PWP-080r

Pricing and Firm Conduct in California's Deregulated Electricity Market

(Forthcoming in Review of Economics and Statistics)

Steven L. Puller

Original Draft: November 2000

This Version: July 2004

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*

July 2004

Abstract

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

JEL Classification: L1, L4, K2, L5

---

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

The restructuring of the electricity industry in California and subsequent meltdown of the market has raised many questions about the feasibility of electricity markets to be competitive. In 1998, California opened electricity generation to competition by restructuring the method of procuring electricity. Incumbent regulated utilities divested many of their plants to private firms, which bid in daily auctions to supply power to the grid. Wholesale prices remained at low to moderate levels during much of 1998-99 but skyrocketed during summer and fall 2000. Figure 1 illustrates prices in the California wholesale market from its inception in April 1998 to its collapse in 2000. By the end of 2000, the incumbent utilities were required to purchase power at high wholesale prices and to sell to customers at substantially lower prices. The utilities eventually lost their creditworthiness, the organized market broke down, and the state government was required to step in to purchase power. This paper investigates the nature of competition that led to skyrocketing wholesale prices.

Studies have found empirical evidence that firms in the California market exercised market power. Adopting a methodology developed by Wolfram [1999], Borenstein et al. [2002] simulate a perfectly competitive market from 1998-2000 and compare those prices to actual prices. They find high price-cost margins during the high demand summer months with the margins becoming very large in 2000. Notably, these margins vary significantly over the three years of the market. Prices during summer months of 1999 and 2000 can be partially explained by the contract positions of the various market participants (Bushnell et al. [2004], Bushnell [forthcoming]). Finally, there is strong evidence of quantity withholding by specific generating firms. Joskow and Kahn [2002] extensively analyze several data sources on California electricity generation during summer 2000 and find evidence of the strategic withholding of capacity by some generating firms.[1]

Although there is evidence of some form of market power, there is less understanding of the type of oligopoly pricing that led to the exercise of market power. Price-cost margins vary due to both demand and supply side factors – demand can become more or less elastic or firms can engage

---

[1] Evidence exists of market power in other restructured electricity markets as well including Australia (Wolak [2000]), Pennsylvania-New Jersey-Maryland (Mansur [2003]), New England, (Bushnell and Saravia [2002]), England and Wales (Wolak and Patrick [1997],Sweeting [2001]), New York(Saravia [2003]), Spain (Fabra and Toro [2002]), and Texas (Hortacsu and Puller [2004]).

in a more or less competitive oligopoly pricing game. For example, the rise in price-cost margins from 1999 to 2000 could have resulted from firms behaving less competitively or firms behaving similarly on a less elastic demand function. Several oligopoly pricing models could apply to this market, including models of unilateral market power and tacit collusion. Individual firms are likely to face relatively inelastic residual demand which provides the ability to unilaterally raise prices.[2] In addition, collusion may be facilitated by the daily repetition of the bidding game between a small set of firms with very accurate information about rivals' costs. This paper analyzes the extent to which higher prices resulted from less competitive pricing behavior rather than less elastic demand or higher costs. I test whether firm-level production behavior is more consistent with unilateral market power or tacit collusion.

This paper decomposes the demand and supply side effects that contribute to the variation in price-cost margins over time. I use hourly firm-level data on output and marginal cost and show that each of the five large generating firms withhold output when price exceeds marginal cost: all of these firms exercise some degree of market power. Next, I compare the prices to two benchmark models of competition – competitive pricing and joint monopoly pricing. I model the market as five large strategic producers competing against other firms who are either relatively small or do not face strong incentives to influence the price. I estimate the supply function of the "competitive fringe" producers and calculate the residual demand for the five large players. Given the estimated residual demand and data on firm-level costs, I calculate prices under competitive pricing and joint monopoly pricing. I find that actual prices are between the two benchmarks.

Such prices can arise from various oligopoly pricing models, including efficient tacit collusion or static non-cooperative behavior. Empirically distinguishing between these modes of behavior has important policy implications, yet is problematic in practice.[3] This paper develops an empirical methodology that separately identifies tacit collusion from static pricing under assumptions about the information structure and the choice variables of the firm. I find that pricing was approximately Cournot for the major firms during much of the 1998-2000 period. Point estimates of the conduct

---

[2] Wolak [2003b] finds that individual firms in California faced residual demand in the real-time market that created potential for substantial unilateral market power.

[3] Corts [1999] demonstrates that standard approaches to estimating firm conduct yield inconsistent estimates if firms are engaged in efficient collusion below joint monopoly prices.

parameter are higher in late 2000, but I do not find evidence of tacit collusion.

Section 2 describes the structure of the California electricity market. Section 3 find the prices were between competitive and joint monopoly levels and motivates the empirical strategy. Section 4 reviews the theoretical models of static and dynamic pricing and specifies a general behavioral model that incorporates static and dynamic market power models as special cases. Section 5 describes the data. Section 6 estimates the behavioral model using a panel of firm-level data. Section 7 concludes.

2 How The California Electricity Market Worked

The electricity industry consists of generation, transmission, and distribution. Historically, these three sectors have been vertically integrated with government regulation of price, entry, and investment. The major producers in California were three 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. Under deregulation, electricity still moves from the generator to the socket the same as always, but the ownership of the infrastructure has changed.

Beginning in the 1990s, policymakers in some countries began to separate the generation side of the industry from the transmission and distribution sectors, and allow firms to compete to supply electrical energy to the network.[4] In California, the restructured market opened in April 1998. The three incumbent utilities divested most of their fossil-fueled powerplants to five private firms that bid into daily auctions to supply power. Southern California Edison divested the vast majority of its plants within a month and a half of the market opening to four different firms: AES-Williams, Dynegy, Reliant, and Thermo Ecotek. Pacific Gas & Electric divested its low cost units to Duke in July 1998 and most of the remaining units to Southern Energy (later spunoff as Mirant) in April 1999. San Diego Gas & Electric divested its plants to Dynegy and Duke in April and May 1999. By the end of the divestiture process, the thermal (fossil-fueled) generation market consisted of roughly

---

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

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 consists of 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. Electricity is also imported from neighboring states in virtually all hours. This paper analyzes the competitive behavior of the five large firms.

California established several institutions to organize the trading of electricity. The three incumbent utilities were still responsible for procuring power for customers in their service territories. The utilities purchased their electricity from a specific day-ahead trading exchange called the Power Exchange (PX). Each day the PX conducted a uniform-price auction for the following day's production. Each firm bidding to supply power submitted an upward-sloping supply schedule for each hour while purchasers (primarily the incumbent utilities) bid downward-sloping demand schedules. The PX aggregated these hourly supply and demand bids to determine the market-clearing price at which all trades were settled. If trades resulted in violation of the transmission line capacity, additional bids were used to ensure that there was no transmission congestion. If the original bids did not cause transmission congestion, PX prices for all of California were identical. If the original schedule was adjusted to meet transmission constraints, PX prices would differ in northern and southern California. During the sample period about 80-90% of all production was sold through the PX, 10% was sold through bilateral trades, and the remainder was sold through hourly real-time "balancing" markets conducted by the grid operator (the Independent System Operator or ISO). The real-time market was also a uniform price auction.

Potential regulatory intervention may have influenced pricing behavior. Market designers set price caps to limit the exercise of market power. The ISO set price caps at $250/MWh until September 1999, raised the caps to $750/MWh in October 1999, but then lowered it to $500 in July 2000 and $250 in August 2000.[5] In addition, state and federal regulators had oversight authority and the ability to impose price caps or order refunds for trades occurring at prices that

---

[5] Although the PX had a higher price cap than the ISO, we expect the cap in the last market to set the effective cap because demand should never bid higher into the PX than the cap in the ISO (Borenstein et al. [2002]).

are not “just and reasonable”. This potential regulatory threat was greatest during the summer and fall of 2000.

The institutions of the California market appear conducive to either unilateral market power or tacit collusion. Because electricity storage is prohibitively costly, firms must produce a quantity equal to demand at all times. Individual firms are likely to face residual demand functions that are inelastic. Inelastic residual demand is rooted in part in a nearly *perfectly* inelastic total demand because consumers do not face hourly wholesale prices. Hence, any elasticity in residual demand arises from elastic supply by other firms. However, other firms are likely to have inelastic supply during periods of high demand. When demand reaches levels near the industry's capacity, if one firm were to withhold capacity to drive up the price, other firms have limited ability to increase output. Hence, firms in this relatively concentrated market are able to raise prices unilaterally.

Repeated interaction also could lead to increased prices through a dynamic pricing game. In general, tacit collusion is facilitated by frequent interaction, up-to-date information on rivals' behavior, and barriers to entry. The California market contains five large firms and a competitive fringe who interact daily in a market where rivals' costs are nearly common knowledge.[6] Firms also observe a great deal of information related to their rivals' competitive behavior. The website of the western U.S. transmission grid coordinator posted real-time generation data for almost all plants until October 2000. Also, the ISO released with a one-day lag each plant's generation that was sold into the real-time market. Several electronic trading exchanges provided electricity traders with the means to observe a record of recent bilateral trades. Demand-side information is also common knowledge; firms observed the ISO's forecast of demand before bidding and the ex post realization of demand. Finally, entry into the market is difficult due to strict environmental siting requirements that can often take more than five years.[7] Hence, the underlying market conditions create a significant potential for restricting output and increasing prices through a variety of oligopoly pricing games.

---

[6] All plants owned by the firms were formerly owned by regulated utilities and are still subject to environmental regulations that make operating characteristics a part of the public record. In particular, firms have good estimates of the fixed and variable costs of rivals' operations with the only uncertainty being whether a plant has had a short-term outage.

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

3 Motivation for Empirical Model

Theory bounds prices between perfectly competitive levels and joint monopoly prices, but a variety of pricing models yield intermediate levels of prices. In this section, I estimate competitive and monopoly prices, and find that actual prices in the PX lie between the two theoretical benchmarks. This suggests that an empirical model to test pricing behavior must incorporate the possibilities of both static oligopoly pricing as well as imperfect collusion.

I estimate the hourly residual demand function faced by the five large firms and use data on the hourly marginal cost of production to compute competitive and joint monopoly prices. I postpone details on the estimation to section 6, but present results here to motivate the structural model. Assume the five large firms face a competitive fringe that supplies at marginal cost. Total residual demand of the five strategic firms (Q^D_{strat}) is the total (perfectly inelastic) market demand (Q^D_{total}) net of supply by the competitive fringe (Q^S_{fringe}):

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

This is diagrammed in Figure 2. The hourly residual demand function faced by the strategic firms is calculated by substracting estimated fringe supply from data on total demand.

