function in electricity differs from most other industries: regulators severely constrain firms' ability to adjust retail prices. Following the restructuring of most U.S. electricity markets, incumbent utilities were required to freeze retail rates for several years.

¹⁴Hendricks and McAfee focus on the role of buyer market power in offsetting the market power of sellers. They develop a modified HHI, or MHI, in which the downstream and upstream concentrations are weighted according to the elasticity for the downstream product. Our findings are analogous to the case where the downstream product is very elastic, as retail firms are very limited in their ability to adjust retail prices. In this case, Hendricks and McAfee find that it is the size of a firm's net position in the upstream market that matters. Large net sellers and net buyers would distort wholesale prices, while firms that are "balanced" (i.e., small net position in the upstream market) will have no incentive to impact upstream prices.

Although entrants were not explicitly bound, the fixed prices of the largest retailers served as caps on all retailers' rates. Customers could always elect to remain with the incumbent. Retailers were vulnerable to wholesale price volatility and responded differentially across the three markets.

In PJM, retailers retained their generation assets. Vertical integration provided a physical hedge against high wholesale prices and dampened wholesalers' incentive to set high 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." For example, PECO was a net seller: in 1999, it owned 17 percent of the generating capacity but only had retail obligations for 9 percent of the market's peak demand. Mansur (2007) examines the relative production decisions of PJM 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 during 1998 as compared to the other, net-buying firms.

In New England, the divestiture of generation from vertically integrated utilities was widespread. To hedge their price exposure, retail utilities signed long-term supply contracts, often with the firms to whom they had divested their generation.¹⁵ In their study

¹⁵The largest producer in New England during our sample period, Northeast Utilities (NU), was in the process of divesting most of its generation during 1999, but these transactions were not finalized until

of the New England electricity market, Bushnell and Saravia (2002) 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 spot markets. During the summer of 1999, there were almost no meaningful long-term arrangements between merchant generation companies and the incumbent utilities. The largest utilities, Pacific Gas & Electric (PG&E) and Southern California Edison, did retain control of nuclear and hydrogeneration capacity. However, this low marginal cost capacity limited the utilities’ ability to exercise monopsony power. 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 (Bushnell,

after September. During the summer of 1999, NU was 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 New England Electric System (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, respectively. The Sithe contract had expired by the summer of 1999, while the Wisvest contract expired the following year.

2004, further discusses the complexity of this claim).

Long-term vertical arrangements have been shown to affect market performance significantly. However, 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.

II Description of Model and Data

We measure the impact of vertical and horizontal market structure on market performance. Our approach abstracts away from the detailed market rules and regulations in each market and examines 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, noncooperative equilibrium. These bounds are, respectively, represented by the Nash-Cournot and perfectly competitive equilibria.

Several models of oligopoly competition in the electricity industry have employed the supply function equilibrium (SFE) concept developed by Paul D. Klemperer and Margaret A. Meyer (1989).¹⁶ 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 rules and local regulatory differences influence market outcomes by helping determine which of the many possible equilibria arise, these impacts can be

¹⁶For example, see Green and Newbery (1992) and Aleksandr Rudkevich, Max Duckworth, and Richard Rosen (1998).

thought of as placing the market price within these bounds.¹⁷

In this section, we briefly describe our equilibrium model and how we apply data from various sources to our calculations. In Appendix A, we describe the data sources on equilibrium quantities and prices, as well as the data on vertical arrangements and long-term contracts. The data sources share features with those used by Borenstein, Bushnell, and Wolak (henceforth BBW, 2002), Bushnell and Saravia (2002), and Mansur (2007) in studying the markets of California, New England, and PJM, respectively. Relative to these other papers, there are several substantial differences in the application of the data to our model. The discussion below focuses on these issues.

A Model

We first consider a general formulation of Cournot competition at the wholesale and retail levels. Strategic firms are assumed to maximize profits according to the Cournot assumption using production quantities as the decision variable. The total production of firm i is represented by $q_{i,t}$. Retail sales are denoted $q_{i,t}^r$.

For each strategic firm $i \in \{1, \dots, N\}$ and time period $t \in \{1, \dots, T\}$ that are assumed to be independent, firm i maximizes profits:

$$(1) \quad \pi_{i,t}(q_{i,t}, q_{i,t}^r) = p_t^w(q_{i,t}, q_{-i,t}) \cdot [q_{i,t} - q_{i,t}^r] + p_{i,t}^r(q_{i,t}^r, q_{-i,t}^r) \cdot q_{i,t}^r - C(q_{i,t}),$$

¹⁷Green 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.

where $q_{-i,t}$ and $q_{-i,t}^r$ are the quantity produced and retail supplied by the other $N - 1$ firms, respectively, and p_t^w and $p_{i,t}^r$ are the wholesale and retail market prices. Wholesale electricity is assumed to be a homogenous commodity with a uniform price. Note that retail commitments could be larger than wholesale production so that $q_{i,t} - q_{i,t}^r$ could be negative, meaning that firm i is a net purchaser on the wholesale market.

In the general formulation, the equilibrium positions of firms would take into account both wholesale and retail demand elasticities as well as production capacity and costs.¹⁸ However, in our context we take advantage of the fact that, by time t , both retail quantity and prices are fixed. Considering that both contract quantity and price are sunk at the time production decisions are made, the second term of (1), $p_{i,t}^r \cdot q_{i,t}^r$, drops out of the equilibrium first-order conditions with respect to actual production, $q_{i,t}$.

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 :

$$(2) \quad \frac{\partial \pi_{i,t}}{\partial q_{i,t}} = p_t^w(q_{i,t}, q_{-i,t}) + [q_{i,t} - q_{i,t}^r] \cdot \frac{\partial p_t^w}{\partial q_{i,t}} - C'_{i,t}(q_{i,t}) \geq 0.$$

The retail position of firm i now plays the same role as a fixed-price forward commitment in its impact on the incentives for wholesale market production. As the forward commitment increases towards the amount produced, the marginal revenue approaches the wholesale price. In other words, the Cournot model with contracts close to $q_{i,t}$ is similar to the

¹⁸For example, Hendricks and McAfee (2006) derive equilibrium conditions for a similar general problem assuming a form of supply function equilibrium.

competitive outcome.

The impact of vertical arrangements becomes more extreme when one considers the possibility that such arrangements could create a negative net position for a supplier. In other words, a supplier's retail commitment can be greater than its wholesale production. In such a circumstance, the supplier would want to drive wholesale prices *below* competitive levels. We do observe this equilibrium on occasion. Under these conditions, a larger degree of market power leads to lower prices. Thus the noncooperative outcomes are still bounded between the Cournot and competitive levels, but the Cournot outcomes constitute the lower bound on prices in the range where retail arrangements exceed wholesale production.

For each market and hour, we simulate three prices: the perfectly competitive equilibrium; the Cournot equilibrium ignoring vertical arrangements; and the Cournot equilibrium that accounts for vertical arrangements. For the Cournot models, we assume firms solve (2). The no-vertical-arrangements case, where we set $q_{i,t}^r = 0$, is the upper bound on the static, noncooperative price outcomes. For the competitive model, the production decision of a nonstrategic firm i at time t is described by the condition:

$$(3) \quad p_t^w(q_{i,t}, q_{-i,t}) - C'_{i,t}(q_{i,t}) \geq 0.$$

Equation (3) is used to calculate the lower bound on static, noncooperative outcomes. Even in the Cournot model, some small firms are assumed to take prices as given and solve (3).

The wholesale market price is determined from the firms' residual demand function

(Q_t) , which equals the market demand (\bar{Q}_t) minus supply from fringe firms whose production is not explicitly represented. We model the supply from imports and small power plants, q_t^{fringe} , as a function of price, thereby providing price responsiveness to Q_t :

$$(4) \quad Q_t(p_t^w) = \bar{Q}_t - q_t^{fringe}(p_t^w).$$

The full solution to these equilibrium conditions is represented as a complementarity problem. Appendix B contains a more complete description of the complementarity conditions implied by the equilibrium and other modeling details, given the functional forms of the cost and inverse demand described below.

B 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 hydroelectric plants, are generally thought to be low-cost technologies. Therefore, as we explain below, the available capacity from units in this second category is treated as inframarginal: the capacity is applied to the bottom of the firm'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 (VO&M) costs. Fuel costs can be calculated by multiplying a unit’s “heat rate,” a measure of its fuel efficiency by the price of fuel, which is updated as frequently as daily. Many units are subject to environmental regulations that require them to obtain nitrogen oxides and sulfur dioxide 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 a generating unit 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.¹⁹ By ordering all the generation units owned by a firm, one can construct a stepwise function for the production cost from that firm’s portfolio. We approximate this step function with a piecewise linear function with five segments, as we describe in Appendix A.

