Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market

By SEVERIN BORENSTEIN, JAMES B. BUSHNELL, AND FRANK A. WOLAK*

We present a method for decomposing wholesale electricity payments into production costs, inframarginal competitive rents, and payments resulting from the exercise of market power. Using data from June 1998 to October 2000 in California, we find significant departures from competitive pricing during the high-demand summer months and near-competitive pricing during the lower-demand months of the first two years. In summer 2000, wholesale electricity expenditures were $8.98 billion up from $2.04 billion in summer 1999. We find that 21 percent of this increase was due to production costs, 20 percent to competitive rents, and 59 percent to market power. (*JEL* L1, L9)

In the spring of 2000, the momentum behind a dramatic restructuring of the electricity industry appeared to be irresistible. There were four regions of the United States with independent system operators running spot markets for wholesale electricity—California, PJM (major parts of Pennsylvania, New Jersey, Maryland, Delaware, Virginia, and the District of Columbia), New England, and New York. Several other states were undertaking initiatives to restructure their electricity sector along similar lines. Beginning in summer 2000, however, soaring wholesale electricity prices in California made international news and threatened the financial stability of the state. The disruptions in California slowed, and threatened to reverse, the movement toward restructured electricity markets in the United States and elsewhere.

In the aftermath of California's electricity crisis, policy makers debated the correct lessons to take from the state's restructuring as well as the proper regulatory response to the crisis. Many of the answers to the questions being debated depend upon a proper diagnosis of the problems that disrupted California's power sector during this period. Were soaring power costs the result of market "fundamentals" such as rising fuel prices, environmental cost, and a scarcity of generating capacity? Or were power suppliers able to exercise significant market power? In this paper we estimate the extent to which each of these factors—input costs, scarcity, and market power—influenced market outcomes in the California power market from 1998 through 2000. We analyze input and output prices, generator variable costs, and actual production quantities to measure the degree to which California wholesale electricity prices exceeded competitive levels. We also address the question of the efficiency impacts of market power in this market.

While market power has been studied and estimated in many industries, there has been

* Borenstein: Haas School of Business, University of California, Berkeley, CA 94720, University of California Energy Institute, and NBER (e-mail: borenste@haas.berkeley.edu); Bushnell: University of California Energy Institute, 2539 Channing Way, Berkeley, CA 94720 (e-mail: bushnell@haas.berkeley.edu); Wolak: Department of Economics, Stanford University, Stanford, CA 94305, and NBER (e-mail: wolak@zia.stanford.edu). Borenstein is a member of the Governing Board of the California Power Exchange. Bushnell was a member of the Power Exchange's Market Monitoring Committee during 1999–2000. He is currently a member of the California Independent System Operator's Market Surveillance Committee, of which Wolak is the Chair. The views expressed in this paper do not necessarily reflect those of any organization or committee. For helpful discussions and comments, we thank Carl Blumstein, Roger Bohn, Peter Cramton, David Genesove, Rob Gramlich, Paul Joskow, Ed Kahn, Alvin Klevorick, Christopher Knittel, Patrick McGuire, Catherine Wolfram, an anonymous referee, and participants in numerous seminars. Jun Ishii, Matt Lewis, Erin Mansur, Steve Puller, Celeste Saravia, and Marshall Yan provided excellent research assistance.

little attention paid to intertemporal variation in the ability to exercise market power. For industries in which the good is storable, such intertemporal variation is necessarily small, because inventories greatly reduce intertemporal supply variation, and possibly, demand variation. In markets for nonstorable goods, including electricity and most services, this is not possible. The problem is exacerbated in electricity because demand is very inelastic in the short run, and supply becomes very inelastic as production approaches the system-generation capacity. Recognizing the dynamics of market power is likely to be important in both determining its causes and crafting remedies as part of the evolving public policy toward electricity restructuring.

Luckily, due to the history of regulation in electricity markets, data exist on the hourly output of all generating units and transmission power flows. In addition, information collected on the technical characteristics of each California generating unit during the regulated monopoly regime allows very accurate estimation of generating unit-level variable costs.

