PRICING AND FIRM CONDUCT IN CALIFORNIA'S DEREGULATED ELECTRICITY MARKET

Steven L. Puller*

Abstract—This paper analyzes the pricing behavior of electricity generating firms in the restructured California market from its inception in April 1998 until its collapse in late 2000. Using detailed firm-level data, I find that conduct is fairly 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 nonutility generators raised prices slightly above unilateral market-power levels in 2000, but fell far short of colluding on the joint monopoly price.

I. Introduction

THE restructuring of the electricity industry in California and the subsequent meltdown of the market raised many questions about the feasibility of competitive electricity markets. 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 averaged $31 per megawatt-hour from 1998 to May 2000 but skyrocketed to $141 during summer and fall 2000, with prices in some hours reaching $750. 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 the competition that led to skyrocketing wholesale prices.

Studies have found empirical evidence that firms in the California market exercised market power. Adopting the Wolfram (1999) methodology, Borenstein, Bushnell, and Wolak (2002) simulate a perfectly competitive market from 1998 to 2000 and compare the resulting prices with 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. Higher prices during the summer months of 1999 and 2000 can be partially explained by the smaller forward contract positions of the various market participants (Bushnell, Mansur, & Saravia, 2005; Bushnell, 2005). Finally, there is strong evidence of quantity withholding by specific generating firms in summer 2000 (Joskow & Kahn, 2002).[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 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 were likely to face relatively inelastic residual demand, which allowed them to raise prices unilaterally.[2] In addition, collusion was possible because electricity was traded through daily repeated auctions between a small set of firms with very accurate information about rivals' costs. Understanding the underlying pricing game is important for the optimal design of restructured electricity markets. Depending upon the pricing game, the market designer can change the competitiveness of the market outcome by altering the structure of ownership, the method and frequency of procurement, and the information available to market participants.

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 was more consistent with unilateral market power or tacit collusion. This paper decomposes the demand- and supply-side effects that contributed 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 withheld output when price exceeded marginal cost: all these firms exercised some degree of market power.

Next, I compare the observed prices to simulated prices under three benchmark models of competition—competitive, Cournot, and joint monopoly pricing. I model the market as five large strategic producers competing against other firms that either are relatively small or do not face strong incentives to influence the price. I estimate the supply

Received for publication July 9, 2004. Revision accepted for publication October 3, 2005.

* Texas A&M University.

I thank Severin Borenstein, James Bushnell, Greg Crawford, Carlos Dobkin, Richard Gilbert, James Griffin, Bronwyn Hall, Ali Hortacsu, Edward Kahn, Erin Mansur, Aviv Nevo, Julio Rotemberg, 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] Evidence exists of market power in other restructured electricity markets, including Australia (Wolak, 2000), Pennsylvania–New Jersey–Maryland (Mansur, forthcoming), New England (Bushnell & Saravia, 2002), England and Wales (Wolak & Patrick, 1997; Sweeting, 2005), New York (Saravia, 2003), Spain (Fabra & Toro, 2005), and Texas (Hortacsu & Puller, 2005).

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

The Review of Economics and Statistics, February 2007, 89(1): 75–87

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 simulate prices under competitive, Cournot, and joint monopoly pricing. I find that actual prices are very close to the prices that would result if all five firms acted as Cournot competitors.

The finding that aggregate prices are near Cournot levels is consistent with firms playing Cournot, but also could result from a subset of firms colluding and the remaining firms behaving competitively. For example, two colluding firms and three competitive firms could yield prices similar to five Cournot firms. Understanding pricing at the firm level has important policy implications for mitigating market power. Therefore, I use firm-level production data to test for different levels of competitive behavior by each of the five large firms. I estimate firm-level supply functions and find modest heterogeneity in behavior but generally fail to reject Cournot pricing for the individual firms. Individual firm behavior appears slightly less competitive in the second half of 2000, but I do not find evidence of tacit collusion by a subset of the firms. This suggests that skyrocketing prices in 2000 resulted from higher costs and less elastic residual demand rather than from tacit collusion.

Section II describes the structure of the California electricity market. In section III, I describe my empirical strategy to distinguish between static and collusive pricing. The data are described in section IV. Section V presents the results, and section VI discusses policy implications and concludes.

II. Institutional Structure of the California Electricity Market

The electricity industry is composed of generation, transmission, and distribution. Historically, these three sectors have been vertically integrated, with regulation of price, entry, and investment. The major producers in California were three investor-owned utilities: Pacific Gas & Electric (PG&E) in northern California, Southern California Edison in south central California, and San Diego Gas & Electric in the southernmost part of the state. Beginning in the 1990s, policymakers in some countries began to separate the generation side of the industry from the transmission and distribution sectors and to allow firms to compete to supply electricity to the network.[3] In California, the restructured market opened in April 1998. The three incumbent utilities gradually divested most of their fossil-fueled power plants to five private firms that bid into daily auctions to supply power: AES-Williams, Reliant, Duke, Southern, and Dynegy.[4] By the end of the divestiture process, the fossil-fueled generation market consisted of roughly five equal-sized firms and two small fringe firms that together owned 54% of the electricity generation capacity in California (see table 1). The remaining in-state capacity consisted 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 was also imported from neighboring states in virtually all hours. This paper analyzes the competitive behavior of the five large firms.

Table 1.—Postdivestiture Market Structure of Fossil-Fueled Generating Units (54% of Total California Capacity)

Firm	Capacity (MW)	Percentage of 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. The 46% of capacity that is not included in this table includes nuclear, hydroelectric, and renewable resources owned largely by PG&E and Southern California Edison.

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). The PX conducted a daily 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 such trades could be supported by the transmission grid, the PX prices were identical throughout California. However, if the trades would violate transmission constraints, the market was separated and prices would differ in northern and southern California. Market designers set a price cap to limit the exercise of market power. The cap was set at $250 per megawatt-hour until September 1999, raised to $750 in October 1999, but then lowered to $500 in July 2000 and $250 in August 2000. During the sample period 80%–90% of all production was sold through the PX, approximately 10% was sold through bilateral trades, and the remainder was sold in an hourly real-time balancing auction conducted by the Independent System Operator (ISO).

The institutions of the California market appear conducive to either unilateral market power or tacit collusion.

