POWER

PWP-080

Pricing and Firm Conduct in California's Deregulated Electricity Market

Steven L. Puller

Original Draft: November 2000

This Version: August 2002

This paper is part of the working papers series of the Program on Workable Energy Regulation (POWER). POWER is a program of the University of California Energy Institute, a multicampus research unit of the University of California, located on the Berkeley campus.

University of California Energy Institute
2539 Channing Way
Berkeley, California 94720-5180
www.ucei.berkeley.edu/ucei

Pricing and Firm Conduct in California's Deregulated Electricity Market

Steven L. Puller*

August 2002

Abstract

This article analyzes the pricing behavior of electricity generating firms in the restructured California market from its inception in April 1998 until its collapse in late 2000. Using detailed firm-level data, I find that conduct is relatively consistent with a Cournot pricing game in 1998-99. In summer and fall 2000, behavior was distinctly less competitive, yet the dramatic rise in prices was more driven by changes in costs and demand than by changes in firm conduct. The five large non-utility generators raised prices slightly above unilateral market power levels in 2000, but fell far short of efficient tacit collusion.

*Department of Economics, Texas A&M University. Email: puller@econ.tamu.edu. I thank Severin Borenstein, James Bushnell, Greg Crawford, Carlos Dobkin, Richard Gilbert, James Griffin, Bronwyn Hall, Edward Kahn, Erin Mansur, Aviv Nevo, Anjali Sheffrin, Steve Wiggins, Catherine Wolfram, an anonymous referee, and seminar participants at various universities for their helpful comments. I am grateful for funding from the California Public Utilities Commission and University of California Energy Institute to support this work.

1 Introduction

In 1998, California opened the electricity generating sector to competitive forces by restructuring the method of procuring electricity. Private electric generating firms bid into a daily auction for the right to supply power to the electrical grid. Policymakers believed that a "deregulated" electricity market would lead to more efficient investment decisions in the long-run and lower prices in the short-run. It remains to be seen if investment is more efficient, but prices were volatile and at times extremely high. Figure 1 illustrates prices in the California wholesale market from its inception in April 1998 to its collapse at the end of 2000. Wholesale prices remained at low to moderate levels during much of 1998-99 but skyrocketed to unprecedented levels during summer and fall 2000. The incumbent utilities were required to purchase power at high wholesale prices and to sell to end-users at substantially lower prices. Eventually, the utilities lost their creditworthiness, the organized market broke down, and the state government was required to step in to purchase power.

Market power may have substantially contributed to the high prices. The market has characteristics that oligopoly theory suggests would favor either unilateral market power or collusion. Several large generating firms have strong incentives to exercise market power. Because demand is relatively inelastic, firms in this concentrated market have individual incentives to withhold capacity to drive up the price. In hours when demand is near industry capacity, individual firms are "pivotal" and will be able to ask for and receive a high price for power. Even in hours when no individual firm is pivotal, firms face fairly inelastic residual demand and are able to raise price above marginal cost. In addition, the market would appear to be especially conducive to tacit collusion. The bidding game is repeated daily between a fixed set of players that have very accurate information about each other's cost structure. The organized spot market is relatively transparent and provides each firm with strong signals about rivals' behavior. The goal of this paper is to estimate whether the spot market in California is best characterized by static non-cooperative oligopoly models or by tacit collusion arising from a dynamic game. I test for evidence of market power and estimate whether firm production behavior is more consistent with unilateral market power or tacit collusion.

Studies have found empirical evidence that firms in the California market exercise market power. Adopting a methodology pioneered by Wolfram [1999], Borenstein et al. [forthcoming] simulate a perfectly competitive market from 1998-2000 and compare those prices to actual prices. They find high price-cost margins during the high demand summer months with the margins becoming very large in 2000. Joskow and Kahn [2002] extensively analyze several data sources on California electricity generation during summer 2000 and find evidence of the strategic withholding of capacity by some generating firms.

Evidence exists of market power in other restructured electricity markets as well. The methodology of comparing actual prices to simulations of competitive prices has yielded evidence of market power in the Pennsylvania-New Jersey-Maryland market (Mansur [2001]), and the New England market (Bushnell and Saravia [2002]). Bushnell and Saravia compare measures of competitiveness in three restructured markets. Market power appears to be a high demand phenomenon. Lerner indices are small during time periods when demand is low relative to capacity. However, prices diverge from marginal cost when demand is large and the market is tight.

Bid data provide some inferences about the underlying strategic behavior. Wolfram [1998] analyzes bids into the England and Wales daily electricity auction and finds that bid markups are higher for generating units owned by a firm likely to have more inframarginal output that will receive the higher price if the unit sets the market price. This is evidence that firms take into account the effect of one production unit on the price earned by other units in production. Wolak and Patrick [1997] analyze bid functions by the two largest generating firms in the England and Wales market. They find evidence that the generators set bid prices close to marginal cost but strategically declared capacity unavailable on a short-term basis to raise the system price.[1]

Although there is evidence of some form of market power, there is not a clear understanding of the firm behavior underlying California electricity prices. For example, the rise in price-cost margins from 1999 to 2000 could have resulted from firms behaving less competitively on a similar demand function or firms behaving identically on a less elastic demand function. Similarly, market power may be prevalent in high demand hours because firms compete differently in those hours or because demand is less elastic. This paper decomposes the movement of prices into changes in input costs, changes in demand, and the exercise of market power.

I measure if firms exercised market power and whether that behavior was more consistent with unilateral market power or tacit collusion. The supply side of the industry is modeled as five strategic quantity-setting firms facing a competitive fringe. I derive theoretical firm supply relations under unilateral market power and efficient tacit collusion. Using detailed firm-level data, I estimate firm conduct with approaches that address Corts' recent critiques to "new empirical industrial organization" methods of estimating conduct (Corts [1999]). Estimating the shape of actual supply relations, I find evidence of static market power for much of 1998-99, and mixed evidence for 2000. Parameter estimates imply that the pricing behavior of electricity generators is approximately Cournot for 1998-99 and less competitive than Cournot during summer and fall 2000. Although firms were distinctly less competitive in 2000, I do not find evidence of efficient tacit collusion.

In Section 2, I describe the structure of the California electricity market. Section 3 reviews the theoretical models of static and dynamic pricing and describes how the existing empirical literature attempts to distinguish between competing models. I specify a general behavioral model that incorporates static and dynamic market power models as special cases. In Section 4, I describe my data. I measure price-cost margins for April 1998-November 2000 and find evidence consistent with large generation owners withholding output to raise the market price. In Section 5, I estimate the behavioral model using a panel of firm-level data. I conclude in Section 6 by discussing general implications for electricity deregulation elsewhere.

2 How The California Electricity Market Worked

Under deregulation, electricity still moves from the generator to the socket the same as it always has. But the ownership of the existing infrastructure has changed. The electricity industry consists of three sectors: generators that produce electricity, a high voltage transmission system that transports electrical energy from the generator to the area of consumption, and a distribution network that delivers electricity to end-users. Historically, these three sectors have been vertically integrated with government regulation of price, entry, and investment. Prior to April 1, 1998 the electricity industry in California was operated by the three major investor owned utilities: Pacific Gas & Electric in northern California, Southern California Edison in south central California, and San Diego Gas & Electric in the southernmost part of the state. These utilities were responsible for generating electricity and supplying customers in their service territories and were regulated by traditional rate-of-return regulation.

Beginning in the 1990s, policymakers in some countries began to separate the generation side of the industry from the transmission and distribution sectors, and allow firms to compete to supply electrical energy to the network.[2] In California, the restructured market opened in April 1998. The three incumbent utilities divested most of their fossil-fueled powerplants to private firms that bid daily to supply power. Southern California Edison divested the vast majority of its plants within a month and a half of the market opening to four different firms: AES-Williams, Dynegy, Reliant, and Thermo Ecotek. Pacific Gas & Electric divested its low cost units to Duke in July 1998 and most of the remaining units to Southern Energy (later spunoff as Mirant) in April 1999. San Diego Gas & Electric divested its plants to Dynegy and Duke in April and May 1999. By the end of the divestiture process, the thermal (fossil-fueled) generation market consisted of roughly five equal-sized firms and two small fringe firms (see Table 1) that together own roughly 54% of the electricity generation capacity in California. The remaining in-state capacity is two nuclear plants jointly owned by the utilities, a large number of hydroelectric units owned primarily by PG&E, and a variety of small independent plants paid under separate contracts. In addition, electricity is imported from neighboring states in virtually all hours. This paper analyzes the competitive behavior of the five large thermal firms.

California established several institutions to organize the trading and dispatch of electricity generation. The vast majority of electricity in California was traded within a day of its production and consumption. The three original utilities were still responsible for procuring power for customers in their service territories. These utilities were required to purchase their electricity from a specific day-ahead trading exchange created under restructuring legislation called the Power Exchange (PX).[3] On the day before delivery, the PX conducted a uniform-price auction for the following day. Each firm bidding to supply power submitted an upward-sloping supply schedule for each hour of the following day. Similarly, firms bidding to purchase power (primarily the incumbent utilities) bid downward-sloping demand schedules for each hour. The PX aggregated the supply and demand bids for each hour, and the market-clearing price defined the price at which all trades were settled. In the event that the accepted production and consumption schedules violated the capacity of the transmission lines, additional bids to increase and decrease output were used to ensure that there was no transmission congestion. California was divided into several zones between which transmission constraints were thought likely to bind frequently. If the original bids did not cause transmission congestion, PX prices for all of California were identical. If the original schedule had to be adjusted to meet transmission constraints, PX prices would differ in each zone. The generators were paid either a northern zonal or a southern zonal price.

During most of my sample period, 80-90% of all production was sold through the Power Exchange. Approximately 10% was sold through bilateral trades. The remaining balance of electricity was sold through an hourly real-time market conducted by the operator of the electricity grid (the Independent System Operator). Just as with the day-ahead market, the real-time market was a uniform price auction. Commitments to buy and sell power were purely financial so market participants were able to arbitrage price differences between the two markets.

Market designers set price caps to limit the exercise of market power. The Independent System Operator (ISO) chose a wholesale price cap that attempted to balance competing goals of encouraging efficient entry and preventing prices from rising significantly over marginal costs (which range from approximately $20-40 per megawatt-hour during 1998-99 and approximately $25-500 during 2000). Prices for electrical energy were capped at $250/MWh from the opening of the market in April 1998 through September 1999. The cap was raised to $750/MWh in October 1999, but then lowered to $500 in July 2000 and to $250 in August 2000.[4]

The institutions of the California market appear conducive to either unilateral market power or tacit collusion. Because storage or inventory of electricity is prohibitively costly, firms essentially must produce a quantity exactly equal to demand at every moment in time. An individual firm can raise the price by withholding some of its available capacity because its residual demand is likely to be inelastic. Several factors contribute to inelastic residual demand. The total demand for electricity is nearly *perfectly* inelastic to the wholesale price because consumers do not face prices that vary with the hourly wholesale price.[5] Any elasticity in residual demand arises from elastic supply by other firms. However, other firms are likely to have inelastic supply relations during periods of high demand. When demand reaches levels near the industry's capacity, if one firm were to withhold capacity to drive up the price, other firms would be unable to replace all of the withdrawn capacity. Hence, firms in this relatively concentrated industry are able to raise price and earn more revenue on all inframarginal output.

In addition, repeated interaction in the relatively transparent market could lead to a form of dynamic pricing. Supergame theory implies that tacit collusion is facilitated by frequent interaction, up-to-date information on rivals' behavior so defection can be detected, and barriers to entry that prevent collusive profits from being eroded. The California market is essentially an auction repeated daily between five large firms and a competitive fringe. The cost structure of the five firms is nearly common knowledge. All plants owned by the firms were formerly owned by regulated utilities and are still subject to environmental regulations that make operating characteristics a part of the public record. In particular, firms have good estimates of the fixed and variable costs of rivals' operations with the only uncertainty being whether a plant has had a short-term outage. Although they do not observe the bids by their rivals, firms do have information that is strongly correlated with rivals' bids. The website of the western U.S. transmission grid coordinator posted real-time generation data for plants comprising about 93% of thermal capacity until October 2000. Also, the ISO released with a one-day lag each plant's generation that was sold into the real-time market. Several electronic trading exchanges provided electricity traders with the means to observe a record of recent bilateral trades. Demand-side information is also common knowledge; firms observed the ISO's forecast of demand before bidding and the ex post realization of demand. Finally, entry into the market is difficult due to strict environmental siting requirements that can often take more than five years. There was no entry during my sample period.[6]

3 Distinguishing Between Static and Dynamic Market Power

This paper measures for the exercise of market power and estimates whether the production behavior of the five large firms is more consistent with unilateral market power or tacit collusion. Studies in the New Empirical Industrial Organization (NEIO) literature have estimated firm conduct by parameterizing the firm's static first-order condition (MR=MC) to allow for price-taking, Cournot competition, and monopoly pricing.[7] Unfortunately, explicitly estimating firm conduct has proven to be unsuccessful. A recent paper by Corts [1999] shows that traditional approaches to estimating conduct from the parameterized static first-order condition can lead to inconsistent estimates of the conduct parameter. He demonstrates that this approach can severely mismeasure the conduct parameter if the true underlying process is not identical on the margin to a conjectural variations game. Corts shows that if firms are engaged in efficient collusion, the traditionally estimated conduct parameter typically will underestimate market power.[8] The root of the problem is that if firms are colluding, the econometrician is estimating the wrong model; one should be estimating the dynamic first-order condition rather than the static first-order condition. The first-order condition of a set of tacitly colluding firms is the solution to maximizing joint profits subject to an incentive compatibility constraint that no firm has an incentive to deviate. As I show below, this dynamic first-order condition is very similar to the static condition with an additional term if firms are engaging in a level of collusion less than perfect price collusion (i.e. the joint monopoly outcome). If firms are engaging in imperfect collusion, the static first-order condition is mis-specified and one obtains inconsistent estimates of firm conduct. As a result, the best one can achieve by estimating the parameterized static first-order condition is to test non-nested hypotheses of perfect competition, Cournot competition, and perfect collusion (see Gasmi et al. [1992] and Nevo [2001]). The existing empirical literature does not to my knowledge suggest methods to estimate conduct when one possible conduct is imperfect collusion.