The data are described more completely in section 5. Briefly, to estimate the fringe supply function, I use data on hourly PX prices and output by the fringe as well as information on fuel input prices, weather, and seasonal variation in hydroelectric and nuclear generation. Based on an assumed functional form of fringe supply $Q^S_{fringe}(p)$, I estimate the fringe supply function for each hour. The choice of functional form is critical and I choose a constant price elasticity model to allow for the shape of fringe supply that electricity market analysts believe is appropriate for this market.[8] To test for sensitivity to functional form, I also estimate fringe supply that is quadratic

---

[8] Based upon the efficiency of the generating units owned by the fringe, the fringe supply should be relatively flat for many levels of output because fossil-fueled units have similar marginal costs. However, supply should become steeper for less fuel efficient peaker units that come online during high demand periods. In addition, much of the fringe is hydroelectric energy which has a marginal cost equal to the opportunity cost of spilling the water at some later hour. Because the opportunity cost is the price in future hours, the fringe supply should gradually rise at prices that are expected to occur in the future, including some very high prices expected during peak hours. A constant price elasticity model allows for such a fringe supply function.

in price and reaches a maximum at a price of \$750.

Given the estimated $Q^S_{fringe_t}(p)$ and data on $Q^D_{total_t}$, I calculate the hourly residual demand function of the five strategic firms. Then, using data on the marginal cost of each strategic firm's generating units that are operating in that hour, I construct a joint marginal cost function for strategic firms.[9] For each hour, I compute the perfectly competitive price by finding where marginal cost intersects residual demand, and calculate the joint monopoly price by finding where marginal revenue equals marginal cost.

Figure 3 shows the monthly average prices in the PX compared to competitive and joint monopoly prices. Actual prices are always between competitive and joint monopoly prices. During the moderately priced years of 1998-1999, prices were very close to competitive levels during the low demand winter months. There was little potential for market power during these periods, as can be seen by the relatively low joint monopoly prices, because demand for electricity from the five large firms is relatively elastic during the winter. During the higher demand summer and fall, actual prices rose to approximately \$10-15 above marginal cost. However, joint monopoly prices under both specifications of fringe supply are substantially higher indicating that residual demand is less elastic, but the five firms are falling far short of perfectly collusive prices. After June 2000, competitive prices rise as input costs increase, but actual prices still average \$20-\$70 above marginal cost. However, the high prices seen in late 2000 are substantially lower than joint monoply prices under either specification.[10]

These computations suggest that an empirical strategy to estimate the pricing behavior must incorporate the possibility of pricing models that predict outcomes between perfect competition and perfect collusion. In particular, the model must allow for the possibility of dynamic pricing that yields prices below the perfectly collusive prices.

---

[9] For reasons I discuss below, I focus on the time period of 5-6pm each day.

[10] The constant elasticity and quadratic specification of fringe supply lead to different slopes of residual demand at high very high prices and therefore different joint monopoly prices. This results from the fact that the joint monopoly price is very sensitive to the shape of residual demand at very high prices which occur rarely in the sample. This suggests that results may be sensitive to fringe supply specification and I address this below.

4 Distinguishing Between Static and Dynamic Market Power

This section derives a model to test if the production behavior of the five large firms is more consistent with unilateral market power or imperfect tacit collusion. Studies in the New Empirical Industrial Organization (NEIO) literature have estimated firm conduct by parameterizing the firm's static first-order condition (marginal revenue = marginal cost) to allow for price-taking, Cournot competition, and joint monopoly pricing.[11] Unfortunately, this approach is inappropriate if firms are engaged in imperfect collusion and pricing below the joint monopoly level. 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. Suppose that firms are colluding but the incentive for one firm to undercut the joint monopoly price requires all firms to lower the collusive price to eliminate the incentive to deviate. In this circumstance of *efficient tacit collusion*, Corts shows that standard methods typically will underestimate market power.[12]

The root of the problem is that if firms are colluding, the econometrician is estimating the wrong model; one should be estimating the dynamic first-order condition rather than the static first-order condition. The first-order condition of a set of tacitly colluding firms is the solution to maximizing joint profits subject to an incentive compatibility constraint that no firm has an incentive to deviate. As I show below, this dynamic first-order condition is very similar to the static condition with an additional term if firms are engaging in a level of collusion less than perfect collusion (i.e. the joint monopoly outcome). If firms are engaging in imperfect collusion, the static first-order condition used in standard methods is mis-specified and one obtains inconsistent estimates of firm conduct. As a result, the best one can achieve by estimating the parameterized static first-order condition is to test non-nested hypotheses of perfect competition, Cournot competition, and perfect collusion (see Gasmi et al. [1992] and Nevo [2001]). The existing empirical literature does not to my knowledge suggest methods to estimate conduct when one possible conduct is imperfect collusion.

---

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

[12] More precisely, he finds that this approach can severely mismeasure the conduct parameter if the true underlying process is not identical on the margin to a conjectural variations game. Comparisons of direct measures of the conduct parameter versus the NEIO estimates have found NEIO methods to understate market power (see Genesove and Mullin [1998] and Wolfram [1999]). In addition, Kim and Knittel [2004] find that traditional NEIO estimates using aggregate data are sensitive to functional form and can mismeasure elasticity-adjusted Lerner indices for the California electricity market.

In this section, I derive empirically tractable models of firm supply relations under both static and dynamic pricing. A full structural model of competition in the California market would incorporate price determination and strategic incentives in both the forward market and the two sequential auctions in which firms bid supply functions.[13] In addition, the model would include the potential threat by state and federal regulators to intervene in the market, adjust price caps, and order refunds. Unfortuntely, data are not available and the empirical specification would require restrictive assumptions to make such estimation tractable.

I model the firms' strategic decisionmaking as a simple quantity-setting game against a competitive fringe. The five firms are assumed to observe both demand and rivals' marginal costs before making their output choice. These assumptions appear to be plausible given data available to market participants (e.g. accurate demand forecasts and website data on rivals' real-time generation). I derive a general first-order condition that represents each firm's supply relation under no market power, Cournot pricing, and efficient tacit collusion. By making reasonable assumptions about the functional form of fringe supply, the model allows me to consistently estimate conduct parameters in a manner that addresses the Corts critique. In Section 6, I use firm-level data to estimate each firm's supply relation and test if the estimated relation is consistent with static or dynamic pricing.

4.1 Static Pricing Model

In static models, firms choose single period quantities or prices to maximize profits without any intertemporal considerations of the effect of current behavior on the future competitive environment. For the California electricity market, a purely price-setting model is not appropriate because capacity constraints prevent any single firm from undercutting and supplying the entire market. Alternatively, a model incorporating capacity constraints in which firms bid supply functions clearly resembles how firms bid into the PX. Because I do not have data on the bids but only the market

---

[13] For example, one could model the market as a stage game in which firms simultaneously choose their contract positions and then with imperfect knowledge of their rivals' contract positions, choose supply functions in two sequential uniform-price multiunit auctions. Theoretical models of contracting followed by some form of oligopoly pricing have been developed by Allaz and Vila [1993] and applied to electricity markets in Newbery [1998]. Models of optimal bidding into a uniform-price electricity auction conditional on contract positions have been empirically estimated in Wolak [2000], Wolak [2003a] and Hortacsu and Puller [2004]

equilibrium quantities, I cannot estimate a supply function model.[14] Therefore, I estimate models of games in which firms choose quantities. The Cournot model is an upper bound of prices supported by a supply function equilibrium so I can sign any bias of the estimates, as I discuss in section 6.[15]

To formalize the model, assume that N firms play a quantity game in which they choose to supply a given (perfectly inelastic) quantity subject to a capacity constraint.[16] Price is determined such that supply equals demand. Denote $P_t(\cdot)$ as inverse demand, $C_{it}(q_{it})$ as firm i's cost of electricity generation in period t, q_{it} as individual firm quantity, and k_{it} as firm capacity. In period t, firm i chooses quantity of output to maximize profit subject to a capacity constraint:

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

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

$$P(q_{it}^* + q_{-it}) - c_{it}(q_{it}^*) + \theta_{it} \cdot P_t' \cdot q_{it}^* - \lambda_{it}^* = 0 \tag{1}$$

where $c_{it}(q_{it})$ is marginal cost and $\theta_{it} \equiv \frac{dQ_t^*}{dq_{it}} = 1 + \sum_{j \neq i} \frac{\partial q_{jt}}{\partial q_{it}}$ is the firm's belief about the effect of increasing its output on total industry output. 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 so one must be cautious about making behavioral interpretations of θ. Nevertheless, θ as a continuous variable is a meaningful index of the general (anti-)competitiveness of the market. Solving for the

---

[14] Klemperer and Meyer [1989] and Green and Newbery [1992] have analytical models of supply function equilibria. See Wolfram [1998], Wolak and Patrick [1997], Wolak [2000, 2003a,b], Sweeting [2001], and Hortacsu and Puller [2004] for empirical analyses of electricity auctions using bid data. An analysis of the California market using bid data is complicated by the fact that generators bid into both the day-ahead and real-time market.

[15] Modeling the market as a single settlement quantity-setting game is a reasonable approximation to the California market because the vast majority of transactions occur in the PX day-ahead market. Although a small fraction of the firms' output was contracted forward, industry analysts suggest forward contracting was relatively small until 2000. I discuss the sign of the potential biases on my estimates in section 6.

[16] 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.

conduct parameter, one finds:

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

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

4.2 Dynamic Pricing Model

Models of dynamic games show that firms repeatedly interacting in an industry with entry barriers can sustain prices higher than one-shot equilibrium levels. Firms that engage in efficient tacit collusion choose production to maximize joint profits subject to the constraint that no firm has an incentive to deviate in order to earn higher one-time profits at the risk of starting a "price war". There are two major classes of tacit collusion models, and the classification hinges on the information available to firms. 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 that trigger price wars. However, if demand and prices are observed ex post, firms always can sustain the collusive regime but the level of collusion will depend upon current and expected future demand (Rotemberg and Saloner [1986], Haltiwanger and Harrington [1991]) and whether firms face capacity constraints (Brock and Scheinkman [1985], Staiger and Wolak [1992]). For example, if current demand is high, the incentives to cut price and earn deviation profits are high, so price must be lowered to check that incentive. Similarly, if demand is expected to rise in the near future, the future collusive profits may be higher, so firms have less incentive to deviate and start a price war today. As a result, for a given level of demand, higher collusive prices can be sustained when demand is rising than when demand is falling. However, these results in general will differ when firms face capacity constraints that affect both deviation and punishment profits.

The major difference between the Green/Porter and Rotemberg/Saloner (and extensions) models stems from what firms are able to observe about their competitors' behavior. Firms in the Rotemberg/Saloner model observe the prices charged by all other firms. Green and Porter firms know only some signal correlated with rival behavior such as market price or their own realized shares. Because firms in the California electricity market have substantial information on their rivals' behavior, I assume a full-information environment similar to the Rotemberg/Saloner model.

Formally, I model the firm optimization problem when the industry is engaged in efficient tacit collusion. The firms choose a joint quantity Q_t^* to maximize joint profits subject to the constraint that no firm has an incentive to deviate. Deviation from the collusive quantity is punished by permanent reversion to a lower profit punishment outcome such as Cournot or price-taking.[17] Assume that demand and cost shocks are observed ex post so that deviating from the collusive regime can be distinguished from exogenous shocks to the environment. Assume that firms are symmetric and that sharing rules specify that each firm produces $\frac{1}{N}$ of the total output.[18] Due to symmetry, maximizing individual firm profit is equivalent to maximizing joint profit.