We find that, due to rising input costs, even a perfectly competitive California electricity market would have seen wholesale electricity expenditures triple between the summers of 1998 and 2000.[1] Market power, however, also played a very significant role. In summer 1998, 25 percent of total electricity expenditures could be attributed to market power, a figure that increased to 50 percent in summer 2000. The increased percentage margins due to market power combined with substantial production cost increases for marginal producers to create a drastic rise in absolute margins and, thus, pushed the market into a crisis later in the summer of 2000.

In Section I, we discuss the issues raised in estimating market power in electricity markets and the consequences of market power. We present an overview of California's electricity market in Section II. In Section III, we describe the estimation technique in detail in the context of the California market, addressing each component of the market and outlining the assumptions made in implementing the analysis. We try to take a conservative approach, interpreting the data in a way that would be likely to understate the degree of market power exercised. In Section IV, we present estimates of premia of actual prices over the competitive levels. In Section V, we attempt to parse changes in competitive revenues between changes in actual costs and changes that reflect rents to inframarginal competitive sellers. We conclude in Section VI.

[1] For the purpose of this analysis, we define the summer to be June through October of each year.

I. Market Power Analysis in the Electricity Industry

During most of the 1990's, regulatory evaluation of short-run horizontal market power in electricity focused on concentration measures, such as the Herfindahl-Hirschman index. Unfortunately, such measures are a poor indicator of the potential for, or existence of, market power in the electricity industry, because the industry is characterized by highly variable price-inelastic demand, significant short-run capacity constraints, and extremely costly storage.[2] It is easy to show that in such circumstances, firms with very small market shares could still exercise significant market power.

We use data collected on the technological characteristics of generating units located in California to construct a competitive market counterfactual that we compare to actual market outcomes. This competitive counterfactual models each firm as a price-taker that would sell power from a given plant so long as the price it received was greater than its incremental cost of production. Of course, the cost of selling a unit of electricity can be greater than the simple production costs if the firm has an opportunity cost that is greater than its production cost, such as the revenue the firm would get from selling power or reserve capacity in a different location or market. On the other hand, a high price in an alternative market can reflect market power in that market, resulting in the transmittal of high

[2] See Borenstein et al. (1999) for a more detailed discussion of the applicability of concentration measures to market power analysis in electricity markets and citations to regulatory decisions that have relied on concentration indices.

prices across markets by the response of competitive suppliers. We discuss these alternative opportunities below and how we account for them in our analysis.

Thus far, we have discussed only situations in which a firm unilaterally exercises market power. Antitrust law is most often concerned with collusive attempts to exercise market power. Unfortunately, many of the attributes that facilitate collusion are present in electricity markets: In most electricity markets, firms play repeatedly, interacting on a daily basis, so there is opportunity to develop subtle communication and collusive strategies. The payoff from cheating on a collusive agreement may be limited due to capacity constraints on production, though for the same reason, the ability to punish defectors may be limited. Finally, the industry has fairly standardized production facilities, so homogeneous costs may make it easier for firms to attain tacit or explicit collusive outcomes. All that said, we have not explored the question of tacit or explicit collusion among firms in the California market as a potential cause of prices in excess of competitive levels.[3] Rather, in this paper we focus on the competitiveness of market outcomes.

In focusing on market outcomes, there are two indicators that clearly distinguish market power, and each leads to a distinct estimation technique. First, in a competitive market, a firm is unable to take any action, including output decisions or offer prices, that significantly affects the price in a market. This suggests a method of estimation that involves studying the bidding and output supply decisions of each firm in the market to detect successful attempts to affect prices. This is the general approach used by Wolak and Robert H. Patrick (1997), Catherine D. Wolfram (1998), Roger Bohn et al. (1999), Bushnell and Wolak (1999), Wolak (2000), and Puller (2001).