In this section, I derive models of firm supply relations under both static and dynamic pricing. I model the firms' strategic decisionmaking as a simple quantity-setting game. I derive a general first-order condition that represents each firm's supply relation under no market power, Cournot pricing, and efficient tacit collusion. The model allows me to consistently estimate conduct parameters in a manner that addresses the Corts critique. In Section 5, I use firm-level data to estimate each firm's supply relation and test if the estimated relation is consistent with static or dynamic pricing.

3.1 Static Pricing Model

In static models, firms choose single period quantities or prices to maximize profits without any intertemporal considerations of the effect of current behavior on the future competitive environment. For the California electricity market, a purely price-setting model is not appropriate because capacity constraints prevent any single firm from undercutting and supplying the entire market. Rather, one can view the strategic decision of the firm as to commit power plants to produce a certain amount of power a day-ahead and then true-up in the real-time market. Therefore, I estimate models of games in which firms choose quantities.[9]

Modeling the static game as a one-shot quantity-setting game is complicated by the fact that there were several sequential markets in California during the period I analyze. Firms sold power to both the day-ahead and real-time markets as well as a limited number of forward contracts. Sequential markets may lead to less market power than a single one-shot market. Power that is sold forward is not considered a part of the firm's inframarginal output when choosing the quantity sold in a real-time market. This will tend to mitigate the exercise of market power in a manner similar to the durable goods monopoly problem. Allaz and Vila [1993] show that if they can commit to observable forward market positions, firms will have an incentive to trade both forward and in real-time a total quantity greater than the one-shot quantity. However, a one-shot model is a reasonable first-order approximation to the California market. One can think of the day-ahead and real-time markets as a single energy market in which firms bid to supply a given quantity of power in the day-ahead market and then make small plant-by-plant adjustments in the real-time market. A final caveat is that some of the firms' output may be contracted forward. However, industry analysts suggest forward contracting was relatively small until 2000. I discuss the potential bias of my estimates in section 5.

To formalize the model, assume that N firms play a quantity game in which they choose to supply a given (perfectly inelastic) quantity subject to a capacity constraint.[10] Price is determined such that supply equals demand. Denote $P(\cdot)$ as inverse demand, c_i as marginal cost, q_i as individual firm quantity, and k_i as firm capacity. In period t, firm i chooses quantity of output to maximize profit subject to a capacity constraint:

$$\max_{q_{it}} \quad [P(q_{it} + q_{-it}) - c_i(q_{it})] \cdot q_{it} \qquad s.t. \quad q_{it} \leq k_{it}$$

This problem yields a first-order condition at the optimal quantity q^*_{it} of:

$$P(q^*_{it} + q_{-it}) - c_i(q^*_{it}) + \theta_{it} \cdot P'_t \cdot q^*_{it} - \lambda^*_{it} = 0 \tag{1}$$

where $\theta_{it} \equiv \frac{dQ^*_t}{dq_{it}} = 1 + \sum_{j \neq i} \frac{\partial q_{jt}}{\partial q_{it}}$ is the firm's belief about the effect of increasing its output on total industry output.[11] The parameter $\theta_{it} = \{0, 1, N\}$ corresponds to perfect competition, Cournot, and monopoly pricing (under symmetry), respectively. There are a limited set of values that θ may take to be either a Nash equilibrium or a consistent conjecture. Nevertheless, θ as a continuous variable is a meaningful index of the general (anti-)competitiveness of the market. Solving for the conduct parameter, one finds:

$$\theta_{it} = \frac{P_t(\cdot) - c_{it}(\cdot) - \lambda^*_{it}}{-P'_t q^*_{it}} \tag{2}$$

The conduct parameter is increasing in the observed difference between price and marginal cost adjusted for the sensitivity of price to an expansion of output (P'_t). The parameter λ^*_{it} is the shadow value of additional capacity when a firm is fully utilizing its existing capacity.

3.2 Dynamic Pricing Model

Models of dynamic games show that firms repeatedly interacting in an industry with entry barriers can sustain prices higher than one-shot equilibrium levels. Firms that engage in efficient tacit collusion choose production to maximize joint profits subject to the constraint that no firm has an incentive to deviate in order to earn higher one-time profits at the risk of starting a "price war". If demand shocks are not observed ex post, Green and Porter [1984] show firms can sustain prices above Cournot levels during periods of high demand but may revert to static equilibrium prices following negative demand shocks. However, if demand and prices are observed ex post, firms always can sustain the collusive regime but the level of collusion will depend upon current and expected future demand (Rotemberg and Saloner [1986], Haltiwanger and Harrington [1991]) and whether firms face capacity constraints (Brock and Scheinkman [1985], Staiger and Wolak [1992]). For example, if current demand is high, the incentives to cut price and earn deviation profits are high, so price must be lowered to check that incentive. Similarly, if demand is expected to rise in the near future, the future collusive profits may be higher, so firms have less incentive to deviate and start a price war. As a result, for a given level of demand, a higher level of collusion can be sustained when demand is rising than when demand is falling. However, these results in general will differ when firms face capacity constraints because those capacity constraints will affect both the deviation and punishment profits.

The major difference between the Green/Porter and Rotemberg/Saloner (and extensions) models stems from what firms are able to observe about their competitors' behavior. Firms in the Rotemberg/Saloner model observe the prices charged by all other firms. Green and Porter firms know only some signal correlated with rival behavior such as market price or their own realized shares. Because firms in the California electricity market have substantial information on their rivals' behavior, I assume a competitive setting similar to the Rotemberg/Saloner model. If they collude, firms are always in an efficient collusive regime and never shift into the non-cooperative "punishment" regime. I discuss the effects of possible regime-switching on my results in section 5.

Formally, I model the firm optimization problem when the industry is engaged in efficient tacit collusion. The firms choose a joint quantity Q_t^* to maximize joint profits subject to the constraint that no firm has an incentive to deviate from the collusive quantity. Deviation from the collusive quantity is punished by permanent reversion to the one-shot Cournot equilibrium.[12] Assume that demand and cost shocks are observed ex post so that deviating from the collusive regime can be distinguished from exogenous shocks to the environment. Assume that firms are symmetric and that sharing rules specify that each firm produces $\frac{1}{N}$ of the total output.[13] Due to symmetry, maximizing individual firm profit is equivalent to maximizing joint profit.

Denote individual firm i profit as π_i. π^*_{is} is optimal collusive profit in future period s. Let $\pi_i^{br}(Q_t)$ represent the individual profit to any firm that unilaterally deviates from the collusive regime by producing its one-shot best response to the collusive quantities of the other firms. Deviation will be punished by reversion to noncollusive profit π_i^p. $E_t[\pi_{is}]$ denotes expectations of future period s profit conditional on information known in period t. Finally δ is the discount factor between periods. Under efficient tacit collusion, firms maximize joint profit subject to the constraint that no firm has an incentive to deviate from the collusive regime:[14]

$$\max_{Q_t} \quad \sum_{i=1}^{N} \pi_i(\frac{Q}{N})$$

$$s.t. \quad \pi_i^{br}(Q_t) + \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi^p_{is}] \leq \; \pi_i(\tfrac{Q_t}{N}) + \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi^*_{is}]$$

A firm will choose not to deviate if current and continuation collusive profits exceed the profits of deviating in the current period and earning noncollusive (Cournot) profits forever afterwards. We can rewrite the optimization problem as:

$$\max_{Q_t,\mu_t} \quad \sum_{i=1}^{N} \pi_i(\frac{Q}{N}) + \mu_t \left[\pi_i(\frac{Q_t}{N}) + \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi^*_{is}] - \pi_i^{br}(Q_t) - \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi^p_{is}] \right]$$

The first-order condition is:

$$(1 + \frac{\mu_t^*}{N}) \cdot \left[P(Q_t^*) - c_i(\frac{Q_t^*}{N}) + P_t' Q_t^* \right] - \mu_t^* \frac{d\pi^{br}}{dQ}_t = 0 \qquad (3)$$

The first-order condition has a natural interpretation. When the incentive compatibility constraint is not binding ($\mu_t^* = 0$), this equation is simply the static joint profit maximization condition of a monopolist. This corresponds to perfect collusion. However, when the constraint is binding, joint $(MR - MC)$ must be lowered (Q raised) so the incentive compatibility constraint is not violated.

Firms can still earn more than one-shot Cournot prices but cannot sustain the joint monopoly price.

Because my analysis uses firm-level data, I transform the first-order condition from the industry level to the firm level. The dynamic first-order condition (3) can be rewritten to show the condition that each firm in a collusive regime is satisfying when choosing the collusive level of output:

$$P(Q_t^*) - c_i(q_{it}^*) + N \cdot P_t' \cdot q_{it}^* - \frac{\mu_t^*}{1+\frac{\mu_t^*}{N}} \frac{d\pi^{br}}{dQ}_t = 0 \tag{4}$$

Again this condition has a simple interpretation. When the incentive compatibility constraint is not binding, the last term is zero and I get the firm-level first-order condition for joint profit maximization. The firm internalizes the effects of price changes on the revenue for all firms' inframarginal output (Nq_{it}^*). When the constraint binds ($\mu_t^* \neq 0$), joint output must rise and price must fall so that no firm deviates to earn best-response profits. This can be seen in the equation above because the incentive compatibility constraint term (including the leading negative sign) is positive which shifts out the first-order condition so that equilibrium output is larger.

3.3 A General (Static and Dynamic) Supply Relation

The static and dynamic first-order conditions I derive above are essentially supply relations of firms engaged in unilateral market power or efficient tacit collusion. I estimate a firm-level model which incorporates as special cases the static (1) and dynamic (4) first-order conditions:[15]

$$P(q_{it}^* + q_{-it}) - c_i(q_{it}^*) - \lambda_{it}^* = -\theta_{it} P_t' q_{it} + \frac{\mu_t^*}{1+\frac{\mu_t^*}{N}} \frac{d\pi^{br}}{dQ}_t \tag{5}$$

H_1: No Market Power: $\theta_{it} = 0$, $\mu_t^* = 0$, $\lambda_{it}^* \geq 0$

H_2: Static Market Power: $\theta_{it} = 1$, $\mu_t^* = 0$, $\lambda_{it}^* \geq 0$

H_3: Dynamic Market Power: $\theta_{it} = N$, $\mu_t^* \geq 0$, $\lambda_{it}^* = 0$

We can view (5) as a general model capturing three alternative explanations for price above marginal cost. First, observed price-cost margins may represent scarcity rents for new production capacity in a perfectly competitive environment ($\lambda^*_{it} > 0$). Firms utilize all capacity with marginal cost less than the price, and margins signal the value of added capacity. Second, margins may result from firms unilaterally withholding current capacity to raise the price and earn higher revenue on their *own* inframarginal units. This corresponds to a model of Cournot competition with capacity constraints. Finally, firms may be jointly withholding capacity to raise the price on *joint* inframarginal units, with this regime maintained by adjusting quantity so that no firm has an incentive to deviate from joint profit maximization.[16] The shapes of these supply relations are illustrated in Figure 2. The vertical axis is the price-cost margin adjusted for scarcity rents and the horizontal axis is the effect of selling one more unit of output on inframarginal revenue. Price-taking firms sell all economical capacity at marginal cost independent of the size of inframarginal sales. Cournot firms drive price above marginal cost when their sales are large and when demand is steep. Finally, tacitly colluding firms internalize the effects of all firms' inframarginal revenue but may have to adjust price downwards to ensure the incentive compatibility constraint is not violated. I measure the behavior of firms in the California electricity market by estimating whether the actual supply relations are more consistent with the theorized supply relations under no market power, unilateral market power, or efficient tacit collusion.

4 Data

In order to estimate the supply relations, I require data on hourly market price as well as each firm's output and marginal costs. Fortunately, restructured electricity markets are subject to data reporting requirements that provide the empirical researcher with rich data on demand, cost structure, and output. I describe an overview of the data in the main text and leave the interested reader to find details and explanations of the assumptions in Appendix A.

4.1 Measures of Price, Marginal Cost and Output

Data on the hourly production of each powerplant are available from EPA's Continuous Emissions Monitoring System (CEMS). CEMS contains hourly output and emission data for all fossil-fueled generation units in the California market except several small generation units. I can reliably calculate marginal cost because the production technology is fairly homogenous, and data are available on the technological capacity of each of the five firms. The powerplants burn natural gas to generate steam or a hot stream of gas that turns a turbine and is converted into electricity. Data are available on the average conversion factors between the heat content of the fuel and the electricity output of each generating unit (Kahn et al. [1997]). Using data on the fuel input cost for each generator, I calculate the marginal fuel costs. Several plants in southern California were required to purchase environmental permits for each pound of nitrogen oxides (NOx) emitted. The hourly marginal permit cost is calculated as the monthly quantity-weighted average price of permits multiplied by the unit's hourly emissions. Adding on an estimate of variable operating and maintenance costs from Borenstein et al. [forthcoming], I estimate the marginal cost for each unit.[17] I assume marginal cost to be constant up to the capacity of the generator.