Denote individual firm i profit as π_i. π_{is}^* is optimal collusive profit in future period s. Let $\pi_i^{br}(Q_t)$ represent the individual profit to any firm that unilaterally deviates from the collusive regime by producing its one-shot best response to the collusive quantities of the other firms. Deviation will be punished by reversion to noncollusive "punishment" profit π_i^p. $E_t[\pi_{is}]$ denotes expectations of future period s profit conditional on information known in period t. Finally δ is the discount factor between periods. Under efficient tacit collusion, firms maximize joint profit subject to the constraint that no firm has an incentive to deviate from the collusive regime:[19]

$$\max_{Q_t} \quad \sum_{i=1}^{N} \pi_{it}\left(\frac{Q_t}{N}\right)$$

$$s.t. \quad \pi_{it}^{br}(Q_t) + \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi_{is}^p] \leq \pi_{it}\left(\frac{Q_t}{N}\right) + \sum_{s=t+1}^{\infty} \delta^{s-t} E_t[\pi_{is}^*] \qquad \forall i$$

A firm will not deviate if current and continuation collusive profits exceed the profits of deviating

---

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

[18] 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. I discuss the effect of asymmetries below.

[19] 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.

in the current period and earning noncollusive (e.g. Cournot) profits forever afterwards. The optimization problem can be rewritten as:

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

The first-order condition w.r.t. Q_t is:

$$(1 + \frac{\mu^*_t}{N}) \cdot [P(Q^*_t) - c_{it}(\frac{Q^*_t}{N}) + P'_t Q^*_t] - \mu^*_t \frac{d\pi^{br}}{dQ}_t = 0 \tag{3}$$

The first-order condition has a natural interpretation. When the incentive compatibility constraint is not binding ($\mu^*_t = 0$), this equation is equivalent to the static joint profit maximization condition of a monopolist. This corresponds to perfect collusion. However, when the constraint is binding, joint ($MR - MC$) must be lowered (Q raised) so the incentive compatibility constraint is not violated. Firms can still earn more than one-shot Cournot prices but cannot sustain the joint monopoly price.

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

$$P(Q^*_t) - c_{it}(q^*_{it}) + N \cdot P'_t \cdot q^*_{it} - \frac{\mu^*_t}{1 + \frac{\mu^*_t}{N}} \frac{d\pi^{br}}{dQ}_t = 0 \tag{4}$$

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

4.3 A General (Static and Dynamic) Supply Relation

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

$$P(q_{it}^{*}+q_{-it})-c_{it}(q_{it}^{*})-\lambda_{it}^{*}=-\theta_{it}P_{t}'q_{it}+\frac{\mu_{t}^{*}}{1+\frac{\mu_{t}^{*}}{N}}\frac{d\pi^{br}}{dQ}_{t} \tag{5}$$

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

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

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

We can view (5) as a general model capturing three alternative explanations for price above marginal cost. First, observed price-cost margins may represent scarcity rents for new production capacity in a perfectly competitive environment ($\lambda_{it}^*>0$). Firms utilize all capacity with marginal cost less than the price, and margins signal the value of added capacity. Second, margins may result from firms unilaterally withholding current capacity to raise the price and earn higher revenue on their *own* inframarginal units. This corresponds to a model of Cournot competition with capacity constraints. Finally, firms may be jointly withholding capacity to raise the price on *joint* inframarginal units, with this regime maintained by adjusting quantity so that no firm has an incentive to deviate from joint profit maximization.[20] The shapes of these supply relations are illustrated in Figure 4. The vertical axis is the price-cost margin adjusted for scarcity rents and the horizontal axis is the effect of selling one more unit of output on inframarginal revenue. Price-taking firms sell all economical capacity at marginal cost independent of the size of inframarginal sales. Cournot firms drive price above marginal cost when their sales are large and when demand is steep. Finally, tacitly colluding firms internalize the effects of all firms' inframarginal revenue but may have to

---

[20] 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.

adjust price downwards to ensure the incentive compatibility constraint is not violated. I measure the behavior of firms in the California electricity market by estimating whether the actual supply relations are more consistent with the theorized supply relations under no market power, unilateral market power, or efficient tacit collusion.

Before moving to the estimation, it is important to consider how asymmetries across firms affect the empirical model. As shown in Table 1, the five strategic firms vary in capacity from 2871 to 3921 MW and output varies across firms as well. The static model allows for heterogeneity in both cost structure and pricing behavior, so asymmetries do not complicate the estimation.

However, the dynamic model has made the simplifying assumption that firms are symmetric in costs and equally split output under joint profit maximization. Joint profit maximization under asymmetric costs is more complicated to model in a manner that can be taken to data. However, I can assess qualitatively how it would affect the empirical results. The complication is modeling how the joint output is split between the asymmetric firms because in general firms may use complex sharing rules. Take a simple sharing rule such as each firm producing $\frac{1}{5}$ of the output even if marginal costs differ. Consider the effect on the estimated θ_{it} in equation (5). For simplicity, suppose the incentive compatibility constraint is not binding ($\mu_t^* = 0$), so that firms are equally dividing the joint monopoly output. A firm with a higher marginal cost function would have a lower price cost margin, yet the same inframarginal revenue term ($P_t' q_{it}$), so the estimated θ_{it} is lower than N. Similarly, a low cost firm would have a θ_{it} greater than N. Therefore, it is possible that efficient tacit collusion under asymmetric costs could yield estimates of θ_{it} where some firms have conduct parameters greater than N while others are less than N, but the parameters "on average" equal N. However, for the differences to be significant, either the marginal costs functions must significantly differ or the sharing rule must be particularly imbalanced. I discuss the possibility of tacit collusion under asymmetric costs in the results section.

5 Data

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

5.1 Measures of Price, Marginal Cost and Output

Data on the hourly production of each powerplant are available from EPA's Continuous Emissions Monitoring System (CEMS). CEMS contains data on the hourly operation status and MW output of fossil-fueled generation units in California. I can reliably calculate the hourly marginal cost for each generating unit because the production technology is fairly homogenous, and data are available on the technological capacity and fuel efficiency of almost all units owned by each of the five firms. Marginal cost is the sum of marginal fuel, emission permit, and variable operating and maintenance costs.[21] Marginal fuel costs are calculated using data on the fuel input cost for each generator combined with data on the average conversion factors between the heat content of the fuel and the electricity output of each generating unit (Kahn et al. [1997]). Several plants in southern California were required to purchase environmental permits for each pound of nitrogen oxides (NOx) emitted so I include the marginal permit cost per MWh of electricity. Variable operating and maintenance costs are from Borenstein et al. [2002]. I assume marginal cost to be constant up to the capacity of the generator. A firm's marginal cost of producing one more MWh of electricity is defined as the marginal cost of the most expensive unit that it is operating and has excess capacity:

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

---

[21] Marginal costs also include the opportunity costs of exporting power to other higher price markets. The potential to export power out-of-state is unlikely to cause me to mismeasure the marginal (opportunity) cost. In-state firms will sell out-of-state if the out-of-state price is greater than the marginal revenue of sales into California. I cannot measure out-of-state prices, however California is virtually never a net exporter during my sample.

I determine if units have excess capacity by comparing observed output from the CEMS data to the unit's capacity. My measure of marginal cost does not include shadow costs of intertemporal adjustment constraints on the rate at which powerplants can increase or decrease output. Therefore, I focus on the period from 5-6pm (hour 18) when those constraints are unlikely to bind. On an average day the total demand nears its peak by 11am and maintains approximately that level until around 9pm. Natural gas powerplants in California can typically ramp from zero to full capacity in times varying from one to three hours. 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. Therefore, I focus on hour 18 and assume any shadow costs of operating constraints to be zero.[22]

The price earned for the observed output is not always known by the econometrician because the power can be sold in the day-ahead market (the PX) or the real-time energy market (the ISO). I use the PX day-ahead energy price because 80-90% of all transactions occurred in the PX during my sample and a simple arbitrage argument suggests that day-ahead and real-time prices should be equal in expectation.[23] Prices can vary by location. When transmission constraints between the north and south bind, there are essentially two different markets clearing at two different prices. Most firms own powerplants in a single transmission zone so I use the PX zonal price.[24]

My measure of output is the total production by each firm's generating units as reported in the CEMS data. These data are fairly complete but a few qualifications are necessary. I may slightly mismeasure the actual amount of generation sold to the energy market (and hence inframarginal output) for several reasons. There are several higher cost peaker units that operate in high demand periods and do not appear in the CEMS data.[25] Thus, I understate output for the firms owning these unit, but this primarily affects Dynegy. In addition, late in the sample period some firms

---

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

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

[24] One firm (Duke) owns generators in both the north and south. During hours when the north and south have different prices, I separate off output from Duke's southern plants and call the firm DukeSouth.

[25] The percentage of each firm's capacity for which CEMS 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.

sold power through an out-of-state third party to avoid the price cap on in-state purchases. In a practice called "megawatt laundering", generators sold power to third parties on the border of California only to sell the power back to California at prices above the cap. Therefore, potential mismeasurement of inframarginal sales may affect my estimates for Dynegy and for all firms late in the sample period. I discuss the sign of the potential bias when I discuss the empirical results.

5.2 Summary Statistics

The observed production behavior suggests firms are not acting in a perfectly competitive manner during hour 18. A price-taking firm will fully utilize capacity with marginal cost less than the price. When a competitive firm is producing below capacity, one expects the marginal cost of the unused capacity to be above the price. Table 2 displays summary statistics of the difference between price and the marginal cost of each firm with unused capacity in hour 18. 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.[26] These margins imply a median Lerner index of 0.13.[27]

Price-cost margins vary considerably over my sample period of April 1998 to November 2000. I calculate the simple average of all firms' margins in each hour. If it is producing at capacity, the firm's margin is set to zero. Figure 5 shows that margins are higher during the third and fourth quarters of each year when total demand for electricity is high in California. Margins during low demand winter and spring months are actually negative in 1998 and hover around zero in 1999 and

---

[26] 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.

[27] 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. As a check for robustness, I consider the possibility that I may understate firms' marginal costs. Separately, I calculate that firms have excess capacity yet observe margins above \$10 in approximately 38% of firm-hours and greater than \$30 in approximately 22% of firm-hours. It is highly unlikely that marginal costs are this severely mismeasured so there is strong evidence that firms are not acting as price-takers. This conclusion is supported by other studies of the California market including Borenstein et al. [2002] and Joskow and Kahn [2002].

most of 2000.[28] I emphasize that these margins are not scarcity rents because these are differences between price and marginal cost when *firms have excess capacity.* These results are consistent with Borenstein et al. [2002] who find the largest divergences between price and marginal cost during the summer months and particularly in 2000. I the next section, I estimate whether the changes in margins resulted from changes in the residual demand faced by the five large firms or from changes in how those firms competed on their residual demand.