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

[4] Southern California Edison divested the vast majority of its plants, within a month and a half of the market opening, to AES-Williams, Dynegy, Reliant, and Thermo Ecotek. PG&E divested its low cost units to Duke in July 1998, and most of the remaining units to Southern Energy in April 1999. San Diego Gas & Electric divested its plants to Dynegy and Duke in April and May 1999. Other generation owners operated units within California but outside the California Independent System Operator's territory, including a large municipal utility, the Los Angeles Department of Water & Power.

Because electricity storage is prohibitively costly, firms had to produce a quantity equal to demand at all times. Individual firms were likely to face residual demand functions—rival supply substracted from total demand—that were inelastic. Total demand was nearly *perfectly* inelastic because consumers did not face hourly wholesale prices. Hence, any elasticity in residual demand arose from elastic supply by other firms. However, other firms were 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 had limited ability to increase output. Therefore, despite the fact that this market is not overly concentrated by antitrust standards, firms were able to raise prices unilaterally.

Repeated interaction also could have led to increased prices through a dynamic pricing game. In general, tacit collusion is facilitated by frequently repeated interaction, up-to-date information on rivals' behavior, and barriers to entry. The California market consisted of five large firms and a competitive fringe that interacted daily in a market where rivals' costs were nearly common knowledge. All power plants were formerly owned by regulated utilities and were still subject to environmental regulations that made operating characteristics part of the public record. In particular, firms had good estimates of the fixed and variable costs of rivals' operations. Firms also observed a great deal of information related to their rivals' competitive behavior. The Web site 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 was also common knowledge; firms observed the ISO's forecast of demand before bidding and observed the ex post realization of demand immediately after the market cleared. Finally, entry into the market was difficult because of strict environmental siting processes that could last more than five years.[5] Hence, the underlying market conditions created a significant potential for restricting output and increasing prices through a variety of oligopoly pricing games.

III. Empirical Strategy to Distinguish between Static and Collusive Pricing

I test pricing behavior in two ways. First, I compare actual prices with simulated prices under three benchmark pricing models. 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, Cournot, and joint monopoly prices. Second, I analyze whether the production behavior of each strategic firm is more consistent with a Cournot model or tacit collusion.

A. *Estimating Strategic Firms' Residual Demand*

The first step is to estimate the hourly demand function of the five strategic firms. Assume the five large firms face a competitive fringe that supplies at marginal cost. The total residual demand of the five strategic firms $[Q^D_{\mathrm{strat}}(p)_t]$ in hour t is the total (perfectly inelastic) market demand $(Q^D_{\mathrm{total}\ t})$ net of supply by the competitive fringe $[Q^S_{\mathrm{fringe}}(p)_t]$:

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

The competitive fringe includes generation from nuclear, hydroelectric, and small independent producers, and imports from outside California. I assume that these suppliers did not bid strategically and model them as a competitive fringe. This assumption appears reasonable. The independent and nuclear units were paid under regulatory side agreements, so their revenues were independent of the price in the energy market. The owners of hydroelectric assets were the same utilities that were also buyers of power and had very dulled incentives to influence the price. Finally, firms importing power into California were likely to behave competitively because most were utilities with the primary responsibility of serving their native demand and then simply exporting any excess generation.[6]

I estimate the supply function by all fringe suppliers for 5–6 p.m. of each day. The fringe supply is a function of hourly PX prices, input prices, weather, and seasonal variation in hydroelectric and nuclear production. Based on an assumed functional form of the fringe supply $[Q^S_{\mathrm{fringe}}(p)]$, 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.[7] I discuss robustness to functional specification below.

To incorporate input cost variation over time, I include the daily price of natural gas as well as month-year and day-of-week dummy variables to capture reservoir levels and nuclear outages. Because the fringe supply includes imports of excess generation from neighboring regions to California, I include differences in neighboring-states' mean

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

[6] Borenstein et al. (2002) make similar assumptions about the behavior of firms owning nuclear, hydroelectric, and import generation.

[7] Based upon the efficiency of the generating units owned by the fringe, the fringe supply should be nearly flat for many levels of output, because fossil-fueled units have similar marginal costs. However, the 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.

daily temperatures from a baseline temperature that requires little heating or cooling (65°F). The model is given by

$$\begin{aligned}\ln Q^S_{\text{fringe } t} &= \beta_0 + \beta_1 \ln P_t + \beta_2 \ln \mathit{GasPrSouth}_t \\ &\quad + \beta_3 \ln \mathit{GasPrNorth}_t \\ &\quad + \beta_4 \ln \mathit{Diff65TempNeigh}_t + \beta_5 \mathit{DAYDUM}_t \\ &\quad + \beta_6 \mathit{MONTHDUM}_t + \upsilon_t.\end{aligned} \tag{1}$$

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

B. *Comparing Observed Prices with Benchmarks for Competitive, Cournot, and Joint Monopoly Pricing*

Given the estimated $Q^S_{\text{fringe}}(p)_t$ and data on $Q^D_{\text{total } t}$, I calculate the hourly residual demand function of the five strategic firms, $Q^D_{\text{strat}}(p)_t$. Then, using data on the marginal cost function of each strategic firm's generating units that are operating in that hour, I compute benchmark prices under alternative pricing regimes. The hourly competitive benchmark is the price at which the five firms' hourly aggregate marginal cost intersects $Q^D_{\text{strat}}(p)_t$. The joint monopoly benchmark is computed as the price that maximizes joint profits given the hourly $Q^D_{\text{strat}}(p)_t$ and aggregate marginal cost function. These two benchmarks bound the prices that can arise from an oligopoly pricing game in this market.

Also, I compute the hourly Cournot equilibrium price for each hour's $Q^D_{\text{strat}}(p)$ and each firm's hourly (asymmetric) marginal cost function. To solve for the equilibrium price, I start with each firm producing one-fifth of the hourly five-firm residual demand function at an intermediate level price. Then I solve numerically for each firm's best-response quantity (where marginal revenue equals marginal cost), assuming that all other firms' quantities are held constant. This process is repeated for each firm and then iterated until no firm has an incentive to change output. If the simulated price exceeds the price cap, the price is set at the cap.[8]