Generators occasionally experience both scheduled and unscheduled downtime for maintenance. Some analysts have suggested that firms exercised market power by shutting down generating units, particularly in 2000. I observe shutdowns but cannot distinguish between true outages and withholding an entire unit to raise the price. In measuring market power, I assume that any plant not operating is unavailable. This could bias downwards my measure of market power if firms shutdown plants to exercise market power. However, an ISO analysis of confidential bid data suggests that this bias may not be too severe in 2000. The Sheffrin [2001] analysis of bid data suggests that all but one firm primarily exercised market power by bidding in available capacity at high prices rather than entirely shutting down available plants.

I need to make several assumptions about a firm's behavior in order to determine the firm's marginal cost in a given hour. If, on a given hour, I look across all of a firm's generating units, I am likely to see the firm operating a lower marginal cost unit at less than full capacity while also operating another higher marginal cost unit. One explanation is that the firm expects that the higher cost unit will be operating in the coming hours (perhaps when total demand is higher) and it needs to keep the higher cost facility operating. Under this scenario it is unclear whether the proper measure of the firm's marginal cost is the lower or higher cost unit that still has available capacity.

If I use the lower cost unit, I ignore the fact that the firm is solving a more complicated dynamic optimization problem and that the true measure of marginal cost should include the shadow values of the operating constraints. If I use the higher cost unit, I ignore that the higher cost unit may be running because it was called under outside reliability contracts by the grid operator.[18] Because the former bias is potentially more severe, I define the firm's marginal cost to be the marginal cost of the most *expensive* unit that is operating and has excess capacity:

$$MC_{it} \equiv \max_j\{MC_{ijt}\} \text{ where } j \text{ indexes firm } i\text{'s units operating in hour } t \text{ with excess capacity}$$

I can determine if units have excess capacity by comparing observed output from the EPA data to the unit's capacity. One problem with this measure of firm-level marginal cost is that there are several small powerplants for which I have no quantity data.[19] Most of the missing units are small high cost units that only operate when demand is very high. Because I have no data on these high cost units, I tend to bias downwards the marginal cost of the firms owning these units when the units are operating. This bias is likely to be most severe for Dynegy.

The price earned for the observed output is not always known by the econometrician because the power can be sold in the day-ahead market (the Power Exchange) or the real-time energy market (the ISO). I use the Power Exchange day-ahead energy price because 80-90% of all transactions occurred in the PX during my sample and a simple arbitrage argument suggests that day-ahead and real-time prices should be equal in expectation.[20] A final issue regarding my measure of price is locational differences in price. As I describe above, prices in the PX are uniform across California unless transmission constraints bind. When transmission constraints between the north and south bind, there are essentially two different markets clearing at two different prices. Most firms own powerplants in a single transmission zone. However, one firm (Duke) owns generators in both the north and south. During hours when the north and south have different prices, I separate off output from Duke's southern plants and call the firm DukeSouth.

My measure of output is the total production by each firm's generating units. This may slightly mismeasure the actual amount of generation sold to the energy market (and hence inframarginal output) for several reasons. It may understate output for the firms that own the small peaker units with no EPA data (e.g. Dynegy). In addition, late in the sample period firms increasingly sold power through an out-of-state third party to avoid the price cap. The price cap applied only to sales by in-state generators and there was no cap on out-of-market purchases. In a practice called "megawatt laundering", generators sold power to third parties on the border of California only to sell the power back to California at prices above the cap. Therefore, potential mismeasurement of inframarginal sales may affect my estimates for Dynegy and for all firms late in the sample period. I discuss the sign of the potential bias in section 5.

In section 5, I estimate supply relations for a particular hour of each day when the first-order conditions derived in section 3 most resemble the firms' actual optimization problem. Recall that my measure of marginal cost includes variable costs of fuel, operating and maintenance, and emission permits. Several conditions could make my measure of marginal costs differ from the actual marginal costs. First, powerplants face shadow costs of intertemporal adjustment constraints on the rate at which they can increase or decrease output. Therefore, I focus on the period from 5-6pm (hour 18) when those constraints are unlikely to bind. Natural gas powerplants in California can typically ramp from zero to full capacity in times varying from one to three hours. On an average day the total demand nears its peak by 11am and maintains approximately that level until around 9pm. By the time 6pm arrives each day, firms have had ample time to ramp up their units while still having the necessary time to ramp down by the time demand begins to fall.[21] Therefore, I focus on hour 18 and assume any shadow costs of operating constraints to be zero.[22] Also, I focus only on generating units that are operating. If it has a unit shut down, a firm incurs startup costs to fire up that unit. In order to deal with startup costs, I analyze the firms' utilization of units that are already operating during the particular hour.

4.2 Summary Statistics

The observed production behavior suggests firms are not acting in a perfectly competitive manner during hour 18. A price-taking firm will fully utilize capacity with marginal cost less than the price. When a competitive firm is producing below capacity, one expects the marginal cost of the unused capacity to be above the price. Table 2 displays summary statistics of the difference between price and the marginal cost of each firm's highest cost operating unit with unused capacity in hour 18. I calculate the price-cost margins using two different assumptions about the true capacity of a unit.[23] Defining capacity as either 80% or 90% of the nameplate rating, I find that firms very often observe price above marginal cost, yet fail to utilize capacity. DukeSouth, Duke, and Reliant crank up to capacity in more hours than AES, Southern, and Dynegy. When they are not producing at capacity, firms vary in their average margins. Southern, Reliant, and DukeSouth enjoy the highest price-cost margins although this result is driven to some extent by the time period in which the firms were in the market.[24] These margins imply a median Lerner index of 0.13.[25] These results are robust to my definition of capacity. As another check for robustness, I consider the possibility that I may understate firms' marginal costs. Separately, I calculate that firms have excess capacity yet observe margins above \$10 in approximately 38% of firm-hours and greater than \$30 in approximately 22% of firm-hours. It is highly unlikely that marginal costs are this severely mismeasured so there is strong evidence that firms are not acting as price-takers. This conclusion is supported by other studies of the California market including Borenstein et al. [forthcoming] and Joskow and Kahn [2002].

Price-cost margins vary considerably over my sample period of April 1998 to November 2000. I calculate the simple average of each firm's margin in each hour. If it is producing at capacity, the firm's margin is set to zero. Figure 3 shows that margins are higher during the third and fourth quarters of each year when total demand for electricity is high in California. Margins during low demand winter and spring months are actually negative in 1998 and hover around zero in 1999 and most of 2000.[26] I emphasize that these margins are not scarcity rents because these are differences between price and marginal cost when *firms have excess capacity*. These results are consistent with Borenstein et al. [forthcoming] who find the largest divergences between price and marginal cost during the summer months and particularly in 2000. I the next section, I estimate whether the changes in margins resulted from changes in the residual demand faced by the five large firms or from changes in how those firms competed on their residual demand.

5 Estimation of the Behavioral Model

I estimate if firms supply power in a manner more consistent with static or dynamic market power. The first-order conditions derived in section 3 are supply relations for firms operating in different types of competitive environments. I apply the model of firm behavior to data and identify parameters of the supply relations that allow me to make inferences about the competitiveness of firms in the California market. Estimated supply relations are compared to the theorized supply relations depicted in Figure 2. First, I estimate the static first-order condition (equation (1)) and find the data are fairly consistent with Cournot pricing for 1998-99. However, behavior is less competitive than Cournot in 2000. Then, I estimate a form of the general first-order condition (equation (5)) and reject the model of efficient tacit collusion.

5.1 Static Model

First, I estimate the static first-order condition equation (1) for each firm in the California market which can be rewritten as:

$$P(q_{it}^* + q_{-it}) - c_i(q_{it}^*) - \lambda_{it}^* = -\theta_{it} P_t' q_{it} \tag{6}$$

This model says that firms exercising market power ($\theta > 0$) observe higher price-cost margins (adjusting for scarcity rents on capacity, λ_{it}^*) when they have more inframarginal output or are operating on price sensitive areas of demand.

Below I detail my approach to modeling the California market and estimating behavior. To summarize, I model the supply side as five large strategic firms and a competitive fringe. Total demand is perfectly inelastic because few customers pay the hourly price of energy. The residual demand of the five strategic players has some degree of elasticity due to supply elasticity by the fringe. Residual demand for power from the five strategic firms is the observed (price inelastic) demand minus the supply by the competitive fringe. I estimate how the five firms compete on the residual demand they face.

Demand Side. Total residual demand of the five strategic firms (Q^D_{strat}) is the total (perfectly inelastic) market demand (Q^D_{total}) net of supply by the competitive fringe (Q^S_{fringe}):

$$Q^D_{strat}(p) \equiv Q^D_{total} - Q^S_{fringe}(p)$$

I estimate the supply function of the competitive fringe and use the negative of the slope of fringe supply to estimate the slope of inverse residual demand P'_t faced by the strategic firms. This is diagrammed in Figure 4.

The competitive fringe includes fringe thermal generators, nuclear generation, hydroelectric and geothermal power, small independent producers, and imports from outside of California.[27] I assume that these suppliers do not bid strategically and can be modeled as a competitive fringe. This assumption appears reasonable. The independent and nuclear units are paid under various regulatory side agreements so revenues are independent of the price in the energy market.[28] The assumption of price-taking supply of hydroelectric and geothermal power is slightly more problematic. It is difficult to directly assess whether hydroelectric power is supplied competitively because measuring the marginal cost of hydroelectric output involves measuring the opportunity costs of using the potential energy of a reservoir in some other period.[29] However, the owners of hydroelectric assets in California are the same utilities that are also buyers of power and have very dulled incentives to influence the price. Finally, firms importing power into California are likely to behave competitively because most are utilities with the primary responsibility of serving their native demand and then exporting "excess generation". Borenstein et al. [forthcoming] make similar assumptions about the behavior by firms owning nuclear, hydroelectric, and import generation.

I estimate the (competitive) supply by all fringe suppliers for hour 18 of each day. Fringe supply is a function of the PX day-ahead electricity price in California as well as cost conditions (e.g. price of natural gas) and seasonal supply variation (e.g. hydroelectric reservoir levels or scheduled nuclear outages). I model fringe supply as having a constant price elasticity so I estimate the model in logs.[30] To incorporate input cost variation over time, I include the price of natural gas as well as month-year and day of week dummy variables to capture reservoir levels and nuclear outages. Fringe supply includes imports of "excess generation" from neighboring regions to California. As a determinant of imports, I include differences in neighboring state mean daily temperatures from a baseline temperature that one would expect to necessitate little heating or cooling (65 degrees).[31] Because price is endogenous to the fringe quantity supplied, I instrument price with the day-ahead forecasted demand (which is independent of price). The model is given by:

$$\ln Q^S_{fringe\,t} = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln GasPrSouth_t + \beta_3 \ln GasPrNorth_t + \\ \beta_4 \ln Diff65TempNeigh_t + \beta_5 DAYDUM_t + \beta_6 MONTHDUM_t + v_t \quad (7)$$

The price elasticity β_1 can be used to calculate the slope of fringe supply which is the same magnitude but opposite sign of the slope of the residual demand faced by the five strategic firms.

Supply Side. In order to estimate the supply relation (6) by the five strategic firms, I need measurements of price, marginal cost, output, and the shadow value of capacity. The data on price, marginal cost, and output are described above. However, I cannot measure the shadow value of additional capacity (λ^*_{it}). The shadow value is zero when capacity constraints are not binding, however the value is unknown when constraints are binding. The proper measure is the difference between marginal cost and the parameterized marginal revenue ($P+\theta P'q$) evaluated at the capacity constrained quantity. The shadow values vary by both firm and time, however adding a separate parameter for each firm-hour when a firm is at capacity would add excessive parameters to the model. Therefore, I add to each supply relation a single dummy variable ($CAPBIND$) equal to 1 if capacity constraints are binding and equal to zero otherwise. The coefficient on $CAPBIND$ is the *average* shadow value of added capacity. To test if this specification problem affects my conduct estimates, I estimate the conduct parameter θ using only observations when the capacity constraints are not binding ($\lambda^*_{it} = 0$). Firms are producing at capacity in only 4.7% of firm-hours in my dataset. The results are similar to those below and are reported in Appendix B.

The static first order condition (6) is in general overparameterized because it allows each firm to have a different behavioral parameter each period. Before examining the possibility of heterogeneous behavior across firms, I assume all firms are strategically choosing quantity in the same manner and restrict the conduct parameter to be equal across all firms in the industry. The supply relation is modeled as:

$$(P-c)_{it} = \lambda \cdot CAPBIND_{it} - \theta \cdot P_t' \cdot q_{it} + \epsilon_{it}$$

In order to relate the estimated fringe supply elasticity to the slope of strategic demand, I use the definition of elasticity $\beta_1 = \frac{P_t}{P_t' Q^S_{fringe\,t}}$ and plug in for P_t':

$$(P-c)_{it} = \lambda \cdot CAPBIND_{it} + \frac{\theta}{\beta_1} \frac{P_t \cdot q_{it}}{Q^S_{fringe\,t}} + \epsilon_{it} \tag{8}$$

This supply relation is identified by shifts in the total California demand. I instrument firm-level output with the total (perfectly inelastic) forecast of demand.