The second empirical approach is at the market level, and this is the one that we adopt here. We examine whether the market as a whole is setting competitive prices given the production capabilities of all players in the market. As such, this approach is less vulnerable to the arguments of coincidence, bad luck, or ignorance that may be directed at analysis of the actions of a specific generator. It is less informative about the specific manifestations of market power, but it is effective for estimating its scope and severity, as well as identifying how departures from competitive outcomes vary over time. This is the general approach used in Wolfram (1999). At least two papers, Erin T. Mansur (2001) and Paul L. Joskow and Edward Kahn (2002), utilize both of these approaches.

A potential drawback of the market-level approach is that it captures all inefficiencies in the market, some of which may not be due to market power. If, for instance, low-cost generators were systematically held out of production simply due to a faulty dispatch algorithm, that would impact the estimate of market power. During the period we study, the California market clearly still had a number of design flaws that may have contributed to inefficient dispatch and market pricing. For the great majority of these, however, the flaw would be benign if firms acted as pure price-takers, rather than exploiting these design flaws to affect the market price. Furthermore, we find that, over substantial periods of time, prices did not significantly differ from our estimates of marginal cost, indicating that there were no systemic inefficiencies raising prices in all periods. Still, our estimates must be taken with the caveat that they include failures to achieve competitive market prices for reasons other than market power, including bad judgment and confusion on the part of some generators or market-making institutions.

The Consequences of Market Power

In analyzing the efficiency consequences of market power in electricity, one must begin from the recognition that short-run electricity demand currently exhibits virtually zero price elasticity. Almost none of the customers in California, or anywhere else in the United States, are charged real-time retail electricity prices that vary hour-to-hour as wholesale prices do. Because the extent of market power varies tremendously on an hourly basis, the absence of

[3] See Steven L. Puller (2001) for an analysis of this issue.

very-short-run elasticity is critical to understanding its consequences.[4] In studying California specifically, one must also consider that, during the 1998–2001 transition period, end-use consumers were insulated from energy price fluctuations by the Competition Transition Charge (CTC). The CTC was implemented along with the restructuring of the industry in order to allow the incumbent utilities to recover their stranded generation costs. Due to the CTC, the vast majority of end-use consumers faced fixed retail rate schedules during the transition period.[5] Thus, the CTC greatly lessened even the monthly elasticity of final consumer electricity demand.

Due to the extreme short-run inelasticity of demand, market power in electricity markets has little effect on consumption quantity or short-run allocative efficiency. As described below, however, generating companies in California vary markedly in their costs and generation capacity, so the exercise of market power by one firm can lead to an inefficient reallocation of production among generating firms: a firm exercising market power will restrict its output so that its marginal cost is below price (and equal to its marginal revenue), while other firms that are price-taking will produce units of output for which their marginal cost is virtually equal to price. Thus, there will be inefficient production on a marketwide basis as more expensive competitive production is substituted for less expensive production owned by firms with market power. This is the outcome Wolak and Patrick (1997) described in the U.K. market, where higher-cost combined-cycle gas turbine generators owned by new entrants provide base-load power that could be supplied more cheaply by coal-fired plants that were being withheld by the two largest firms. Joskow and Kahn (2002) find evidence of withholding by large firms in the California market.

In addition, several recent analyses have demonstrated that the exercise of market power in an electricity network can greatly increase the level of congestion on the network.[6] This increased congestion impacts negatively both the efficiency and the reliability of the system. Market power can also lead firms to utilize their hydroelectric resources in ways that decrease overall economic efficiency.[7]

Lastly, electricity prices influence long-term decision-making in a way that can seriously impact the economy and generation investment. While it has been pointed out that high prices should spur new investment and entry in electricity production, these investments may not be efficient if motivated by high prices that are caused by market power, which may indicate a need not for new capacity, but for the efficient use of existing capacity. Artificially high prices also can lead some firms *not* to invest in productive enterprises that require significant use of electricity, or to inefficiently substitute to less electric-intensive production technologies.