C. *Estimating Firm-Level Pricing*

The simulated benchmark prices allow me to compare the *average* level of behavior of the five strategic firms relative to benchmark prices from oligopoly pricing models. However, measuring the average level of pricing does not necessarily identify the individual firm-level behavior. Market prices under five-firm Cournot pricing could be consistent with other pricing models, such as two colluding firms and three competitive firms. Moreover, one should not necessarily expect identical behavior by firms in a newly restructured electricity market—the five large generators might use different pricing strategies in the first few years of this new market. Hortacsu and Puller (2005) find that firms in the new Texas electricity market exhibit a wide range of bidding behavior that demonstrates varying levels of strategic sophistication. Therefore, I test pricing individually for each of the five large firms. In this section I derive an empirically tractable model of firm-level pricing to test if behavior is more consistent with competitive pricing, Cournot, or tacit collusion.[9]

I model competition as firms choosing hourly quantities to supply to the 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 equilibrium quantities, I cannot estimate a supply-function model. However, 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 VC.[10] Denote by $P_t(\cdot)$ the inverse demand in period t, by $C_{it}(q_{it})$ firm i's cost of electricity generation, by q_{it} the firm quantity, and by k_{it} the firm capacity. N strategic firms are assumed to observe both demand and rivals' marginal costs before choosing output. These assumptions appear to be plausible, because market participants had access to accurate demand forecasts and Web site data on rivals' real-time generation.

Firm i chooses quantity of output in period t to maximize profit subject to a capacity constraint:

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

The first-order condition characterizing an interior solution at the optimal quantity q^*_{it} is

$$P(q^*_{it} + q_{-it}) - c_{it}(q^*_{it}) + \theta_{it} P'_t q^*_{it} - \lambda^*_{it} = 0, \tag{2}$$

[8] A similar procedure is used in Borenstein and Bushnell (1999). Because residual demand is a smooth function, this process is unlikely to suffer from the complication of multiple equilibria. Note that these simulations assume that the entire California market is integrated and transmission constraints between the north and south do not bind. This assumption holds for 80% of the hours in my sample.

[9] 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. 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. Unfortunately, data are not available, and the empirical specification would require restrictive assumptions to make such estimation tractable.

[10] Klemperer and Meyer (1989) and Green and Newbery (1992) have analytical models of supply-function equilibria. Industry analysts note that firms often submit supply schedules that resemble hockey sticks, with low prices for most output and high prices for the last few units of output. Such bidding can lead to low prices if a firm defects from the collusive price (Klemperer, 2002). A quantity-setting model is an extreme form of hockey stick bidding with only a vertical section. See Wolfram (1998), Wolak and Patrick (1997), Wolak (2000, 2003a,b), Sweeting (2005), and Hortacsu and Puller (2005) 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 markets and may engage in arbitrage across markets.

where $c_{it}(q_{it}^*)$ is the marginal cost and $\theta_{it} \equiv dQ_t^*/dq_{it} = 1 + \sum_{j \neq i} \delta q_{jt}/\delta q_{it}$ 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 joint monopoly pricing (under symmetry), respectively. There is 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.

This model makes clear how price-cost margins respond to changes in the strategic firm residual demand $Q_{\text{strat}}^D(p)$. Equation (2) is easily transformed into the Lerner index:

$$\frac{P(q_{it}^* + q_{-it}) - c_{it}(q_{it}^*) - \lambda_{it}^*}{P(q_{it}^* + q_{-it})} = \theta_{it} \frac{s_{it}}{\eta_{\text{strat } t}^D},$$

where $\eta_{\text{strat } t}^D = (dQ_{\text{strat}}^D(p)_t/dP)P_t/Q_{\text{strat}}^D(p)_t$ and $s_{it} = q_{it}/Q_{\text{strat}}^D(p)_t$. Conduct parameters associated with less competitive forms of behavior raise margins. In addition, increasing residual demand raises margins. To see this, consider the effect of increasing total electricity demand and holding constant the fringe supply. Because it is perfectly inelastic, an increase in total demand shifts residual demand parallel to the right. As a result, residual demand is less elastic [$dQ_{\text{strat}}^D(p)_t/dP$ is unchanged and $Q_{\text{strat}}^D(p)_t$ is larger for any given price], and price-cost margins are higher.

Equation (2) captures three alternative explanations for pricing above marginal cost. First, observed price-cost margins may represent scarcity rents for new production capacity in a perfectly competitive environment ($\theta_{it} = 0$ and $\lambda_{it}^* > 0$). Firms utilize all capacity that has a marginal cost less than the price, and margins signal the value of added capacity. Second, margins may reflect firms unilaterally withholding output to raise the price and earn higher revenue on their own inframarginal units. This corresponds to a model of Cournot competition with capacity constraints ($\theta_{it} = 1$ and $\lambda_{it}^* \geq 0$). Finally, firms may be jointly withholding output to raise the price on *joint* inframarginal units to achieve joint profit maximization ($\theta_{it} = N$ and $\lambda_{it}^* \geq 0$).

This approach of parameterizing the first-order condition of a static game to infer conduct may be inappropriate if firms are engaging in imperfect collusion and are pricing below the joint monopoly level. Corts (1999) shows that traditional approaches to estimate conduct from the parameterized static first-order condition can lead to inconsistent estimates of the conduct parameter θ. The root of the problem is that if firms are colluding, the researcher is estimating the wrong model; the researcher should be estimating the first-order condition of a dynamic game rather than a static game. 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 (for example, see Rotemberg & Saloner, 1986). As shown in Puller (2006), this dynamic first-order condition is very similar to equation (2), with an additional term if the incentive compatibility constraint is binding and firms collude at a price less than the joint monopoly price. If firms are engaging in imperfect collusion, the static first-order condition is misspecified and one obtains inconsistent estimates of firm conduct. Puller (2006) derives and estimates a more general model that uses firm-level data to compute consistent estimates of the conduct parameter in a manner that addresses the Corts critique. The results from estimating the more general model for the California market yield estimates very similar to the static model derived above.

IV. Data

I use data on hourly market price as well as each firm's hourly output and marginal cost function. Restructured electricity markets are subject to data reporting requirements that provide the researcher with rich data on the demand, cost structure, and output. I describe an overview of the data in the main text and include details in the Appendix.

Data on the hourly production of each power plant are available from the EPA's Continuous Emissions Monitoring System (CEMS). CEMS contains data on the hourly operation status and power output of fossil-fueled generation units in California. I can reliably calculate the hourly marginal cost for each generating unit because data are available on the technological capacity and fuel efficiency of almost all units owned by the five firms. The marginal cost is the sum of marginal fuel, emission permit, and variable operating and maintenance costs. The marginal fuel cost is calculated using data on the daily fuel input cost and each unit's average conversion factor between the heat content of the fuel and electricity output (Kahn, Bailey, & Pando 1997). Several plants in southern California were required to purchase environmental permits for each pound of nitrogen oxides (NO_x) emitted, so I include the marginal permit cost per megawatt-hour of electricity. Variable operating and maintenance costs are from Borenstein et al. (2002). I assume the marginal cost function to be constant up to the capacity of the generator.