6 Estimation of the Behavioral Model

I estimate if firms supply power in a manner more consistent with static or dynamic market power. I apply the model of firm behavior from section 4 to data and identify parameters of the supply relations that allow me to make inferences about the competitiveness of firms in the California market. Estimated supply relations are compared to the theorized supply relations depicted in Figure 4. First, I estimate the static first-order condition (equation (1)). Conditional on the modeling assumption of a full-information quantity-setting game against a constant elasticity fringe supply, I find the data are consistent with Cournot pricing for most of the sample period. Then, I estimate the general first-order condition (equation (5)) and reject the model of efficient tacit collusion. I discuss the direction of bias if certain assumptions are violated.

6.1 Static Model

In order to estimate the static first-order condition equation (1) for each firm in the California market, I model the supply side as five large strategic firms and a competitive fringe. I estimate how the five firms compete on the residual demand they face.

Demand Side. As diagrammed in Figure 2, total residual demand of the five strategic firms (Q^D_{strat}) is the total (perfectly inelastic) market demand (Q^D_{total}) net of supply by the competitive

---

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

fringe (Q^S_{fringe}): $Q^D_{strat}(p) \equiv Q^D_{total} - Q^S_{fringe}(p)$. I estimate the supply function of the competitive fringe and use the negative of the slope of fringe supply to estimate the slope of inverse residual demand P'_t faced by the strategic firms.

The competitive fringe includes generation from fringe thermal, nuclear, hydroelectric, small independent producers, 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 regulatory side agreements so revenues are independent of the price in the energy market.[29] The owners of hydroelectric assets are the same utilities that are also buyers of power and have very dulled incentives to influence the price. Finally, firms importing power into California are likely to behave competitively because most are utilities with the primary responsibility of serving their native demand and then exporting "excess generation". Borenstein et al. [2002] make similar assumptions about the behavior of firms owning nuclear, hydroelectric, and import generation.

I estimate the supply by all fringe suppliers for hour 18 of each day. Fringe supply is a function of the PX day-ahead electricity price in California as well as cost conditions (e.g. price of natural gas) and seasonal supply variation (e.g. hydroelectric reservoir levels or scheduled nuclear outages). For reasons discussed in section 3, I model fringe supply as having a constant price elasticity so I estimate the model in logs. 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. Because fringe supply includes imports of "excess generation" from neighboring regions to California, I include differences in neighboring state mean daily temperatures from a baseline temperature that requires little heating or cooling (65 degrees). The model is given by:

$$\begin{aligned}\ln Q^S_{fringe\,t} &= \beta_0 + \beta_1 \ln P_t + \beta_2 \ln GasPrSouth_t + \beta_3 \ln GasPrNorth_t + \\ &\quad \beta_4 \ln Diff65TempNeigh_t + \beta_5 DAYDUM_t + \beta_6 MONTHDUM_t + v_t \qquad (6)\end{aligned}$$

The price elasticity β_1 can be used to calculate the slope of fringe supply which is the same

---

[29] 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.

magnitude but opposite sign of the slope of the residual demand faced by the five strategic firms.

Supply Side. In order to estimate the supply relation (1) by the five strategic firms, I need measurements of price, marginal cost, output, and the shadow value of capacity. The data on price, marginal cost, and output are described above.[30] However, I cannot measure the shadow value of additional capacity (λ^*_{it}). The shadow value is zero when capacity constraints are not binding, however the value is unobserved when constraints are binding. The shadow values vary by both firm and time, however adding a separate parameter for each firm-hour when a firm is at capacity would add excessive parameters to the model. Therefore, I add to each supply relation a single dummy variable ($CAPBIND$) equal to 1 if capacity constraints are binding and equal to zero otherwise. The coefficient on $CAPBIND$ is the *average* shadow value of added capacity.

The static first order condition (1) is in general overparameterized because it allows each firm to have a different behavioral parameter each period. Before examining the possibility of heterogeneous behavior across firms, I assume all firms are strategically choosing quantity in the same manner and restrict the conduct parameter to be equal across all firms in the industry. The supply relation is modeled as:

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

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

$$(P - c)_{it} = \lambda \cdot CAPBIND_{it} + \frac{\theta}{\beta_1} \frac{P_t \cdot q_{it}}{Q^S_{fringe\,t}} + \epsilon_{it} \tag{7}$$

The θ parameter can be identified by substituting a consistent estimate of β_1 from the demand side.

---

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

The econometric errors ϵ_{it} and v_t represent shocks to marginal cost observed by the firm. For example, suppose that after the PX price has been determined one-day ahead, an unanticipated weather shock increases the total demand for electricity. The ISO real-time prices will rise above the PX price and firms will sell more output and have a higher marginal cost than they would if the PX price (i.e. the price measure used in the model) had prevailed in the real-time market as well. Due to the correlation between actual output and the econometric error, I instrument output with the day-ahead forecast of total (perfectly inelastic) demand.[31] I simultaneously estimate the system of fringe supply (6) and each firm's supply relation (7) via the generalized method of moments. The error term in each supply relation is modeled as heteroskedastic, contemporaneously correlated with the errors in the other supply relations, and serially correlated with its own error for the past 7 days.

It is important to emphasize that the estimates of the conduct parameter are conditional on the assumed functional form of fringe supply. The slope of (inverse) residual demand is the negative slope of fringe supply. If the estimated fringe (inverse) supply is flatter than the true fringe supply, firms are experiencing the same measured price-cost margin for a residual demand that is steeper than I measure it to be. Therefore, I would overestimate θ. This can be seen in equation (2) – if P_t' is biased downwards, the estimated θ is biased upwards.[32] I experiment with two functional forms to assess for sensitivity – constant elasticity and quadratic. I plot fitted values for various days and visually compare the estimated shapes to the expected shape of marginal cost of generation from nuclear, hydroelectric, thermal and imports. For prices below \$50, the slope is sensitive to functional form, so I choose a constant elasticity specification because it better matches the typical marginal cost function of electricity generators. For higher prices, the average slopes are very similar under both specifications. Perhaps most importantly, the slope of residual demand for the

---

[31] Although not ideal, this instrument appears reasonably valid. For my sample, the day-ahead forecast error is not correlated with the forecast except at high levels of forecasted demand. $CAPBIND$ is potentially endogenous to ϵ_{it} if large demand shocks increase the real-time price and induce firms to produce at capacity. To test if this affects my conduct parameter estimates, I estimate the conduct parameter θ using only observations when the capacity constraints are not binding ($\lambda_{it}^* = 0$). Firms are producing at capacity in only 4.7% of firm-hours in my dataset. The results are similar to those below and are reported in Appendix B.

[32] As seen in Figure 3, the functional form of fringe supply yields different joint monopoly prices. This occurs because the slope of fringe supply at *very high* prices is sensitive to functional form. However, at prices closer to those observed in the PX, this is not always the case.

range of prices observed in the market is not sensitive to fringe supply specification during the interesting price runup period of June-November 2000.

Before showing formal estimation results, I illustrate the shape of the supply relation by strategic firms. Note that when $CAPBIND$ is zero (95% of firm-hours), the static model reduces to a simple bivariate (instrumental variables) regression: $(P-c)_{it} = \theta(-P_t')q_{it} + \epsilon_{it}$. I graph the data in two dimensions and compare to the theorized supply relations of Figure 4. Figure 6 plots the price-cost margins against the fitted values of $-P_t'q_{it}$ so that the slope of this relationship is the estimated conduct parameter. The static model implies that if behavior (θ) is constant, then the margins are linear in $-P_t'q_{it}$: firms have higher margins when (1) they have more inframarginal quantity and (2) they are operating on less price sensitive areas of demand. The top panel of Figure 6 plots the kernel regression estimate and the data for the complete sample of July 1998-November 2000.[33] The relationship has roughly a constant slope up to approximately $-P_t'q_{it} = 25$ which comprise 70% of firm-hours. The relationship approximately passes through the origin as one expects from a conjectural variations game. 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 after June 2000. This suggests the market may have been less competitive after June 2000.[34]

Next, I show formal results from jointly estimating the system of fringe supply (6) and the strategic firm supply relations (7).[35] I break the sample into a period during which there were four firms in the market from July 1998-April 1999 and another period with five firms from April 1999-November 2000. Results are shown in Table 3 and are similar for both time periods. Fringe supply is relatively inelastic in both periods (0.18 and 0.19). Given the relative size of the fringe

---

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

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

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

and total demand, this suggests that the strategic firms face a total residual demand elasticity of approximately -2.38 during the 4-firm period and -0.97 during the 5-firm period.[36] Therefore, identical competitive behavior would lead to higher price-cost margins in the later period.

The estimates of the strategic firm supply are consistent with Cournot pricing in both periods. 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 a $\hat{\theta} = 0.97$ that is not statistically different from unity. For the period from mid-April 1999 to November 2000, the results are similar and I obtain an identical conduct parameter $\hat{\theta} = 0.97$ statistically indistinguishable from unity. Although pricing behavior is similar in both periods, margins were higher during the 5-firm period when strategic firms faced less elastic residual demand. Finally, the estimates of λ imply that firms operating at capacity are willing to pay \$21.52 and \$41.20 on average for an additional MW of capacity during the 4-firm and 5-firm periods, respectively.

Finally, for the period of the price runup of 2000, I can decompose the high price-cost margins averaging \$74.05/MWh into "demand side" and "supply side" factors. Margins were higher because firms faced very inelastic residual demand (-0.69) due to an unusually hot summer that increased demand in California and reduced imports from western states. Also, low levels of snowfall the previous winter reduced hydroelectric imports from the northwest. These results imply that the state was not only increasingly dependent upon the strategic firms' generation, but the strategic firms may have supplied less competitively. The estimated conduct parameter of $\hat{\theta} = 1.54$ confirms the apparent pivot in the supply relation after June 2000 in Figure 6. Although this finding suggests less competitive pricing, I withhold any behavioral interpretation until presenting the results from the dynamic model.

The restriction that all firms behave identically may hide important differences across firms. The five large generators may use different pricing strategies in the first few years of this new market. Alternatively, some subset of the firms could be engaged in tacit collusion. Finally, any firm-specific misspecification or mismeasurement could significantly bias the overall conduct parameter if θ is

---

[36] The coefficients on other explanatory variables in the fringe supply equation are consistent with theory. Note that the coefficients on gas prices are of the opposite sign which is to be expected given the strong collinearity. In unreported regressions, I restrict the coefficients to be equal and they are negative and statistically significant as expected.

restricted to be equal across firms. In fact, the models with equal θs are overidentified and J-tests reject the overidentifying restrictions imposed by the structural models and functional form assumptions. One possible cause is that the conduct parameter varies by firm.

Therefore, I re-estimate equations (6) and (7) allowing each firm to have a different conduct parameter (θ) and shadow value of capacity (λ). Conduct parameter estimates are reported in Table 4. I find a modest degree of heterogeneity across firms, but generally fail to reject $\theta = 1$. During both the periods with four and five firms in the market, AES and Duke have lower point estimates than Southern and Reliant. But with a few exceptions, I fail to reject pricing that is consistent with Cournot behavior. Dynegy has a particularly large conduct parameter estimate during the four firm market that decreases but remains high in the second part of the sample. These high conduct parameter estimates may result from the fact that I have incomplete quantity data for some of Dynegy's small peaker units.[37] When I focus on the period of the price runup in June-November 2000, firms' conduct parameter estimates are almost uniformly larger. Dynegy (with data caveats) and Southern have conduct parameter estimates statistically larger than unity while AES, Duke, and Reliant's parameters are consistent with Cournot pricing. I discuss behavioral interpretations of these estimates in section 6.3. In Appendix B, I report conduct estimates under alternative assumptions.

