POWER

PWP-080

Pricing and Firm Conduct in California's Deregulated Electricity Market

Steven L. Puller

Original Draft: November 2000

This Version: November 2001

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*

November 2001

Abstract

Deregulating wholesale electricity markets has the potential to increase the long-run efficiency of the industry but poses the risk that firms may exercise market power. This paper 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. The wholesale market was organized by a uniform price multiunit auction that was repeated daily. Oligopoly theory provides several static and dynamic pricing models that may capture behavior in this market. I use detailed firm-level data to test for both the unilateral exercise of market power and tacit collusion. Direct measures of price-cost margins suggest that, to differing extents, all large merchant generating firms exercised market power from 1998-2000. Over the three years, conduct varied moderately with a general strengthening of competition during summer 1999 and a weakening of competition during most of 2000. In 1998-99, I estimate conduct to be consistent with a Cournot *static* pricing game. Results are less clear for 2000. Although behavior was distinctly less competitive, the dramatic rise in prices was more driven by changes in costs and demand than by changes in firm conduct.

*Department of Economics, Texas A&M University. Email: puller@econ.tamu.edu. I thank Severin Borenstein, James Bushnell, Carlos Dobkin, Richard Gilbert, James Griffin, Bronwyn Hall, Edward Kahn, Erin Mansur, Aviv Nevo, Steve Wiggins, Catherine Wolfram, 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

Electricity markets are being restructured in many jurisdictions in the United States and across the world. Traditional rate-of-return regulation is being replaced with private investment decisions and the competitive procurement of electric power. Restructuring (often referred to as deregulation) offers the promise of more efficient decisions on the quantity and technology of investment in new electric generation capacity. However, deregulated markets also risk that firms bidding to supply power will have the ability to exercise a substantial degree of market power. California restructured its market by divesting a substantial portion of its generating assets from regulated utilities to private firms. These firms bid into a daily auction to supply power. The market operated from April 1998 until it collapsed in late 2000 under the pressure of skyrocketing wholesale prices.

The market has characteristics that oligopoly theory suggests would favor either unilateral market power or collusion. 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 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 has characteristics that may facilitate dynamic collusion. The bidding game is repeated daily between a fixed set of players that have substantial information about each other's behavior. The relative transparency of the market and the short-run barriers to entry contribute to a competitive environment that is potentially susceptible to collusion.

This paper uses detailed firm-level data on demand, costs, and output to analyze the pricing behavior of electricity generating firms in California from 1998-2000. I analyze whether firms exercised market power and whether that behavior was more consistent with unilateral market power or tacit collusion. In addition, I measure the change in behavior over time to estimate whether the skyrocketing prices in 2000 resulted from increases in factor input costs, higher demand, or less competitive behavior.

Previous work has identified non-price-taking behavior of firms in deregulated electricity markets. Direct measures of price-cost margins in the UK and the California markets find prices higher than those associated with Bertrand pricing but lower than levels associated with joint profit maximization. Wolfram [1999] measures price-cost margins in the England and Wales market and finds prices to be significantly higher than marginal costs but not as high as would be predicted by standard oligopoly models. Borenstein et al. [2000] simulate a perfectly competitive market from 1998-2000 in California and find actual prices to be higher than the perfectly competitive prices. Joskow and Kahn [2001a,b] 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. This paper focuses on identifying the pricing behavior that underlies the observed evidence of market power.[1]

I empirically test whether observed behavior in the California electricity market is more consistent with static/unilateral market power or dynamic/multilateral market power. I use detailed firm-level hourly production and cost data from the restructured electricity market in California to distinguish between alternative explanations for positive price-cost margins. I find strong evidence that firms do not produce competitively, but rather exercise market power throughout 1998-2000. I find evidence of static market power for much of 1998-99, and mixed evidence for 2000. In 2000, firms were distinctly less competitive but 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. Section 5 measures price-cost margins and finds evidence consistent with large generation owners withholding output to raise the market price. In Section 6, I estimate the behavioral model using a panel of firm-level data from April 1998-November 2000. Parameter estimates imply that the pricing conduct of electricity generators is approximately Cournot for 1998-99 and less competitive than Cournot in 2000. I conclude in Section 7 by discussing general implications for

[1] Other empirical papers that address competition in deregulated electricity markets include Wolfram [1998], Wolak [2000], and Wolak [2001].

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 components: generators that produce electricity, a high voltage transmission system that transports the electricity from the generator to the area of consumption, and a distribution network that delivers electricity to end-users. Generated energy is a homogenous product in the sense that electricity injected into the grid by one firm is identical to energy injected by another firm.[2]

The electricity industry has undergone many structural changes over the past decade. Prior to recent developments, the industry had been considered a natural monopoly that required government regulation of price, entry, and investment decisions. Vertically integrated firms were obligated to provide energy services at regulated prices and were allowed to earn specified rates of return on investments deemed prudent by the regulator. Recently, policymakers in some jurisdictions have broken off the generation side of the industry from the transmission and distribution sectors, and allowed deregulated firms to compete to supply electrical energy to the network. Competitive generation of various forms has been introduced in many regions of the world including England and Wales, Norway, New Zealand, Australia, Alberta, California, Texas, and the Northeast United States.

This restructuring has the potential to increase the long-run efficiency of the electricity industry but also poses several risks. Privately owned generation provides strong incentives for more efficient operation of capital assets as well as improvements in labor productivity. In addition, market prices may offer better signals for the efficient quantity and technology of new investment than the policies of regulators. However, vertical separation of the generation sector may forego complementarities existing between the generation and transmission sectors that a centralized system could otherwise

[2]The only form of differentiation is spatial when the transmission system is congested or transport over long distances leads to losses through the dissipation of heat.

exploit. Finally, deregulated firms owning the generation assets may possess significant horizontal market power that can lead to inefficient production and high prices. This paper focuses on the final issue and analyzes the horizontal market power in one of these restructured electricity markets.

Prior to April 1, 1998 the electricity generation 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 vertically integrated utilities were responsible for generating electricity and supplying customers in their service territories. Each was regulated by traditional rate-of-return regulation.

In California, the industry was restructured in 1998 with the intent of separating the generation component of the industry from the transmission and distribution segments. The three original utilities divested many of their powerplants to private firms that bid daily for the right to supply electricity to the market. California established several institutions to organize the trading and dispatch of electricity generation. Electricity was traded through a centralized pool for markets that cleared one-day ahead of delivery and in real-time. These institutions managed the electricity market from the beginning of deregulation in April 1998 until the market collapsed in late 2000. During the second half of 2000, large increases in input prices and possibly increased market power caused wholesale electricity prices to skyrocket. The utilities were required to purchase power at high wholesale prices but to sell to end-users at substantially lower prices. Eventually, the utilities lost their creditworthiness and the state government was required to step in and purchase power. This paper analyzes the competitiveness of the California market from its inception until the market began to collapse in late 2000.

From 1998-2000, firms owning power plants bid to supply electricity into either a day-ahead or a real-time market. The three original utilities that were still responsible for serving their customers were required to purchase their electricity from a specific day-ahead trading exchange (the Power Exchange).[3] As a result, the Power Exchange's day-ahead market had the largest volume of electricity trades in the market during my sample period. I use the price in this market

[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 did begin to purchase up to 10% of their power with forward contracts.

as my benchmark price.

The day before electricity was scheduled to be produced, firms that produced electricity and utility distribution companies that "consumed" electricity bid hourly supply and demand into the Power Exchange (PX). The bid schedules were strictly monotonic piece-wise linear functions. The PX intersected the aggregate supply and demand schedules for each hour to arrive at an hourly market clearing price. This price represents the bid of the last megawatt of electricity that was called upon in the market and it is this price at which *all* trades were settled. On the day that electricity was scheduled to be produced, an hourly real-time market was conducted by the operator of the electricity grid (the Independent System Operator) to ensure that supply and demand exactly match. Just as with the day-ahead market, the real-time market was a uniform price auction. The state was divided into either two or three transmission zones during my sample, and the prices would vary between the zones if the transmission of electricity reached the capacity of the transmission line.

The design of the restructured market attempted to reduce the extent to which firms could exercise 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 vary between approximately $20 and $40 per megawatt-hour during 1998-99 and approximately $25 to $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]

In addition, some areas of the state have large demand and limited transmission capacity that puts them at risk for local market power. For example, San Francisco has large demand but has a limited number of transmission lines into the city because it sits on the end of a peninsula. The local generation capacity is very expensive and owned by two firms. If they were paid the price that clears the total California market, the firms may find it uneconomical to produce in many

[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.

hours and the lights would go out. However, if San Francisco were separated off into a separate market with its own market clearing price, the two firms would face limited import competition and could significantly raise prices. Recognizing the potential for local monopoly power in certain areas, the ISO signed outside contracts to purchase energy from these Reliability Must-Run (RMR) units when their bids were not accepted in the energy market.

As part of the restructuring process, the three original utilities were required to divest their power plants that generate electricity using fossil fuels. Southern California Edison divested the vast majority of its units 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 in April 1999.[5] San Diego Gas & Electric divested its units 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 production behavior of the five large thermal firms.