A firm's marginal cost of producing one more megawatt-hour of electricity is defined as the marginal cost of the most expensive unit that it is operating and that has excess capacity. I determine if units have excess capacity by comparing observed output from the CEMS data with the unit's capacity. I cannot measure the shadow costs of intertemporal adjustment constraints on the rate at which power plants can increase or decrease output. Therefore, I focus on an hour of sustained peak demand from 5 to 6 p.m. (hour 18) each day, when those constraints are unlikely to bind.[11]

[11] On an average day the total demand nears its peak by 11 a.m. and maintains approximately that level until around 9 p.m. Natural-gas power plants in California can typically ramp from 0 to full capacity in 1 to 3 hours. By the time 6 p.m. 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 that demand begins to fall. Therefore, I focus on hour 18 and assume any shadow costs of operating constraints to be 0. Note that price-cost margins are higher on average in high-demand hours than off-peak hours. However, higher margins do not imply less competitive

TABLE 2.—HOUR 18 PRICE-COST MARGINS WHEN FIRMS ARE NOT AT CAPACITY

Firm	% hours not at capacity	Price-Cost Margin ($/MWh) Mean	Median	St. Dev.	Min	Max	Median Lerner
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

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 that is operating and has excess capacity.
Notes:
1. The Lerner index $\equiv$ (*price* $-$ *MC*)/*price* is presented as a general measure of market power.
2. 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.
3. The large negative margin for AES represents a single day in which a unit was operating but in the process of starting up so that the emission costs were high.

The exact price earned for observed output is not known by the researcher, because it could have been 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, and a simple arbitrage argument suggests that day-ahead and real-time prices should be equal in expectation.[12] Prices vary by location when transmission constraints between the north and south bind. Most firms own power plants in a single transmission zone, so I use the PX zonal price.[13]

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 the 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.[14] Thus, I understate output for the firms owning these units, but this primarily affects Dynegy. In addition, late in the sample period some firms sold power through an out-of-state third party to avoid the price cap on in-state purchases. In this practice, called *megawatt laundering*, generators sold power to third parties on the border of California only to have them sell the power back to California at prices above the cap. Therefore, potential mismeasurement of inframarginal sales may affect my estimates for Dynegy and for all firms late in the sample period. I discuss the sign of the potential bias in section VC.

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 when the marginal cost is 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 marginal cost for each firm with unused capacity in hour 18. Firms very often observed a price above marginal cost, yet fail to utilize their capacity. DukeSouth, Duke, and Reliant fully utilize 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. These margins imply a median Lerner index of 0.13.[15]

Price-cost margins vary considerably over my sample period of April 1998 through 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 0. Figure 1 shows that margins are higher during the third and fourth quarters of each year, when total demand for electricity is high in California. Margins during low-demand winter and spring months are actually negative in 1998 and hover around 0 in 1999 and most of 2000.[16] I emphasize that these margins are not scarcity rents, because they are differences between price and marginal cost *when firms have excess*

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.

[12] See Borenstein et al. (2005) for an analysis of the PX-ISO arbitrage condition over time.

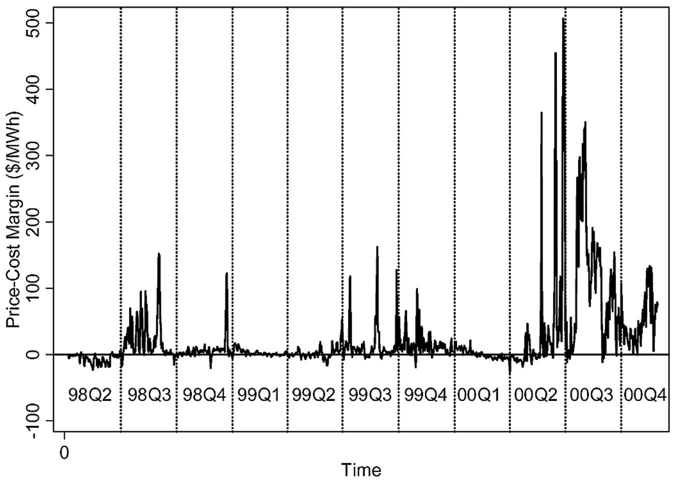
[13] One firm (Duke) owns generators in both the north and south. During hours when the north and south have different prices, I separate output from Duke's southern plants and call the firm DukeSouth.

[14] The percentages of the firms' 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.

[15] The margins are not interpreted as measures of profitability, because firms incur other ongoing 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 start-up costs yet fail to utilize capacity when price is above marginal cost. I perform various checks for robustness. First, I find that average firm-level margins in other peak hours to be very similar. Also, 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 above $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 pricetakers. This conclusion is supported by other studies of the California market, including Borenstein et al. (2002) and Joskow and Kahn (2002).

[16] 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 & Wolak, 1999). This became less of an issue over time as the original RMR contracts were amended.

FIGURE 1.—AVERAGE PRICE-COST MARGINS IN HOUR 18



Note: Figure represents the average price-cost margin across strategic firms. When a firm is operating at capacity, the margin is set equal to 0. During several hours in 2000, an AES generating unit had an unusually large NO_x emission rate, causing the AES margin to be significantly negative despite all other firms having positive margins. These AES observations are used in the remaining analysis, but are excluded from this calculation.

capacity. In 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.