There are several categories of generation for which it is impractical to model explicitly marginal production costs. Much of this energy is produced by conventional generation

¹⁹This approach to modeling unit availability is similar to Catherine D. Wolfram (1999). However, it is a departure from the methods used by BBW, Bushnell and Saravia (2002), and Mansur (2007). 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.

sources, but there is also a substantial amount of production from energy-limited (primarily hydroelectric) resources. Because such generation features either extremely low marginal cost or “take-or-pay” contractual requirements, we refer to such production as “must-run.” Our treatment of these resources is described in detail in Appendix A. For purposes of constructing firm-level production functions, we take the amount of must-run generation actually produced as given for each hour and apply that production to the cost function of each firm.

Thus a firm’s estimated marginal cost function consists of a piecewise linear function of fossil fuel production costs, where each segment of the piecewise 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. The function is perfectly inelastic at full available capacity. This piecewise linear function is shifted rightward by an amount equal to the quantity of electricity produced by that firm from must-run (hydroelectric and nuclear) resources. The aggregate production capacity of a firm can therefore change from hour to hour if that firm has volatile hydroelectric production.

C 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 firms. Recall that the derived demand in wholesale electricity markets is completely inelastic; therefore, the residual demand curve will equal market demand less the elastic supply of net imports (imports minus exports) and other fringe plants not modeled. In California, this supply includes net imports and must-take plants.²⁰ In New England, net imports from New York and production from small firm generation comprise this supply.²¹ We estimate only net import supply in PJM. For all markets, the sample period is June to September, 1999.

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 that are exporting energy into the restructured markets take prices as given 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

²⁰BBW discuss must-take plants and why they are not modeled directly in measuring firm behavior. These plants include nuclear and independent power producers.

²¹Canadian 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.

price differences.

For each hour t , we proxy regional prices using daily temperature in bordering states ($Temp_{st}$),²² 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 wholesale market price ($\ln(p_t^w)$),²³ 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):

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

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.

²²For 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 above 65° F) and heating degree days (degrees daily mean below 65° F). As such, $Temp_{st}$ has four variables for each bordering state. These data are state averages from the National Oceanic & Atmospheric Administration (NOAA) web site daily temperature data.

²³This functional form is smooth, defined for all net imports, and accounts for the inelastic nature of imports nearing capacity. For robustness, we also estimate linear, square root, and cube root models of fringe supply. We discuss how these alternative functional forms impact our results in the next section. Note that a constant elasticity (log-log) model would drop observations with negative net imports, a substantial share of the data in some markets. Appendix A describes the sources of the price data.

For each market, Table 2 reports the 2SLS β coefficients and standard error estimates that account for serial correlation and heteroskedasticity.²⁴ First-stage regressions suggest that the instruments are strong.²⁵ The first column reports the results for our main linear-log specification. California has the most price-responsive import and fringe supply, with a $\beta = 5392$ (and a standard error of 704). In New England, β is 1391 (s.e. 162). Finally, in PJM, we estimate β as 861 (s.e. 118). We also report the implied elasticities of the residual demand (i.e., load minus fringe supply) at the sample mean. For robustness, we also estimate several polynomial functional form specifications of the fringe supply function: linear, square root, and cube root models. With these models, we find qualitatively similar elasticity estimates (see Table 2).²⁶

The coefficient estimates for our main specification, shown in (5), 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:

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

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:

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

²⁴We test the error structure for autocorrelation (Breusch-Godfrey LM statistic) and heteroskedasticity (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 heteroskedasticity.

²⁵More specifically, the coefficients on load for each of the markets and functional form models are significant at the 1-percent level.

²⁶In the web appendix, we test how these alternative models impact our Cournot and competitive simulations. We show that our main findings are robust.

III Results

We first set the retail commitment, $q_{i,t}^r$ in (2), equal to zero for all firms. This provides us with counterfactual equilibria whereby the incentive effects of vertical arrangements and long-term contracts are ignored. We can then test the impact of horizontal market structure alone. Next, we test the importance of vertical arrangements by comparing these outcomes with those when we set $q_{i,t}^r$ equal to the approximate levels that we have been able to determine from public data sources. These commitments will not affect the behavior of firms taking prices as given. Therefore, the competitive prices—the “lower” bound—are the same in both the case with contracts and the case without contracts.²⁷

These phenomena 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 that 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 treat-

²⁷While 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 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.

ment 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 nonconvex costs and intertemporal 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.

A further caveat concerns the measurement of price. As noted in Appendix A, the PJM market may have many “nodal” prices for a given hour while we use a weighted average. The average price will be a noisy measure of the price each firm faces. However, prices did not vary substantially by location in the summer of 1999.²⁸

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 overproduce in order to drive down their wholesale cost of power purchased for retail service. In terms of (2), when $q_{i,t}^r > q_{i,t}$ 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.

²⁸In the summer of 1999, prices varied across locations in 19 percent of the hours. However, these differences were small: the largest difference across locations was less than \$1 during 90 percent of the hours. Furthermore, Mansur (2007) notes that, *ex ante*, congestion has an indeterminate effect on the average price. Even the effect of market power is ambiguous: while congestion reduces the elasticity of residual demand, PJM regulations cap bids in congested areas.

A Market Level Results

Table 3 summarizes the prices for the Nash-Cournot equilibrium with and without vertical arrangements, as well as the competitive 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 hydroproduction. Functionally, this means that there is no meaningful difference between a “no vertical arrangements” and “with vertical arrangements” case in California.²⁹

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

²⁹As 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 hydroelectric 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 firm in a perfectly competitive market). Any attempts to do so by PG&E, the large hydroproducer 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 vertical arrangements” 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 hydroelectric resources is beyond the scope of this paper. See Bushnell (2003) for an examination of the potential impacts of strategic hydroproduction in the Western U.S.

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 between 11 am and 8 pm on weekdays.

In all three markets, actual prices appear to be consistent with Cournot prices in comparison with competitive prices during the peak hours of the day. The average prices in California were \$43 for actual prices and \$45 for Cournot prices, while competitive prices were only \$35. In New England, actual and Cournot (with vertical arrangements) both averaged \$55, which is \$13 above the average competitive price simulation. In PJM, actual prices were \$97 on average. While this is greater than the Cournot price average of \$87, it is nearly triple the competitive average of \$35.

Our off-peak competitive price estimates exceed actual prices in all markets. For California and PJM, the low prices do not appear to be caused by monopsony behavior, as the Cournot prices exceed the competitive prices even at low demand. By contrast, the nega-

tive price-cost margins during off-peak hours in New England are, in fact, consistent with strategic behavior to some degree. Over the entire sample of off-peak hours, the median Cournot equilibrium price is slightly below the median competitive price. However, in the September off-peak hours, the median Cournot price in New England was \$27.21/MWh in comparison to an estimated competitive price of \$31.02/MWh. The median of the actual off-peak September prices was \$25.55/MWh. 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 extramarginal resources. This allows these firms to produce at a loss, on the margin, thereby reducing the equilibrium price.

Kernel Regression Results

Figure 2 plots actual hourly prices in California from June 1 to September 30, 1999. We estimate a nonparametric kernel regression of the relationship between the actual hourly prices and the ratio of current demand to summer peak demand.³⁰ In Figure 2, 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.

In both the New England and PJM markets, Cournot prices that do not account for

³⁰We use the 100 nearest neighbor estimator, namely the Stata command “knnreg.”

vertical commitments far exceed actual prices for even moderate demand levels. Figure 3 presents the Cournot prices without the vertical arrangements for New England. Once the quantity demanded reaches 60 percent of the summer’s peak demand, prices increase substantially. The results for Cournot pricing without contracts in PJM are most startling. In Figure 4, we show that for any level of residual demand above 50 percent of installed capacity, the Cournot price would have been at the price cap of \$1000/MWh had firms divested as in California. From these results, one may be led to conclude that the New England and PJM markets did not exhibit conventional oligopoly behavior, and were competitive relative to California despite their less favorable horizontal structures.

However, once we account for the vertical arrangements, Cournot prices in the two markets are similar to the actual prices at high demand levels. Figure 5 presents the 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 slightly below the competitive prices. This is consistent with the monopsony overproduction strategy previously discussed. Figure 6 illustrates the same analysis for PJM. Again, Cournot prices are quite close to actual prices at higher demand levels and actually exceed Cournot prices at the very highest levels of demand. In a web appendix, we show that the conclusions drawn from Figures 2, 5, and 6 are robust to the errors in measuring the $\hat{\beta}$ coefficient in (5).

Testing Market Performance

We examine the relative goodness-of-fit of the two estimated price series—Cournot with

vertical arrangements and competitive-to actual prices. For each market and simulation, we measure the difference between actual hourly prices (p_t^{actual}) and the simulated hourly prices (p_t^{sim}). We then compute a variation on the traditional R^2 to measure each model's fit. Here, we define R^2 as one minus the ratio of the sum of the squared errors over the sum of the squared actual prices:

$$(8) \quad R^2 = 1 - \frac{\sum_{t=1}^T (p_t^{actual} - p_t^{sim})^2}{\sum_{t=1}^T (p_t^{actual})^2},$$

where p_t^{sim} equals either the hourly Cournot price (p_t^{cour}) or the hourly competitive price (p_t^{comp}).