Entry into the California generation market is difficult and time-consuming. Strict environmental siting requirements and local communities afflicted with "Not In My Backyard (NIMBY)" have often stretched out the siting process to more than five years. In addition, power plants involve substantial fixed costs and require several years to construct. Financing the capital costs can be difficult if firms plan to sell power only to a daily trading pool rather than sign long-term fixed price contracts. Given these barriers to entry, collusion would not be threatened by entry in the short-run and incumbents only would consider the effect of entry a number of years down the road.

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

3 Distinguishing Between Static and Dynamic Market Power

3.1 Review of the Theory

Oligopoly theory provides several static and dynamic pricing models that may capture aspects of pricing behavior in the California electricity market. In static models, firms choose single period quantities or prices to maximize profits without explicit consideration of the effect of behavior in one period on the competitive environment in other periods. In quantity-setting games, firms can sustain prices above marginal costs in equilibrium. Also, positive price-cost margins are sustainable in price-setting games where firms face capacity constraints. Finally, if they can choose supply functions, firms in general can sustain prices anywhere between Cournot and Bertrand levels.[6]

Models of dynamic interaction show that firms 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]). 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 and 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. Brock and Scheinkman show that these results in general will differ when firms face capacity constraints because those capacity constraints will affect both the deviation and punishment profits.

[6] See Klemperer and Meyer [1989].

The California electricity market of 1998-2000 did not precisely correspond to any of these models but it did resemble several. A purely price-setting static model is not appropriate because capacity constraints prevented any single firm from undercutting and supplying the entire market. As a result, one might expect to observe market power during periods of high demand. Because storage or inventory is extremely costly, firms essentially must produce a quantity exactly equal to demand at every moment in time. When demand reaches levels near the industry's capacity, if one firm were to withhold capacity to try to drive up the price, other firms would not be able to replace all of the withdrawn supply. Hence, firms in this relatively concentrated industry are able to raise price and earn more revenue on all inframarginal output. This strategy is powerful because demand is very inelastic and customers in California do not face prices that vary with the real-time price of energy.[7] 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.[8]

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.[9] 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

[7] 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.

[8] 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. Because econometrically identifying supply function equilibria is not tractable with my data, I estimate models of games in which firms choose quantities.

[9] Data are not publicly available on the forward contract positions of any of the electricity generators. Industry analysts claim the volume of such contracts grew in 2000 and that contracts often took the form of contracts for differences.

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. Finally, applying this model to the market is inaccurate to the extent that some of the observed output may be contracted forward. However, industry analysts suggest forward contracting was relatively small until 2000.

The competitive environment in California contains properties of both the Green/Porter and Rotemberg/Saloner dynamic pricing models. The major differences between the Green/Porter and the Rotemberg/Saloner (and extensions) models stem 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 as well as demand shocks. Green and Porter firms know only some signal correlated with behavior such as prices or their own realized shares.

Firms in the California market have some intermediate level of information. A firm could partially observe the hourly production behavior of its rivals through several mechanisms. The website of the western U.S. transmission grid coordinator posted real-time generation data for all plants greater than 250 MW until October 2000.[10] 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. Other observable signals correlated with rival behavior include the market price and the forecast and actual realization of demand. Therefore, for many but not all of the shocks to residual demand, the five new generation owners could distinguish whether the shocks resulted from changes in rival behavior or other factors.

3.2 Review of the Empirical Literature

The literature contains a variety of empirical studies that make inferences about firm behavior and the pricing model that prevails in a particular industry. The challenge to the econometrician is to use some combination of detailed data on the industry and reliable structural assumptions. For

[10] These data covered approximately 93% of thermal capacity.

industries with relatively rich data, studies have exploited comparative static relationships implied by the various pricing models. Borenstein and Shepard [1996] analyze a panel of price-cost margins in the retail gasoline market and find that current margins are rising in expected future demand and falling in expected future costs. This evidence suggests some form of dynamic pricing behavior and is consistent with tacit collusion as in the Haltiwanger and Harrington model. 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 are engaging in some form of static pricing where they take into account the effect of one production unit on the price earned by other units in production. Finally, 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.

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.[11] Porter [1983] studies the Joint Executive Committee railroad cartel and finds evidence of switches between cooperative and noncooperative pricing behavior. Ellison [1994] extends the Porter analysis and finds no evidence that pricing regimes are associated with Rotemberg/Saloner collusion adjustments but does provide evidence that secret price cuts affect pricing.

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 mis-measure the conduct parameter if the true underlying process is not identical on the margin to a conjectural variations game. Corts demonstrates that if firms are engaged in efficient collusion, the traditionally estimated conduct parameter typically will underestimate market power.[12] The root of the problem is that if firms are colluding, the

[11]See the survey articles Bresnahan [1997] and Bresnahan [1989].

[12]Comparisions of direct measures of the conduct parameter versus the NEIO estimates have found NEIO methods

econometrician is estimating the wrong model; (s)he 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 to maximize joint profits subject to the incentive compatibility constraint that no firm has an incentive to deviate. As I show below, this dynamic first-order condition has an additional term that is non-zero 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 we obtain inconsistent estimates of firm conduct. As a result, the best one can expect to achieve by estimating the parameterized static first-order condition is to test non-nested hypotheses of perfect competition, Cournot, 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. Below I derive a general model that incorporates static market power and imperfect collusion as special cases, and I exploit the panel structure of my data to estimate a measure of conduct consistent under static and dynamic pricing.

3.3 Model of Firm Behavior Under Static and Dynamic Pricing

I derive a general first-order condition that I can estimate to make inferences about firm conduct. The model allows me to consistently estimate conduct parameters in settings where the firms may be engaged in static pricing or imperfect tacit collusion. I model the firms' strategic decisionmaking as a simple quantity-setting game. As I describe above, I believe modeling quantity as the choice variable is a good first-order approximation to production in this industry.

3.3.1 The Optimization Problem Under the Static Game

The key difference between static and dynamic pricing is that firms in the industry are solving different optimization problems. In the static model, firms are choosing to supply electricity as a function of contemporary supply and demand conditions without any intertemporal considerations

to understate market power (see Genesove and Mullin [1998] and Wolfram [1999]).

of the effect of current behavior on future competitive conditions of the market. Assume that in a static optimization setting, N firms play a quantity game in which they choose to supply a given (perfectly inelastic) quantity subject to a capacity constraint.[13] 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:

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

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.[14] 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'). I interpret λ_{it}^* as the shadow value of additional capacity when a firm is fully utilizing its capacity.

[13] 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.

[14] 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 will describe in section 4.

3.3.2 The Optimization Problem Under the Dynamic Game

Next I model the firm optimization problem when the industry is engaged in efficient tacit collusion.[15] The firms in the industry 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.[16] Assume that demand and cost shocks are observed ex post so that deviating from the collusive regime can be distinguished from 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.[17] Due to symmetry, maximizing individual firm profit is equivalent to maximizing joint profit.

Denoting the individual firm profits $\pi_i(\cdot)$, the optimization problem is to maximize joint profits subject to the constraint that no firm has an incentive to deviate from the collusive regime:

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

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

where $\pi_i^{br}(Q_t)$ is firm i's best (deviation) response to the joint collusive quantity Q_t, $E_t[\pi_{is}^{p}]$ are expectations of future period s noncollusive "punishment" profits, and $E_t[\pi_{is}^{*}]$ are current expectations of future period s collusive profits.[18] A firm will choose not to deviate if current and continuation collusive profits exceed the profits of deviating in the current period and earning non-collusive (Cournot) profits forever afterwards. We can rewrite the optimization problem as:

$$\max_{Q_t,\mu_t} L \equiv \sum_{i=1}^{N} \pi_i(\tfrac{Q}{N}) + \mu_t \left[\pi_i(\tfrac{Q_t}{N}) + \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi_{is}^{*}] - \pi_i^{br}(Q_t) - \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi_{is}^{p}]\right]$$

[15] This model assumes firms collude to achieve the maximum joint profits that are sustainable. Clearly, other collusive outcomes are equilibria as well by the "folk theorem".

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

[17] 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.

[18] 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 will affect the best-response profits of all the firms.

The first-order condition becomes:

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

When the incentive compatibility constraint is not binding ($\mu_t^*=0$), this equation is simply the static joint profit maximization problem of a monopolist. This would correspond 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.

3.3.3 A General (Static and Dynamic) First-Order Condition

I analyze a firm-level model which incorporates as special cases the static and dynamic first-order conditions. 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:

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

This equation can be generalized to incorporate the firm-level first-order conditions for both static (1) and dynamic market power (4):

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

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: $\theta_{it}=N$, $\mu_t^*\geq 0$

We can view (5) as a general model capturing various explanations for price above marginal cost. First, observed margins may represent scarcity rents for new production capacity in a perfectly competitive environment ($\lambda_{it}^*>0$). Second, margins may also result from firms unilaterally withholding current capacity to raise the price and earn higher revenue on their own current inframarginal units.

Finally, firms may be jointly withholding capacity to raise the price on joint inframarginal units, with this regime kept together by adjusting quantity so that no firm has an incentive to deviate from joint profit maximization.[19]

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 ($\mu^* = 0$).