I simultaneously estimate the system of fringe supply (7) and each firm's supply relation (8) via the generalized method of moments. The error term in each supply relation is modeled as heteroskedastic, contemporaneously correlated with the errors in the other supply relations, and serially correlated with its own error for the past 7 days. This estimation would correspond to three-stage least squares except that I allow for heteroskedasticity and individual serial correlation.

Before I show formal estimation results, let me illustrate the shape of the supply relation by strategic firms. Note that when $CAPBIND$ is zero (95% of firm-hours), the static model reduces to a simple bivariate (instrumental variables) regression: $(P-c)_{it} = \theta(-P_t')q_{it} + \epsilon_{it}$. I can graph the data in two dimensions and compare it to the theorized supply relations of Figure 2. Figure 5 plots the price-cost margins against the fitted values of $-P_t' q_{it}$.[32] The slope of this relationship is an estimate of the conduct parameter. Recall that the static model says that if behavior (θ) is constant, then the margins are linear in $-P_t' q_{it}$: firms have higher margins when (1) they have more inframarginal quantity and (2) they are operating on less price sensitive areas of demand. In addition, the relationship should go through the origin because firms with no inframarginal output have no incentive to price above marginal cost.

Figure 5 depicts the shape of the supply relations. The top panel plots the kernel regression estimate and the data for the complete sample of July 1998-November 2000.[33] The relationship has roughly a constant slope up to approximately $-P_t'q_{it} = 25$ after which the relationship is steeper and non-monotonic. The relationship approximately passes through the origin as one expects from a conjectural variations game. Observations with instrumented $-P_t'q_{it} < 25$ comprise about 70% of firm-hours and roughly correspond to hours when firm output is less than 1400 MW and the slope of residual inverse demand is flatter than -\$0.02/MW. The steeper supply relation for observations above $-P_t'q_{it} = 25$ may imply less competitive behavior during high demand hours. When total demand is larger, strategic firms have less elastic residual demand and can raise price further above marginal cost (holding conduct constant). Figure 5 suggests that not only is residual demand less elastic when demand is high, but conduct may be less competitive as well. These two factors would contribute to make prices substantially higher during peak periods. Alternatively, the steeper supply relation may result from my overestimating the fringe supply elasticity (and hence the demand elasticity) in these peak hours. Finally, I do observe periods when firms are operating with small negative margins. These observations are low demand hours mostly occurring in April and May of each year during which firms operate only a small fraction of their capacity.

The bottom panel of Figure 5 illustrates the supply relation before and after June 2000 when the California market experienced dramatically higher prices. The supply relations for pre-June 2000 and post-June 2000 are both approximately rays through the origin, as predicted by the static pricing model. However, the relationship is almost uniformly steeper after June 2000. This suggests the market was less competitive after June 2000.[34]

Next, I show formal results from jointly estimating the system of fringe supply (7) and the strategic firm supply relations (8).[35] I break down the sample into a period during which there were four firms in the market from July 1998-April 1999 and another period with five firms from April 1999-November 2000.[36] Results are shown in Table 3 and are similar for both time periods. Fringe supply is relatively inelastic in both periods (0.15 and 0.19). Given the relative size of the fringe and strategic players, this suggests that the strategic firms face a total residual demand elasticity of approximately -1.97 during the 4-firm period and -0.98 during the 5-firm period. Therefore, identical competitive behavior would lead to higher price-cost margins in the later period. Higher natural gas input prices leads to less supply by the fringe, and out of state temperatures significantly affect fringe supply in the second period but not the first.

The estimate of the supply relation by the strategic firms suggests the firms are behaving approximately Cournot. In the first period from July 1998 to April 1999, the coefficient on $\frac{P_t \cdot q_{it}}{Q^S_{fringe\,t}}$ and the estimate of β_1 imply $\hat{\theta} = 1.00$ with a standard error of 0.02. In the second period of my sample from mid-April 1999 to November 2000, the results are similar and I obtain a slightly larger estimated conduct parameter $\hat{\theta} = 1.13$ with a standard error of 0.03. I fail to reject Cournot pricing during the 4-firm period and observe pricing statistically higher than Cournot levels in the 5-firm period.[37] Although pricing behavior was relatively similar in both periods, margins were higher during the 5-firm period when strategic firms faced less elastic residual demand. Finally, the estimates of λ imply that firms operating at capacity are willing to pay \$25.25 and \$57.51 on average for an additional MW of capacity during the 4-firm and 5-firm periods, respectively.

Next, I estimate how conduct varies over my sample period. Recall from Figure 3 that direct measures of price-cost margins are highest during the high demand summer months and lowest during the relatively low-demand winter. In particular, margins peaked during the second half of 2000, and also were sustained at high levels during the summer of 1998 and fall of 1999. However, periods of high price-cost margins do not necessarily reflect periods with less competitive conduct. High margins result from less competitive behavior and/or less elastic residual demand. I expect residual demand to be more elastic during the winter when total demand is low and supply by non-strategic players (e.g. hydroelectric) satisfies a substantial fraction of that demand. In contrast, residual demand is less elastic during the summer and fall when demand peaks and water reservoirs are low. Therefore, a priori it is unclear whether the high summer margins imply that behavior is less competitive. My estimation decomposes the margins into the effects of firm conduct and residual demand elasticity. I estimate conduct for the peak and off-peak months of each year where peak months are defined to be June-November and off-peak months are December-May.[38] The first panel of Table 4 reports conduct parameters in which conduct is restricted to be equal across firms but is allowed to vary by time period. I also report estimates of the average strategic residual demand elasticity and the average price-cost margins that resulted from the firms' conduct on that residual demand.

Margins were largest during the peak demand months of each year. These margins resulted from some combination of firms behaving less than competitively on a residual demand function that is not perfectly elastic. During peak months of 1998, firm production was very consistent with Cournot behavior ($\hat{\theta} = 1.03$) and led to margins averaging \$16.43 when faced with relatively elastic residual demand. By the peak period of 1999, residual demand was less elastic due to further divestiture from fringe to strategic firms. Nevertheless, average margins were comparable to 1998 because behavior was more competitive than Cournot ($\hat{\theta} = 0.87$). The peak months of 2000 generated hour 18 margins five times larger than in 1998-99. The estimates suggest these large margins resulted from both less elastic residual demand and less competitive behavior ($\hat{\theta} = 1.38$).

During the offpeak months of 1998-2000, hour 18 margins averaged less than \$5/MWh. From December 1998 to May 1999, I divide the sample into the periods before and after Southern acquired three large powerplants from PG&E and entered the market. During the first subperiod I estimate conduct to be significantly less competitive than Cournot ($\hat{\theta} = 2.36$) largely due to a very large estimated demand elasticity. It is difficult to draw conclusions for these months, however the large conduct parameter is driven by one particular month. The model is not a good fit for data in December. When I exclude December, I obtain a conduct estimate of $\hat{\theta} = 1.06$. In late April and May 1999, the estimated conduct parameter is very low with residual demand estimated to be very elastic. In the second offpeak season from December 1999 to May 2000, margins were low despite fairly inelastic demand because conduct was relatively competitive ($\hat{\theta} = 0.66$).

Overall, the market appears to have become more competitive in 1999 and less competitive in 2000. The results suggest that the skyrocketing margins during the summer and fall of 2000 resulted from firms operating less competitively on a less elastic residual demand function. Residual demand was substantially less elastic (-0.57) due to an unusually hot summer that increased demand in California and reduced imports from other western states also experiencing high demand. Also, low levels of snowfall the previous winter reduced imports from the northwest. I estimate behavior to be distinctly less competitive ($\hat{\theta} = 1.38$) than in previous peak demand months. These estimates formally confirm the apparent pivot in the supply relation after June 2000 seen in Figure 5.

However, various institutional changes in 2000 may bias my conduct parameter estimates. In late 2000, the utilities began to face financial crises that could prevent them from paying for power purchased on the wholesale market. When skyrocketing wholesale prices threatened the creditworthiness of the utilities, the risk of non-payment may have increased marginal costs of supplying power beyond the simple production costs. My measure of marginal cost may understate the true cost of supplying power in late 2000 and bias upwards my conduct estimates. However, there are several factors that may lead me to understate the true conduct parameter as well. The most severe concern is that firms forward-contracted some of their production and that I mis-measure the output sold to the PX/ISO energy market. There is widespread belief that in 2000 several firms forward-contracted some fraction of their production. Firms only have an incentive to raise the price on the amount they produce beyond the contract position because the price earned on the contracted quantity is already locked-in. I assume all observed production is sold in the PX/ISO energy market. If some of the observed generation is sold forward, firms were enjoying the same profit margins for smaller quantities sold through the energy market. This would imply that I understate the conduct parameter θ. A final potential bias in 2000 is that some transactions through the PX/ISO market did not occur at the PX/ISO prices. The fall and winter saw some out-of-market transactions above the price cap ("megawatt laundering"). The bias from risk premia is only a concern during the last few weeks of my sample whereas the bias from contracts and out-of-market transactions likely exists for much of the summer and fall. Therefore, my conduct estimates are likely biased downwards in 2000.

In some seasons, the conduct parameter estimates reject all theorized equilibrium values of games of static pricing and perfect tacit collusion (i.e. $\theta = \{0, 1, N\}$). Although one would ideally hope to reject all but one theorized behavioral parameter, it is not surprising that I am unable to identify a equilibrium pricing model for each period. In each peak and offpeak season, I estimate some average measure of conduct. Conduct may vary over a season either because firms play different equilibrium outcomes or because firms are not always in equilibrium as they learn to compete in the newly deregulated market. Alternatively, one may view conduct estimates statistically higher than Cournot levels as an equilibrium of a repeated game (by a folk theorem result) falling short of efficient tacit collusion. Nevertheless, I can make some inferences about the overall competitiveness of the market. If I treat the estimate of θ as a continuous measure of competitiveness, the market displays levels of competition that varied substantially less than the price-cost margins. From 1998-2000, conduct varied moderately with a general strengthening of competition during summer 1999 and a weakening of competition during most of 2000. The dramatic variations in prices (from Figure 1) were more driven by changes in costs and residual demand elasticity than by the conduct of the firms.[39]

Finally, I allow the conduct parameters to vary by firm and estimate each firm's individual competitive behavior. Conduct parameter estimates are reported in the bottom panel of Table 4. I find a modest degree of heterogeneity in behavior. During the period with four strategic firms in the market from July 1998-April 1999, Reliant's behavior is statistically indistinguishable from Cournot while AES is more competitive and Duke less competitive than Cournot. Dynegy has a particularly large parameter estimate that decreases but remains high during the five-firm period.[40] The high conduct parameter estimate may result from the fact that I have incomplete data for some of Dynegy's small peaker plants. During the period from April 1999-November 2000 with five strategic firms, conduct is more competitive than Cournot for AES and less competitive for Dynegy, Southern, Duke, and Reliant. For the four firms in the market the entire sample, conduct is relatively similar across the two periods with the exception of Dynegy.[41] When I focus on the period of the price runup in June-November 2000, firms are uniformly less competitive. Dynegy (with data caveats) and Southern appear to be the least competitive and AES is most competitive. In Appendix B, I report conduct estimates under alternative assumptions about my data.

5.2 Dynamic Model

Results from the static behavioral model are consistent with static pricing over much of my sample period. However, estimating conduct using the static first-order condition can lead to inconsistent conduct parameter estimates as shown by Corts [1999]. In this section, I estimate the dynamic first-order condition to check for this potential mis-specification.

Before formally estimating the model, I provide informal evidence against dynamic pricing.

The shape of the estimated supply relation in Figure 5 fails to suggest efficient tacit collusion. Corts shows that conduct parameter estimates are not consistent if the true underlying game is not equivalent on the margin to a conjectural variations game. The supply relation for a conjectural variations game is a ray through the marginal cost intercept with higher θ parameters corresponding to rotations in the supply relation. Figure 5 suggests the firms are engaged in a conjectural variations game: the supply relation appears to be a ray through the origin. If we had observed positive margins for very small levels of output, we may believe firms are engaged in some other (non-conjectural variations) game and that the static model would not yield consistent estimates of conduct.

The results of the static behavioral model estimate conduct parameters slightly higher than Cournot levels for several time periods and firms. If the underlying game is efficient tacit collusion, conduct parameters from estimating the static model are biased as discussed in section 3.3. Therefore, I estimate the general first-order condition (equation (5)) to test for dynamic pricing. The dynamic first-order condition is simply the first-order condition of a joint monopolist with an added term (the last term in equation (5)). This extra term captures the additional output (relative to joint monopoly output) required to prevent any individual firm from deviating and earning best-response profits. Although I do not have data on this term, note that this term is constant across all firms during a given period. Therefore, I can condition out this term by including time fixed-effects. Including fixed effects for *every* time period (i.e. every day) would add many parameters and absorb much of the variation in margins. Instead, I assume that the level or "bindingness" of the incentive compatibility constraint is constant for similar periods of time, and I add fixed effects for those time periods.