6.2 Dynamic Model

Results from the static behavioral model are consistent with static pricing over much of the 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 test 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 does not suggest efficient tacit collusion.

---

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

Corts shows that conduct parameter estimates are consistent if and only if the true underlying game is 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 conjectural variations game: the supply relation appears to be a ray through the origin. If this were not the case, firms may be engaged in some other (non-conjectural variations) game and the static model would not yield consistent estimates of conduct.

Formally, I estimate the general first-order condition (equation (5)) to test for dynamic pricing. The dynamic first-order condition is the first-order condition of a joint monopolist with an added term capturing the additional output (relative to joint monopoly output) required to prevent any individual firm from deviating and earning best-response profits. Although I do not have data on it, this extra term is constant across all firms during a given period so I condition it out by including time fixed-effects. Including fixed effects for *every* time period (i.e. every day) would add many parameters and absorb much of the variation in margins. Instead, I assume that the level or "bindingness" of the incentive compatibility constraint is equal for similar periods of time, and I add fixed effects for those time periods.

The level of the incentive compatibility constraint depends upon several factors. First, the size of residual demand relative to individual firm capacity determines incentives to deviate and earn best response profits (Staiger and Wolak [1992]). Individual firm capacity does not change substantially during the months exhibiting market power because nearly all plants are online. Therefore, I essentially need to capture the size of the residual demand function. I observe the quantity produced in equilibrium on the residual demand function, but this quantity is endogenous to price. To mitigate this endogeneity, I use quartiles of equilibrium residual demand to capture these effects on the incentive compatibility constraint. Second, the level of the incentive compatibility constraint varies over the year depending upon whether demand is expected to rise or fall in coming months (Haltiwanger and Harrington [1991]). I include month fixed effects to incorporate seasonal adjustment in the incentive to deviate. These month and demand quartile fixed effects capture the incentive compatibility adjustments ($\hat{IC}_t$) depicted in Figure 4.

I estimate the dynamic model just as I have estimated the static model (7) above, except that I add month and demand quartile fixed effects to the strategic firm supply relations to capture the incentive compatibility term. If the firms are efficiently tacitly colluding, this specification yields a consistent estimate of the conduct parameter $\theta = N$. However, if the firms are following the static pricing model, the estimators of $\hat{\theta}$ from both the static model (7) and this more general model (5) are both consistent. Hence, Cournot pricing should yield $\hat{\theta}$ statistically indistinguishable from unity in this more general specification. Table 5 shows estimated conduct parameters for different specifications of the dynamic model. I find the estimated conduct parameters to be significantly lower than $N =$ 2,3,4, or 5 in all specifications. I reject efficient tacit collusion in all periods.

Because the model is overidentified, I report the outcome of J-tests of over-identifying restrictions implied by the structural model and functional form of fringe supply. I reject the models that include only demand quartile fixed effects. However, the data do not reject the models that include month fixed effects. If only month effects are included, I obtain conduct parameter estimates close to those obtained in the static model of section 6.1.[38] For the four firm period, both the static and general model fail to reject Cournot pricing with conduct parameters of 0.97 and 1.07, respectively. In the five firm period, the static model yields $\hat{\theta} = 0.97$ while the general model yields $\hat{\theta} = 1.28$.

The period of the price runup from June-November 2000 is particularly important to analyze because there are widespread allegations of price manipulation and my estimation of the static model yielded conduct parameters greater than unity. Results are reported in Table 6. The general model yields results similar to the static model. Estimated conduct parameters are above Cournot levels but substantially below levels implied by efficient tacit collusion. These results provide strong evidence against *efficient* tacit collusion.

6.3 Interpretation

The static model yields conduct parameter estimates that are consistent with Cournot pricing *on average* during the 4-firm period of July 1998-April 1999 and the 5-firm period of April 1999-

---

[38] During the four firm period, I obtain conduct parameter estimates close to zero for specifications including demand quartile effects and near one for the specification including only month effects. Apparently, margins vary by total residual demand but do not vary by individual firm output after controlling for total residual demand.

November 2000. The more general empirical specification that allows for efficient tacit collusion in addition to Cournot pricing yields similar results. When I allow for firms to have individual conduct parameters, I typically fail to reject Cournot pricing. This suggests that the large variation in both prices (Figure 1) and price-cost margins (Figure 5) was not driven by changes in the oligopoly pricing game but rather by changes in the size of residual demand faced by the five large firms with incentives to exercise market power.

Perhaps the most interesting period to analyze is the second half of 2000 when skyrocketing prices added substantial debt to the incumbent utilities. The dynamic model easily rejects *efficient* tacit collusion by two or more of the firms. In other words, firms were pricing far below the maximum prices that could be sustained subject to the incentive compatibility constraint. However, the empirical model does not test for *inefficient* tacit collusion or another form of dynamic pricing that sustains prices between the Cournot and efficient collusive prices. In fact, both the static and general model yield estimates of the conduct parameter in the 1.5 to 1.7 range. Several models are consistent with this result but are not testable within my empirical framework. For example, these conduct parameter estimates are consistent with a static game with an evolving set of beliefs about the slope of fringe supply or rival behavior. Alternatively, the folk theorem implies that firms can sustain any collusive level of pricing between Cournot and joint monopoly prices. To the extent that the California market is viewed as an infinitely repeated game with a discount factor between days very close to one, any level of pricing behavior between one-shot Cournot and joint monopoly levels can be sustained in equilibrium. Such equilibrium behavior would be measured by a conduct parameter between unity and the number of colluding firms. However, the behavioral interpretation of the conduct parameter would change. An estimate of $1 \leq \hat{\theta}_i \leq N$ does not consistently estimate how rivals' output changes in response to an increase in firm i's output. For example, an increase in firm i's output could trigger the higher output of the "punishment" regime, but $\hat{\theta}_i$ is not a consistent estimate of this regime shift. Rather $\hat{\theta}_i$ estimates behavior *in equilibrium*.[39] Nevertheless, even if firms engaged in a form of dynamic pricing, the economic significance and subsequent welfare loss is not substantially larger than that of Cournot pricing.

---

[39] For a good discussion of interpreting (and misinterpreting) θ_i, see Reiss and Wolak [2002].

When I allow for heterogeneous conduct parameters from June-November 2000, I fail to reject Cournot pricing for AES, Duke and Reliant. However, Southern and Dynegy's conduct parameters, $\hat{\theta} = 1.46$ and 2.39 respectively, are statistically higher than unity. I cannot rule out the possibility that these two firms are engaging in some form of tacit collusion between firms with asymmetric cost structures as discussed at the end of section 4.3. However, there are strong reasons to believe that Dynegy's conduct parameter is biased upwards because I do not have data on the output from its peaking units.

It is important to note the sensitivity of the results to modeling assumptions and data. I assume a quantity-setting model whereas firms actually bid more complex supply functions into both the Power Exchange and the ISO's real-time market. To the extent that a supply function model is more realistic, my conduct parameters are biased downwards. The sign of the bias can be easily understood. In a quantity-setting model, a firm's residual demand gets all its slope from the total demand because its rivals are "bidding" vertical supply functions. If its rivals actually bid non-vertical supply functions, the true residual demand obtains elasticity from both total demand *and* rival supply. Because firms in my data are enjoying the same observed price-cost margin on a residual demand function that is flatter that I measure it to be, the θ estimated by my model is biased downwards.

Also, the estimated conduct parameters depend critically on the slope of residual demand which is generated by the functional form of fringe supply. I choose a functional form that fits the typical shape of marginal costs for the fringe generators. It is noteworthy that both functional forms I test yield very similar results for the period of the late 2000 price runup.

Finally, several institutional factors that changed in 2000 deserve special caveats. In late 2000, the utilities began to face financial crises that could prevent them from paying for power purchased on the wholesale market. Because the risk of non-payment may have increased marginal costs of supplying power, my measure of marginal cost may understate the true cost of supplying power and bias upwards my conduct estimates. Analysts believe this is most applicable to November 2000. However, there are several factors that may lead me to understate the true conduct parameter as well. The most severe concern is that firms forward-contracted some of their production and that I overstate the output sold to the PX/ISO energy market. There is widespread belief that in 2000 Duke forward-contracted some of its 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. 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 in the fall did not occur at the PX/ISO prices but a higher prices via "megawatt laundering". Overall, the bias from risk premia is only a concern during the last few weeks of the sample whereas the bias from contracts and out-of-market transactions likely exists for much of the summer and fall. Therefore, my conduct estimates are likely biased downwards in 2000.

7 Conclusions

A variety of states and countries have designed restructured electricity markets so that a large fraction of transactions occur in daily spot markets. These spot markets may appear conducive to tacit collusion due to the repeated nature of the auctions and the high level of information available to market participants. However, I find that the California market does not exhibit evidence of efficient tacit collusion. Nevertheless, there is strong evidence of market power. Price-cost margins varied substantially over time with higher margins during the higher demand third and fourth quarters of each year. I estimate the extent to which high margins resulted from less competitive conduct and/or less elastic demand which affords firms more opportunity to exercise market power. I find that the large variation in price-cost margins was primarily driven by changes in residual demand elasticity. During 1998-2000, I generally fail to reject Cournot pricing for the sample as a whole.

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 slightly less competitive (Figure 6) but the shift was not as dramatic as prices would suggest. Primary factors contributing to price increases were higher input costs and less elastic residual demand. Nevertheless, the five large in-state non-utility generators raised prices slightly above unilateral market power levels in 2000, but fell far short of efficient tacit collusion.

It is difficult to form a specific conclusion about firm behavior in 2000. I reject the hypothesis that all firms were pricing at Cournot levels, but the observed prices are much closer to Cournot prices than efficient collusive prices. The observed prices in 2000 are consistent with a variety of other possible behavior: some other form of dynamic pricing, some average of various non-equilibrium behavior, or a static game with an evolving set of beliefs about the shape of fringe supply or rival behavior. Distinguishing between each possible behavior is a formidable empirical task. For this reason, it would be problematic to use this type of methodology for antitrust purposes.

Nevertheless, two important points do emerge about the market's competitiveness. First, firms in this daily repeated auction fell far short of efficient tacit collusion. Second, whatever the underlying behavior, prices in 2000 were not substantially above the maximum prices sustainable in a full-information static game.

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.[40] 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 or the amount of real-time information

---

[40] In fact, other markets that have required forward contracting or vesting contracts do not exhibit evidence of substantial market power except at large levels of demand (Bushnell et al. [2004]).

made available to market participants. Some work has suggested that collusion is less likely under discriminatory auctions than uniform-price auctions.[41] 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.