4 Data

I estimate models using very detailed firm-level data for power plants in California from April 1998 until the market's meltdown in November 2000. Restructured electricity markets are subject to data reporting requirements that provide the empirical researcher with rich data on demand, cost structure, and output. Hourly output data are available from EPA's Continuous Emissions Monitoring System (CEMS). CEMS contains hourly output data for all fossil-fueled generation units in the California market except several small capacity 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 firm. To generate electricity in California, fossil fuel (primarily natural gas) is burned 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 fac-

[19] 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.

tors 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.[20] Several plants in southern California were required to purchase environmental permits for each pound of nitrogen oxides (NOx) emitted.[21] The hourly marginal permit cost is calculated as the monthly quantity-weighted average price of permit trades multiplied by the unit's emission rate in the hour from the CEMS data.[22] Adding on an estimate of variable operating and maintenance costs from Borenstein et al. [2000], I estimate a marginal cost for each unit.[23] I assume this marginal cost measure to be constant up to the capacity of the generator.[24,25]

Generators occasionally experience both scheduled and unscheduled downtime for maintenance. Some analysts have suggested that firms exercise 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

[20]All of the units for which I have 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.

[21]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.

[22]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.

[23]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 mis-measure 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.

[24]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. In the static model, the results are not sensitive to defining capacity as 80%, 90%, and 95% of 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.

[25]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.

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 I should take as the measure of the firm's marginal cost the lower or higher cost unit that still has available capacity. If I use the lower number, I would be ignoring 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 take the higher number, I may ignore the fact that the higher cost unit is running because it was called under outside contracts for Reliability Must Run (RMR).[26] Because I believe the former bias to be 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 my measure of the unit's capacity.[27] One problem with this measure of firm-level marginal cost

[26] However, given that they turn on the 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.

[27] 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. 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.

is that there are several small powerplants for which I do not have quantity data.[28] 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 by the firm in a given hour is not always known by the econometrician because transactions can be settled through the day-ahead market (the Power Exchange) or the real-time energy market (the ISO).[29] I use the Power Exchange day-ahead energy price for the zone of the state in which the firm generates. This price is most appropriate 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.[30] One firm (Duke) owns generators in different transmission zones. When transmission constraints are binding, I separate off the output attributable to the south generators and call the firm DukeSouth.

The measure of output I use in the empirical analysis is the total generation by each firm's thermal generating units. This may mismeasure the actual amount of generation sold to the energy market (and hence inframarginal output) for several reasons. It may understate output for the

[28]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.

[29]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 [2001b] 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.

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

firms that own small peaker units with no EPA data (e.g. Dynegy). The measure can be too high if some of the observed output is sold outside of the energy market either forward or under Reliability Must-Run contracts. I reduce the mismeasurement by focusing on peak hours of the day when less energy is sold under these RMR agreements. 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. My sample period ends in November 2000 when the uniform price auction ends and out-of-market purchases became very large. Therefore, potential mismeasurement of output may affect my market power estimates for Dynegy and for all firms late in the sample period. I discuss the sign of the potential bias below.

5 Motivating Empirical Evidence

The wealth of rich data in the California electricity market allow researchers to test directly for evidence of market power. Borenstein et al. [2000] find observed prices to be higher than those prices simulated by perfectly competitive behavior of the thermal generating firms. The price-cost margins are higher in periods when overall thermal demand is high. These results are consistent with the exercise of some form of market power. I formally estimate a model of behavior in the next section, but I begin by investigating evidence of market power by each of the five large generating firms. In this section, I consider two measures that suggest firms are exercising market power. First, I directly measure price-cost margins and find that firms frequently fail to utilize capacity when price is above marginal cost. Second, I find that a given power plant is utilized systematically less when it is owned by a new generation owner rather than a regulated utility.

5.1 Data Exploration

The observed production behavior of the firms suggests that they are not acting in a perfectly competitive manner. 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. I focus on the hour of the day with the highest average demand (5-6pm or hour 18). I choose this particular hour so that these results are comparable to my empirical analysis below which is best suited for this hour.[31] Industry analysts have suggested that the reported nameplate capacity overstates the true capacity of a unit. Therefore, I calculate the price-cost margins using two different assumptions about the true capacity. 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.[32] These margins imply a median Lerner index of 0.17.[33] 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 37% of firm-hours and greater than \$30 in approximately 22% of firm-hours. It is highly unlikely that marginal costs are this severely mis-measured so there is strong evidence that firms are not acting as price-takers.

One cost-based explanation for less than full utilization is that firms may face intertemporal adjustment constraints such as the rate at which a unit can increase or decrease output. If adjacent hours are not economical, a price-taking firm may utilize an economical generating unit at less than full capacity because it cannot ramp to full capacity in hour 18. For the thermal generators in California, units can typically ramp from zero to full capacity in times varying from one to

[31] In short, my analysis below estimates a structural model of firm output. This hour has the desirable property that unobservable intertemporal shadow values are likely to be small in this hour. 6pm (hour 18) is the highest average demand hour largely because offices have not shut down yet many people have gone home and begun to use lights and appliances.

[32] 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.

[33] 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.

three hours.[34] At 6pm (hour 18) on the average day in California, both demand and prices have been near their peak for several hours and will continue to be high for two to three more hours. Therefore, units often have sufficient time to ramp up and ramp down to high levels of output if they are economical in hour 18.[35]

Next, I explore how these price-cost margins vary 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 1 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 months (January-May) are actually negative in 1998 and hover around zero in 1999 and most of 2000.[36] The bottom panel breaks down margins by firm and demonstrates that all firms enjoyed positive margins. I emphasize that these margins are not scarcity rents because these are differences between price and marginal cost when *firms have excess capacity.*

Margins were sustained at consistently high levels during much of the late summer and early fall 2000. To some extent, high prices were mitigated by the price cap. The price cap in the ISO should effectively act as a cap on PX prices because demanders would have no incentive to bid above the ISO cap. This cap was not binding until the summer of 2000. As the ISO lowered the cap twice in 2000Q3, the cap began to play a significant role in the ability to exercise market power. Figure 2 shows the frequency with which the price hit the cap. Interestingly, prices and margins both rose as the price cap was lowered.

Figure 1 makes clear that market power existed in 1998-99 but was much less substantial than in 2000. In 2000, margins first spiked in June and then remained consistently high throughout July-September. These dramatic increases in margins were coincident with several shocks to the cost of inputs. The top panel of Figure 3 shows the steady rise in natural gas prices throughout 2000 and

[34] California Energy Commission rate hearing data.

[35] An earlier version of the paper also analyzed each firm's utilization of economical capacity. For hour 18, ramping time is not a significant determinant of utilization. However, firms that are "exogenously larger" withhold more economical capacity than they would otherwise provide in a perfectly competitive market.

[36] 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. This became less of an issue over time as the original RMR contracts were amended.

the rapid tripling of price in November. NOx permit prices were negligible at the beginning of the year but steadily rose beginning in July to $40/lb by November so that permit costs alone could add over $100/MWh to the cost of electricity. The bottom panel shows the consequent effect on the marginal cost for the five firms. The marginal cost averaged across the five firms gradually rose during 2000 and became more volatile during the latter months. This panel also shows the five-firm average price-cost margin. There is no discernible pattern between cost shocks and changes in the margins.[37]

These summary statistics focus on hour 18 (5-6pm) when demand on average is at its peak in California. I choose this hour to reduce mis-measurement in my structural analysis below. However, I can assess how representative this hour is of market power across all hours. Table 3 shows the yearly average firm price-cost margins for different periods of the day. The thermal generating firms enjoy higher margins during high demand hours of each day but negative margins during off-peak hours. One explanation is that during low demand hours, lower cost generation such as nuclear power satisfies a large fraction of demand so that the residual demand faced by the thermal firms is very small or perhaps zero. As a result, thermal gas-fired units are less likely to be economical or in a position to set the price. Hour 18 has margins higher than the average peak demand hour so this paper will focus on periods with relatively high margins. Figure 4 shows the daily pattern of margins during two high margin months (August 1998 and 2000) and a low margin month (April 1999). In all three months, margins follow a daily cycle between positive margins in high demand hours and slightly negative margins during the night. Theory suggests firms would not operate with negative margins, but an explanation for the observed behavior is that firms continue uneconomical operation during night hours to avoid the cost of shutting down and restarting the unit the next day. It is important to note than lower margins during off-peak hours does not necessarily imply more competitive behavior. Even if conduct were the same during off-peak hours, one expects to see lower margins because the residual demand for the five thermal firms is more elastic.

[37]Arguably the firm with the *highest* marginal cost is most relevant because it may set the price and affect all firms' margins. However, a similar analysis using the highest marginal cost rather than the average marginal cost also fails to identify a clear pattern.

5.2 Capacity Utilization and Market Power

My exploration of the data suggests that firms are withholding output that would be economical to produce if they were price-takers. This would appear to be fairly strong evidence that firms exercise market power. However, the evidence is even stronger if the withholding of economical capacity occurs when incentives exist to exercise market power. In this section, I analyze the utilization of capacity during periods with strong exogenous incentives to withhold capacity.