The level of the incentive compatibility constraint depends upon several factors. First, the size of residual demand relative to individual firm capacity determines incentives to deviate and earn best response profits (Staiger and Wolak [1992]). Individual firm capacity does not change substantially during the months exhibiting market power because nearly all plants are online. Therefore, I essentially need to capture the size of the residual demand function. I observe the quantity produced in equilibrium on the residual demand function, but this quantity is endogenous to price. To mitigate this endogeneity, I use quartiles of equilibrium residual demand to capture these effects on the incentive compatibility constraint.

Second, because electricity demand follows a seasonal cycle, the level of the incentive compatibility constraint varies over the year depending upon whether demand is expected to rise or fall in the future (Haltiwanger and Harrington [1991]). I include month fixed effects to incorporate seasonal adjustment in the incentive to deviate. These month and demand quartile fixed effects capture the incentive compatibility adjustments ($\hat{IC_t}$) depicted in Figure 2. I estimate the dynamic model just as I have estimated the static model above, except that I add month and quartile fixed effects to the strategic firm supply relations to capture the incentive compatibility term.

One can also view this estimation as directly addressing the Corts critique. Corts shows that estimates of conduct parameters using the static first-order condition are not consistent if the true supply relation is not a ray through the marginal cost intercept (or that $P - c$ is not a ray through the origin). The specification of the static first order condition that I estimate above has no intercept and forces the estimated relationship to go through the origin. This specification frees up that restriction and allows the supply relation to have non-zero intercepts.

If the firms are efficiently tacitly colluding, this specification yields a consistent estimate of the conduct parameter $\theta = N$. However, if the firms are following a static pricing model, the estimators of $\hat{\theta}$ from both the static model (equation(6)) and this more general model (equation (5)) are both consistent. Cournot pricing should yield $\hat{\theta}$ statistically indistinguishable from one in this specification.

Table 5 shows estimated conduct parameters for different specifications of the dynamic model. I allow for the incentive compatibility terms to be captured by residual demand quartile and/or month fixed effects. I find the estimated conduct parameters to be significantly lower than N (either 4 or 5) in all specifications.[42] I reject efficient tacit collusion in all periods.

For the most part, I obtain conduct estimates in this general specification very similar to those in the static specification. During the period with four firms in the market, I obtain conduct parameter estimates close to zero for specifications including demand quartile effects and near one for the specification including only month effects. Apparently, margins vary by total residual demand but do not vary by individual firm output after controlling for total residual demand. However, if only month effects are used to add intercepts to the supply relations, I obtain a conduct parameter estimate very close to the estimate obtained in the static model (0.97 vs. 1.00). In the five firm period, I obtain conduct estimates again generally consistent with Cournot pricing. Conduct parameter estimates are similar to the parameter estimate under the static model ($\hat{\theta} = 1.13$). Under all specifications, I find pricing slightly higher than Cournot levels but substantially smaller than efficient tacit collusion levels. Finally, for the period of the price runup in June-November 2000, the dynamic model yields results similar to the static model. Estimated conduct parameters are above Cournot levels but substantially below levels implied by efficient tacit collusion.

These results provide strong evidence against *efficient* tacit collusion. However, the conduct parameter estimates above one for June-November 2000 suggest that pricing may have been less competitive than unilateral market power. Recall that the general behavioral model (equation (5)) contained three alternative hypotheses of no market power, unilateral market power, and efficient tacit collusion corresponding to $\theta = \{0, 1, N\}$, respectively. However, there exist behavioral interpretations consistent with the estimated conduct parameter in late 2000 of $\hat{\theta} \geq 1$. As discussed above, firms also can sustain any collusive level of pricing $1 \leq \theta \leq N$ by a folk theorem result. To the extent that the California market is viewed as an infinitely repeated game with a discount factor between days very close to one, any level of pricing behavior between one-shot Cournot and joint monopoly levels can be sustained in equilibrium. A conduct estimate slightly above one is consistent with pricing above unilateral market power levels but not raising prices to the maximum sustainable under tacit collusion. The incentive compatibility constraint places an upper bound on how close firms can come to joint monopoly prices. If they are not *efficiently* tacit colluding, the firms can still price above Cournot levels but not at the upper bound of sustainable prices. Therefore, the conduct estimates for late 2000 are consistent with some form of dynamic pricing. Nevertheless, even if firms engaged in a form of dynamic pricing, the economic significance and subsequent welfare loss is not substantially larger than that of Cournot pricing.

6 Conclusions

A variety of states and countries have designed restructured electricity markets so that a large fraction of transactions occur in the spot market. Although spot electricity markets appear particularly conducive to tacit collusion, the California market does not exhibit evidence of efficient tacit collusion. However, there is strong evidence of market power. Price-cost margins varied substantially over time with higher margins during the higher demand third and fourth quarters of each year. I estimate the extent to which high margins resulted from less competitive conduct and/or less elastic demand which affords firms more opportunity to exercise market power. During 1998-99, I generally fail to reject Cournot pricing but find that the market was slightly more competitive in 1999 than in 1998. Although there was some variation in conduct, the large variations in margins were largely driven by changed in the residual demand faced by firms with incentives to price strategically.

An important policy question is whether the rapid increase in prices during the second half of 2000 was more related to increases in input costs, higher demand, or less competitive behavior by generators. Results suggest behavior was distinctly less competitive (Figure 5) but the shift was not as dramatic as prices would suggest. Other factors contributing to price increases were higher natural gas and emission costs and less elastic residual demand. Nevertheless, the five large in-state non-utility generators raised prices slightly above unilateral market power levels in 2000, but fell far short of efficient tacit collusion. I should emphasize that my estimates of anticompetitive behavior for 2000 are likely understated to the extent that firms forward contracted some of their output.

These findings bear on a set of issues that arise in designing deregulated electricity markets in other states and countries. Many jurisdictions are currently in the process of deregulating the generation sector of the electricity industry, and this paper confirms earlier work that market power is a concern. Policymakers must consider the magnitude and source of market power when considering market design issues such as divestiture of power plants, trading institutions, and bidding rules. Prescriptions for mitigating market power can depend upon the underlying pricing game. If market power is a unilateral/static phenomenon, then increasing the number of players in the game through further divestiture or new entry can make the market more competitive. Alternatively, if they are required to forward contract a large fraction of their output, firms will have less incentive to withhold output to drive up the price in the spot market.[43] However, if there is evidence that firms begin to engage in some form of dynamic pricing, regulators may wish to focus on the design and frequency of the auction. Some work has suggested that collusion is less likely under discriminatory auctions than uniform-price auctions.[44] Also, market designers could reduce the frequency of interaction by auctioning the right to sell electricity every week or month rather than every day. Finally, an asymmetric divestiture process that divides the industry into a large and several small firms may make tacit collusion more difficult to coordinate and sustain.

References

B. Allaz and J.L. Vila. Cournot Competition, Forward Markets and Efficiency. Journal of Economic Theory, 1993.

S. Borenstein, J.B. Bushnell, C. Knittel, and C. Wolfram. Trading Inefficiencies in California's Electricity Markets. POWER Working Paper PWP-086, University of California Energy Institute, 2001.

S. Borenstein, J.B. Bushnell, and F.A. Wolak. Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market. American Economic Review, forthcoming.

T.F. Bresnahan. Empirical Studies of Industries with Market Power. In Handbook of Industrial Organization. Elsevier, 1989.

T.F. Bresnahan. Testing and Measurement in Competition Models. In Advances in Economics and Econometrics: Theory and Applications. Cambridge University Press, 1997.

W.H. Brock and J.A. Scheinkman. Price Setting Supergames with Capacity Constraints. Review of Economic Studies, 1985.

J.B. Bushnell and C. Saravia. An Empirical Assessment of the Competitiveness of the New England Electricity Market. CSEM Working Paper 101, 2002.

J.B. Bushnell and F.A. Wolak. Regulation and the Leverage of Local Market Power: Reliability Must-Run Contracts in the California Electricity Market. POWER Working Paper PWP-070, University of California Energy Institute, 1999.

K.S. Corts. Conduct Parameters and the Measurement of Market Power. Journal of Econometrics, 1999.

Energy Information Administration. Form EIA-759 "Monthly Power Plant Database". 1998-2000.

Energy Information Administration. Form EIA-900 "Monthly Nonutility Power Plant Database". 1999.

N. Fabra. Uniform-Price Auctions Facilitate Collusion: The Case of Electricity Markets. Mimeo, European University Institute, 2000.

F. Gasmi, J.J. Laffont, and Q. Vuong. Econometric Analysis of Collusive Behavior in a Soft-Drink Market. Journal of Economics and Management Strategy, 1992.

D. Genesove and W.P. Mullin. Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914. RAND Journal of Economics, 1998.

E.J. Green and R.H. Porter. Noncooperative Collusion Under Imperfect Price Information. Econometrica, 1984.

J. Haltiwanger and J.E. Harrington. The Impact of Cyclical Demand Movements on Collusive Behavior. RAND Journal of Economics, 1991.

S.M. Harvey and W.H. Hogan. On the Exercise of Market Power Through Strategic Withholding in California. Mimeo, 2001.

T.A. Johnsen, S.K. Verma, and C.D. Wolfram. Zonal Pricing and Demand-Side Bidding in the Norwegian Electricity Market. POWER Working Paper PWP-063, University of California Energy Institute, 1999.

P. Joskow. Deregulation and Regulatory Reform in the U.S. Electric Power Sector. In S. Peltzman and C. Winston, editors, Deregulation of Network Industries: The Next Steps. Brookings Press, 2000.

P. Joskow. California's Electricity Crisis. Oxford Review of Economic Policy, 2001.

P. Joskow and E. Kahn. Identifying the Exercise of Market Power: Refining the Estimates. Mimeo, 2001.

P. Joskow and E. Kahn. A Quantitative Analysis of Pricing Behavior In California's Wholesale Electricity Market During Summer 2000: The Final Word. Mimeo, 2002.

E. Kahn, S. Bailey, and L. Pando. Simulating Electricity Restructuring in California: Interactions with the Regional Market. Resource and Energy Economics, 1997.

J. Klein. The Use of Heat Rates in Production Cost Modeling And Market Modeling. Electricity Analysis Office, California Energy Commission, 1998.

P.D. Klemperer. What Really Matters in Auction Design. Mimeo, 2000.

P.D. Klemperer and M.A. Meyer. Supply Function Equilibria in Oligopoly Under Uncertainty. Econometrica, 1989.

E.T. Mansur. Pricing Behavior in the Initial Summer of the Restructured PJM Wholesale Electricity Market. POWER Working Paper PWP-083, University of California Energy Institute, 2001.

Natural Gas Intelligence. NGI's Daily Gas Price Historical Data. 1998-2000.

A. Nevo. Measuring Market Power in the Ready-to-Eat Cereal Industry. Econometrica, 2001.

Pacific Gas & Electric Company. Various natural gas tariff schedules. 1998-2000.

J.J. Rotemberg and G. Saloner. A Supergame-Theoretic Model of Price Wars during Booms. American Economic Review, 1986.

A. Sheffrin. Empirical Evidence of Strategic Bidding in California ISO Real-time Market. Report of California Independent System Operator, March 2001.

Southern California Gas Company. Various natural gas tariff schedules. 1998-2000.

R.W. Staiger and F.A. Wolak. Collusive Pricing with Capacity Constraints in the Presence of Demand Uncertainty. RAND Journal of Economics, 1992.

A. Sweeting. Market Outcomes and Generator Behaviour in the England and Wales Wholesale Electricity Market 1995-2000. mimeo, MIT, 2001.

R. Wilson. Architecture of Power Markets. Econometrica, 2002.

F.A. Wolak. An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market. Mimeo, Stanford University, 2000.

F.A. Wolak. Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets. Mimeo, Stanford University, 2001.

F.A Wolak and R.H. Patrick. The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market. University of California Energy Institute, POWER Working Paper PWP-047, 1997.

C.D. Wolfram. Strategic Bidding in a Multiunit Auction: an Empirical Analysis of Bids to Supply Electricity in England and Wales. RAND Journal of Economics, 1998.

C.D. Wolfram. Measuring Duopoly Power in the British Electricity Spot Market. American Economic Review, 1999.

Notes

[1]Other papers analyzing bidding behavior include Wolak [2000, 2001] and Sweeting [2001].

[2]For a detailed discussion of the history and goals of restructuring in the electricity industry, see Joskow [2000].

[3]For much of my sample period, utilities were not allowed to hedge by contracting to buy power more than a day-ahead. However, beginning in June 1999, utilities were allowed to purchase up to 10% of their power with forward contracts.

[4]Although the Power Exchange had a higher price cap than the ISO, we expect the cap in the last market to set the effective cap because demand should never bid higher into the PX than the cap in the ISO Borenstein et al. [forthcoming].

[5]Retail electricity rates were frozen for the vast majority of customers during the period I analyze. For those customers not under the rate freeze, the prices are not the hourly wholesale price but rather the average price over some extended period of time.

[6]See Joskow [2001] for a complete history of the California restructuring experiment from 1994-2001. For a discussion of market design in restructured markets, see Wilson [2002].

[7]See the survey articles Bresnahan [1997, 1989].

[8]Comparisons of direct measures of the conduct parameter versus the NEIO estimates have found NEIO methods to understate market power (see Genesove and Mullin [1998] and Wolfram [1999]).

[9]A model incorporating capacity constraints in which firms choose supply functions (of which Cournot competition is a special case) closely resembles how firms bid into the market. In general, supply function models generate multiple equilibria. Some researchers use restrictive forms of the Klemperer and Meyer [1989] supply function model to simulate outcomes in electricity markets. Because econometrically identifying supply function equilibria is not tractable with my data, I estimate models of games in which firms choose quantities. Future work with actual bid data could estimate supply function models.