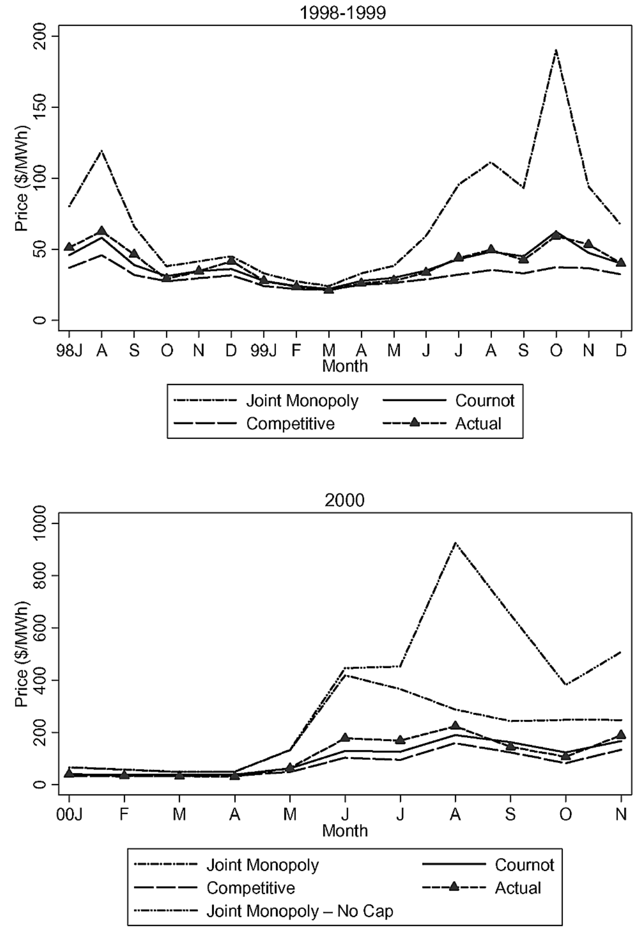
V. Results

A. *Actual Prices Compared with Benchmark Prices*

Figure 2 shows the monthly average actual prices compared with simulated competitive, Cournot, and joint monopoly prices.[17] During the moderate-price years 1998–1999, competitive prices vary slightly and tend to be higher during the high-demand summer months. In the summer there is potential to exercise market power as exhibited by the relatively large joint monopoly and Cournot prices. However, during the low demand winter months, there is little potential for market power, because Cournot and joint monopoly prices are relatively close to competitive prices. The firms have more potential market power during the summer because total electricity demand is larger relative to the fringe supply than during the winter. During 1998–1999, actual prices closely match the simulated Cournot prices.

After June 2000, competitive prices rise substantially as input costs increase. The potential to exercise market power is high, as can be seen by the very large joint monopoly price if there were no price cap. However, the lowering of the price cap to $250 in August 2000 greatly reduces the effective joint monopoly price. The actual prices in 2000 are substantially lower than the joint monopoly prices but nevertheless average $20–$75 above the competitive price.

FIGURE 2.—ACTUAL PRICES VERSUS COMPETITIVE, COURNOT, AND JOINT MONOPOLY PRICES



Monthly average 5–6-p.m. 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 and Cournot prices are computed numerically based upon the marginal cost functions of the five strategic firms and an estimated constant-elasticity fringe supply. In 2000, the price cap is binding for competitive prices on 10 days, Cournot prices on 13 days, and joint monopoly on 112 days. Before 2000, the price cap is binding three times for the joint monopoly price.

It appears that actual pricing is most consistent with the five-firm Cournot benchmark.

I formally test whether the means of daily actual and benchmark prices are statistically equal. To do so, I estimate a model of the form

$$P_t^{\text{Actual}} - P_t^{Benchmark} = \alpha + \epsilon_t \quad \text{for}$$

$$Benchmark \in \{\text{Competitive, Cournot, JointMonopoly}\}.$$

Under the null that actual prices and benchmark prices have the same mean, $\alpha = 0$.[18] To allow for heteroskedasticity and serial correlation in the shocks, I compute Newey-West standard errors with a 7-day-lag moving-average structure. Actual prices differ from Cournot prices on average by $3.28, but this difference is not statistically different from 0

[17] These computations are based upon my estimate of the fringe supply relationship that is described in section III A.

[18] The econometric error represents mismeasurement of the benchmark price because the computations use fringe supply parameters that are estimated with error.

($p = 0.18$). However, actual prices differ from competitive prices by \$16.22, and from joint monopoly prices by −\$52.03; both mean differences are statistically different from 0 at the 1% level. This conclusion does not change if the sample is divided into various periods, including the periods with either four or five strategic firms in the market. In addition, for the price run-up of June–November 2000, I fail to reject Cournot pricing and reject both competitive and joint monopoly pricing. One noteworthy exception is the subperiod June–August 2000, when I find prices to be statistically above Cournot levels but nevertheless far below the joint monopoly prices.

B. *Firm-Level Pricing*

In order to test for heterogeneity in behavior by the strategic firms, I jointly estimate the firm-level supply relations (2) and fringe supply (1). In order to estimate the supply relation (2) for 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.[19] However, I cannot measure the shadow value of additional capacity (λ^*_{it}). The shadow value is 0 when capacity constraints are not binding. When capacity constraints bind, the shadow value is the difference between the measured price-cost margin and the inframarginal revenue term $\theta_{it}P_t{}'q_{it,}$ which is a function of the unknown conduct parameter. Although shadow values vary by both firm and time, adding a separate parameter for each firm-hour when a firm is at capacity would add excessive parameters to the model. Therefore, I add a single dummy variable (*CAPBIND*) equal to 1 to each supply relation if capacity constraints are binding and equal to 0 otherwise. The coefficient on *CAPBIND* is the *average* shadow value of added capacity.

I assume that a firm's behavior is constant over time and model its supply relation as

$$(P - c)_{it} = \lambda_i \cdot CAPBIND_{it} - \theta_i P'_t q_{it} + \epsilon_{it}.$$

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

$$(P - c)_{it} = \lambda_i \cdot CAPBIND_{it} + \frac{\theta_i}{\beta_1}\frac{P_t q_{it}}{Q^S_{\text{fringe }t}} + \epsilon_{it}. \qquad (3)$$

The parameter θ_i is identified by substituting a consistent estimate of β_1 from the demand side.

The econometric errors ϵ_{it} and υ_t represent shocks to marginal cost that are 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 (that is, 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.[20] I simultaneously estimate the system of the fringe supply (1) and each firm's supply relation (3) via the generalized method of moments. The error 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 the fringe supply. The slope of the (inverse) residual demand is the negative slope of the 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. I experiment with two functional forms to assess for sensitivity: (1) constant price elasticity, and (2) quadratic in price, reaching a maximum at \$750. I plot fitted values for various days and visually compare the estimated shapes with the expected shape of the 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 the two specifications. Perhaps most importantly, the slope of residual demand for the range of prices observed in the market is not sensitive to fringe supply specification during the price run-up period of June–November 2000.

First, I estimate the five firms' supply relations, imposing θ_i and λ_i to be equal across firms so that these results can be compared with the simulation results. I break the sample into a period with four firms in the market from July 1998 to April 1999, and a period with five firms, from April 1999

[19] 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. The presence of a price cap should not affect production behavior when the cap is not binding.