I compare the utilization rates of power plants owned by the five new generation owners to the utilization rates of the same plants when owned by the old utilities. I exploit the fact that after deregulation began on April 1, 1998, many of the fossil-fueled units were owned by the old utilities for some period of time before they were divested to the new generation owners. I observe the same power plant operated by the old utility and by the new generation owner. This quasi-experiment allows me to observe the same unit's production behavior in a deregulated environment when owned by a utility and by a merchant generating firm. Theory suggests the new generation owners have incentives to reduce output below competitive levels in order to raise the price. However, utility companies have very dampened incentives to influence prices. I test whether these differing incentives to suppress output manifest themselves in the data. For generating units that were divested from the utilities to the unregulated firms, I analyze the capacity utilization rates before and after divestiture as a function of the profit margins.

I estimate a kernel regression of utilization rates on price-cost margins for the same units pre- and post- divestiture for hours 18 in which the units are operating. Figure 5 plots the fraction of a unit's capacity that is utilized for a given profit margin. I also plot a kernel density estimate of the price-cost margins. Under both forms of ownership, units tend to produce more when the profit margins are higher. However, for the same margin, units tend to utilize less capacity when owned by deregulated firms ("post-divestiture"). In fact, focusing on the range of margins that occurred most frequently, the new generation owners almost uniformly utilized less capacity.[38] This

[38] This assumes the generating units had the same efficiency rating before and after divestiture. This appears to be a reasonable assumption because the new generation owners contracted for the utilities' engineers to continue to maintain the plants for two years after divestiture. Therefore, the only significant difference pre- and post-divestiture is the agent who was bidding the plants' output.

is consistent with the five large deregulated generation owners suppressing output to exercise some form of market power.

6 Estimation of the Behavioral Model

In this section, I impose more theoretical structure on the data to estimate if firms supply power in a manner more consistent with static or dynamic market power. The first-order conditions in section 3.3 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. First, I estimate the static first-order condition and find the data are fairly consistent with Cournot pricing, however I find behavior less competitive than Cournot in 2000. Then, I estimate a form of the general first-order condition (equation (5)) and reject the model of optimal tacit collusion in all quarters of my sample.

6.1 Static Model

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

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

This model says that firms exercising market power ($\theta > 0$) will observe higher price-cost margins (adjusting for scarcity rents on capacity) 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. 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. Therefore, 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 their residual demand.

Demand Side. The first step to estimating the static model is to estimate the demand parameter P_t'. This parameter is the amount by which the market price will fall when any of the five strategic firms choose to produce another unit of output. Total demand is assumed to be perfectly inelastic. Few customers face the hourly wholesale price of power and even if more had the opportunity to respond, demand for electricity is very inelastic. The residual demand of the five strategic players has some degree of elasticity due to supply elasticity by other fringe suppliers. These other suppliers include the fringe thermal generators (PG&E, SCE, SDG&E, Thermo Ecotek), nuclear generation, hydroelectric and geothermal power, small independent producers,[39] and imports from outside of California. 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.[40] 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.[41] 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".

I estimate the (competitive) supply by all fringe suppliers for Hour 18. To the extent that any of these fringe firms exercise market power, my estimate of the fringe supply function may be biased because fringe supply would not be merely a function of costs but also a function of the behavior of the five strategic firms. Total residual demand of the five strategic firms (Q^D_{strat}) is the total (perfectly inelastic) market demand net of supply by the competitive fringe:

[39]For example, some oil refineries self-provide electricity and are qualified to sell surplus generation to the grid under the Public Utility Regulatory Policies Act of 1978.

[40]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.

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

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

I estimate the supply function of the competitive fringe and use the negative of the slope of fringe supply as my estimate of the slope of strategic firm demand P_t'.

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.[42] 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 excess generation out of state, 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).[43] 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:

$$\begin{aligned} \ln Q^S_{fringe\,t} = {} & \beta_0 + \beta_1 \ln P_t + \beta_2 \ln GasPrSouth_t + \beta_3 \ln GasPrNorth_t + \\ & \beta_4 \ln Diff65TempNeigh_t + \beta_5 DAYDUM_t + \beta_6 MONTHDUM_t + v_t \end{aligned} \tag{7}$$

β_1 can be used to calculate the slope of fringe supply which equals the opposite of the slope of the 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 value of scarcity rents on capacity. The measures of price and output are discussed in section 4. My measure of marginal costs includes fuel

[42] 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.

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

and operating costs but no production adjustment costs. My model is a simple model of quantity choice that abstracts from the more complex dynamic programming dispatch problem that the firm actually is solving. I analyze firm behavior during periods when my simple model and the more complex model are least likely to differ. Several conditions could make my measure of marginal costs differ from the actual marginal costs. For example, if it has a unit shut down, a firm would incur startup costs to fire up that unit. In order to deal with startup costs, I only analyze the firm behavior by units that are already operating during the particular hour. Also, firms face constraints on the rate at which they can ramp units and these constraints show up as shadow values in the firm's dynamic programming problem. I focus on a particular hour of the day when ramping shadow costs are likely to be low: 6pm (hour 18). 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.[44] Therefore, I focus on hour 18 and assume any shadow costs of operating constraints to be zero. Accordingly, a price-taking firm has incentives to fully utilize all of its capacity with marginal cost below the market price.

I cannot measure the scarcity rents on capacity (λ_{it}^*). The shadow value of capacity is zero when capacity constraints are not binding, however the value is unknown when constraints are binding.[45] Firms are producing at capacity in only 4.4% of firm-hours in my dataset. I assume the scarcity rents to be constant across firms and time and estimate the scarcity rents by including a dummy variable ($CAPBIND$) equal to 1 if capacity constraints are binding and equal to zero otherwise.[46] The coefficient on $CAPBIND$ is the (average) shadow value of added capacity.

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 hetero-

[44]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.

[45]The proper shadow value is the difference between marginal cost and the parameterized marginal revenue ($P + \theta P'q$) evaluated at the capacity constrained quantity. If firms are price-takers ($\theta = 0$) then the shadow value is the difference between price and marginal cost. However, this overstates the shadow value if firms are not perfectly competitive. For example, firms may be acting as Cournot competitors ($\theta = 1$) yet still produce up to full capacity when demand is very high. I face the problem that the unobserved shadow value is a function of the unknown conduct parameter θ.

[46]I also estimate the conduct parameter θ using only observations in which the capacity constraints are not binding ($\lambda_{it}^* = 0$), and the results are very similar.

geneous 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.[47] I simultaneously estimate the fringe supply (7) and each firm supply relation (8) via three-stage least squares.

I show estimation results below but first I illustrate the shape of the supply relation by the strategic firms. When $CAPBIND$ is zero (96% of the observations), the static model reduces to a simple bivariate (instrumental variables) regression. Figure 6 plots the price-cost margins against the fitted values of $-P_t' q_{it}$.[48] 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 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 6 suggests that the overall supply relation is relatively consistent with static pricing. The top panel plots the kernel regression estimate and the data for the complete sample of July 1998-

[47] I use the day-ahead forecast of demand rather than realized demand because an unexpected demand shock raises output and hence marginal cost, but does not raise the day-ahead price. Therefore, actual demand would not be a valid instrument.

[48] Below I show results of estimating the fringe supply elasticity during two periods of my sample. I obtain very similar fringe supply elasticity estimates and use an intermediate value in constructing this figure.

November 2000.[49] 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 we expect from a static pricing game. This larger conduct parameter for observations above $-P_t' q_{it} = 25$ may imply a form of dynamic pricing during high demand hours, or may simply result from my overestimating the fringe supply elasticity (and hence the demand elasticity) in these peak hours. I do observe periods when firms are operating with 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 6 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 for various levels of inframarginal output in the period after June 2000. Although this evidence does not necessarily suggest a shift from static to some form of dynamic pricing, it does suggest the market was less competitive after June 2000.[50]

Next, I show results from jointly estimating the system of fringe supply (7) and the strategic firm supply relations (8).[51] 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 occuring in August. I estimate the conduct parameter by ignoring days when the price hit the cap.[52] I break down the sample into a period during which there were four firms in the market and another period later in the sample with five firms.

[49] I exclude 1998Q2 because negative margins are inconsistent with any reasonable static pricing model. Industry analysts believe the firms were selling power under alternative regulatory (RMR) agreements rather than actually selling to a market with price less than marginal cost.

[50] 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.

[51] 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.

[52] 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.

Results are shown in Table 4 and are similar for both time periods. Fringe supply is relatively inelastic in both periods (0.15 and 0.20). Given the relative size of the fringe and strategic players, this suggests that the strategic firms face a total residual demand elasticity of approximately 1.32 during the 4-firm period and 0.79 during the 5-firm 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 $\hat{\theta} = 1.07$ with a standard error of 0.09. 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 $\hat{\theta} = 1.19$ with a standard error of 0.06. I fail to reject Cournot pricing during the 4-firm period and observe pricing statistically higher than Cournot levels in the 5-firm period.[53]

Next, I estimate how conduct varies over my sample period. Recall from Figure 1 that direct measures of price-cost margins are highest during the second half of 2000, and also are sustained at high levels during the summer of 1998 and fall of 1999. Whether periods of high margins correspond to levels of less competitive conduct will depend on the residual demand elasticity the firms face. Less competitive behavior may not yield higher margins if residual demand is more elastic. I expect the residual demand of the five thermal generators to be more elastic in the low demand winter and spring months when nuclear and hydroelectric generation satisfy a substantial fraction of total demand. The first panel of Table 5 reports conduct parameters in which conduct is restricted to be equal across firms but is allowed to vary by quarter. I also report estimates of the 5-firm residual demand elasticity evaluated at mean output levels and the average price-cost margins that resulted from the firms' conduct on that residual demand.