In all three markets, the Cournot price simulations have greater measures of R^2 than the competitive price simulations. For California, the R^2 is 0.94 for the Cournot estimates and 0.92 for the competitive prices. In New England, the R^2 is 0.82 for Cournot and 0.69 for competitive. In PJM the values are 0.78 and 0.18 for the Cournot and competitive prices respectively. During just the peak hours, the goodness-of-fit measures are similar.³¹

A more formal test can examine whether these values are in fact meaningfully different. The empirical model is that actual price equals either 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 Russell Davidson and James G. MacKinnon (1993,

³¹Specifically, for California on-peak, the R^2 is 0.95 for the Cournot estimates and 0.91 for the competitive prices. In New England, the values are 0.84 and 0.69. In PJM, they are 0.78 and 0.16.

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 Cournot and competitive prices:

$$(9) \quad p_t^{actual} = \gamma_1 p_t^{cour} + \gamma_2 p_t^{comp} + u_t.$$

We estimate this equation using ordinary least squares (OLS). The standard errors are adjusted using the correction for heteroskedastic and autocorrelated errors developed by Whitney K. Newey and Kenneth D. West (1987) (we assume a 24-hour lag structure). Note that the prices p_t^{cour} and p_t^{comp} are imputed from the $\hat{\beta}$ coefficient in (5), which is estimated with error. Therefore, we must correct the variance-covariance matrix from estimating (9) to account for this first-stage uncertainty. We use the method described in equation (15') of Kevin M. Murphy and Robert H. Topel (1985).³² In our tests, we impose a more rigorous standard on the model. In particular, in order for the actual prices to be consistent with a Cournot model, we require: (1) that we cannot reject $\gamma_1 = 1$ and $\gamma_2 = 0$; and (2) that we can reject $\gamma_1 = 0$ and $\gamma_2 = 1$. In other words, that the model is consistent with Cournot pricing but is not consistent with competitive pricing. Note that this is a strong test and that in many supply function equilibria, actual prices will fall between

³²After estimating (9) using OLS, the correction requires three steps. First we approximate how a small change in $\hat{\beta}$ impacts each of the hourly imputed prices. To do this, we re-estimate the Cournot and competitive prices using $\hat{\beta}^*$, where $\hat{\beta}^*$ equals $\hat{\beta} * 1.001$. The change in the Cournot price, $dp_t^{cour}/d\hat{\beta} \equiv f_t^{cour}$, equals $(p_t^{cour}(\hat{\beta}^*) - p_t^{cour}(\hat{\beta})) / (\hat{\beta}^* - \hat{\beta})$. The change in the competitive price, f_t^{comp} , is similarly defined. Then, we compute $F_t^* = \hat{\gamma}_1 f_t^{cour} + \hat{\gamma}_2 f_t^{comp}$ and regress it on both of the imputed prices: p_t^{cour} and p_t^{comp} . We call the estimated coefficients $\hat{\delta}_1$ and $\hat{\delta}_2$. Finally, we calculate the adjusted standard errors. Let the initial estimated standard errors on $\hat{\gamma}_1, \hat{\gamma}_2$, and $\hat{\beta}$ be $\hat{\sigma}_{\gamma_1}, \hat{\sigma}_{\gamma_2}$, and $\hat{\sigma}_{\beta}$, respectively. The corrected standard error on $\hat{\gamma}_i$ equals $\sqrt{\hat{\sigma}_{\gamma_i}^2 + \hat{\delta}_i^2 \hat{\sigma}_{\beta}^2}$, for $i=1$ and 2 . This method assumes independence of the errors in the fringe supply, (5), and non-nested test, (9), regressions.

these extremes.

For all three markets, the tests suggest that Cournot prices are a better fit for actual peak-hour prices than the competitive prices. In California, we find that the coefficient on Cournot price is 0.89 with a standard error of 0.24. In contrast, the competitive price is insignificant at the 5-percent level with a coefficient of 0.13 (s.e. of 0.25). Wald tests reject the competitive market hypothesis but do not reject the Cournot pricing hypothesis. The encompassing test for New England results in similar findings as in California. The Cournot price coefficient equals 1.69 with a standard error of 0.44. In contrast, the competitive price coefficient is -0.65 (s.e. of 0.34). Finally, the test in PJM implies similar results: the Cournot price coefficient is 1.02 (s.e. of 0.15) and the competitive price coefficient is 0.40 (s.e. of 0.90). In all three markets, these estimates fail to reject the Cournot model but do reject the competitive model. During the off-peak hours when firms have less incentive to exercise market power, the non-nested tests are somewhat less definitive.³³

In addition to examining predicted prices, we compare the quantity decisions of firms. In the web appendix, we examine the firm-level production implied by our three cases—perfect competition and Cournot with and without vertical obligations—as well as the actual production recorded by the EPA’s Continuous Emissions Monitoring System (CEMS) dataset. These results reinforce a conclusion that firms’ behavior reflected strategies more complex than Cournot, at least outside of California. As with prices, actual firm-level pro-

³³For these hours, only New England and PJM reject the competitive model but do not reject the Cournot model. The encompassing test for California rejects both models. In California, the Cournot coefficient is 1.51 (0.25) and the competitive coefficient is -0.82 (0.26). In New England, the Cournot coefficient is 2.40 (0.82) and the competitive coefficient is -1.43 (0.71). In PJM, the Cournot coefficient is 1.50 (0.60) and the competitive coefficient is -0.93 (0.62).

duction quantities were far less extreme than those produced in the “no vertical arrangements” case.

B Social Welfare Impacts

Utilizing our measures of firms’ cost functions, we can construct estimates of total production costs for the various cases. This allows us to produce an estimate of the welfare impacts of the vertical arrangements. Recall that end-use demand is effectively inelastic in these markets, so any social welfare impacts of strategic behavior will be reflected solely in an increase in production costs from perfectly competitive levels. Also recall that, in PJM, the no-vertical-arrangements case produces many extreme outcomes with a large number of potential equilibria resulting in prices at the capped level. In PJM, we therefore focus on the welfare impacts of those hours in which the price did not reach \$1000/MWh in the case of no vertical arrangements. This provides a lower bound on the welfare impact, as the results will be less extreme in these relatively low-demand hours.

In PJM, the production costs under the case with vertical contracts are 59 percent lower than costs without vertical contracts. Extrapolating to the full four-month sample, the costs of producing electricity for the PJM market from June 1 to September 30, 1999 would have increased by \$2.1 billion if there were no vertical arrangements. In New England, production costs for our simulations are 32 percent, or \$327 million, lower with vertical contracts. In other words, had New England and PJM regulators impeded the formation of vertical arrangements (as in California), production costs would have increased by over

45 percent in each market.

In both cases, the cost impacts of these estimates are dominated by the increased costs of imports. In other words, the bulk of the welfare impact comes from an increase in high-cost imports, which are assumed to be competitively priced and which replace relatively low-cost production that is withheld to a far greater extent in the scenarios with no vertical arrangements due to increased market power.

These results reinforce the perception that the horizontal market structure in the Eastern markets, particularly in PJM, is not competitive, and that vertical arrangements are playing a critical role in mitigating the exercise of market power in the spot market.

C Stability of Vertical Structure

The above results highlight the impact of the substantial vertical obligations held by some of the largest generation firms in the two Eastern markets. The 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 better understood and can reasonably be considered to be exogenous. However, a natural question to ask is, are these vertical obligations likely to be stable or would we expect these firms to change their vertical position?

While this is a very complex issue, with many factors influencing the organization of these firms, we can examine this question in the context of its impact on market competitiveness and firm profits. In doing so, we adopt the framework of Allaz and Vila (1988),

who model the impact of forward contracts on competition. As we have argued above, in electricity markets the decision to take on a retail customer shares essential features with the decision to sign a forward contract. Most importantly, retailers commit to a fixed price for a subscription period, in advance of wholesale market outcomes. This “advance” retail rate that a wholesale firm agrees to is therefore a key parameter in addressing the optimal vertical structure of a firm.

For the purposes of this exercise, we adopt the extreme assumption that retail margins will be near zero, and that retail rates will therefore be based upon expectations of wholesale prices.³⁴ This is comparable to the perfect arbitrage assumption adopted by Allaz and Vila in their study of forward markets.³⁵ Given that retailing has no direct profit, one might imagine that wholesale firms would have no interest in participating at all. However, by establishing a vertical position, a producer also makes a credible commitment to produce more in subsequent wholesale markets, thereby reaping advantages similar to those enjoyed by the leader in a Stackelberg game. This is the logic demonstrated by Allaz and Vila, who find that Cournot producers find it optimal to sign fixed-price forward contracts.³⁶

³⁴In many vertical contexts the retail markup over wholesale prices is more important than wholesale profits. It is not clear what the likely retail margins in electricity markets will be, but the ease of entry into retailing, along with continued regulatory oversight of retail pricing, argue that such margins will be relatively small.