---

[41] See Klemperer [2000] and Fabra [2003].

References

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

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

S. Borenstein, J.B. Bushnell, and F.A. Wolak. Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market. American Economic Review, 2002.

T.F. Bresnahan. Empirical Studies of Industries with Market Power. In R. Schmalensee and R.D. Willig, editors, Handbook of Industrial Organization. Elsevier, 1989.

T.F. Bresnahan. Testing and Measurement in Competition Models. In D.M. Kreps and K.F. Wallis, editors, Advances in Economics and Econometrics: Theory and Applications. Cambridge University Press, 1997.

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

J.B. Bushnell. Looking for Trouble: Competition Policy in the U.S. Electricity Industry. In J. Griffin and S.L. Puller, editors, Electricity Deregulation: Choices and Challenges. University of Chicago Press, forthcoming.

J.B. Bushnell, E.T. Mansur, and C. Saravia. Market Structure and Competition: A Cross-Market Analysis of U.S. Electricity Deregulation. CSEM Working Paper 126, 2004.

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

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

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

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

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

N. Fabra. Tacit Collusion in Repeated Auctions: Uniform versus Discriminatory. Journal of Industrial Economics, 2003.

N. Fabra and J. Toro. Price Wars and Collusion in the Spanish Electricity Market. Mimeo, 2002.

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.

R.J. Green and D.M. Newbery. Competition in the British Electricity Spot Market. Journal of Political Economy, 1992.

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.

A. Hortacsu and S.L. Puller. Testing Strategic Models of Firm Behavior in Restructured Electricity Markets: A Case Study of ERCOT. CSEM Working Paper 125, 2004.

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

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

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

P. Joskow and E. Kahn. A Quantitative Analysis of Pricing Behavior In California's Wholesale Electricity Market During Summer 2000. Energy Journal, 2002.

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

D. Kim and C.R. Knittel. Biases in Static Oligopoly Models?: Evidence from the California Electricity Market. CSEM Working Paper 131, 2004.

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

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

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

E.T. Mansur. Vertical Integration in Restructured Electricity Markets: Measuring Market Efficiency and Firm Conduct. CSEM Working Paper 117, University of California Energy Institute, 2003.

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.

D.M. Newbery. Competition, contracts, and entry in the electricity spot market. RAND Journal of Economics, 1998.

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

P.C. Reiss and F.A. Wolak. Structural Econometric Modeling: Rationales and Examples from Industrial Organization. Mimeo, Stanford, 2002.

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

C. Saravia. Speculative Trading and Market Performance: The Effect of Arbitrageurs on Efficiency and Market Power in the New York Electricity Market. CSEM Working Paper 121, 2003.

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

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

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

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

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

F.A. Wolak. An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market. International Economic Journal, 2000.

F.A. Wolak. Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets. In Advances in economics and econometrics: Theory and applications. 2003a.

F.A. Wolak. Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market, 1998-2000. American Economic Review, 2003b.

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 Market Structure of Fossil-Fueled Generating Units

(54% of Total CA Capacity)

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

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

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

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

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

Notes:

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

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

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

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

	4 Firm Market[†]		**5 Firm Market**[††]		**June-Nov 2000**	
	Fringe	Strategic	Fringe	Strategic	Fringe	Strategic
Dependent Variable:	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$
$\frac{P\cdot q}{Q^s_{fringe}}$	–	5.457	–	5.041	–	5.765
	–	(0.323)	–	(0.228)	–	(0.280)
λ (\$/MW)	–	21.52	–	41.20	–	98.05
	–	(0.95)	–	(6.65)	–	(6.69)
Ln(Price) β_1	0.178	–	0.192	–	0.266	–
	(0.029)	–	(0.020)	–	(0.031)	–
Ln(GasPrSouth)	-0.464	–	-0.187	–	-0.367	–
	(0.172)	–	(0.213)	–	(0.223)	–
Ln(GasPrNorth)	0.081	–	0.067	–	0.242	–
	(0.196)	–	(0.225)	–	(0.238)	–
Ln(Diff65TempNeigh)	0.009	–	-0.023	–	0.001	–
	(0.027)	–	(0.013)	–	(0.036)	–
Constant	9.910	–	9.541	–	9.059	–
	(0.106)	–	(0.082)	–	(0.166)	–
						–
Obs.	268		573		163	
$\hat{\eta}^D_{strat}$	-2.38		-0.97		-0.69	
Average Margin	\$10.88		\$26.96		\$74.05	
J-statistic p-value	0.00		0.00		0.00	
$\hat{\theta}$	0.97		0.97		1.54	
	(0.16)		(0.09)		(0.18)	

Fringe represents equation (6) and Strategic represents equation (7). Although the system contains a supply relation for each firm, the coefficients are restricted to be equal in this model so I only report one set of parameters for the strategic supply relations. The instruments entering (7) are day-ahead forecast demand and CAPBIND, and the instruments entering (6) are log of day-ahead forecast demand and all regressors except log price. Standard errors, constructed using the optimal GMM weighting matrix, account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). Day and month-year dummies are included in the fringe supply equation but are not reported here. I exclude hours (in 2000) when the price cap is hit. 8% of hour 18 observations in 2000 hit the price cap with the majority occuring in August.

[†] 7/1/98-4/15/99.

[††] 4/16/99-11/30/00.

Table 4: Static Conduct Parameters by Firm for Hour 18

	4 Firm Market[†]		5 Firm Market[††]	
Firm	Estimate	Std Error	Estimate	Std Error
Southern	–	–	1.21	0.11
Reliant	1.48	0.32	1.01	0.09
Duke	1.02	0.18	0.81*	0.08
AES	0.99	0.20	0.82	0.12
Dynegy	5.15*	1.14	1.75*	0.19

	June-November 2000	
Firm	Estimate	Std Error
Southern	1.46*	0.17
Reliant	1.19	0.14
Duke	1.15	0.15
AES	0.96	0.17
Dynegy	2.39*	0.29

Estimates of θ_i from estimation of system of (6) and (7) where each strategic firm supply relation contains a firm-specific parameter for conduct θ_i and shadow value of capacity λ_i. The instruments entering (7) are day-ahead forecast demand and CAPBIND, and the instruments entering (6) are log of day-ahead forecast demand and all regressors except log price. Standard errors, constructed using the optimal GMM weighting matrix, account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). I exclude hours (in 2000) when the price cap is hit. 8% of hour 18 observations in 2000 hit the price cap with the majority occuring in August.

* Reject $H_o : \theta_i = 1$ at 5% level.

[†] 7/1/98-4/15/99

[††] 4/16/99-11/30/00

Table 5: Dynamic Conduct Parameter Estimates

IC_t Specification	4 Firm Market[†]	5 Firm Market[††]	Jun-Nov 2000
Demand Quartile+Month Effects	0.05	1.02	1.56
	(0.02)	(0.06)	(0.07)
	**	**	**
Only Demand Quartile Effects	0.12	1.03	1.47
	(0.02)	(0.09)	(0.20)
Only Month Effects	1.07	1.28	1.69
	(0.17)	(0.07)	(0.09)
	**	**	**

** J-test of overidentifying restrictions *fails* to reject at 5% level.

Table reports the estimates of θ from GMM estimation of fringe supply (6) and strategic supply (7) including fixed effects for the IC constraint term. The standard errors from the GMM estimates account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). I exclude hours (in 2000) when the price cap is hit.

[†] 7/1/98-4/15/99.

[††] 4/16/99-11/30/00.

Table 6: June-November 2000: General Specification for Static and Dynamic Market Power

	Fringe	Strategic
Dependent Variable:	$\ln Q^S_{fringe\,t}$	$(P-c)_{it}$
Ln(Price)	0.205	
	(0.009)	
Ln(GasPrSouth)	-0.331	
	(0.103)	
Ln(GasPrNorth)	0.178	
	(0.109)	
Ln(Diff65TempNeigh)	-0.011	
	(0.015)	
Constant	9.400	
	(0.051)	
$\frac{P\cdot q}{Q^s_{fringe}}$		7.596
		(0.153)
λ (\$/MW)		104.692
		(1.562)
Demand Quartile2		8.714
		(10.839)
Demand Quartile3		-1.777
		(3.237)
Demand Quartile4		6.934
		(4.207)
July		-18.872
		(4.195)
August		-15.889
		(4.469)
September		-43.263
		(3.439)
October		-37.491
		(3.189)
November		-14.801
		(3.625)
θ	1.56	
	(0.07)	
J-statistic probability	0.822	
Obs.	163	

Fringe represents equation (6) and Strategic represents equation (7). Although the system contains a supply relation for each firm, the coefficients are restricted to be equal in this model so I only report one set of parameters for the strategic supply relations. The instruments entering (7) are day-ahead forecast demand, CAPBIND, and the month and demand quartile fixed effects. The instruments entering (6) are log of day-ahead forecast demand and all regressors except log price. Standard errors, constructed using the optimal GMM weighting matrix, account for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7). Day and month-year dummies are included in the fringe supply equation but are not reported here. I exclude hours when the price cap is hit. 8% of hour 18 observations in 2000 hit the price cap with the majority occuring in August.

APPENDIX A: Data

The marginal fuel cost for each genearting unit is calculated from daily natural gas spot prices and average heat rates. All of the units for which I have generation data burn natural gas as their primary fuel. I use the daily spot price of natural gas (Natural Gas Intelligence [1998-2000]) for the PG&E Citygate and California-Arizona border hubs plus the distribution cost charged to those units by the natural gas utility (Southern California Gas Company [1998-2000] and Pacific Gas & Electric Company [1998-2000].) Although some firms may have contracted for natural gas at a different price, the spot price is the proper measure of the opportunity cost of fuel. Average heat rates are from datasets collected by the California Energy Commission and Southern California Gas Company. These heat rates also have been used in Borenstein et al. [2002] and Kahn et al. [1997].

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

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

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

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

I need to make several assumptions about a firm's behavior in order to determine the firm's marginal cost of producing one more unit in a given hour. If, on a given hour, I look across all of a firm's generating units, I am likely to see the firm operating a lower marginal cost unit at less than full capacity while also operating another higher marginal cost unit. One explanation is that the firm expects that the higher cost unit will be operating in the coming hours (perhaps when total demand is higher) and it needs to keep the higher cost facility operating. Under this scenario it is unclear whether the proper measure of the firm's marginal cost is the lower or higher cost unit that still has available capacity. If I use the lower cost unit, I ignore the fact that the firm is solving a more complicated dynamic optimization problem and that the true measure of marginal cost should include the shadow values of the operating constraints. If I use the higher cost unit, I ignore that the higher cost unit may be running because it was called under outside reliability contracts by the grid operator. However, given that they turn on the Reliability Must Run (RMR) units to meet RMR contracts, competitive firms should still increase production in these units if marginal cost is lower than price. In practice, the RMR units are not always higher cost units and when they are, the costs are at most a few dollars higher than other units. Because the former bias is potentially more severe, I define the firm's marginal cost to be the marginal cost of the most *expensive* unit that is operating and has excess capacity.