The point estimates suggest that the market was more competitive in 1999 than in either

[53]Note that the supply relation I estimate has no intercept because the theory suggests the relationship is a ray through the marginal cost intercept. If I include an intercept, I find a very small (and statistically zero) intercept of \$0.24 in the 4-firm period and a small (yet statistically non-zero) intercept of -\$4.64 in the 5-firm period. The corresponding slopes of the supply relation are 1.05 and 1.31, respectively.

1998 or 2000. During 1998-99, seasonal patterns appear in the estimates of both conduct and residual demand elasticity. During the highest demand months of July-September, I fail to reject Cournot pricing in either year. Nevertheless, Cournot behavior leads to high margins because residual demand is less elastic. Conduct is less competitive in the fourth quarters yet does not yield substantially higher margins because demand is more elastic than during summer months. Finally, in the lower demand first and second quarters of 1999, I find conduct statistically indistinguishable from Cournot in quarter 1 and statistically lower than Cournot in quarter 2.

In 2000, I estimate behavior to be distinctly less competitive. The point estimates of the conduct parameters are substantially higher in 2000Q2-Q4 with the estimates statistically higher than Cournot from April to September. These estimates formally confirm the apparent pivot in the supply relation after June 2000 as seen in Figure 6. Not only is the market less competitive, but the relatively inelastic residual demand during this period contributes to the very high price-cost margins seen in Figure 1. Finally, note that I find implausible estimates for the low margin first quarter of 2000. I find the competitive fringe to supply more energy when the price is lower which forces the fringe supply elasticity and the conduct parameter to be negative.

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 which 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 substantial fractions 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.[54] I assume all observed production

[54]This is true whether the forward contract is a hedge contract (contract for differences) or a forward contract for

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"). Overall, I believe that the bias from ignoring risk premia is dominated by the biases from overstating output and understating prices. Therefore, my conduct estimates are likely biased downwards in 2000.

In some quarters, the conduct parameter estimates reject all theorized equilibrium values of games of static pricing and perfect tacit collusion. 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 quarter. In each quarter I estimate some average measure of conduct. Conduct may vary over a quarter 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). 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. Over the ten quarters, 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 margins (from Figure 1) were more driven by changes in costs (Figure 3) and residual demand elasticity than by the conduct of the firms.

Finally, I allow the conduct parameters to vary by firm and estimate each firm's competitiveness during the first and second half of my sample. I find a modest degree of heterogeneity in firm behavior. During the period with four strategic firms in the market from July 1998-April 1999, I find conduct parameters above one for Reliant and Duke and below one for AES, but I cannot reject Cournot for any of these firms. Dynegy has a particularly large parameter estimate that decreases but remains high during the five-firm period.[55] This high conduct parameter estimate

delivery.

[55]Given the unusually high conduct estimates for Dynegy, one may be concerned that conduct estimates above are

may result from my missing data for some of Dynegy's small peaker plants. During the period from April 1999-November 2000 with five strategic firms, I reject Cournot pricing for four of the five firms. For the four firms in the market for the entire sample, conduct is relatively similar with the exception of Dynegy. When I focus on the period of the price runup in June-November 2000, firms are uniformly less competitive with four of the five firms exhibiting "super-Cournot" pricing. Southern and Dynegy (with data caveats) appear to be the least competitive and AES appears to be most competitive.

6.2 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 6 fails to suggest 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 6 suggests the firms are engaged in a (static) 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 such as dynamic pricing.

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 dynamic, conduct parameters from estimating the static model are biased as discussed in section 3.3.3. Therefore, I estimate the general first-order condition (equation (5)) to test for dynamic pricing. I do not

substantially driven by Dynegy behavior. I re-estimate the static models above allowing Dynegy to have a different conduct estimate, and find that neither the estimates nor the inferences substantially change.

have data on the effect of output on the best-response profits $(\frac{d\pi^{br}}{dQ}_t)$. However, note that this term is constant across all firms during a given period. I can condition out this effect by including time fixed-effects and estimating conduct off of the between firm variation in margins and output. Rather than include fixed effects for every different time period (i.e. every day), I assume that the "bindingness" of the incentive compatibility constraint is constant for "similar" periods of time. The level of the incentive compatibility constraint depends upon the size of the capacity of the five strategic firms relative to residual demand of those firms. Thus the value of the constraint is largely determined by overall demand for electricity in California. I condition out the IC constraint term by including fixed effects for each demand quartile. If pricing follows the dynamic first-order condition and firms are symmetric, the conduct parameter $\theta = N$. Table 6 shows the estimated conduct parameters to be significantly lower than N (either 4 or 5). I reject tacit collusion for all quarters.

These results also have implications for the interpretation of the static model. If the firms are following the static pricing model, the estimators of $\hat{\theta}$ from both the static model (equation(6)) and more general model (equation (5)) are both consistent. Yet the estimates of the conduct parameter θ reported in Tables 5 and 6 differ substantially in several of the quarters of my sample. In particular, during the low margin quarters of 1999Q1-Q2 and 2000Q1, the estimates from the fixed effects model are closer to zero.[56] The demand quartile fixed effects allow each quartile to have a different intercept so that I estimate conduct off the within-quartile variation in margins and output. This should not affect my conduct estimates if the supply relation is linear ($P - c$ is linear in $\hat{P}'q$ in Figure 6). However, if I mis-estimate the slope of residual demand ($\hat{P}'$) in a way that varies with demand, then the estimates in Table 6 provide consistent estimates of conduct. An alternative explanation is that demand levels capture much of the variation in observed margins during the quarters in which the estimates vary.

[56]The high margin quarter of 1998Q3 also has a small conduct estimate and this is driven Dynegy. When I allow Dynegy to have separate parameters, the conduct estimate for the other three firms is nearly the same as in the static model.

7 Conclusions

Oligopoly theory identifies several plausible pricing models that may apply to the California electricity market. Despite the fact that it is a commodity, electricity has special characteristics such as non-storability and very inelastic demand which can allow individual firms to withhold output and raise the price. In addition, the market is cleared through a daily auction between a small number of players with substantial information about one another. Such repeated interaction would appear to be an environment ripe for collusion. This paper investigates the pricing behavior in the California wholesale electricity market from its inception in April 1998 until its collapse in late 2000. I use firm-level production and cost data to test whether firms exercised market power and make inferences about the underlying pricing behavior. I focus on a particular hour of the day (hour 18) for which my measure of conduct is likely to yield the most accurate estimates. I directly measure market power and test if the observed price-cost margins are more consistent with static or dynamic pricing.

I find direct evidence of market power: units systematically have unused capacity when price is above marginal cost. To differing extents, all five large new generation owners exercised market power from 1998-2000. In addition, I compare the same generating units' capacity utilization rates when owned by the old utilities and the new generation owners and find evidence that the new generation owners suppress output.

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 and find that much of the variation in margins is driven by changes in the residual demand that the five firms face. In addition, I find that the market was slightly more competitive in 1999 than in 1998.

Conclusions about firm behavior in 2000 are less clear. 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 6) 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. Finally, 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.

Market power is likely to have larger effects on prices than on welfare. Because total demand for electricity is nearly perfectly inelastic, there is unlikely to be a large amount of substitution away from electricity towards other less efficient sources of energy. The largest welfare effect of market power is likely to be inefficient production of electricity. Low cost generation withheld by the large strategic firms is replaced by higher cost generation by fringe firms.

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. 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.[57] 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.

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

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. Diagnosing Market Power in California's Deregulated Wholesale Electricity Market. POWER Working Paper PWP-064, University of California Energy Institute, 2000.

S. Borenstein and A. Shepard. Dynamic Pricing in Retail Gasoline Markets. RAND Journal of Economics, 1996.

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.

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

G. Ellison. Theories of Cartel Stability and the Joint Executive Committee. RAND Journal of Economics, 1994.

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 and E. Kahn. A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000. Mimeo, 2001a.

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

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.

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.

R.H. Porter. A Study of Cartel Stability: The Joint Executive Committee, 1880-1886. Bell Journal of Economics, 1983.

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.

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.

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%

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	77.42	16.47	133.31	-29.05	695.02	.26
Southern	99	38.34	12.10	81.56	-21.21	1046.57	.27
Reliant	94	32.25	7.91	76.68	-25.41	686.50	.22
Dynegy	100	26.00	3.51	73.42	-31.23	689.49	.10
AES	99	22.20	3.66	93.30	-1514.56	685.44	.11
Duke	88	19.98	4.15	45.73	-20.16	474.59	.13
		Capacity ≡ 80% Nameplate					
DukeSouth	79	69.44	16.76	115.96	-29.05	690.50	.24
Southern	94	37.16	11.93	81.82	-21.21	1048.27	.27
Reliant	94	32.90	8.44	77.26	-25.41	686.50	.23
Dynegy	99	26.25	3.47	73.88	-31.23	689.49	.10
AES	94	19.90	3.28	94.01	-1514.56	687.35	.10
Duke	80	16.94	3.53	38.95	-20.16	392.10	.11

This table represents summary statistics 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 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: Average Firm Price-Cost Margins by Time of Day ($/MWh)

	Year		
	1998	1999	2000*
Hour 18	15.00	10.88	62.28
Peak (9am-10pm)	9.27	6.91	47.56
Off-Peak (11pm-8am)	-7.47	-7.02	-6.01

* January-November 2000.