Beyond the efficiency considerations, market power has potentially large and important redistributional effects. The California electricity crisis of 2000–2001 illustrates both the immense potential size of these effects and the difficulty of analyzing them. The transitional retail rate freeze associated with the CTC meant that the utilities bore the brunt of the wholesale price increases. The utilities' eventual response was to declare bankruptcy in one case, and threaten to in another, so the ratepayers or taxpayers

[4] In California and elsewhere, time-of-use rates are common for large users. These price schedules generally have preset peak, shoulder, and off-peak rates, which are changed only twice per year. They do not distinguish, for instance, a weekday afternoon with extremely high wholesale prices from a more moderate weekday afternoon. Borenstein (2001, 2002) argues that time-of-use rates are an extremely poor substitute for real-time electricity pricing and that real-time pricing would greatly mitigate wholesale price volatility. Patrick and Wolak (1997) estimate the within-day price responsiveness of industrial and commercial customers facing real-time half-hourly energy prices in the England and Wales electricity market. They argue that an electricity market would be much less susceptible to the exercise of market power if even one-quarter of peak demand had the average level of price responsiveness that they estimate.

[5] Even "direct access" consumers, who bought energy from some source other than their incumbent utility, were insulated from wholesale energy price fluctuations in the short run by the CTC. This is because the stranded cost component paid by all consumers was calculated in a way that moved inversely to the energy price: the higher the energy price, the lower the CTC payment for that hour.

[6] See Judith B. Cardell et al. (1997), Bushnell (1999), Borenstein et al. (2000), and Joskow and Jean Tirole (2000).

[7] See Bushnell (2003).

ultimately became the bearers of much of these costs. At this writing, it is still unclear who will bear what share of the expense, and how much of the revenues paid to generators will be refunded to buyers under orders from federal regulators.

II. The California Electricity Market

Through December 2000, the two primary market institutions in California were the Power Exchange (PX) and the Independent System Operator (ISO). The PX ran a day-ahead and day-of market for electrical energy utilizing a double-auction format. Firms submitted both demand and supply bids, then the PX set the market-clearing price and quantity at the intersection of the resulting aggregate supply and demand curves. In the PX day-ahead market, which was by far the largest market in California, firms bid into the PX offers to supply or consume power the following day for any or all of the 24 hourly markets. The PX markets were effectively financial, rather than physical; firms could change their day-ahead PX positions by purchasing or selling electricity in the ISO's real-time electricity spot market.[8]

The PX was not the only means for buyers and sellers to transact electricity in advance of the actual hour of supply. A buyer and seller could make a deal bilaterally. All institutions that scheduled transactions in advance, including the PX, were known as "scheduling coordinators" (SCs).[9] Because SCs use the transmission grid to complete some transactions, they are required to submit the generation and load schedules associated with these transactions to the ISO.

The ISO is responsible for coordinating the usage of the transmission grid and ensuring that the cumulative transactions, or schedules, do not constitute a reliability risk, i.e., are not likely to overload the transmission system.[10] As the institution responsible for the real-time operation of the electric system, the ISO must also ensure that aggregate supply is continuously matched with aggregate demand. In doing so, the ISO operates an "imbalance energy" market, which is also commonly called the real-time, or spot, energy market. In this market, additional generation is procured in the event of a supply shortfall, and generators are relieved of their obligation to provide power in the event that there is excess generation being supplied to the grid. Like the PX, this market is run through a double-auction process, although of slightly different format. Firms that deviate from their formal schedules are required to purchase (or sell) the amount of their shortfall (or surplus) on the imbalance energy market. During our sample period, no further penalties were assessed for deviating from an advance schedule. The imbalance energy market therefore serves as the de facto spot market for energy in California. During our sample period, the ISO imbalance energy market constituted less than 5 percent of total energy sales with the PX accounting for about 85 percent and the remainder taking place through bilateral trades.