[20] 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 forecast 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 ($\lambda_{it}^* = 0$), and the results are similar. Also note that I use day-ahead forecast demand as an instrument in each strategic supply relation which restricts how firm-level output varies in total demand. Unfortunately, unique instruments are not available for each supply relation.

TABLE 3.—ESTIMATES OF FRINGE SUPPLY AND STRATEGIC SUPPLY RELATIONS FOR HOUR 18

	4-Firm Market*		5-Firm Market†		June–Nov. 2000	
Dependent Variable	Fringe $\ln Q^S_{\text{fringe } t}$	Strategic $(P - c)_{it}$	Fringe $\ln Q^S_{\text{fringe } t}$	Strategic $(P - c)_{it}$	Fringe $\ln Q^S_{\text{fringe } t}$	Strategic $(P - c)_{it}$
$\frac{Pq}{Q^S_{\text{fringe}}}$	—	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	
$\hat{\eta}^D_{\text{strat}}$	−2.38		−0.97		−0.69	
Average margin	$10.88		$26.96		$74.05	
$\hat{\theta}$	0.97		0.97		1.54	
	(0.16)		(0.09)		(0.18)	

"Fringe" represents equation (1), and "strategic" represents equation (3). Although the system contains a supply relation for each strategic 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 equation (3) are day-ahead forecast demand and *CAPBIND,* and the instruments entering equation (1) are log of day-ahead forecast demand and all regressors except log price. Standard errors, constructed using the optimal GMM weighting matrix, allow for firm-level heteroskedasticity, contemporaneous cross-equation error correlation, and individual serial correlation of MA(7) in fringe and strategic supply relations. 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, the majority occurring in August). The average elasticity of strategic demand is computed using the estimated fringe supply elasticity and the average size of fringe and strategic supply.

*July 1, 1998 to April 15, 1999.

†April 16, 1999 to November 30, 2000.

to November 2000.[21] Results are shown in table 3 and are similar for the two time periods. Fringe supply is relatively inelastic in both periods (0.18 and 0.19). Given the larger size of fringe output than that of strategic output, the strategic firms face a total residual demand elasticity of −2.38 during the four-firm period and −0.97 during the five-firm period.[22] Therefore, identical competitive behavior would lead to higher price-cost margins in the latter period.

The estimates of the strategic firm supply are consistent with Cournot pricing in both periods. For July 1998 through mid-April 1999, the coefficient on $P_t q_{it}/Q^S_{\text{fringe } t}$ and the estimate of β_1 imply a $\hat{\theta} = 0.97$ that is not statistically different from unity. For mid-April 1999 through November 2000, I obtain an identical conduct parameter $\hat{\theta} = 0.97$ that is statistically indistinguishable from unity. Although the pricing behavior is very similar in the two periods, margins are higher during the five-firm period when strategic firms face less elastic residual demand.

For the period of the price run-up of 2000, I can decompose the high price-cost margins averaging $74.05/MWh into *demand-side* and *supply-side* factors. Margins are higher because firms face 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. The conduct parameter estimate $\hat{\theta} = 1.54$ confirms the simulation results that find that late 2000 prices average slightly higher than Cournot levels. These results imply not only that the state was increasingly dependent upon the strategic firms' generation in 2000, but that the strategic firms supplied slightly less competitively.

Next, I estimate equations (1) and (3), allowing each firm to have a different conduct parameter (θ_i) and shadow value of capacity (λ_i). 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

[21] 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.

[22] 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.

TABLE 4.—CONDUCT PARAMETER ESTIMATES 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 equations (1) and (3) where each strategic firm supply relation contains a firm-specific parameter for conduct θ_i and shadow value of capacity λ_i. The instruments and estimation techniques are described in the note to table 3.
*Reject H_0:$\theta_i = 1$ at 5% level.
†July 1, 1998 to April 15, 1999.
‡April 16, 1999 to November 30, 2000.

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.[23] When I focus on the period of the price run-up between June and November 2000, firms' conduct parameter estimates are almost uniformly larger. Dynegy (with data caveats) and Southern have conduct parameter estimates statistically larger than unity, whereas AES, Duke, and Reliant's parameters are consistent with Cournot pricing.

C. *Interpretation*

Both empirical approaches suggest that the level of prices is consistent with a Cournot model. Therefore, the large variation in price-cost margins (figure 1) was not likely to have been 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 and there were widespread allegations of price manipulation. The benchmark simulations find prices substantially lower than joint monopoly levels and statistically indistinguishable from five-firm Cournot prices. The firm-level estimation suggests three firms are pricing at Cournot levels and two are less competitive than Cournot. 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 greater than unity. Several factors could explain the conduct parameter estimates' being greater than unity, but the factors are not testable within my empirical framework. For example, these conduct parameter estimates are consistent with a static game that includes an evolving set of beliefs about the slope of fringe supply or rival behavior. Alternatively, the folk theorem implies that colluding firms can sustain any level of prices between Cournot and efficient collusion prices.[24] To the extent that the California market is viewed as an infinitely repeated game with a discount factor between days very close to 1, any level of pricing behavior between one-shot Cournot and efficient collusion prices can be sustained in equilibrium. Such equilibrium behavior would be measured by a conduct parameter between unity and the number of colluding firms.[25] I cannot rule out the possibility that these two firms are engaging in some form of tacit collusion and they have asymmetric cost structures. However, there are strong reasons to believe that Dynegy's conduct parameter is biased upward because I do not have data on the output from its peaking units. Nevertheless, even if these two firms are engaged in some form of dynamic pricing, the resulting market prices are not substantially larger than Cournot prices, as we see in figure 2.

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 PX and the ISO's real-time market. To the extent that a supply function model is more realistic, my conduct parameters are biased downward. The sign of the bias can be easily understood. In a quantity-setting model, all of a firm's residual demand elasticity arises from the total demand, because its rivals are bidding vertical supply functions. If its rivals actually bid nonvertical supply functions, the true residual demand would obtain elasticity from both total demand and rival supply. Because firms are enjoying the same observed price-cost margin on a residual demand function that is flatter than I measure it to be, the θ estimated by my model is biased downward. For similar reasons, the simulated Cournot price would overstate price from a static pricing game. Another important assumption is that transmission congestion within California does not cause the slope of residual demand to differ for firms in the north and south. For example, if transmission is constrained from the north to the south, residual demand is likely to be steeper for strategic firms in the south. When I estimate the

[23] Given the unusually high conduct estimates for Dynegy, one may be concerned that estimates in which conduct is restricted to be equal across firms are substantially driven by Dynegy. I reestimate the model allowing Dynegy to have a different conduct parameter, and find that neither the estimates nor the inferences change substantially.