Table 4: Estimate of Fringe Supply and Strategic Supply Relations for Hour 18†

	4 Firm Market*		**5 Firm Market****	
	Fringe	Strategic ***	Fringe	Strategic ***
Dependent Variable:	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$
$\frac{P \cdot q}{Q^s_{fringe}}$	–	7.183	–	5.883
	–	(0.170)	–	(0.108)
λ (\$/MW)	–	25.947	–	59.423
	–	(1.490)	–	(2.526)
Log(Price)	0.149	–	0.203	–
	(0.013)	–	(0.009)	–
Log(GasPrSouth)	-0.262	–	-0.140	–
	(0.095)	–	(0.113)	–
Log(GasPrNorth)	-0.087	–	-0.079	–
	(0.060)	–	(0.122)	–
Log(Diff65TempNeigh)	0.012	–	-0.025	–
	(0.016)	–	(0.008)	–
Constant	9.865	–	9.440	–
	(0.051)	–	(0.055)	–
N	268		573	
R^2	0.70		0.52	
$\hat{\theta}$	1.07		1.19	
	(0.09)		(0.06)	

Fringe represents equation (7) and Strategic represents equation (8). Standard errors are in parentheses.

Note: 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.

*** Although the system contains a supply relation for each firm, the coefficients are restricted to be equal in this model.

Table 5: Static Conduct Parameters by Quarter and Firm for Hour 18[†]

By Quarter: Conduct Parameter, Avg Margin, and Resid Demand Elasticity

Quarter	Estimate	Std Error	Margin	$\hat{\eta}^{D}_{strat}$
98Q3	1.25	0.19	23.46	-1.30
98Q4	1.77	0.38	8.87	-3.02
99Q1*	1.27	0.31	2.59	-5.09
99Q2*	0.76	0.12	3.28	-1.63
99Q3	0.74	0.14	14.97	-0.49
99Q4	1.49	0.20	15.67	-1.26
00Q1	-1.02	0.46	1.12	+3.76
00Q2	1.91	0.15	37.25	-1.11
00Q3	1.72	0.20	89.76	-0.65
00Q4	1.38	0.27	43.16	-0.89

By Firm and Time Period: Conduct Parameter

	4 Firm Market**			5 Firm Market***		
Firm	Estimate	Std Error	R^2	Estimate	Std Error	R^2
Southern	–	–	–	1.51	0.08	0.84
Reliant	1.12	0.10	0.82	1.20	0.06	0.83
Duke	1.16	0.11	0.87	1.21	0.07	0.85
AES	0.87	0.08	0.80	0.96	0.09	0.35
Dynegy	4.70	0.46	0.75	2.38	0.13	0.80

	June-November 2000		
Firm	Estimate	Std Error	R^2
Southern	1.74	0.14	0.86
Reliant	1.40	0.11	0.85
Duke	1.40	0.12	0.91
AES	1.12	0.17	0.36
Dynegy	2.81	0.24	0.83

[†] 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.

*Due to a new strategic firm entering April 16, 1999, I extend 99Q1 through April 15 and begin 99Q2 on April 16.

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

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

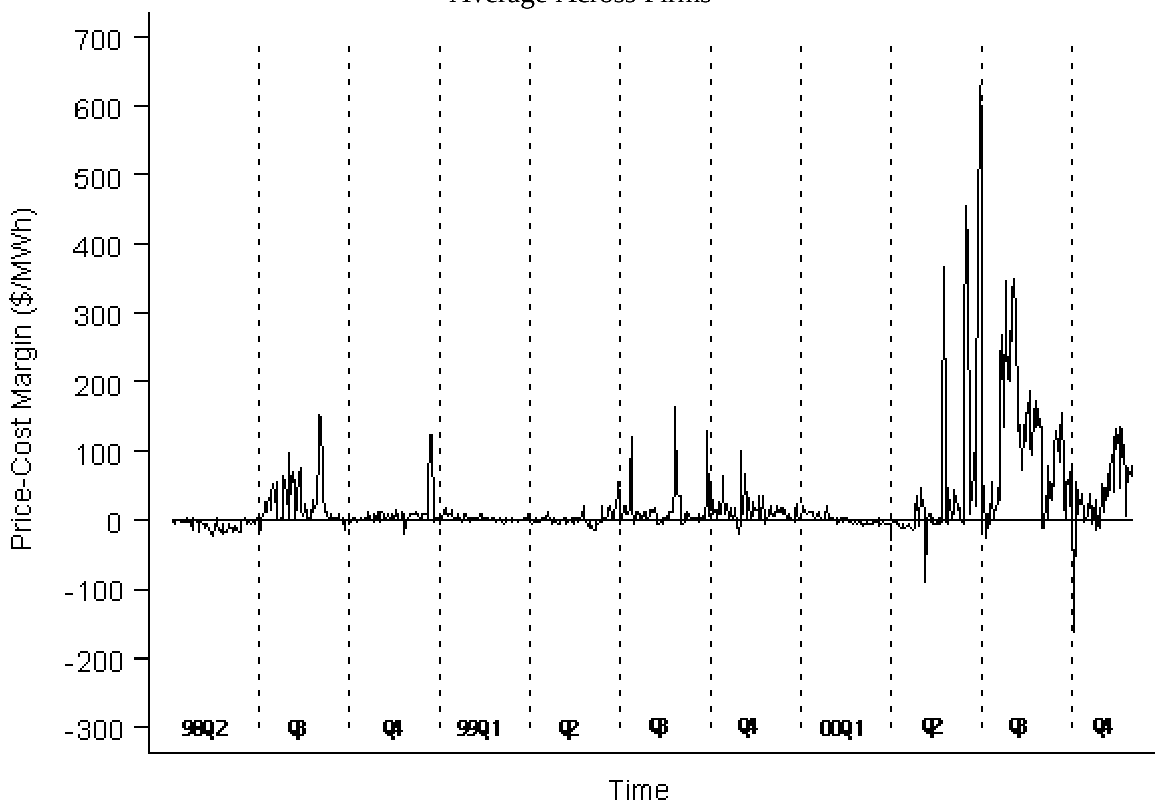
Table 6: Dynamic Conduct Parameters by Quarter

Quarter	Estimate	Std Error
98Q3	0.15	0.05
98Q4	1.07	0.25
99Q1*	0.02	0.02
99Q2*	0.11	0.04
99Q3	0.62	0.13
99Q4	0.58	0.11
00Q1	-0.34	0.16
00Q2	2.19	0.18
00Q3	1.92	0.24
00Q4	0.62	0.18

*Due to a new strategic firm entering April 16, 1999,
I extend 99Q1 through April 15 and begin 99Q2 on April 16.