The ISO also operates markets for the acquisition of reserve, or stand-by, capacity. Reserve capacity is used to meet unexpected demand peaks and to adjust production at different points on the grid in order to relieve congestion on the transmission grid while still meeting all demand. These reserves, known as "ancillary services," are purchased through a series of auctions that determine a uniform price for the *capacity* of each reserve purchased. Most of the reserve capacity is still available to provide imbalance energy in real time, and therefore will impact the spot price. A production unit committed to provide reserve capacity during an hour would therefore earn a capacity payment for being available and, if called upon in real time, would earn the imbalance energy price for actually providing energy.

"Regulation reserve," the most short-term re-

[8] Though the transaction costs of trading in the PX and ISO differed, these differences were negligible relative to the costs of the underlying commodity, electrical energy.

[9] In January of 2001, the PX ceased operation and the California Department of Water Resources assumed responsibility for the bulk of wholesale purchases on behalf of all investor-owned utilities in California, negotiating bilateral purchases and operating as its own SC.

[10] Unlike the PX, the ISO continued to function in approximately its original role through the 2000–2001 electricity crisis.

serve, is treated differently. Regulation reserve units are directly controlled by the ISO and adjusted second-by-second in order to allow the ISO to continuously balance supply and demand, and to avoid overloading of transmission wires. For this reason, we treat it differently in our analysis as described later.

A. *Market Structure of California Generation*

The California electricity generation market at first glance appears relatively unconcentrated. The former dominant firms, Pacific Gas & Electric (PG&E) and Southern California Edison (SCE), divested the bulk of their fossil-fuel generation capacity in the first half of 1998 and most of the remainder in early 1999. Most of the capacity still owned by these utilities after the divestitures was covered by regulatory side agreements, which prescribed the price the seller was credited for production from these plants independent of the PX or ISO market prices. These divestitures left the generation assets in California more or less evenly distributed between seven firms. The generation capacity of these firms that was located within the ISO system during July 1998 and July 1999 is listed in Table 1. "Fossil" includes all plants that burn natural gas, oil, or coal to power the plant, but over 99 percent of the output from these plants is fueled by natural gas. The vast majority of capacity listed as owned by "other" firms was composed of small independent power projects. The market structure during 2000 was largely unchanged from that of 1999.

As can be seen from Table 1, PG&E was the largest generation company during the summer of 1998. The seemingly dominant position of PG&E is offset to a large extent by its other regulatory agreements. All of its nuclear generation in California, for instance, is treated under rate agreements that do not depend on market prices. More importantly, the incumbent utilities in California were the largest buyers of electricity during this time period.[11]

TABLE 1—CALIFORNIA ISO GENERATION COMPANIES (MW)

July 1998—online capacity				
Firm	Fossil	Hydro	Nuclear	Renewable
AES	4,071	0	0	0
Duke	2,257	0	0	0
Dynegy	1,999	0	0	0
PG&E	4,004	3,878	2,160	793
Reliant	3,531	0	0	0
SCE	0	1,164	1,720	0
SDG&E	1,550	0	430	0
Other	6,617	5,620	0	4,267
July 1999—online capacity				
Firm	Fossil	Hydro	Nuclear	Renewable
AES	4,071	0	0	0
Duke	2,950	0	0	0
Dynegy	2,856	0	0	0
PG&E	580	3,878	2,160	793
Reliant	3,531	0	0	0
SCE	0	1,164	1,720	0
Mirant	3,424	0	0	0
Other	6,617	5,620	430	4,888

Source: California Energy Commission (www.energy.ca.gov).

B. *Analyzing Market Power in California's Electricity Market*

Critical to studying market power in California is an understanding of the economic interactions between the multiple electricity markets in the state. Participants moved between markets in order to take advantage of higher (for sellers) or lower (for buyers) prices. For instance, if the ISO's real-time imbalance energy price was consistently higher than the PX day-ahead price, then sellers would reduce the amount of power they sell in the PX and sell more in the ISO imbalance energy market. These attempts to arbitrage the PX/ISO price difference would cause the PX price and ISO imbalance energy price to converge. For this reason, it is not useful to study the PX market, or any other of the California markets, in isolation. The strong forces of financial arbitrage mean that any change in one market that affects

[11] The utilities had no incentive to raise market prices because they were net buyers of electricity and the revenue from power that they sold into the PX was just netted out from their power purchase costs. In fact, the CTC mechanism paid the three investor-owned utilities the difference between fixed wholesale price (implicit in their frozen retail rate) and the hourly wholesale price per unit of energy consumed in their distribution service territory.

that market price will spill over into the other markets.[12]

This interaction of the different California electricity markets means that we must study the entire California energy market in order to analyze market power in the state. For this reason, in the analysis below we look at all generation in the ISO control area regardless of whether the power from a generating plant is being sold through the ISO, the PX, or some other scheduling coordinator.