[24] Puller (2006) rejects that pricing in 2000 is consistent with efficiently colluding at prices below the joint monopoly levels.

[25] However, the behavioral interpretation of the conduct parameter would change. An estimate of $1 < \hat{\theta}_i < N$ 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.* For a good discussion of interpreting (and misinterpreting) θ_i, see Reiss and Wolak (2005).

model using only uncongested hours, the conduct parameter estimates tend to be smaller, but the qualitative conclusions do not change.

Finally, several institutional factors that changed in 2000 warrant special caveats. In late 2000 the utilities began to face financial crises that could have prevented them from paying for power purchased on the wholesale market. Because the risk of nonpayment may have increased the marginal costs of supplying power, my measure of marginal cost may understate the true cost of supplying power. Analysts believe this is most applicable to November 2000. However, several factors might 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 a widespread belief that in 2000 Duke forward-contracted some of its production. Firms had an incentive to raise the price only on the amount they produced beyond the contract position, because the price earned on the contracted quantity was already locked in. The presence of unobserved contract positions would lead me to overstate the simulated Cournot price. Also, my conduct parameter estimates would be biased. If some of the observed generation were sold forward, then firms were achieving the same profit margins for smaller quantities sold through the energy market, and I would understate the conduct parameter θ. If data on contract positions became available, one could correct this bias by adjusting inframarginal sales by the amount that was forward-contracted. A final potential bias in 2000 is that some transactions in the fall did not occur at the PX/ISO prices but at 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 downward in 2000.

This potential bias suggests that pricing may have been above Cournot levels for part of the summer of 2000. Figure 2 illustrates that the actual prices during June–August 2000 averaged \$43 above my estimate of Cournot prices. After August, prices were very close to Cournot. The electricity crisis was garnering significant political attention by the fall of 2000, and the increased threat of regulatory intervention may have mitigated pricing behavior at the end of 2000. Nevertheless, prices throughout 2000 were close to Cournot levels and substantially below the joint monopoly prices.

VI. Conclusions

A number of states and countries have designed restructured electricity markets so that a large fraction of transactions occurs 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 did not exhibit evidence of efficient tacit collusion. Nevertheless, there is strong evidence of unilateral market power. Price-cost margins varied substantially over time, with higher margins during the high-demand periods of each year. The large variation in price-cost margins was primarily driven by changes in residual demand elasticity rather than less competitive behavior. I generally fail to reject Cournot pricing for 1998–2000.

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. Primary factors contributing to price increases were higher input costs and less elastic residual demand. Nevertheless, the five large in-state nonutility 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 firms' behavior in 2000. I reject the hypothesis that all firms were pricing at Cournot levels, but the observed prices were much closer to Cournot prices than to efficient collusion prices. The observed prices in 2000 are consistent with a variety of other possible behaviors: some other form of dynamic pricing, some average of various nonequilibrium behavior, or a static game with an evolving set of beliefs about the shape of fringe supply or rival behavior. Distinguishing among the possible behaviors is a formidable empirical task. For this reason, it would be problematic to use this type of methodology to determine antitrust activity.

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. This paper confirms earlier work that market power is a concern. Policymakers must consider the source of market power when considering market design issues such as the 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 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.[26] However, if there is evidence that firms begin to engage in some form of dynamic pricing, regulators should focus on the design and frequency of the auction or the amount of real-time information made available to market participants. Some work has suggested that collusion is less likely under discriminatory auctions than

[26] In fact, other markets that have required forward contracting or vesting contracts do not exhibit evidence of substantial market power except at high levels of demand (Bushnell et al., 2005).

under uniform-price auctions.[27] 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 one large and several small firms may make tacit collusion more difficult to coordinate and sustain.

REFERENCES

Borenstein, Severin, and James B. Bushnell, "An Empirical Analysis of the Potential for Market Power in California's Electricity Industry," *Journal of Industrial Economics* 47 (September 1999), 285–324.

Borenstein, Severin, James B. Bushnell, Christopher R. Knittel, and Catherine D. Wolfram, "Inefficiencies and Market Power in Financial Arbitrage: A Study of California's Electricity Markets," CSEM working paper no. 138, University of California Energy Institute (2005).

Borenstein, Severin, James B. Bushnell, and Frank A. Wolak, "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market," *American Economic Review* 92 (December 2002), 1376–1405.

Bushnell, James B., "Looking for Trouble: Competition Policy in the U.S. Electricity Industry," in James M. Griffin and Steven L. Puller (Eds.), *Electricity Deregulation: Choices and Challenges* (University of Chicago Press, 2005).

Bushnell, James B., Erin T. Mansur, and Celeste Saravia, "Vertical Arrangements, Market Structure and Competition: An Analysis of Restructured U.S. Electricity Markets," CSEM working paper no. 126 (2005).

Bushnell, James B., and Celeste Saravia, "An Empirical Assessment of the Competitiveness of the New England Electricity Market," CSEM working paper no. 101 (2002).

Bushnell, James B., and Frank A. Wolak, "Regulation and the Leverage of Local Market Power: Reliability Must-Run Contracts in the California Electricity Market," POWER working paper no. PWP-070, University of California Energy Institute (1999).

Corts, Kenneth S., "Conduct Parameters and the Measurement of Market Power," *Journal of Econometrics* 88 (February 1999), 227–250.

Energy Information Administration, "Monthly Power Plant Database," Form EIA-759 (1998–2000).

——— "Monthly Nonutility Power Plant Database," Form EIA-900 (1999).

Fabra, Natalia, "Tacit Collusion in Repeated Auctions: Uniform versus Discriminatory," *Journal of Industrial Economics* L1 (September 2003), 271–293.

Fabra, Natalia, and Juan Toro, "Price Wars and Collusion in the Spanish Electricity Market," *International Journal of Industrial Organization* 23 (April 2005), 155–181.

Green, Richard J., and David M. Newbery, "Competition in the British Electricity Spot Market," *Journal of Political Economy* 100 (October 1992), 929–953.