Figure 1: Price-Cost Margins in Hour 18

Average Across Firms



By Firm

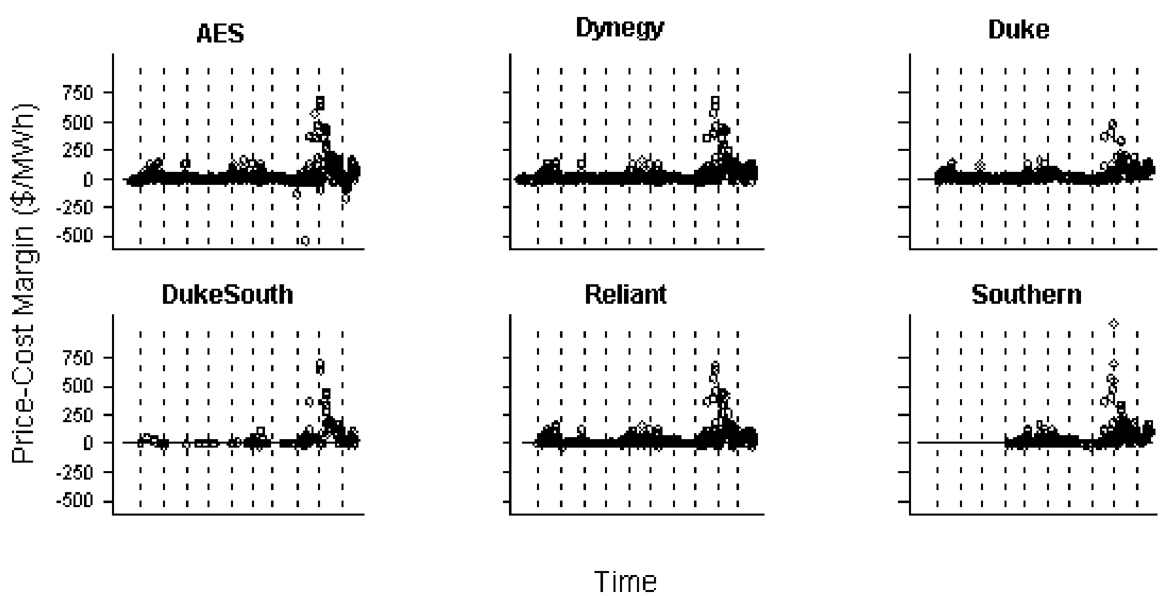


Figure 2: Hour 18 Prices and the Price Cap in 2000

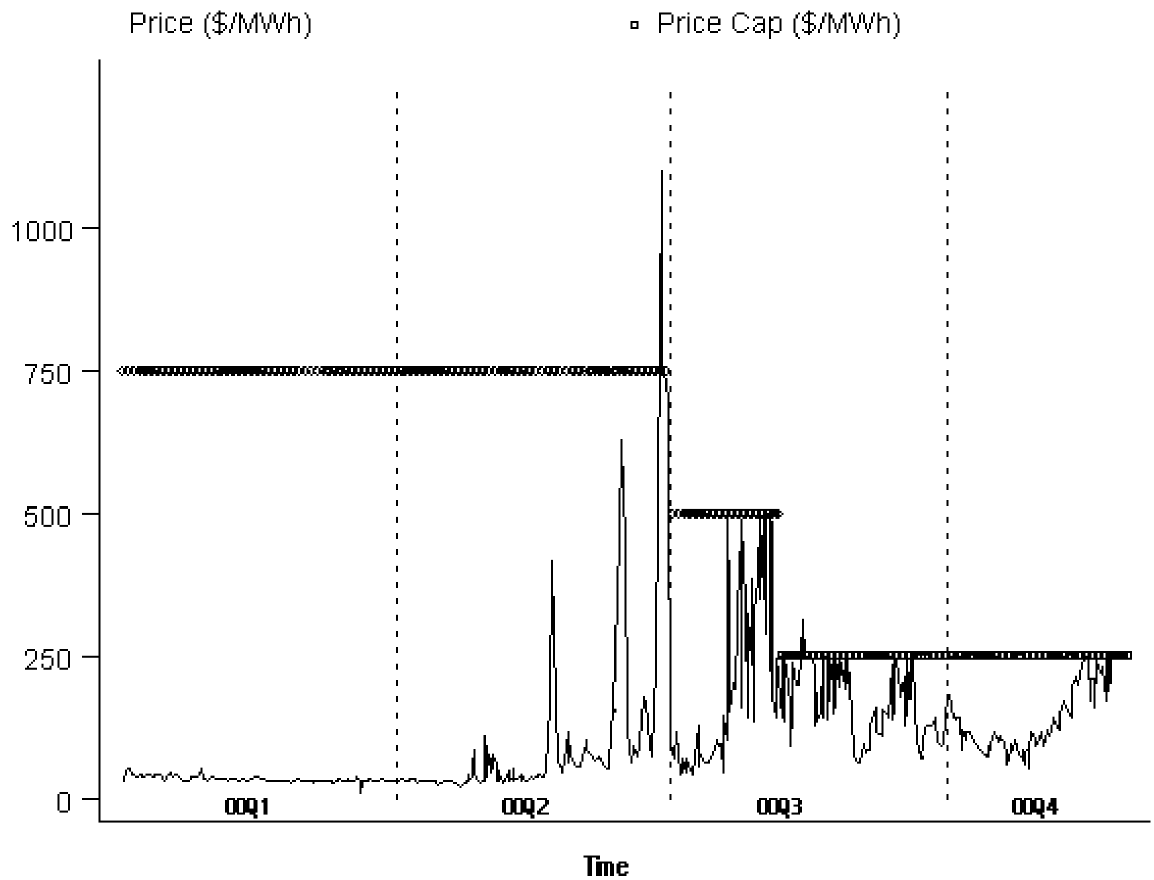
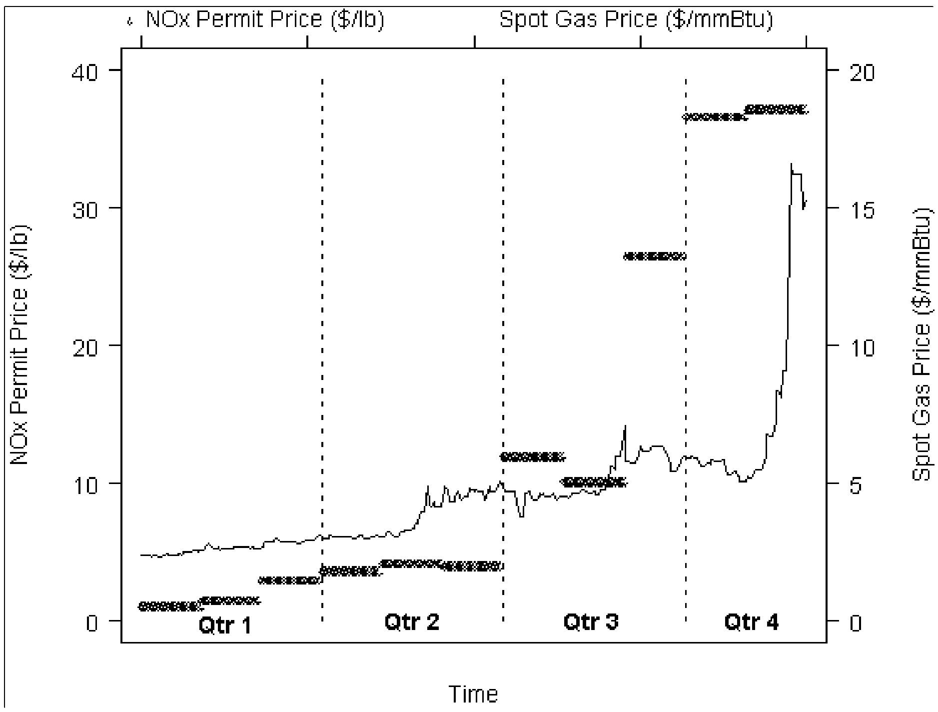
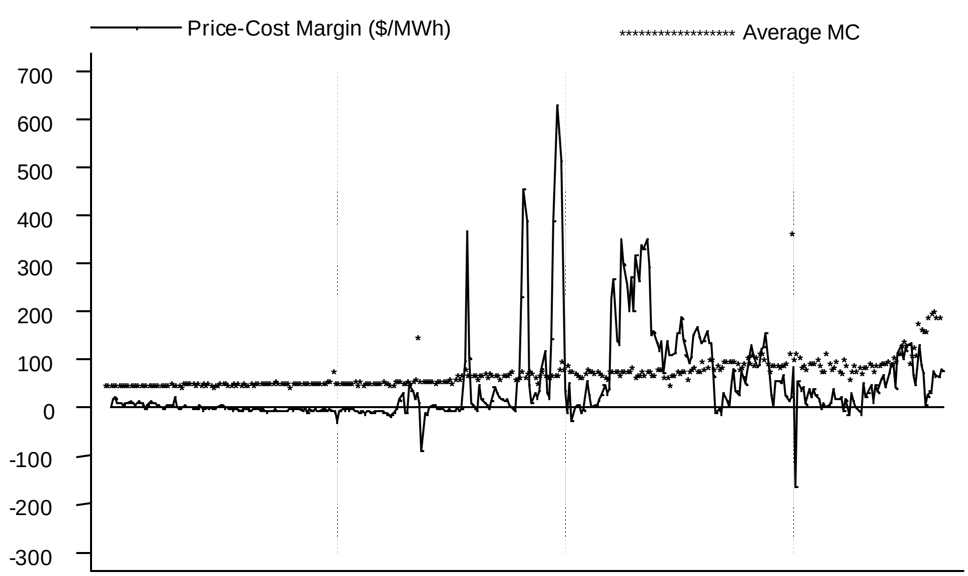


Figure 3: Input Cost Shocks in 2000

Input Prices



Marginal Cost vs. Price-Cost Margins



Hours 18 in 2000

Figure 4: Average Firm Price-Cost Margins in All Hours ($/MWh)