[10]I assume that the firms are taking industry structure as given and not choosing output to strategically influence entry into the market. Limit pricing seems unlikely in this market because information on individual firm costs is publicly available.

[11]This assumes $c(\cdot)$ is constant for small changes in q so that there is no $c'q_i$ term. I believe this is reasonable given the assumptions about unit level marginal costs I describe in section 4.

[12]This can be generalized to other punishment strategies (such as finite-period Nash reversion) without affecting my estimation results below. My estimation strategy requires only that the level of the incentive compatibility constraint be equal across firms in a given time period.

[13]Although firms in the California market do not have identical cost structures, Table 1 shows symmetry among the five largest firms is a somewhat reasonable characterization.

[14]I do not include a capacity constraint because I assume that the capacity constraint of the

group of collusive firms is never hit. In my data, there is no period in which all firms produce at capacity. However, capacity constraints do affect the best-response profits of all the firms.

[15]Equation (5) has potentially important implications for market power studies of industries in which imperfect collusion is possible. Even if marginal cost is observed, estimating the static first-order condition can yield biased estimates of the conduct parameter (equation (2)) if the exercise of market power has a dynamic component. From equation (5) we can see that market power studies that estimate the static first-order condition (which *excludes* the IC constraint term) will obtain biased estimates of conduct if the IC constraint is ever binding ($\mu^* > 0$) and the best-response profits are non-linear in q ($\frac{d\pi^{br}}{dQ}_t$ is correlated with q_{it}). This is another interpretation of Corts [1999] and illustrates how one can mischaracterize market power by estimating conduct parameters for industries with any form of dynamic interaction short of perfect collusion ($\mu^* = 0$).

[16]Studies in the empirical literature have addressed whether markups change over the business cycle. In collusion models such as Rotemberg/Saloner, firms never change their conduct over the business cycle – they are always colluding. Rather, firms change their pricing to keep collusion sustainable. My dynamic first-order condition would capture such behavior by estimating a θ that is constant over time with "incentive compatibility" adjustments reflected in the IC term.

[17]Marginal costs also include the opportunity costs of exporting power to other higher price markets. The potential to export power out-of-state is unlikely to cause me to mismeasure the marginal (opportunity) cost. In-state firms will sell out-of-state if the out-of-state price is greater than the marginal revenue of sales into California. I cannot measure out-of-state prices, however California is virtually never a net exporter during my sample. Finally, my measure of marginal cost is complicated by the cost of starting up a unit. A unit that is not operating will incur a start up cost that is typically approximated by three hours of fuel burn. To avoid the endogeneity of shut down decisions and costs, I restrict my analysis to plants that are already operating.

[18]However, given that they turn on the Reliability Must Run (RMR) units to meet RMR contracts, competitive firms should still increase production in these units if marginal cost is lower than price. In practice, the RMR units are not always higher cost units and when they are, the costs are at most a few dollars higher than other units.

[19]The percentage of each firm's capacity for which EPA has data are: AES 100%, Reliant 99%, Duke 95%, Southern Energy 87%, and Dynegy 68%. These percentages are lower bounds for the completeness of the data because some of the missing units were shut down during significant portions of my sample.

[20] See Borenstein et al. [2001] for an analysis of the PX-ISO arbitrage condition in this market over time.

[21]Of course, firms have the incentive to ramp up production only if price is above marginal cost during the ramping hours. On average, prices peak between 2pm and 7pm.

[22]Price-cost margins are higher on average during the high demand hours 18 than during other hours. However, higher margins do not imply less competitive behavior. Even if conduct were the same during lower demand hours, one expects to see lower margins because the residual demand for the five firms is more elastic.

[23]Industry analysts suggest that the reported nameplate capacity overstates the true capacity of

a unit. To test for robustness, I estimate the static models defining capacity as 80%,90%, and 95% of nameplate and get very similar results.

[24]Recall that the firm "DukeSouth" represents the generating units owned by Duke in the southern part of the state when transmission capacity constraints are binding. Transmission constraints tend to bind when demand (and perhaps the potential to exercise market power) are high.

[25]The margins are not interpreted as measures of profitability because firms incur other on-going costs such as the cost of starting up a generator. Rather, these positive margins are measures of non-price-taking behavior because the units I analyze have already incurred the startup costs yet fail to utilize capacity when price is above marginal cost.

[26]Industry analysts believe the market observed negative margins in the second quarter of 1998 because many firms were not selling their power into the (unprofitable) energy market but rather were selling power under alternative profitable RMR regulatory side agreements (Bushnell and Wolak [1999]). This became less of an issue over time as the original RMR contracts were amended.

[27]The fringe thermal generators include a few units still owned by the utilities PG&E, SCE, and SDG&E as well as two small powerplants owned by Thermo Ecotek. Small independent producers include, for example, oil refineries that self-provide electricity and are qualified to sell surplus generation to the grid.

[28]Although the nuclear generation is partially owned by the utilities owning other generation assets, nuclear units operate under very strict regulations that preclude operators from adjusting output to influence the price earned by the utilities' thermal generation units.

[29]See Johnsen et al. [1999] for a paper that uses a difference in differences approach to measure market power in a hydro system.

[30]A constant elasticity supply function by the fringe can capture the shape many industry analysts envision, and it also fits the data well. I estimate fringe supply to be relatively flat at prices below \$100 but progressively steeper for higher prices. Some have suggested that the estimated supply relationship should be vertical at high levels of demand (e.g. when transmission constraints are binding). However, it is important to keep in mind that my fringe includes not only imports but also hydroelectric and expensive gas peaking units in California.

[31]Daily temperature data come from the National Climatic Data Center website.

[32]This figure is constructed using slopes of residual inverse demand $\hat{P}_t'$ estimated below.

[33]I exclude the first quarter of the market's operation 1998Q2 because much of the ownership transfer had yet to take place. Also, industry analysts believe firms were selling substantial amounts of power under alternative regulatory (RMR) agreements rather than into the energy market.

[34]To confirm that this result is not picking up seasonal differences in supply relations, I compare June-November 2000 to the same months in 1998 and 1999 and find very similar results.

[35]Duke has its units divided into two markets during periods of transmission congestion (approximately 9% of hours in 1998, 12% in 1999, and 44% in 2000). The capacity in the South is separated into a firm named DukeSouth only during congested hours. Therefore, I exclude DukeSouth to make the system estimable. As a result, I only partially characterize Duke's behavior

during congested hours.

[36]Recall that prices began to hit the price cap in summer 2000. During hours of 2000 when the price cap is binding, the first-order condition underlying the supply relation does not hold with equality because the cap creates a discontinuity in marginal revenue. This affects 7.8% of hour 18 observations in 2000 with the majority occurring in August. I estimate the conduct parameter by ignoring days when the price hit the cap. Under static pricing, the presence of a price cap should not affect production behavior when the cap is not binding. This may not be the case under dynamic pricing.

[37]Note that the supply relation I estimate has no constant because theory suggests the relationship is a ray through the marginal cost intercept. If I include a constant, I find small (yet statistically non-zero) intercept terms of -\$0.25 in the 4-firm period and -\$5.57 in the 5-firm period. The corresponding slopes of the supply relation are 1.00 and 1.26, respectively.

[38]The strategic firms supply the largest fraction of total hour 18 demand, roughly 15-25%, from June-November.

[39]These results speak to the issue of how overall market conditions in 2000 differed from previous years. Some fundamentals were left unchanged. As noted in Borenstein et al. [forthcoming], Lerner indices were similar after controlling for equilibrium residual demand. However, the combination of the residual demand function being less elastic and conduct being slightly less competitive contributes to higher price-cost margins.

[40]Given the unusually high conduct estimates for Dynegy, one may be concerned that conduct estimates restricted to be equal across firms reported above are substantially driven by Dynegy. I re-estimate the static models above allowing Dynegy to have a different conduct parameter, and find that neither the estimates nor the inferences substantially change.

[41]Although the conduct parameter estimates differ statistically, the behavior is economically similar across time periods.

[42]One complication is that in a small fraction of hours one but not all firms hit their capacity constraint, and that is not consistent with symmetric collusion. To deal with this complication, I include a λ_{it} for those firm-hours.

[43]In fact, other markets that have required forward contracting or vesting contracts do not exhibit evidence of substantial market power except at large levels of demand (Bushnell and Saravia [2002]).

[44]See Klemperer [2000] and Fabra [2000].

Table 1: Post-Divestiture Thermal Market Structure

(54% of Total CA Capacity)

Firm	Capacity (MW)	Pct Capacity
AES	3921	22%
Reliant	3698	21%
Duke	3343	19%
Southern	3130	18%
Dynegy	2871	16%
PG&E*	570	3%
Thermo Ecotek	274	2%

* PG&E reached an agreement by which it would retain ownership of two old plants until they could be retired.

Table 2: Hour 18 Price-Cost Margins When Firms **Not** at Capacity

Firm	% hours NOT at capacity	Price-Cost Margin ($/MWh) Mean	Median	St Dev	Min	Max	Median Lerner
Capacity ≡ 90% Nameplate							
DukeSouth	88	61.43	13.97	100.98	-29.67	443.19	.23
Southern	98	37.71	11.55	81.97	-22.60	1045.94	.26
Reliant	94	31.70	7.31	76.71	-26.05	686.36	.21
Dynegy	100	25.20	2.60	73.61	-32.43	688.68	.08
AES	99	22.42	2.96	78.51	-524.76	684.50	.09
Duke	87	19.75	3.69	45.67	-20.80	475.79	.11
Capacity ≡ 80% Nameplate							
DukeSouth	78	58.50	13.62	99.38	-29.67	443.19	.23
Southern	92	36.34	11.30	82.11	-22.60	1047.63	.25
Reliant	93	32.43	7.82	77.50	-26.05	686.36	.22
Dynegy	99	25.54	2.60	74.06	-32.43	688.68	.08
AES	93	20.19	2.52	78.54	-524.76	684.50	.08
Duke	79	16.56	3.06	38.91	-20.80	391.83	.10

This table contains summary statistics of hours when firms are not operating at capacity and can increase output. The price-cost margin is the difference between price and the marginal cost of the highest marginal cost unit which is operating and has excess capacity. The manufacturer (or nameplate) rated capacity of a generator may overstate the actual capacity if the unit degrades over time. To account for possible nameplate capacity degrading, I define capacity as both 80% and 90% of nameplate capacity.

Notes:

(1) The large negative margin for AES represents a day in which a unit was operating but in the process of starting up so that the emission costs were high.

(2) The Lerner index $\equiv \frac{price-MC}{price}$ is presented as a general measure of market power. I use the median rather than the mean because the Lerner index does not treat negative and positive margins as symmetric. For example, if price is $10 and marginal cost is $1, the Lerner index is $\frac{10-1}{10} = 0.9$. However, if price is $1 and marginal cost is $10, the Lerner index is $\frac{1-10}{1} = -9$. Therefore, the mean of the Lerner index may not be a meaningful measure of average competitiveness in the presence of negative margins.

(3) The firm "DukeSouth" represents the generating units owned by Duke in the southern part of the state when transmission capacity constraints are binding. Transmissions constraints tend to bind when demand (and perhaps the potential to exercise market power) are high.

Table 3: Static Model: Estimates of Fringe Supply and Strategic Supply Relations for Hour 18

	4 Firm Market*		**5 Firm Market****	
	Fringe	Strategic	Fringe	Strategic
Dependent Variable:	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$
$\frac{P \cdot q}{Q^s_{fringe}}$	–	6.801	–	5.851
	–	(0.035)	–	(0.021)
λ ($/MW)	–	25.251	–	57.508
	–	(0.149)	–	(0.619)
Log(Price)	0.147	–	0.193	–
	(0.004)	–	(0.004)	–
Log(GasPrSouth)	-0.250	–	-0.156	–
	(0.028)	–	(0.042)	–
Log(GasPrNorth)	-0.095	–	-0.060	–
	(0.012)	–	(0.047)	–
Log(Diff65TempNeigh)	0.016	–	-0.022	–
	(0.005)	–	(0.004)	–
Constant	9.963	–	9.628	–
	(0.018)	–	(0.015)	–
R^2	0.72		0.60	
Obs.	268		573	
$\hat{\eta}^D_{strat}$	-1.97		-0.98	
Average Margin ($/MWh)	10.88		26.96	
$\hat{\theta}$	1.00		1.13	
	(0.02)		(0.03)	

Fringe represents equation (7) and Strategic represents equation (8). Although the system contains a supply relation for each firm, the coefficients are restricted to be equal in this model so I only report one set of parameters for the strategic supply relations. Standard errors are in parentheses. The standard errors from the GMM estimates account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). Day and month-year dummies are included in the fringe supply equation but are not reported here. I exclude hours (in 2000) when the price cap is hit. 8% of hour 18 observations in 2000 hit the price cap with the majority occuring in August.

* 7/1/98-4/15/99.