Harvey, Scott M., and William W. Hogan, "On the Exercise of Market Power through Strategic Withholding in California," Kennedy School of Government, Harvard University, mimeograph (2001).

Hortacsu, Ali, and Steven L. Puller, "Understanding Strategic Bidding in Restructured Electricity Markets: A Case Study of ERCOT," NBER working paper no. 11123 (2005).

Joskow, Paul L., "Deregulation and Regulatory Reform in the U.S. Electric Power Sector," in Sam Peltzman and Clifford Winston (Eds.), *Deregulation of Network Industries: The Next Steps* (Brookings Press, 2000).

——— "California's Electricity Crisis," *Oxford Review of Economic Policy* 17 (Autumn 2001), 365–388.

Joskow, Paul L., and Edward Kahn, "Identifying the Exercise of Market Power: Refining the Estimates," Kennedy School of Government, Harvard University, mimeograph (2001).

——— "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market during Summer 2000," *Energy Journal* 23 (2002), 1–35.

Kahn, Edward, Shawn Bailey, and Luis Pando, "Simulating Electricity Restructuring in California: Interactions with the Regional Market," *Resource and Energy Economics* 19 (March 1997), 3–28.

Klein, Joel B., "The Use of Heat Rates in Production Cost Modeling and Market Modeling," Electricity Analysis Office, California Energy Commission (1998).

Klemperer, Paul D., "What Really Matters in Auction Design," *Journal of Economic Perspectives* 16 (Winter 2002), 169–189.

Klemperer, Paul D., and Margaret A. Meyer, "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica* 57 (November 1989), 1243–1277.

Mansur, Erin T., "Upstream Competition and Vertical Integration in Electricity Markets," *Journal of Law and Economics* 50:1 (forthcoming).

Natural Gas Intelligence, "NGI's Daily Gas Price Historical Data" (1998–2000).

Pacific Gas & Electric Company, Various natural gas tariff schedules (1998–2000).

Puller, Steven L., "Consistent Estimation of Competitive Conduct When Firms Are Efficiently Colluding: Addressing the Corts Critique," Texas A&M University mimeograph (2006).

Reiss, Peter C., and Frank A. Wolak, "Structural Econometric Modeling: Rationales and Examples from Industrial Organization," Stanford University mimeograph (2005).

Rotemberg, Julio J., and Garth Saloner, "A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review* 76 (June 1986), 390–407.

Saravia, Celeste, "Speculative Trading and Market Performance: The Effect of Arbitrageurs on Efficiency and Market Power in the New York Electricity Market," CSEM working paper no. 121 (2003).

Sheffrin, Anjali, "Empirical Evidence of Strategic Bidding in California ISO Real-Time Market," report of California Independent System Operator (2001).

Southern California Gas Company, Various natural gas tariff schedules (1998–2000).

Sweeting, Andrew, "Market Power in the England and Wales Wholesale Electricity Market 1995–2000," Faculty of Economics, University of Cambridge, working paper no. 0455 (2005).

Wilson, Robert, "Architecture of Power Markets," *Econometrica* 70 (July 2002), 1299–1340.

Wolak, Frank A., "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market," *International Economic Journal* 14 (Summer 2000), 1–40.

——— "Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications,* Vol. II (2003a).

——— "Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market, 1998–2000," *American Economic Review* 93 (2003b), 425–430.

Wolak, Frank A., and Robert H. Patrick, "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market," POWER working paper no. PWP-047, University of California Energy Institute (1997).

Wolfram, Catherine D., "Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales," *RAND Journal of Economics* 29 (Winter 1998), 703–725.

——— "Measuring Duopoly Power in the British Electricity Spot Market," *American Economic Review* 89 (September 1999), 805–826.

APPENDIX

Data

The marginal cost function includes the daily marginal fuel, permit, and variable operating and maintenance costs of all units that are operating during hour 18. The marginal fuel cost for each generating unit is calculated from daily natural gas spot prices and average heat rates. All of

[27] See Klemperer (2002) and Fabra (2003).

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; Pacific Gas & Electric Company, 1998–2000). To test for the possibility of natural gas prices being endogenous to the fringe's production, I also use prices at a larger, more distant hub (Henry Hub in Louisiana) and the results are very similar. 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 data sets 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). 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 was virtually never a net exporter during my sample.

Several generators in southern California were required to purchase permits for emissions of NO_x. The hourly marginal permit cost is calculated as the monthly quantity-weighted average price of permits multiplied by the unit's hourly emissions. Permit costs were negligible until mid-2000 because total emissions were less than the number of allocated permits, 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 & 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 the 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.

The CEMS 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 nameplate capacity; however, the results are robust to defining capacity as 80% of nameplate. I assume that each unit's marginal cost function is constant up to the capacity of the generator, and this assumption is supported by Klein (1998). Unfortunately, I cannot take account of the very occasional partial outages that temporarily reduce the operating capacity of a unit.

Because a firm incurs start-up costs to fire up a unit, I analyze the firms' utilization of units that are already operating during the particular hour. Some analysts have suggested that firms exercise market power by shutting down generating units, and this is particularly the case for 2000. I observe shutdowns but cannot distinguish between true outages and withholding an entire unit to raise the price. This could bias downward my measure of market power if firms shut down plants to exercise market power. However, an ISO analysis of confidential bid data suggests that this bias is not severe in 2000. Sheffrin's (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. In addition, I find a strong level of persistence in the number of units that a firm is operating from one hour 18 to the next. This is consistent with Sheffrin's finding.

I need to make several assumptions about a firm's dispatch decisions in order to determine the firm's marginal cost of producing one more unit of output 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 still should increase production in these units if marginal cost is lower than the price. In practice, the RMR units are not always higher-cost units, and when they are, their 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 intrazonal 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 PX and total demand forecasts are from the PX and ISO Web sites, respectively (available at www.ucei.berkeley.edu). 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. 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 Web site.

Focusing on prices in the PX and ISO energy markets introduces a slight complication. 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 affects firm behavior in the energy market. This 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 in the real-time market's bid stack. Therefore, exercising 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 ISO. 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 by the fossil-fueled units I am analyzing. Although it is unknown how much regulation reserve is satisfied with thermal generating units, Joskow and Kahn's (2001) analysis of summer 2000 assumes that an additional 3% of thermal demand is purchased as reserves. This mismeasurement is mitigated by the fact that the quantity of regulation reserve bought during hour 18 is typically smaller than quantities bought at other hours of the day.