³⁵In other words, the retail and wholesale markets in our context play the role of the forward and spot markets in the Allaz and Vila context.

³⁶However, recent papers have challenged the view that forward contracting is necessarily procompetitive. In particular, producers could enter the forward market as *buyers*, thereby increasing their net position (see Philippe Mahenc and François Salanié, 2004), or use the forward market to facilitate tacit collusion (see Matti Liski and Juan-Pablo Montero, 2006). These papers raise interesting issues in the context of the vertical arrangements discussed in this section. However, these papers do not challenge the notion that, given forward obligations that reduce the net positions of firms, the static equilibrium will

Thus, there are strategic reasons why the firms in these markets would take on vertical obligations at fixed prices. Their self-imposed obligations commit them to behave more competitively in the wholesale market, but also force their competitors to respond by reducing their own production in the wholesale market. But are the positions held by the firms in the markets studied here near the optimal size? Bushnell (forthcoming) studies this question in an equilibrium context.³⁷

While a full equilibrium model with asymmetric firms and nonlinear costs is beyond the scope of this paper, we can examine the unilateral optimality of a firm's 1999 vertical position. We do so by marginally changing a given firm's vertical position, while holding the position of all other firms constant. Specifically, we assume that each firm's vertical position increases by an amount equal to 1 percent of that firm's production capacity.

The results are reported in Table 4. This table describes the change in profits to each large firm in a market from an increase in forward position by the firm listed in the first column. Thus the diagonal entries are of the most interest, as they describe the impact on profits of a given firm from its own increase in forward position. The results indicate that, as expected, firms in California would want to increase their positions above zero,

be more competitive. Further, in the case of vertical integration, there is no analog to going short in the forward market. A pure producer cannot make itself less integrated.

³⁷Bushnell derives a modified version of the Allaz and Vila model and applies parameters derived from these three markets to that model. He finds that the number of firms in the market is the key parameter to determining the extent of forward (vertical) commitments in the market. This simplified model indicates that the overall level of vertical commitments in the PJM markets is roughly equal to the equilibrium levels, while 1999 levels are a little below equilibrium in New England and are obviously far below equilibrium levels in California. Bushnell finds that the equilibrium level of forward commitments in these three markets are 81, 71, and 55 percent for PJM, NE, and California respectively. This is contrasted to the actual levels estimated here to be 85, 50, and zero percent for the same three markets. However, one key restriction of that model is that it features symmetric firms, and it is therefore unable to address the question of the equilibrium position of actual individual firms.

while all the largest firms in PJM would prefer to reduce their positions. The two firms in New England with major vertical commitments would also prefer to reduce their positions, while the merchant producers in that market such as Mirant would prefer to add forward commitments. This latter result is consistent with the observation that the two large firms in New England produced quantities consistent with more balanced vertical positions than their public obligations provided.

This analysis indicates that, from the perspective of the wholesale profits of an integrated firm, generation suppliers will continue to seek long-term retail commitments. However, from the wholesale perspective, those commitments should not be as extensive as those that were created as an artifact of the transition from state-level regulation. It is important to emphasize that many other considerations will factor into such decisions, not least among them the size of retail margins. Thus, this analysis sheds light on only one aspect of the question of the optimal vertical structure of electricity firms in restructured markets.

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 retail arrangements. A critical question will be the extent to which retailers continue to offer relatively long-term price commitments to customers, since these price commitments are such a mitigating force on the wholesale market behavior of integrated firms.

IV Conclusions

Within the U.S., experiences with electricity restructuring have varied dramatically. While the consequences of such initiatives have been 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. Much of the debate has centered on the relative influence of market structure and market design.

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 perfectly competitive position. While other noncooperative 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 examine 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. These findings support Wolak's (2000) analysis of long-term contracts in the Australian electricity market as well as Fabra and Toro's (2005) analysis of the Spanish market. The concentration 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 similarly 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.

It is worth noting that our analysis has focused on only one, albeit central, aspect of electricity 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.

We conclude that our results do carry important implications for both electricity restructuring and antitrust policies. The horizontal structure of the markets is important, but similar horizontal structures can produce dramatically different outcomes under different vertical arrangements. The extent to which these arrangements constitute firm price

commitments also plays a strong role in the impact of vertical structure on the market outcomes. While we observe the central impact of vertical arrangements that have been exogenously determined, it is a far more difficult task to predict the type of vertical structure that may eventually evolve in an industry. For electricity markets, the question of 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): 1-16.
- [2] **Borenstein, Severin, and James B. Bushnell.** 1999. "An Empirical Analysis of the Potential for Market Power in California's Electricity Market." *Journal of Industrial Economics*, 47(3): 285-323.
- [3] **Borenstein, Severin, James B. Bushnell, and Christopher R. Knittel.** 1999. "Market Power in Electricity Markets: Beyond Concentration Measures." *Energy Journal*, 20(4): 65-88.
- [4] **Borenstein, Severin, James B. Bushnell, and Frank A. Wolak.** 2002. "Measuring Market Inefficiencies in California's Deregulated Wholesale Electricity Market." *American Economic Review*, 92(5): 1376-1405.
- [5] **Bushnell, James B.** 2003. "A Mixed Complementarity Model of Hydro-Thermal Electricity Competition in the Western U.S." *Operations Research*, 51(1): 80-93.
- [6] **Bushnell, James B.** 2004. "California's Electricity Crisis: A Market Apart?" *Energy Policy*, 32(9): 1045-1052.
- [7] **Bushnell, James B.** Forthcoming. "Oligopoly Equilibria in Electricity Contract Markets." *Journal of Regulatory Economics*.

- [8] **Bushnell, James B., and Celeste Saravia.** 2002. "An Empirical Analysis of the Competitiveness of the New England Electricity Market." CSEM Working Paper WP-101, University of California Energy Institute.
- [9] **Czyzyk, Joseph, Michael P. Mesnier, and Jorge J. Moré.** 1998. "The NEOS Server." *IEEE Journal on Computational Science and Engineering*, 5: 68-75.
- [10] **Davidson, Russell, and James G. MacKinnon.** 1993. *Estimation and Inference in Econometrics*, Oxford University Press.
- [11] **Dirkse, Steven P., and Michael C. Ferris.** 1995. "The PATH solver: Anon-monotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5: 123-156.
- [12] **Fabra, Natalia, and Juan Toro.** 2005. "Price Wars and Collusion in the Spanish Electricity Market." *International Journal of Industrial Organization*, 23(3-4): 155-181.
- [13] **Gilbert, Richard, and Justine Hastings.** 2005. "Vertical Integration in Gasoline Supply: An Empirical Test of Raising Rivals' Costs." *Journal of Industrial Economics*, 53(4): 469-492.
- [14] **Green, Richard J.** 1999. "The Electricity Contract Market in England and Wales." *Journal of Industrial Economics*, 47(1): 107-124.

- [15] **Green, Richard J., and David M. Newbery.** 1992. "Competition in the British Electricity Spot Market." *Journal of Political Economy*, 100(5): 929-53.
- [16] **Hendricks, Kenneth, and R. Preston McAfee.** 2006. "A Theory of Bilateral Oligopoly." Unpublished.
- [17] **Hobbs, Benjamin.** 2001. "Linear Complementarity Models of Nash-Cournot Competition in Bilateral and POOLCO Power Markets." *IEEE Transactions on Power Systems*, 16(2): 194-202.
- [18] **Hortaçsu, Ali, and Steven L. Puller.** Forthcoming. "Testing Strategic Models of Firm Behavior in Restructured Electricity Markets: A Case Study of ERCOT." *RAND Journal of Economics*.
- [19] **Joskow, Paul J.** 2003. "The Difficult Transition to Competitive Electricity Markets in the U.S." In *Electricity Restructuring: Choices and Challenges*, ed. James M. Griffen and Steven L. Puller. Chicago: University of Chicago Press.
- [20] **Joskow, Paul J., and Edward P. Kahn.** 2002. "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000." *Energy Journal*, 23(4): 1-35.
- [21] **Klemperer, Paul D., and Margaret A. Meyer.** 1989. "Supply Function Equilibria in Oligopoly Under Uncertainty." *Econometrica*, 57(6): 1243-77.