** 4/16/99-11/30/00.

Table 4: Static Conduct Parameters by Season and Firm for Hour 18

By Time Period: Conduct Parameter, Avg Margin, and Resid Demand Elasticity

Time Period	Estimate	Std Error	Margin	$\hat{\eta}^D_{strat}$
4-firms (Jul 98-Apr 99)	1.00	0.02	10.88	-1.97
5-firms (Apr 99-Nov 00)	1.13	0.03	26.96	-0.98
1998 Peak (Jul-Nov)	1.03	0.22	16.43	-1.36
Offpeak4* (Dec-Apr)	2.36	0.11	4.99	-5.07
Offpeak5* (Apr-May)	0.27	0.11	0.78	-2.33
1999 Peak (Jun-Nov)	0.87	0.01	13.99	-0.72
Offpeak (Dec-May)	0.66	0.11	4.28	-0.76
2000 Peak (Jun-Nov)	1.38	0.15	74.06	-0.57

By Firm and Time Period: Conduct Parameter

	4 Firm Market**		5 Firm Market***	
Firm	Estimate	Std Error	Estimate	Std Error
Southern	–	–	1.38	0.03
Reliant	1.02	0.03	1.12	0.03
Duke	1.15	0.04	1.16	0.03
AES	0.80	0.03	0.94	0.02
Dynegy	3.88	0.15	2.15	0.05

	June-November 2000	
Firm	Estimate	Std Error
Southern	1.60	0.24
Reliant	1.31	0.16
Duke	1.32	0.12
AES	1.15	0.06
Dynegy	2.60	0.27

The standard errors from the GMM estimates account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). I exclude hours (in 2000) when the price cap is hit.

*Due to a new strategic firm (Southern) entering April 16, 1999, I estimate conduct separately for the 4-firm and 5-firm subperiods.

** 7/1/98-4/15/99

***4/16/99-11/30/00

Table 5: Dynamic Conduct Parameter Estimates

IC_t Specification	4 Firm Market*	5 Firm Market**	Jun-Nov 2000
Quartile+Month Effects	0.05	1.04	1.73
	(0.02)	(0.02)	(0.18)
Only Quartile Effects	0.10	1.12	1.53
	(0.01)	(0.02)	(0.07)
Only Month Effects	0.97	1.25	1.73
	(0.02)	(0.03)	(0.03)

The standard errors from the GMM estimates account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). I exclude hours (in 2000) when the price cap is hit.

* 7/1/98-4/15/99.

** 4/16/99-11/30/00.

APPENDIX A: Data

The marginal fuel cost for each genearting unit is calculated from daily natural gas spot prices and average heat rates. All of the units for which I have generation data burn natural gas as their primary fuel. I use the daily spot price of natural gas (Natural Gas Intelligence [1998-2000]) for the PG&E Citygate and California-Arizona border hubs plus the distribution cost charged to those units by the natural gas utility (Southern California Gas Company [1998-2000] and Pacific Gas & Electric Company [1998-2000]. Although some firms may have contracted for natural gas at a different price, the spot price is the proper measure of the opportunity cost of fuel. Average heat rates are from datasets collected by the California Energy Commission and Southern California Gas Company. These heat rates also have been used in Borenstein et al. [forthcoming] and Kahn et al. [1997].

Several generators in the South Coast Air Quality Management District were required to purchase permits for emissions of NOx. The hourly marginal permit cost is calculated as the monthly quantity-weighted average price of permits multiplied by the unit's hourly emissions. I use the weighted average of trade prices rather than the highest trade price because large outliers in trade prices make it difficult to believe that the highest price is a good measure of the marginal cost of a permit. Permit costs were negligible until mid-2000 because total emissions were less than the number of allocated permits. The cost of a permit rose above $1/lb (approximately $1-2/MWh) in January 2000, so I include permit costs beginning in 2000. In addition several plants faced annual emission limits that were binding for six units in 2000 (Harvey and Hogan [2001]). However, this will not alter my results because I observe capacity withholding by other unaffected units owned by the same firms in each hour of my sample.

Data on hourly production of each unit are from EPA's Continuous Emissions Monitoring System (CEMS). The CEMS output data available are the gross output which includes electricity generated for sale as well as electricity used at the plant for station operations. I use independent data sources (Energy Information Administration [1998-2000], Energy Information Administration [1999]) containing data on net generation to calculate plant-level scale factors that convert gross generation to net generation sold to the grid.

Data on each unit's capacity are also from the CEMS data. The EPA data contain measures of the manufacturer rated (nameplate) capacity of each unit. Analysts familiar with the industry claim that firms typically do not view their capacity to be as large as the EPA nameplate capacity. Therefore, I somewhat arbitrarily define capacity to be 90% of the EPA capacity. One potential problem with this definition is that I cannot observe the very occasional partial outages that temporarily reduce the operating capacity of a unit. If a firm suffers a partial outage and produces up to its temporary capacity, I consider that firm to have excess capacity. I assume that each unit's marginal cost is constant up to the capacity of the generator. Klein [1998] analyzes heat rates (inverse of fuel efficiency) and estimates marginal cost functions for many of the units in California. For the vast majority of units, the marginal cost is nearly constant from one-quarter to full capacity. Therefore, my assumption of constant marginal cost up to capacity appears very reasonable for units that are producing more than minimal levels of output.

I measure market power by observing whether firms withheld capacity of a unit with marginal cost less than the price. In theory, if a unit is not operating some capacity, the firm placed a bid for that capacity higher than the market clearing price. This may not hold precisely due to several operating procedures of the grid operator. Occasionally firms are instructed by the ISO to reduce output to avoid intra-zonal transmission congestion. To the extent that firms bid to supply full capacity but were instructed to cut output, I will overstate market power. Also, the ISO has the discretion to skip over lower priced units that are more flexible in favor of higher priced units in case increases in power are needed on short notice.

Data on prices in the Power Exchange and total demand forecasts are from the PX and ISO websites, respectively. Those data also can be downloaded from http://www.ucei.org/datamine/datamine.htm. I use the PX day-ahead zonal price as my benchmark price because the vast majority of transactions occurred in the PX. The ISO log of real-time transactions shows that typically less than 10% of the power sold by the five large firms was traded in the real-time market. A notable exception was the period beginning in September 2000 when the firms began to shift between one-quarter and one-half of their sales to the real-time market. During this later period of my sample, real-time ISO prices were on average higher than the PX price. To the extent that firms earned the ISO price, I will tend to understate margins late in my sample.

There is a slight complication posed by focusing on prices in the PX and ISO energy markets. Generators not only compete in the market to supply electrical energy, but they also compete in "ancillary services" markets to provide stability and reliability services to the system operator. I do not explicitly model the ancillary services market, however the opportunity cost of selling into this alternative market will affect firm behavior in the energy market. The presence of an ancillary services market only slightly complicates my analysis. For most of the ancillary services market, firms bid a "standby" payment and a "production" payment. All bids for the production payments are placed into the real-time market's bid stack. Therefore, an exercise of market power in these ancillary services markets will manifest itself as market power in the real-time market. For one form of ancillary services (regulation reserve), units essentially turn over control of some fraction of their unit to the Independent System Operator. Because the ISO seeks to always have some units with excess capacity standing by, these units are essentially being paid not to produce. If some of the units that I measure to be withholding capacity are actually selling this capacity to the ISO as regulation reserve, I may overstate the firm's price-cost margin. I do not have data on each unit's sales to regulation reserve, however anecdotal evidence suggests that most regulation reserve is sold by hydroelectric units rather than the fossil-fueled units I am analyzing. Although it is unknown how much regulation reserve is satisfied with thermal generating units, the Joskow and Kahn [2001] analysis of summer 2000 assumes that an additional 3% of thermal demand is purchased as reserves. This mismeasurement will be mitigated by the fact that the quantity of regulation reserve bought during the hour of the day I analyze below (hour 18) is typically lower than other hours of the day.

APPENDIX B: Conduct Parameter Estimates Under Alternative Assumptions

4-Firm Market

Firm	(1)	(2)	(3)	(4)	(5)
All	0.94	1.03	–	0.99	–
	(0.05)	(0.02)	–	(0.02)	–
Southern	–	–	–	–	–
	–	–	–	–	–
Reliant	–	–	1.05	–	1.01
	–	–	(0.03)	–	(0.03)
Duke	–	–	1.21	–	1.13
	–	–	(0.04)	–	(0.04)
AES	–	–	0.84	–	0.79
	–	–	(0.03)	–	(0.03)
Dynegy	–	–	4.00	–	3.85
	–	–	(0.14)	–	(0.15)

5-Firm Market

Firm	(1)	(2)	(3)	(4)	(5)
All	0.90	1.16	–	1.13	–
	(0.05)	(0.02)	–	(0.03)	–
Southern	–	–	1.43	–	1.38
	–	–	(0.04)	–	(0.04)
Reliant	–	–	1.14	–	1.11
	–	–	(0.03)	–	(0.03)
Duke	–	–	1.17	–	1.14
	–	–	(0.03)	–	(0.03)
AES	–	–	1.04	–	0.92
	–	–	(0.03)	–	(0.02)
Dynegy	–	–	2.21	–	2.14
	–	–	(0.05)	–	(0.06)

June-November 2000

Firm	(1)	(2)	(3)	(4)	(5)
All	1.45	1.47	–	1.41	–
	(0.06)	(16.89)	–	(0.13)	–
Southern	–	–	1.70	–	1.63
	–	–	(0.72)	–	(0.46)
Reliant	–	–	1.41	–	1.32
	–	–	(0.68)	–	(0.39)
Duke	–	–	1.36	–	1.37
	–	–	(0.78)	–	(0.29)
AES	–	–	1.24	–	1.14
	–	–	(0.50)	–	(0.30)
Dynegy	–	–	2.79	–	2.63
	–	–	(1.12)	–	(0.75)

Definitions of Alternative Specifications:

(1) Using only observations when no firms operate at capacity ($\lambda^*_{it} = 0$), capacity==90% nameplate
(2) All observations, capacity==80% nameplate, conduct restricted equal across firms
(3) All observations, capacity==80% nameplate, conduct unrestricted across firms
(4) All observations, capacity==95% nameplate, conduct restricted equal across firms
(5) All observations, capacity==95% nameplate, conduct unrestricted across firms

The standard errors from the GMM estimates account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). I exclude hours (in 2000) when the price cap is hit.

Figure 1: Wholesale Prices in California Electricity Market

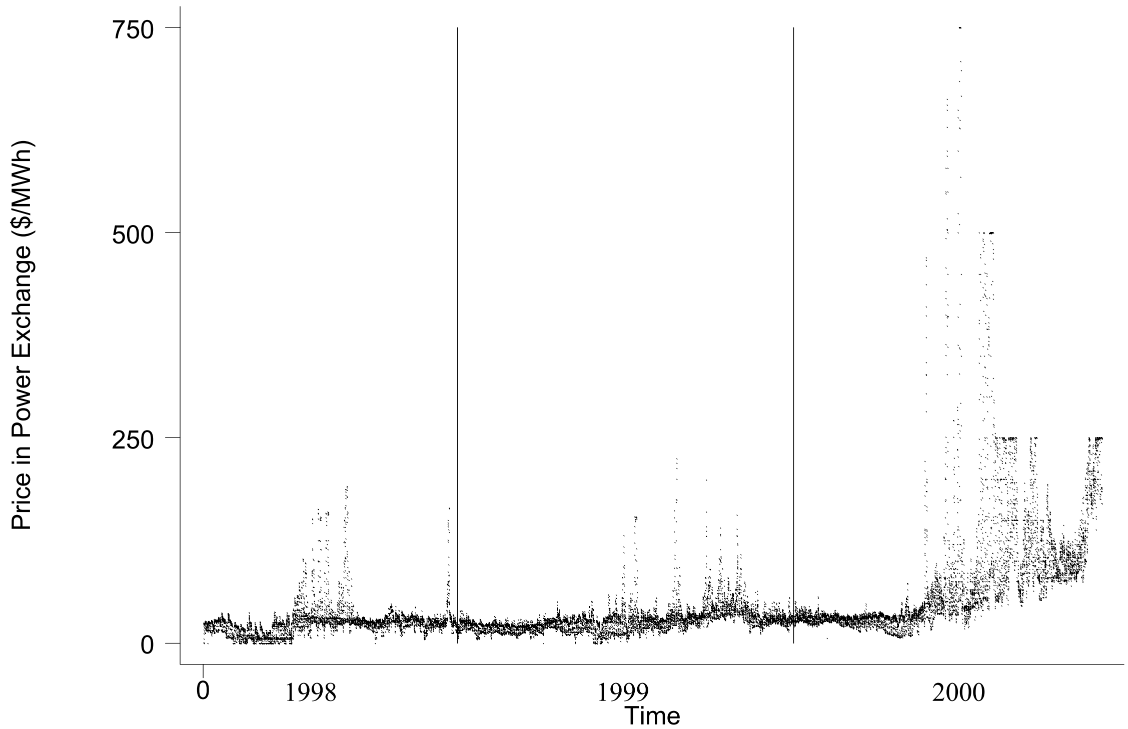
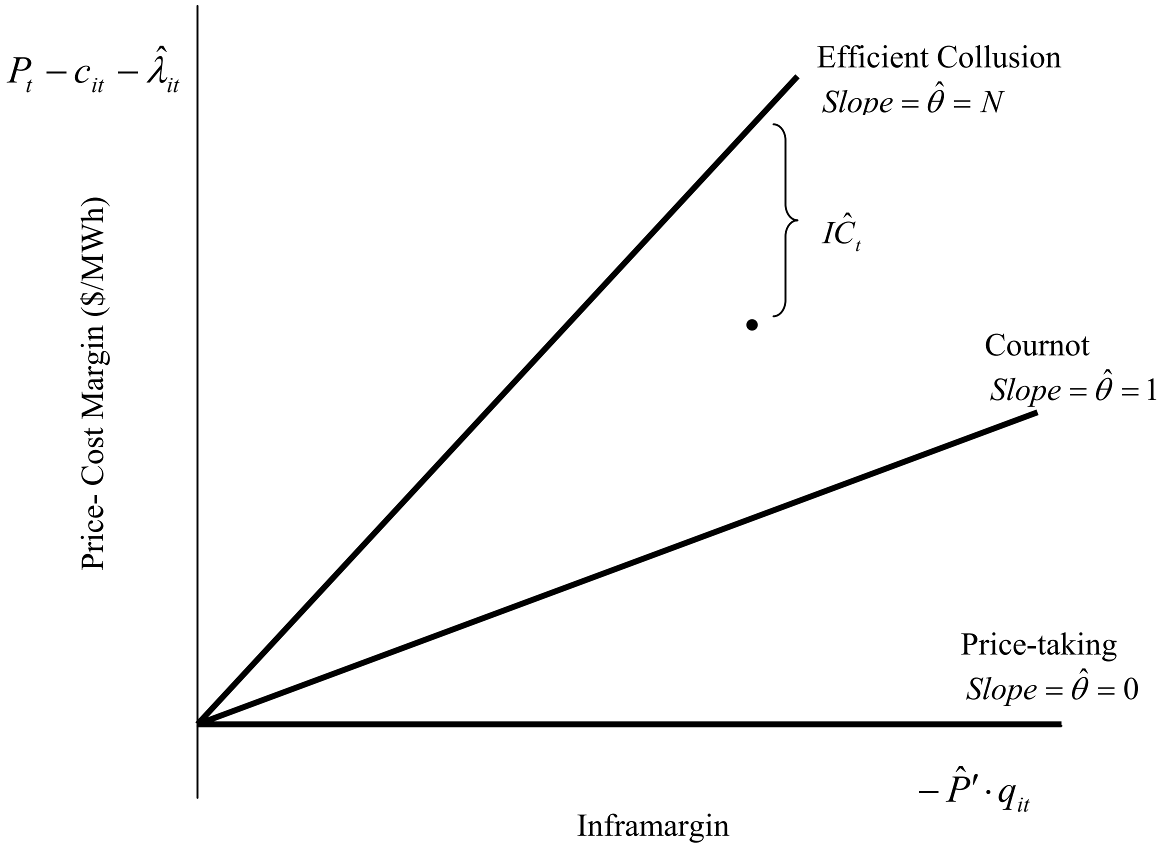