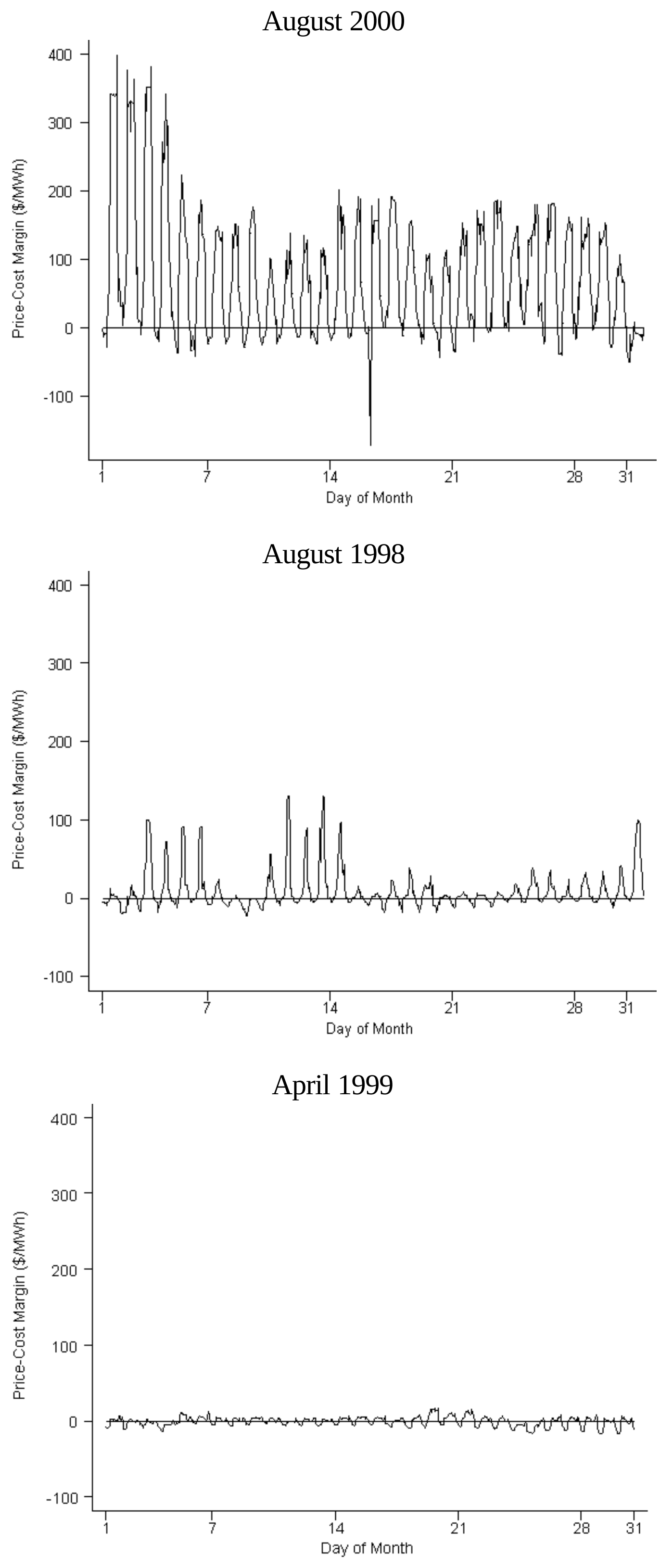
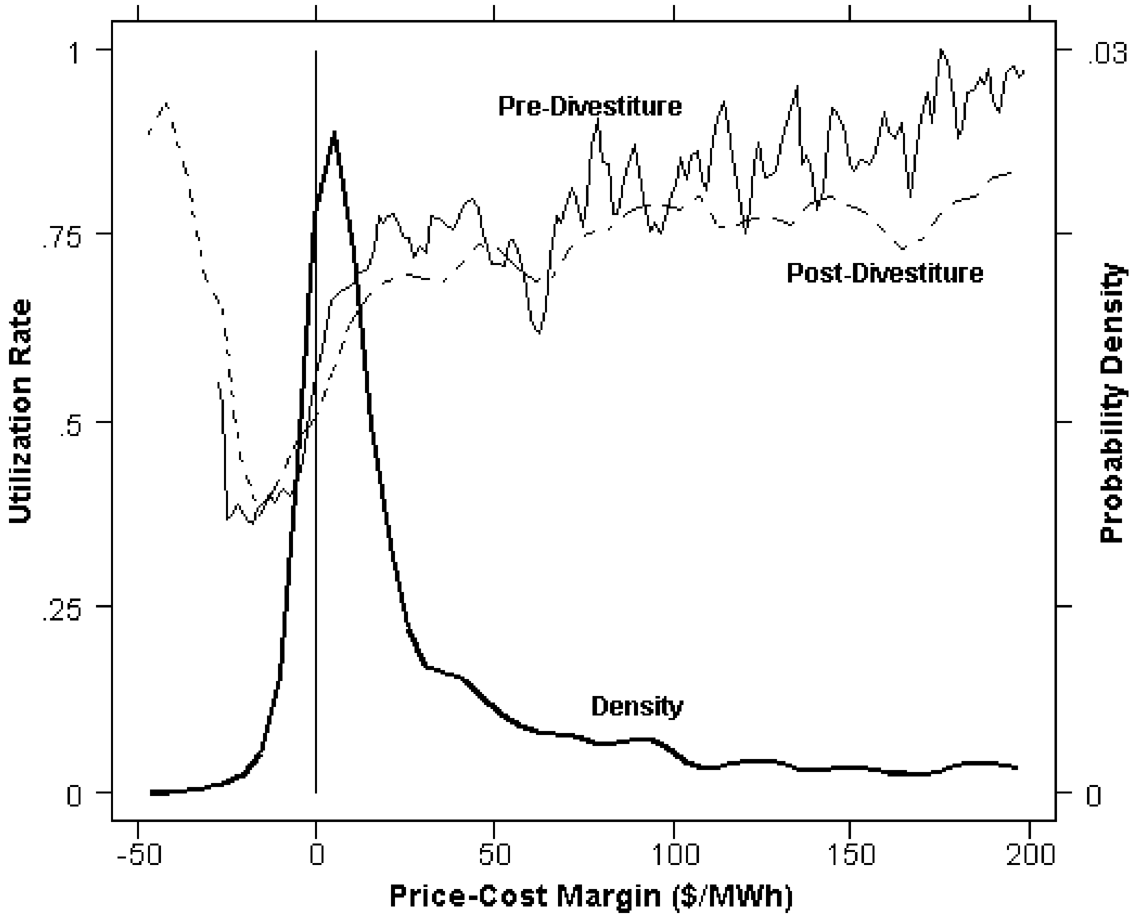
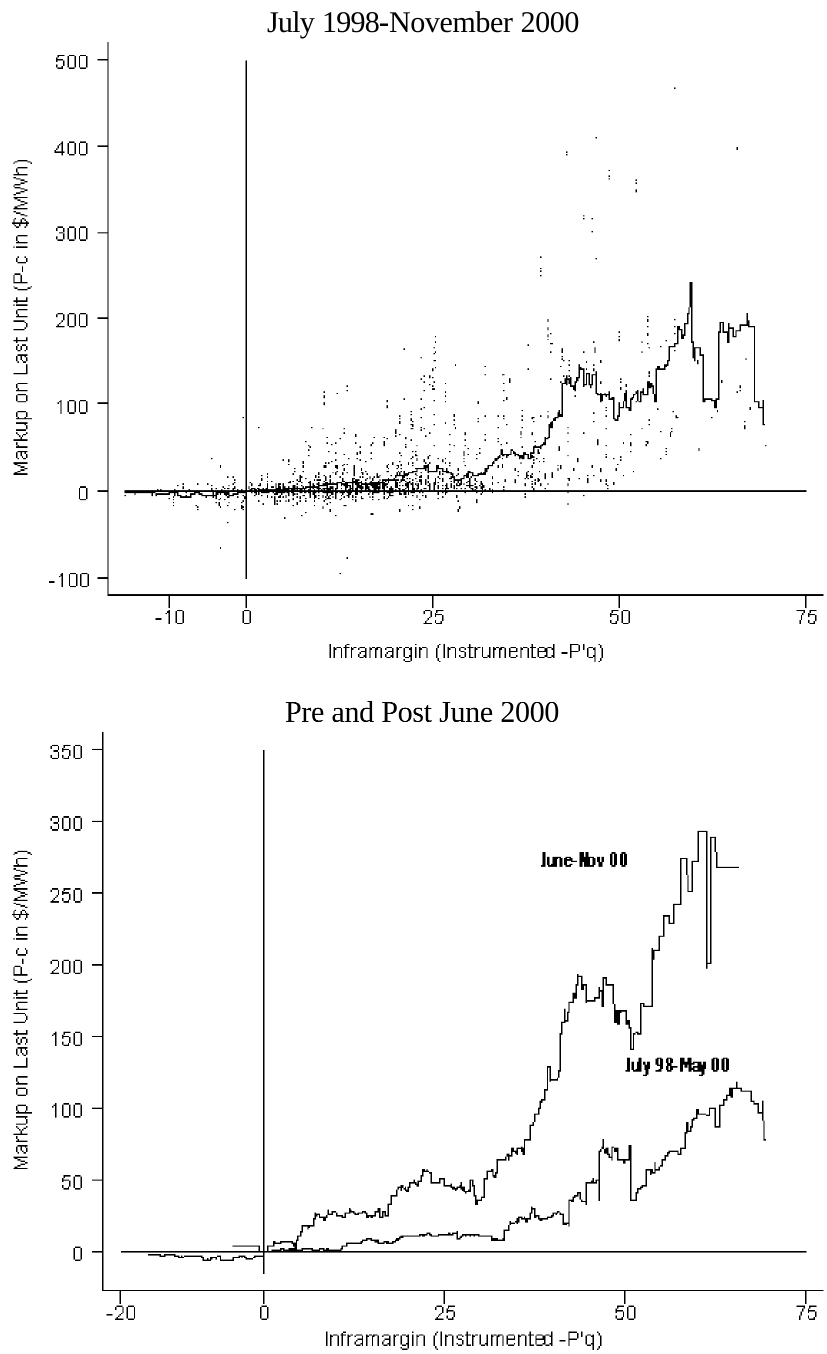


Figure 5: Unit Utilization Rates Pre- and Post-Divestiture



Kernel regression of utilization rates on profit margins of units operating in hour 18. Post-divestiture refers to units when owned by the new generation owners and pre-divestiture refers to units when owned by the original utilities. The estimated relationship for pre-divestiture units is less stable because there were very few high margin hours during the period before divestiture.

Figure 6: Static Behavioral Model



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, nine outlier observations with large and small margins are excluded.