- [22] **Kolstad, Charles D., and Lars Mathiesen.** 1991. "Computing Cournot-Nash Equilibria." *Operations Research*, 39(5): 739-48.
- [23] **Krattenmaker, Thomas G., and Steven C. Salop.** 1986. "Competition and Cooperation in the Market for Exclusionary Rights." *American Economic Review Papers and Proceedings*, 76(2): 109-13.
- [24] **Liski, Matti, and Juan-Pablo Montero.** 2006. "Forward Trading and Collusion in Oligopoly." *Journal of Economic Theory*, 131(1): 212-30.
- [25] **Mahenc, Philippe, and François Salanié.** 2004. "Softening Competition through Forward Trading." *Journal of Economic Theory*, 116: 282-93.
- [26] **Mansur, Erin T.** 2007. "Upstream Competition and Vertical Integration in Electricity Markets." *Journal of Law and Economics*, 50(1): 125-56.
- [27] **Mansur, Erin T.** Forthcoming. "Measuring Welfare in Restructured Electricity Markets." *Review of Economics and Statistics*.
- [28] **Murphy, Kevin M., and Robert H. Topel.** 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics*, 3(4): 88-97.
- [29] **Newey, Whitney K., and Kenneth D. West.** 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3): 703-08.

- [30] **Puller, Steven L.** 2007. "Pricing and Firm Conduct in California's Deregulated Electricity Market." *Review of Economics and Statistics*, 89(1): 75-87.
- [31] **Rey, Patrick, and Jean Tirole.** Forthcoming. "A Primer on Foreclosure." In *Handbook of Industrial Organization, Volume 3*, ed. Mark Armstrong and Robert H. Porter. New York: Elsevier.
- [32] **Rudkevich, Aleksandr, Max Duckworth, and Richard Rosen.** 1998. "Modeling Electricity Pricing in a Deregulated Generation Industry: The Potential for Oligopoly Pricing in a Poolco." *Energy Journal*, 19(3): 19-48.
- [33] **Schmalensee, Richard, and Bennett W. Golub.** 1984. "Estimating Effective Concentration in Deregulated Wholesale Electricity Markets." *RAND Journal of Economics*, 15(1): 12-26.
- [34] **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.
- [35] **Wolfram, Catherine D.** 1999. Measuring Duopoly Power in the British Electricity Spot Market. *American Economic Review*, 89(4): 805-26.

Appendix A: Data

This appendix discusses the data used in our analysis. First we discuss how we measure the market clearing prices and quantities. Second, we describe our method of measuring a firm's vertical arrangements. Next, we outline the sources of data for calculating a firm's marginal costs and residual demand functions. Then, we discuss how we construct monthly piecewise linear approximations of a firm's marginal cost curves. Finally, we discuss how we model production from nuclear, cogeneration, and energy-limited resources.

Market Clearing Quantities and Prices

Data on market quantities and prices are available from the ISO websites: www.caiso.com, www.pjm.com, and www.iso-ne.com. 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 a 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 3 percent to overall demand.

In California, the market price is the day-ahead unconstrained price (UCP) from the Power Exchange (PX). About 85 percent 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 England ECP reflects any geographic variation in response to transmission constraints.

In contrast, the PJM market uses a nodal pricing system and reports no single "generic" market-wide price. Under nodal pricing, transmission congestion may result in the market having thousands of different locational prices at a given moment. As our estimates of competitive and Cournot equilibria ignore these congestion issues, we want to use a measure of actual prices that are also not likely to be substantially effected by congestion. The effect of congestion on average price is unclear *ex ante*, even though total costs must increase. We use the hourly load-weighted average nodal price.

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 utility's share of market demand was calculated by taking the ratio of the utility's peak demand to the peak of the overall PJM demand. This ratio was applied to all hours. We therefore assume that the relative demand of utilities in the system is constant.³⁸

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.

Data Sources for Marginal Cost and Residual Demand Functions

The California Market

The California data sources are identical to those used in BBW. As in BBW, hourly residual demand is derived by subtracting hourly production from imports and fringe producers designated as must-take by the California ISO from total hourly demand. The

³⁸Hourly 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.

hourly requirement for regulation—or automated generation control (AGC)—reserves is added to this energy demand when calculating the demand met by modeled generation. Thus, by identity, the hourly residual demand is the sum of the actual production from modeled (non-must-take) generation plus the hourly requirement for AGC. The market price used in the results is the unconstrained day-ahead price in the PX. Fuel and environmental costs are the same as used in BBW.

The New England Market

In the case of New England, public data from the Energy Information Administration (EIA) and the Environmental Protection Agency (EPA) are utilized. Specifically, we utilize generation level output data from the EPA’s Continuous Emissions Monitoring System (CEMS) for the large thermal plants.³⁹ Monthly hydro- and nuclear production is taken from the EIA’s Form 906.⁴⁰ Hydroproduction is distributed amongst hours within a month using a peak-shaving heuristic in which the monthly energy from each firm is applied to high demand hours subject to a unit’s maximum capacity limits.

The residual demand is constructed by combining the aggregated hourly production of

³⁹The EPA reports gross thermal output. To obtain net energy production, the gross output of all modeled plants was reduced to .95 times the reported gross output.

⁴⁰There were two large pump storage plants in the data set. The Northfield Mountain plant reported its net and gross energy output. The Bear Swamp plant only reported gross production, which is negative for a pump storage plant. We assumed an operating efficiency 66 percent for the Bear Swamp plant. That is to say that net production is 2/3 of gross energy consumed. Roughly half of the Florida Power & Light (FPL) hydrogeneration capacity failed to report production to the EIA in the summer of 1999. To compensate, we adjusted total hydroproduction from the FPL plants according to the ratio of production seen from those plants during 1998, for which full data are available. Plants that did report in 1999 comprised 56 percent of total production from all the plants. Total production was derived by applying this same ratio to the production of reporting plants in 1999. In other words, the production of reporting plants was multiplied by 1.8 to obtain an estimated total production from all FPL plants during 1999.

modeled generation taken from CEMS, the estimated hourly production from nuclear and hydro units, and the hourly AGC reserve requirement in the ISO-NE system. By identity, this is equivalent to subtracting the hourly production of imports and small thermal units from total ISO-NE system demand. The market price used is the hourly ISO-NE system Energy Clearing Price (ECP). Generation unit characteristics and fuel costs are taken from the Platts POWERDAT dataset. The data sources on SO₂ and NO_x permit prices are discussed below.

The PJM Market

Like Mansur (2007), we assume that nuclear and hydroelectric generation will not depend on the competitiveness of the market. In the summer of 1999, nuclear plants operated near full capacity, so we assume constant production within a month. We use data on monthly hydroelectric production from EIA Form 759 and hourly bid data from PJM's web site. Hydroelectric generation is bid into PJM differently than other sources of generation, allowing us to approximate total hydroelectric hourly bids. They bid what are called "zero-priced" bids. Note that these bids were not binding in 1999, so they may be inconsistent with actual output. We assume that hourly hydroelectric generation varies consistently with the scheduled "zero-priced" bids. Hourly production is scaled for each firm so that total output matches the EIA monthly production. We measure the efficiency rate of pumped storage units using EIA Form 759 data on monthly consumption and net generation. A firm's hydroelectric output equals the sum of run of river production and the implied gross production of the pumped storage.

The PJM web site, www.pjm.com, reports data on load and imports (net of exports). Generation plant characteristics and fuel prices are the same as in Mansur (2007). We use the average of two monthly SO₂ price indices of permits from the brokerage firms Cantor Fitzgerald and Fieldston. We use Cantor Fitzgerald data on NO_x prices. The NO_x regulation ended in September and the price was approximately \$1000 per ton at that time. We use this average for the entire summer.

Piecewise Linear Approximation

We measure the expected output that would be generated if the unit attempts to produce, which equals the unit's capacity times 1 minus the forced outage factor. This is also known as the derated capacity. For each market, we use the unit-level marginal cost and derated capacity data to construct a firm's marginal cost curve that varies daily. For computational reasons, we approximate this step function with a monthly Piecewise linear function with five segments. First, we calculate the monthly average marginal cost for each unit and compare the monthly average marginal cost curves for the fossil units in each of the three markets in June of 1999. In all three markets, marginal costs of the fossil units are relatively flat, increasing from \$20/MWh to \$40/MWh over 90 percent of the capacity. The cost curves then quickly increase for the remaining capacity.

Then, for each firm and month, we sort units by marginal cost and calculate the available operating capacity that is the total derated capacity with average marginal cost less than or equal to price. For each firm and month, we determine the quintiles of average

marginal cost. We then construct a Piecewise linear cost function with five segments for each month and firm. The quantity produced for each segment is based on the available operating capacity at each quintile. Marginal costs of these linear segments are simply measured by connecting the successive quintiles of costs. The costs are tied to the lowest and highest measured monthly average marginal costs for each firm and month.

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 impact significantly our results for the summer months, as these are high-demand periods when few units traditionally perform scheduled maintenance. This is one reason why we limit our comparisons to summer months.