Figure 2: Supply Relations Under No, Static, and Dynamic Market Power



$IC_t = \frac{\mu_t^*}{1+\frac{\mu_t^*}{N}} \cdot \frac{d\pi^{br}}{dQ}_t$ is the adjustment from perfect collusion (the joint monopoly outcome) to respect the incentive compatibility constraint. Variables with hats are estimated and all other variables are measured.

Figure 3: Average Price-Cost Margins in Hour 18

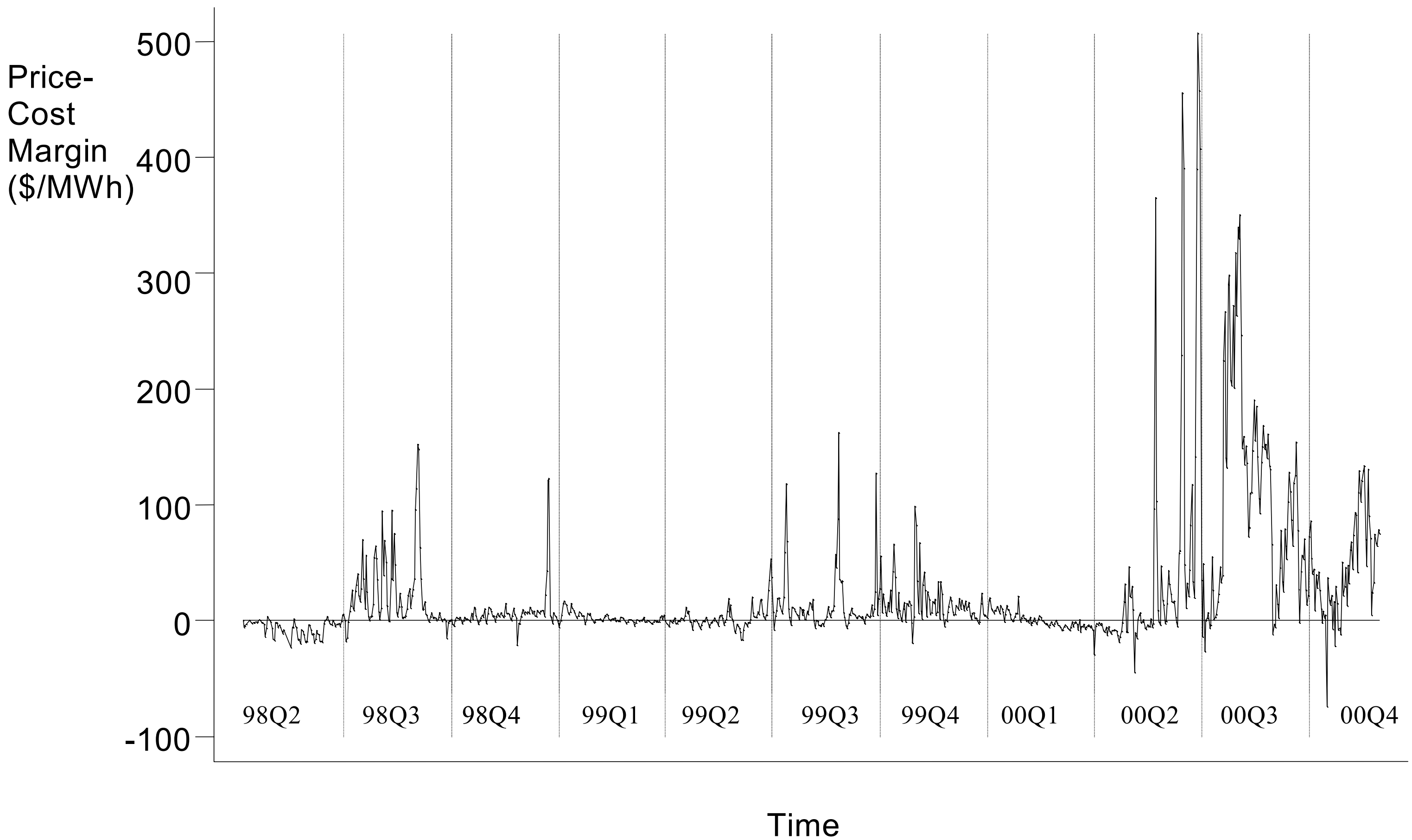


Figure 4: Demand for Electricity from Strategic Firms

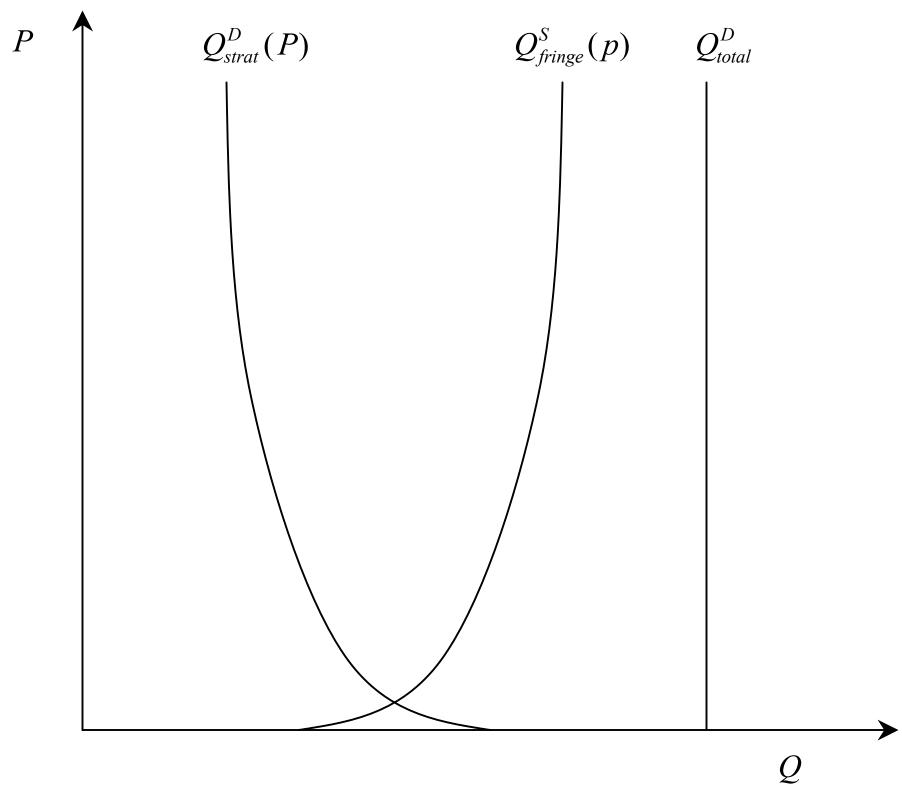
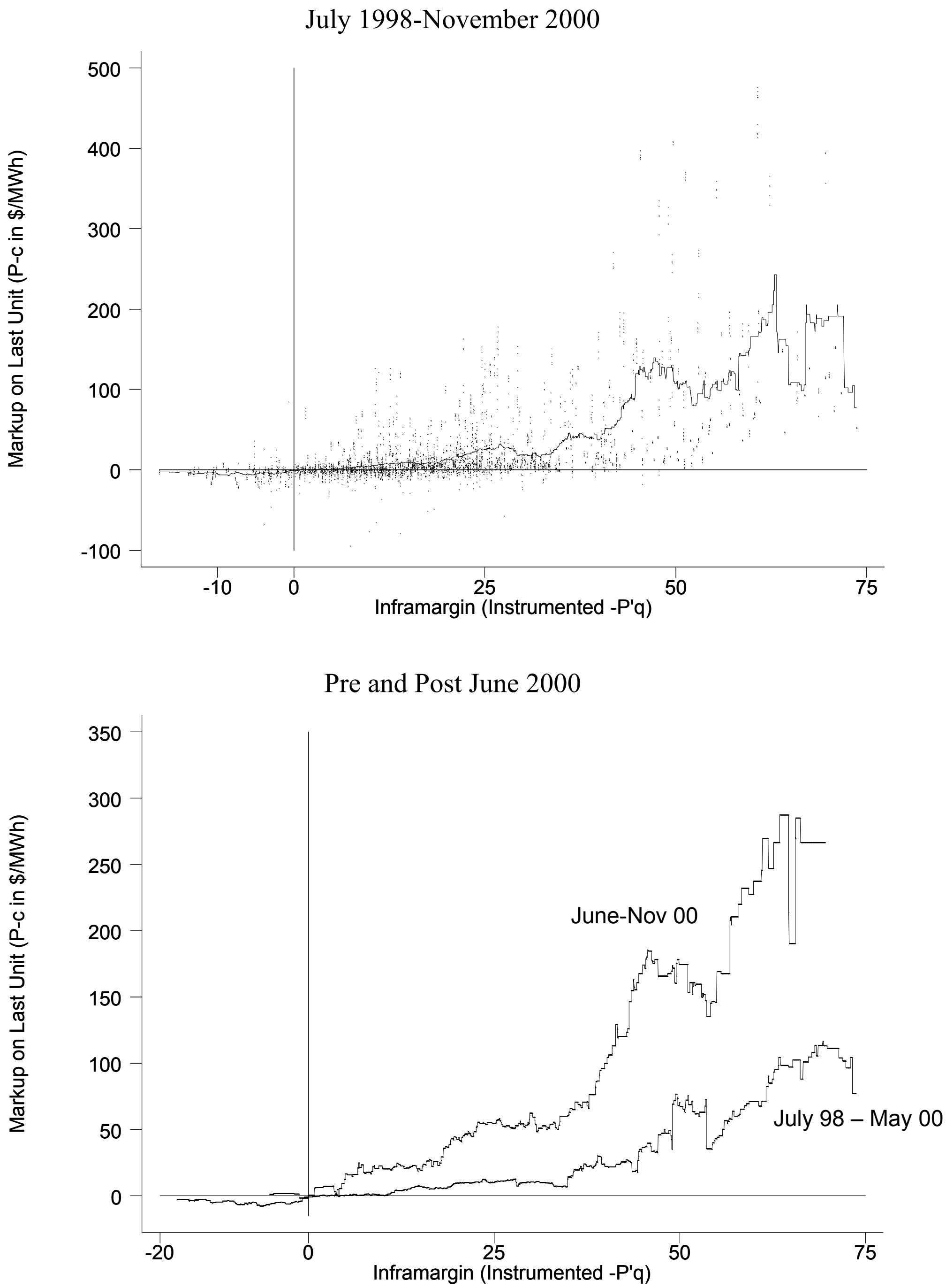


Figure 5: Static Behavioral Model for Hour 18



Kernel regression of price-cost margins on instrumented $-P'q$ for firm-hours when the capacity constraint is not binding (lambda=0). The slope of the relationship is an estimate of the conduct parameter under static pricing. In the top panel, 14 outlier observations with very large or small margins are excluded from the figure to maintain scale.