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

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

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

APPENDIX B: Conduct Parameter Estimates Under Alternative Assumptions

4-Firm Market

Firm	(1)	(2)	(3)
All	0.92	1.06	0.24
	(0.13)	(0.18)	(0.05)
Southern	–	–	–
	–	–	–
Reliant	–	–	–
	–	–	–
Duke	–	–	–
	–	–	–
AES	–	–	–
	–	–	–
Dynegy	–	–	1.02
	–	–	(0.22)

5-Firm Market

Firm	(1)	(2)	(3)
All	0.75	0.89	1.19
	(0.07)	(0.09)	(0.03)
Southern	–	–	–
	–	–	–
Reliant	–	–	–
	–	–	–
Duke	–	–	–
	–	–	–
AES	–	–	–
	–	–	–
Dynegy	–	–	2.20
	–	–	(0.05)

June-November 2000

Firm	(1)	(2)	(3)
All	1.53	1.66	1.66
	(0.19)	(0.18)	(0.11)
Southern	–	–	–
	–	–	–
Reliant	–	–	–
	–	–	–
Duke	–	–	–
	–	–	–
AES	–	–	–
	–	–	–
Dynegy	–	–	3.10
	–	–	(0.19)

Definitions of Alternative Specifications:

(1) Robustness to CAPBIND: Static Model, only observations when no firms operate at capacity ($\lambda_{it} = 0$)
(2) Robustness to Definition of Capacity: Static Model, Capacity ==80% Nameplate, conduct restricted across firms
(3) Robustness of Dynamic model to Dynegy: Dynamic Model with demand quartile and month effects, onduct restricted equal across firms except for Dynegy

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

Figure 1: Wholesale Prices in California Electricity Market

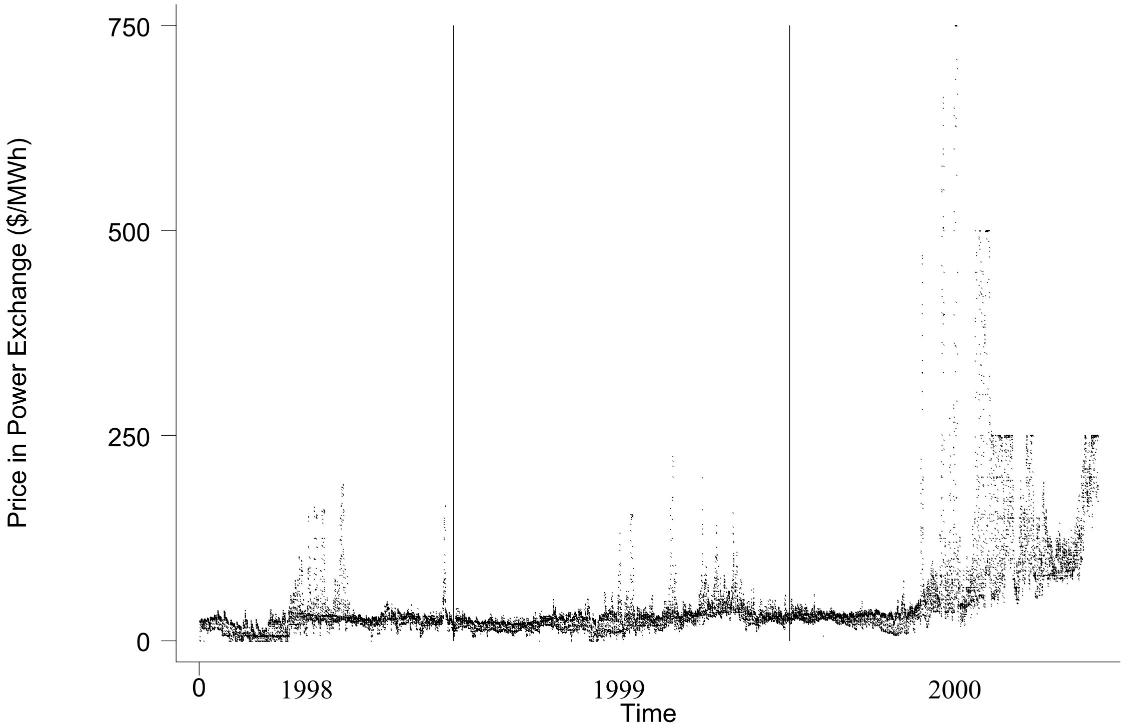


Figure 2: Demand for Electricity from Strategic Firms

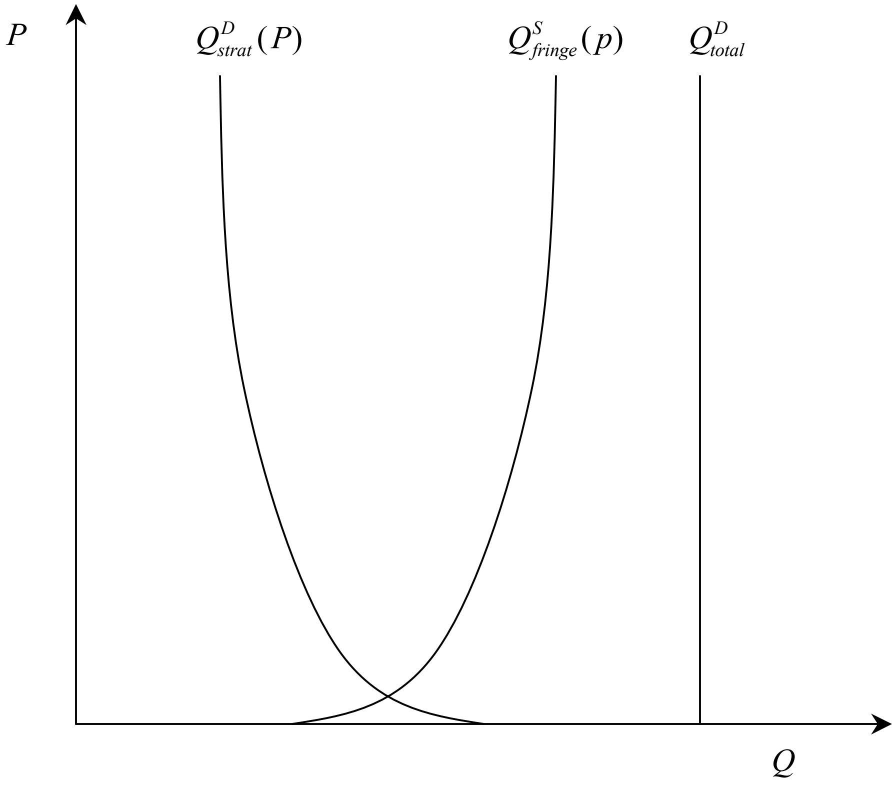
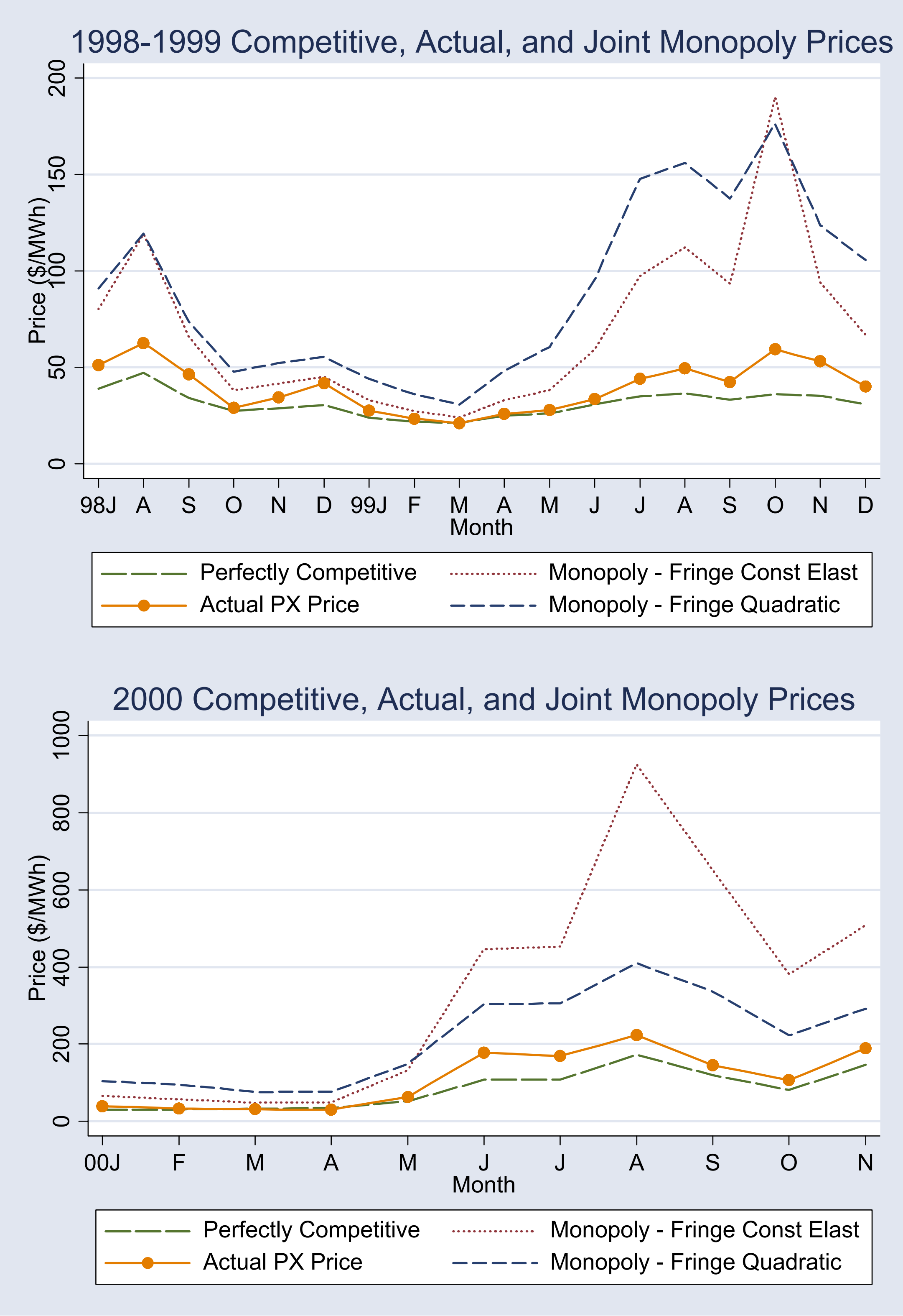
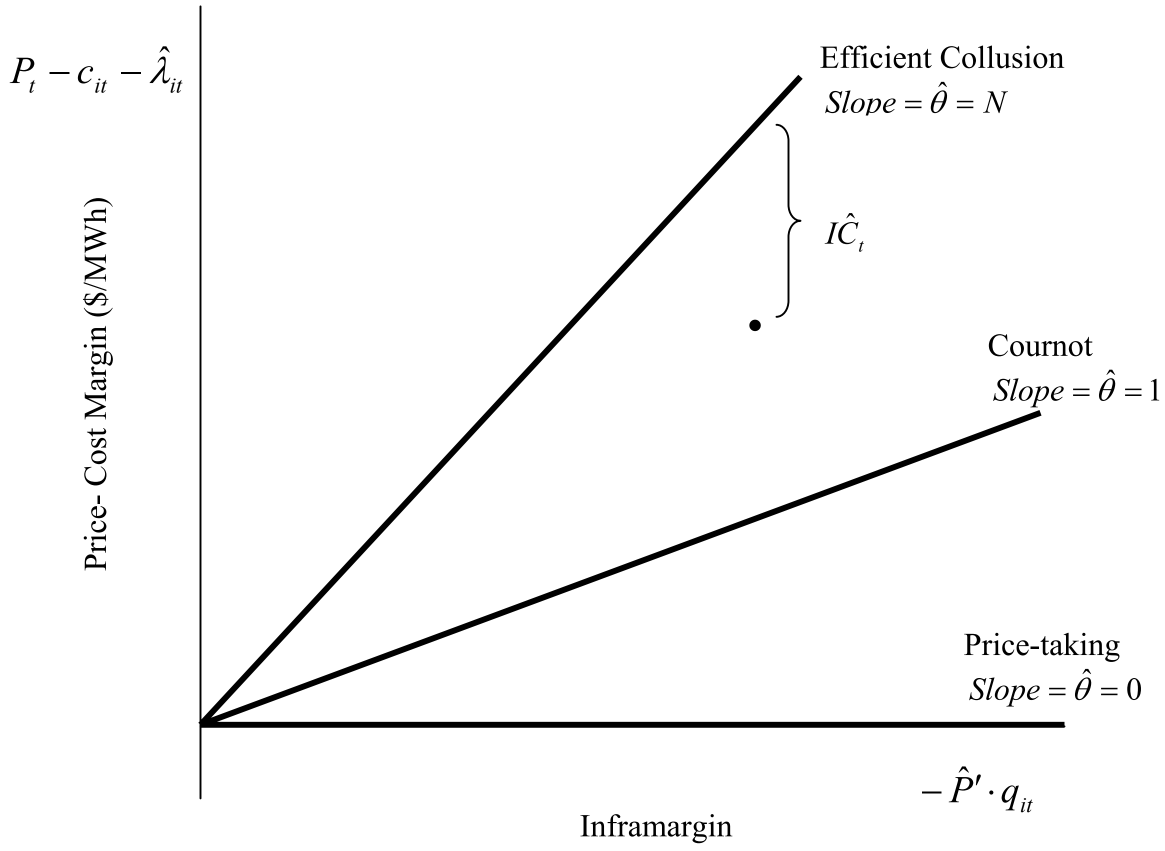