In order to measure the goodness-of-fit of this approximation, we use the Piecewise linear functions to predict the marginal cost of each unit and hour. This fitted marginal cost is compared with the initial one by regressing the initial marginal cost on the fitted value. In all three markets, the approximation captures most of the variance of the initial cost data. In California, the coefficient on the fitted marginal cost is 0.97 and the R^2 is 0.93. In New England, the coefficient on the fitted marginal cost is 1.00 and the R^2 is 0.95. Finally, in PJM, the coefficient on the fitted marginal cost is 0.98 and the R^2 is 0.96. The goodness-of-fit of the Piecewise linear approximations for the top quartile of marginal costs are slightly worse.⁴¹

⁴¹For the top quartile in California, the coefficient is 0.93 (R^2 is 0.85). In New England, the coefficient is 0.72 (R^2 is 0.71). In PJM, the coefficient is 0.80 (R^2 is 0.86).

Nuclear, Cogeneration, and Energy-Limited Resources

There are several categories of generation for which it is impractical to model explicitly 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 hydroelectric) resources. Most of this generation is produced by firms considered to be nonstrategic. Because the production decisions for firms controlling energy-limited resources are quite different from those controlling conventional resources, we treat the two categories differently.

Most production from these conventional nonmodeled sources is controlled by firms considered to be nonstrategic. 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 some PJM and New England firms.⁴² While nuclear production is an extreme inframarginal resource, and unlikely to be strategically withheld from the market for both economic and technical reasons, the substantial amount of inframarginal production could likely have a significant impact on the amounts that nuclear firms may choose to produce from the other plants in their portfolios. We therefore take the hourly

⁴²It should also be noted that a large amount of production in California comes 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 1980s and early 1990s. 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.

production from nuclear resources as given and apply that production quantity as a zero-cost resource at the bottom of its owner's cost function.

Energy-limited units (i.e., hydroelectric units) present a different challenge than other units in the nonmodeled category since the concern is not over a *change* in output relative to observed levels but rather a *reallocation* over time of the limited energy that is available. The production cost of hydroelectric units does 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, 2003). 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 hydroproduction 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.⁴³

⁴³Specific data on hydroproduction are available for California. For the PJM and New England markets, monthly hydroproduction was applied using the peak-shaving heuristic described above.

Appendix B: Complementarity Formulation of Cournot-Nash Equilibrium

We assume that the market demand $Q_t = a_t - b \ln(p_t)$, or $p_t = \exp((a_t - Q_t)/b)$. 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 nondecreasing marginal cost curve.

The thermal capacity of fringe firms is aggregated into a single, competitive 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 2.A reduce to the following set of mixed linear complementarity conditions.

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

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

(11) For $q_{ft}^{Th,j}, \forall j, t$:

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

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

where the symbol \perp indicates complementarity. The first condition (10) is the standard condition equating marginal revenue to marginal cost for a Cournot producer. The second condition, which applies to firms that take prices as given, equates price to marginal cost. The total quantity, q_{it} , produced by firm i at time t , includes the sum of production from all thermal segments as well as any “must-run” production from hydro- or nuclear sources.

$$q_{it} = q_{it}^{must_run} + \sum_j q_{it}^{Th,j}$$

Simultaneously solving for the dual and primal variables $\{q_{it}^{Th,j}, \psi_{it}^j\}$ for all i, j produces an equilibrium for a single period t . For n producers (including the fringe), J segments to the thermal marginal cost curve, the above conditions for a single independent time period t produce $2nJ$ complementarity conditions 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 pseudoconcave and strictly concave at the point where the first-order condition (A1) 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 (Charles D. Kolstad and Lars Mathiesen, 1991).

These complementarity conditions are modeled in AMPL, a mathematical programming environment, and solved for using the PATH algorithm (Steven P. Dirkse and Michael C. Ferris, 1995). We utilized the NEOS server for optimization, a multiinstitution service centered around the Argonne National Laboratory that provides support for a wide variety of optimization solvers (Joseph Czyzyk, Michael P. Mesnier, and Jorge J. Moré, 1998). This server allows remote submission of optimization problems (see <http://neos.mcs.anl.gov/neos/>).

Figures and Tables

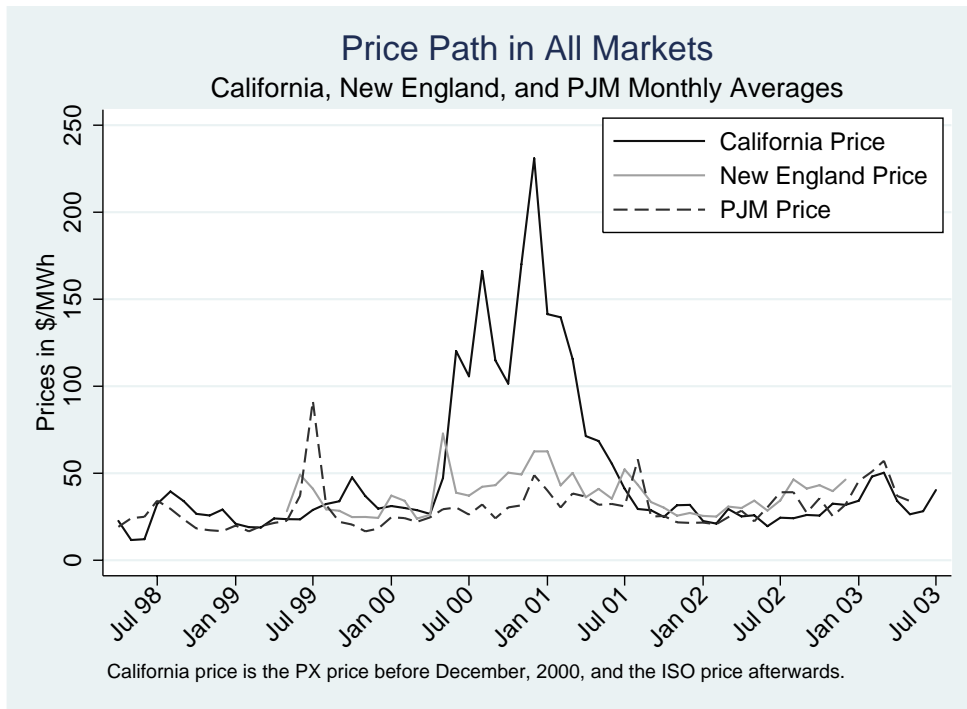


Figure 1:

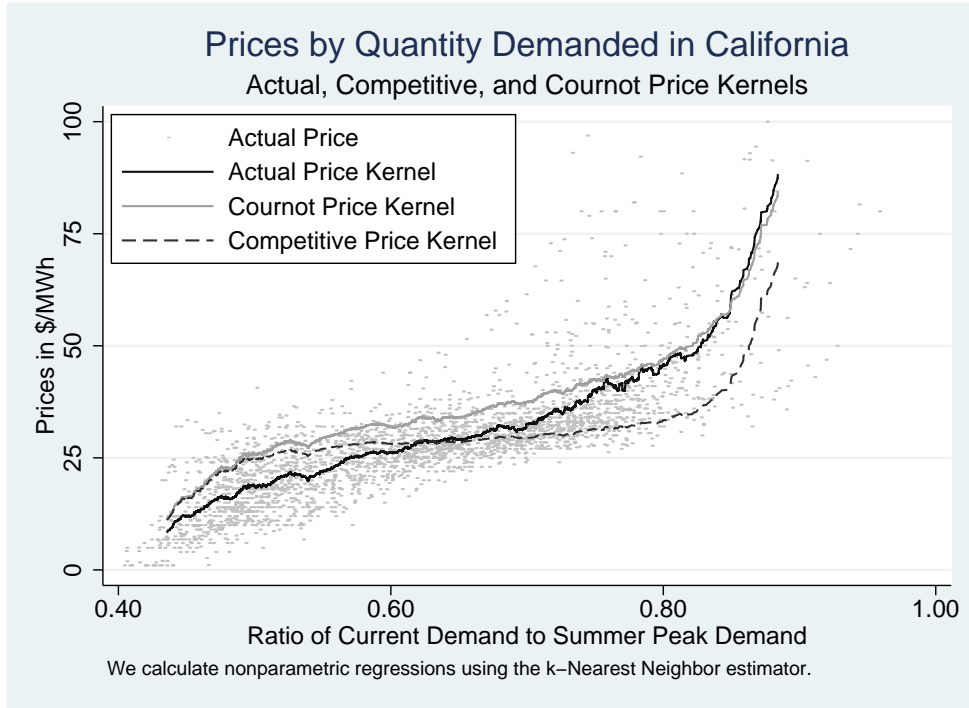


Figure 2:

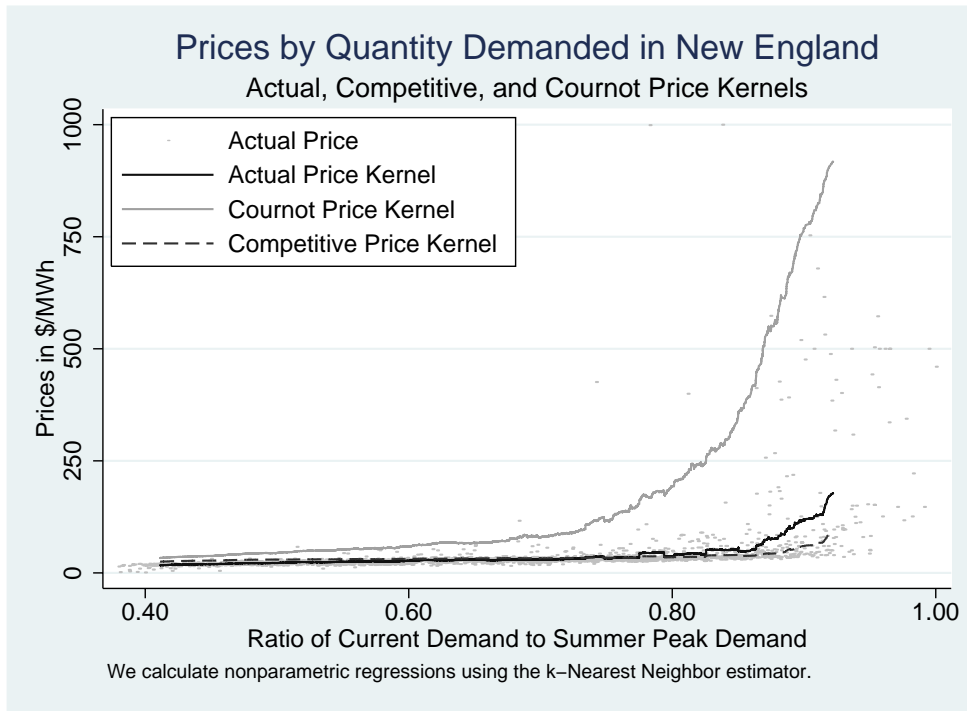


Figure 3:

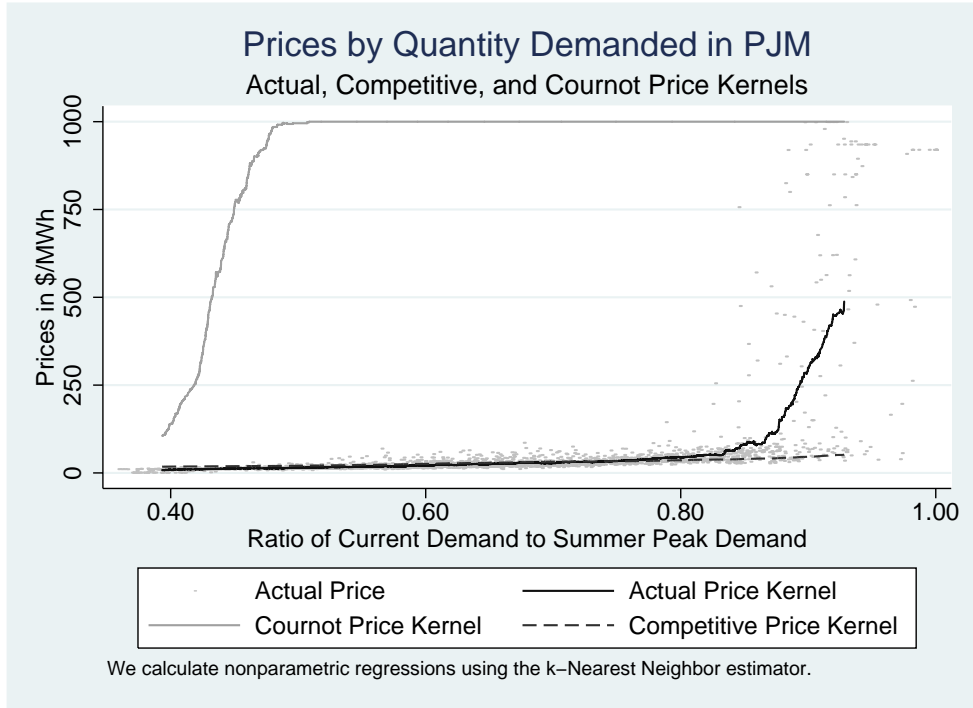


Figure 4:

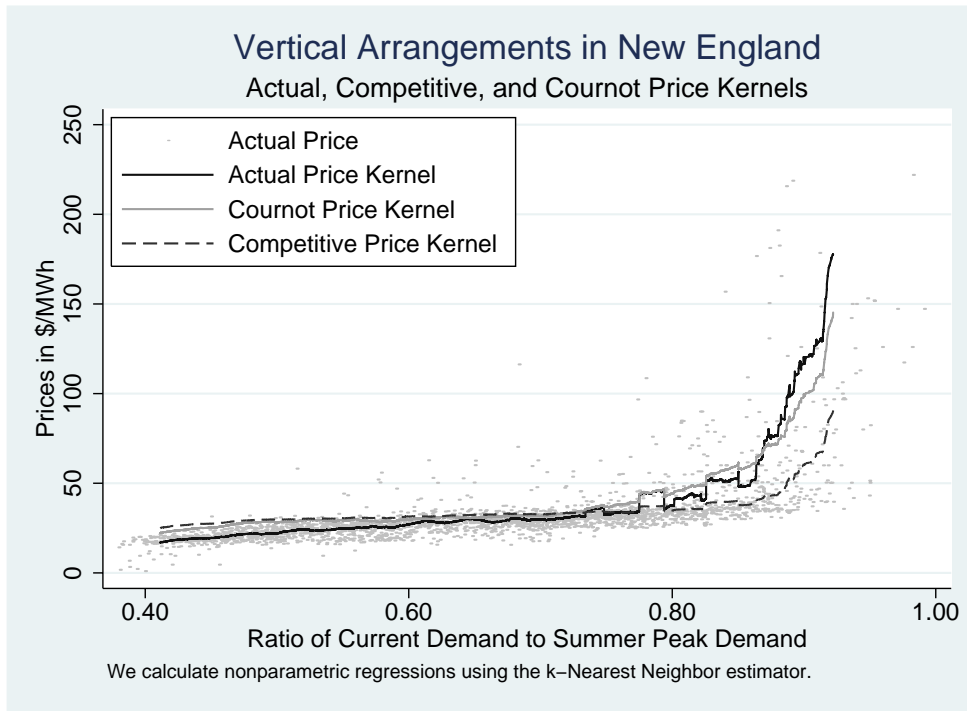


Figure 5:

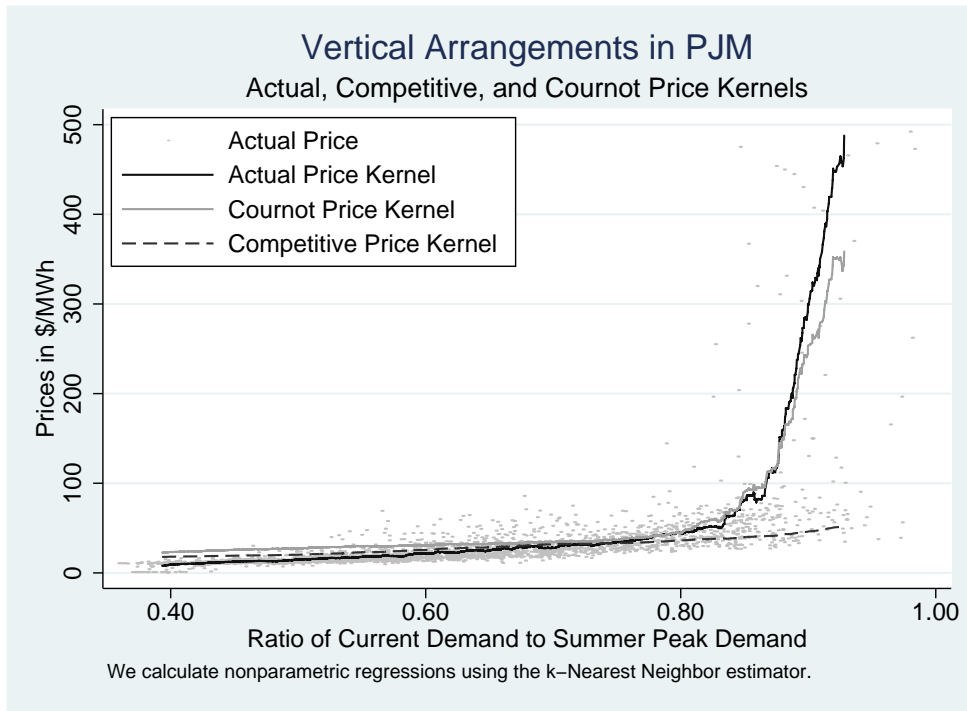


Figure 6:

Table 1—Firm Characteristics for Each Market: Summer 1999

Panel A: California Firm Characteristics.