Figure 3: Actual Prices vs. Competitive and Joint Monopoly Prices



Average monthly 5-6pm prices. Actual prices are the unconstrained prices in the Power Exchange. Competitive prices are calculated using marginal cost of all operating units owned by the five strategic firms. Joint monopoly prices are calculated under different functional form assumptions on the shape of fringe supply.

Figure 4: Supply Relations Under No, Static, and Dynamic Market Power



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

Figure 5: Average Price-Cost Margins in Hour 18

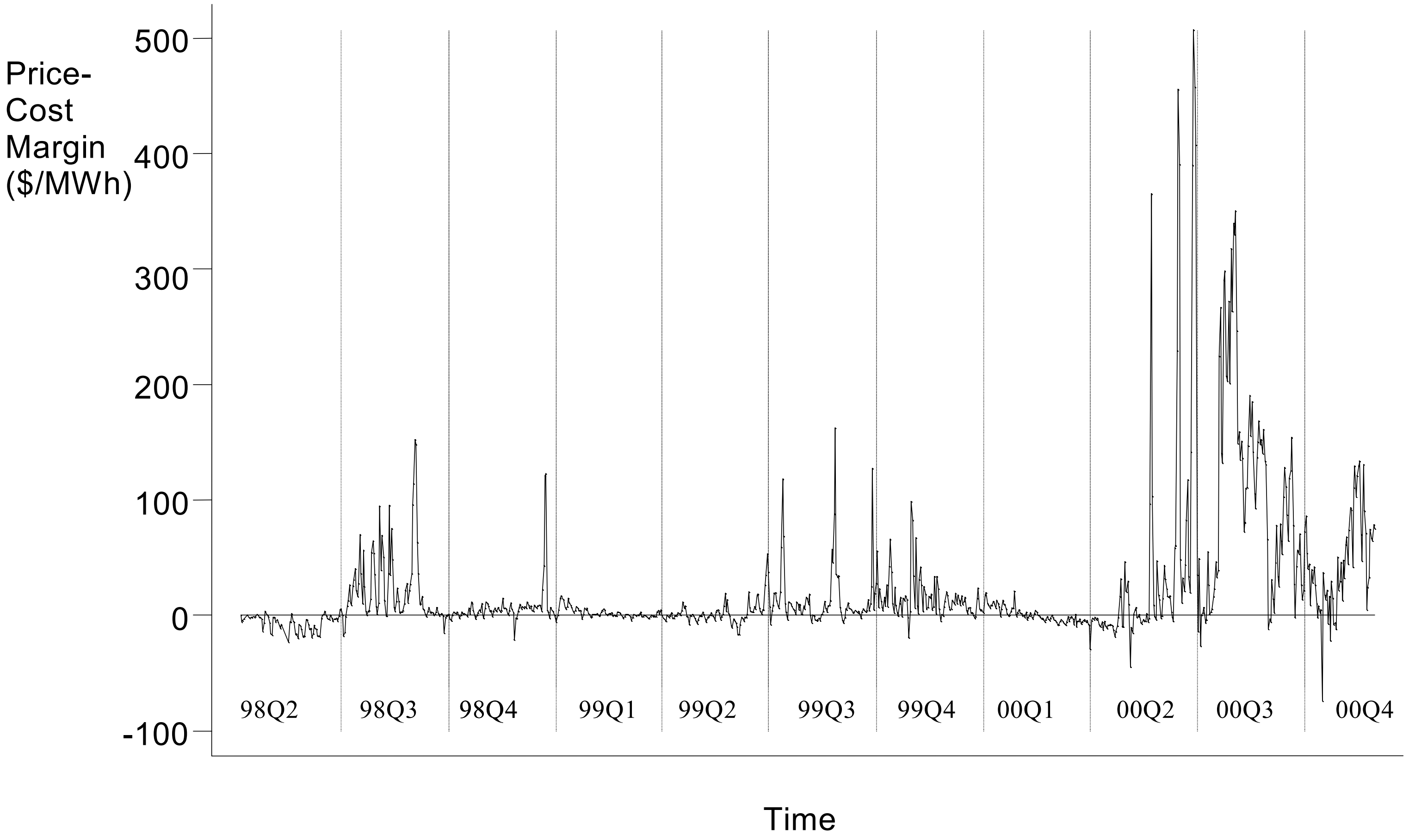
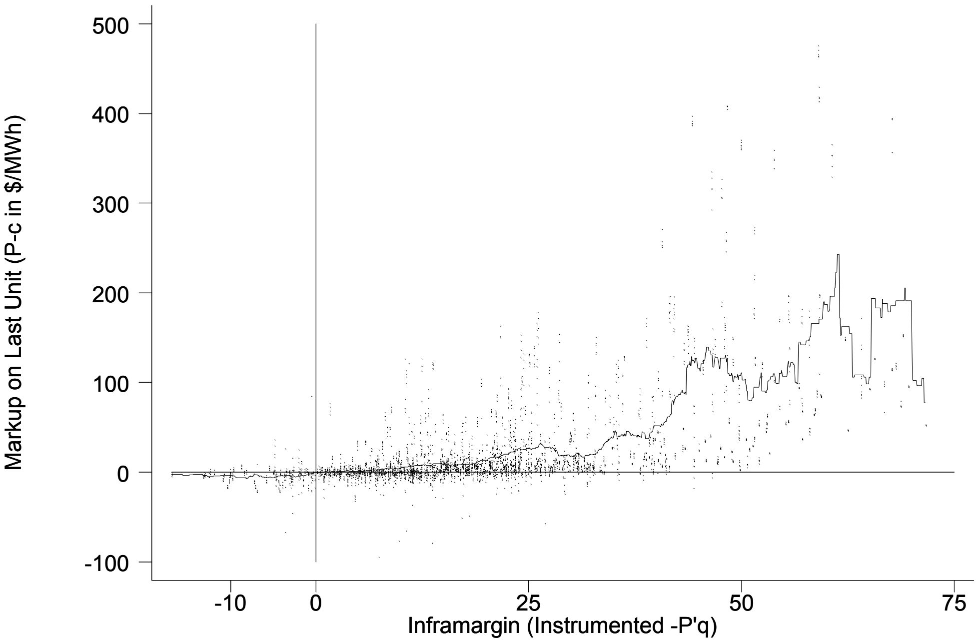
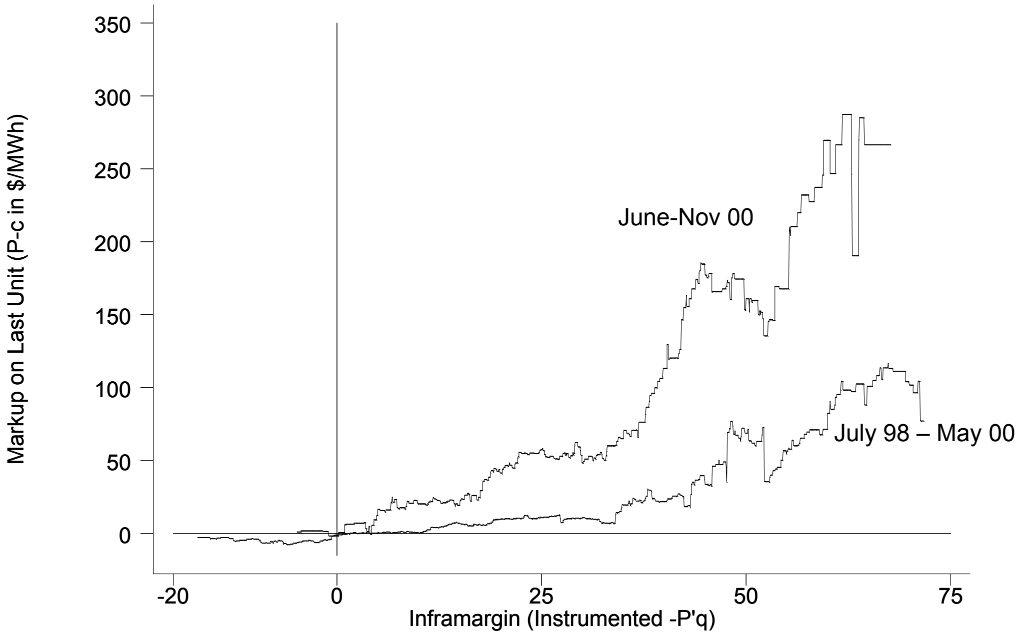


Figure 6: Static Behavioral Model for Hour 18

July 1998-November 2000



Pre and Post June 2000



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