Firm	Fossil	Water	Nuclear	Other	Output	Output	Load	Load
					Max	Share	Max	Share
PG&E	570	3,878	2,160	793	7,400	0.17	17,676	0.39
AES/Williams	3,921	-	-	-	3,921	0.09	-	-
Reliant	3,698	-	-	-	3,698	0.08	-	-
Duke	3,343	-	-	-	3,343	0.08	-	-
SCE	-	1,164	2,150	-	3,314	0.08	19,122	0.42
Mirant	3,130	-	-	-	3,130	0.07	-	-
Dynegy/NRG	2,871	-	-	-	2,871	0.06	-	-
Other	6,617	5,620	-	4,267	16,504	0.37	9,059	0.20
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	Output	Load	Load
					Max	Share	Max	Share
Northeast Util.	3,250	1,406	2,116	175	6,947	0.27	7,440	0.33
PG&E N.E.G.	2,736	915	-	165	3,816	0.15	4,440	0.20
Mirant	1,219	-	-	16	1,235	0.05	-	-
Sithe	1,810	-	-	-	1,810	0.07	-	-
FP&L Energy	965	365	-	-	1,330	0.05	-	-

Wisvest	979	-	-	-	979	0.04	1,200	0.05
Other	4,722	1,095	2,495	1,319	9,595	0.37	9,281	0.42
Total	15,681	3,781	4,611	1,675	25,712		22,361	

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	0.18	8,947	0.17
PECO	3,682	1,274	4,534	9,490	0.17	4,551	0.09
GPU, Inc.	7,478	454	1,513	9,445	0.17	7,602	0.15
PP&L Inc.	6,102	148	2,304	8,554	0.15	5,120	0.10
Potomac Electric	6,507	-	-	6,507	0.11	5,378	0.10
Baltimore G&E	3,945	-	1,829	5,774	0.10	5,792	0.11
Delmarva P&L	2,458	-	-	2,458	0.04	3,103	0.06
Edison	2,012	-	-	2,012	0.04	-	-
Atlantic City							
Electric	1,309	-	-	1,309	0.02	2,224	0.04
Other	428	439	-	867	0.02	8,998	0.17
Total	40,681	2,315	13,690	56,686		51,715	

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

Dependent variable is hourly fringe supply by market.

	$\ln(\text{Price})$	Price	$\sqrt{\text{Price}}$	$\sqrt[3]{\text{Price}}$
California	5392.4*	124.8*	1890.6*	5164.3*
	(704.2)	(11.4)	(128.3)	(360.1)
	[-0.672]	[-0.463]	[-0.642]	[-0.665]
New England	1391.1*	10.8*	308.5*	1006.5*
	(162.3)	(3.2)	(53.4)	(148.2)
	[-0.168]	[-0.048]	[-0.113]	[-0.135]
PJM	860.7*	8.5*	220.2*	687.7*
	(118.3)	(2.4)	(42.9)	(117.1)
	[-0.027]	[-0.012]	[-0.023]	[-0.026]

Notes: This table presents 2SLS coefficients for various functional form specifications of price. Each coefficient represents a separate regression. 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-percent level and (#) at the 10-

percent level. For each regression, we report in brackets the price elasticity of fringe supply, at the sample averages of price and fringe supply. The regressions include fixed effects for month of year, day of week, and hour of day. Weather variables for bordering states are also 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 stages, we regress a given functional form specification of price on the exogenous variables and an instrument of hourly load (MWh) in each market. The load variables have the same functional form specification as price. For example, in the first regression, we regress the log of price on the log of load.

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: Peak Hours (11 am to 8 pm Weekdays).

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
<i>California</i> Actual	43.15	34.52	27.0	17.2	225.0
Competitive	35.01	30.88	19.8	24.8	233.8
Cournot	45.17	40.19	21.0	25.2	233.8
<i>New England</i> Actual	55.05	33.16	82.9	17.7	753.2
Competitive	41.72	35.04	33.9	29.7	333.3
Cournot	54.63	40.44	52.0	26.5	454.7
Cournot n.v.a.	280.47	145.86	298.3	50.3	1000.0
<i>PJM</i> Actual	97.31	33.17	210.2	11.2	999.0
Competitive	35.08	33.27	9.1	20.8	75.6
Cournot	87.05	36.00	171.8	22.7	1000.0
Cournot n.v.a.	1000.00	1000.00	0.0	1000.0	1000.0

Panel B: Off-Peak Hours.

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
<i>California</i> Actual	23.90	24.99	9.9	1.0	96.9
Competitive	26.10	27.44	6.4	1.2	50.3

Cournot	30.00	31.25	9.4	1.2	70.3
<i>New England Actual</i>	29.18	26.61	37.9	1.0	1000.0
Competitive	31.73	31.14	11.9	4.7	356.9
Cournot	32.63	30.54	18.7	4.7	481.3
Cournot n.v.a.	86.16	55.82	105.4	4.7	1000.0
<i>PJM Actual</i>	23.84	18.10	30.9	0.1	677.5
Competitive	25.42	23.78	6.3	16.4	52.7
Cournot	32.73	30.00	16.6	15.5	316.7
Cournot n.v.a.	900.57	1000.00	261.2	31.2	1000.0

Panel C: All Hours.

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
<i>California Actual</i>	29.69	27.99	19.1	1.0	225.0
Competitive	28.78	28.60	12.8	1.2	233.8
Cournot	34.56	33.60	15.6	1.2	233.8
<i>New England Actual</i>	36.96	28.52	56.6	1.0	1000.0
Competitive	34.73	32.06	21.6	4.7	356.9
Cournot	39.24	31.95	34.0	4.7	481.3
Cournot n.v.a.	144.56	67.28	206.0	4.7	1000.0
<i>PJM Actual</i>	45.92	20.99	122.8	0.1	999.0
Competitive	28.32	26.80	8.5	16.4	75.6
Cournot	49.06	31.27	98.4	15.5	1000.0

Cournot n.v.a.	930.45	1000.00	223.1	31.2	1000.0
----------------	--------	---------	-------	------	--------

* Note: There are 2,928 hourly observations: 880 peak and 2,048 off-peak.

Cournot n.v.a. means no vertical arrangements.

Table 4—Stability of Vertical Structure

Panel A: Major California Firms

	Base	Percent Change				
	Profits	AES	Reliant	Duke	Mirant	Dynergy
AES/Williams	43.3	5.8	-4.4	-3.0	-3.4	-3.6
Reliant	46.3	-4.7	5.5	-2.9	-3.3	-3.6
Duke	28.9	-5.0	-4.4	4.2	-3.4	-3.7
Mirant	37.7	-4.9	-4.4	-3.0	5.0	-3.7
Dynergy/NRG	48.8	-4.4	-4.0	-2.8	-3.2	5.0

Panel B: Major New England Firms

	Base	Percent Change				
	Profits	NU	PG&E	Mirant	Sithe	FP&L
Northeast Util.	332.5	-0.4	-0.1	-0.3	-0.5	-0.2
PG&E N.E.G.	122.6	-0.4	-0.3	-0.4	-0.8	-0.3
Mirant	24.3	-0.3	0.0	0.5	-1.2	-0.5
Sithe	22.6	-0.2	0.0	-0.5	0.6	-0.5
FP&L Energy	32.2	-0.2	0.0	-0.4	-1.0	0.3

Panel C: Major PJM Firms

	Base	Percent Change					
	Profits	PSEG	PECO	GPU	PP&L	Potomac	BG&E
Public Service Elec.	553.8	-1.7	-1.0	-2.3	-2.2	-0.9	-0.3

PECO	577.7	-1.0	-0.4	-1.4	-1.3	-0.6	-0.2
GPU, Inc.	522	-1.6	-0.9	-1.7	-2.1	-0.9	-0.3
PP&L Inc.	427.2	-1.4	-0.8	-2.0	-1.1	-0.8	-0.2
Potomac Electric	265.8	-2.2	-1.3	-3.1	-2.9	-1.2	-0.3
Baltimore G&E	399.4	-1.5	-0.8	-2.0	-1.9	-0.9	-0.2

Notes: Percent deviation from the base scenario profits, in millions of dollars earned by each firm during the Summer of 1999.

* *Bushnell*: University of California Energy Institute, Research Director, 2547 Channing Way, Berkeley, CA 94720-5180 (email: bushnell@haas.berkeley.edu). *Mansur*: Yale School of Management and NBER, Associate Professor of Economics, 135 Prospect Street, P.O. Box 208200, New Haven, CT 06520-8200 (email: erin.mansur@yale.edu). *Saravia*: University of California Energy Institute, Research Assistant, 2547 Channing Way, Berkeley, CA 94720-5180. We thank Soledad Arellano, Severin Borenstein, Joe Bowring, Judy Chevalier, Phil Haile, Stephen Holland, Jun Ishii, Wally Mullin, Steve Puller, Peter Schott, Fiona Scott Morton, Raphael Thomadsen, Larry White, Matt White, Frank Wolak, Catherine Wolfram, two anonymous referees, and seminar participants at Colby College, International Industrial Organization Conference, Massachusetts Institute of Technology, New York University, University of California Energy Institute, University of California at Irvine, University of Toulouse, Wesleyan University, and Yale University for helpful comments.