Recognition that the California power market is effectively an integrated market due to arbitrage forces yields two other important insights. First, although the California market has some large buyers of electricity directly purchasing from the transmission network who may respond to hourly wholesale prices, the large utility distribution companies (UDCs) cannot control the level of end-use demand of their customers because these customer face price schedules that do not vary with the hourly wholesale price. The UDCs cannot therefore reduce end-use consumption in a given hour in order to lower overall power purchase costs. They did have some limited freedom as to *which* market they used for purchase of their required power, choosing between buying day-ahead in the PX and spot purchases from the ISO imbalance energy market. Nonetheless, because sellers could move between markets as well, ultimately the buyers had no ability to exercise monopsony power, because they could not reduce their hourly demand for energy.

The second insight from a recognition of market integration involves the impact of price caps in the various markets. Because the ISO imbalance energy market was the last in a sequence of markets, the level of the price cap in the imbalance energy market fed back to form an implicit cap on prices in the other advance markets. That is, knowing that the maximum one might have to pay for power in real time was capped at $250 per megawatt-hour (MWh), for example, no buyer would be willing to pay more than $250 for purchases in advance. Thus, the aggregate demand curve in the day-ahead PX market became near horizontal at prices approaching the level of the price cap in the ISO imbalance energy market.[13]

Many of the suppliers that compete in the ISO or PX also are eligible to earn capacity payments for providing ancillary services, as well as energy payments for generating in real time, if they bid successfully into one of the ancillary services markets. Ancillary services therefore represent an alternative use of much of the generation capacity in California. It is therefore necessary to consider the interaction between the energy and ancillary services markets. In the case of the California market, the relevant consideration is that the provision of ancillary services in most cases does not preclude the provision of energy in the real-time market. Thus, for the bulk of generation, the decisions to sell into ancillary service capacity markets and real-time energy markets are not mutually exclusive.

It is important to recognize that the pool of suppliers available to ancillary services markets is very similar to that available to the energy markets. The main difference is that some generators are physically unable to provide certain ancillary services. Thus, there are fewer potential suppliers for some ancillary services than there are for energy. We therefore would expect that the energy market would be at least as competitive as the ancillary services markets, and probably more so. In fact the ancillary services markets, for a variety of reasons, appear to have been significantly less competitive than the energy market during the time period of our study.[14]

The other prominent opportunity for the usage of California generation is the supply of power to neighboring regions. Higher prices for electricity outside of California could produce a result in which generators within California were able to earn prices above their marginal cost, even if all generators behaved as price-takers. For this to be the case, however, the California ISO region would have to be a net exporter of power. During our sample period,

[12] Borenstein et al. (2001) finds that although significant price differences between the PX and ISO did occur during individual months, overall, there was no consistent pattern of the PX price being higher or lower than the ISO price.

[13] See Bohn et al. (1999).

[14] See Wolak et al. (1998).

such conditions arose in only 17 hours out of the 22,681 hours in the sample. Even in these hours, the maximum net quantity of energy exported out of the ISO control area in any hour was a modest 608 MWh. Therefore, if we assume that trading in these markets was efficient, the net export opportunities for producers within California were very limited relative to the California market.

Even if this were not the case, however, our analysis would fully account for the opportunity cost of exports, because under the California market structure firms from other states had the option to purchase power through markets run by the PX and ISO. Thus, power exported to Arizona, for example, would raise the quantity demanded in the California market, and therefore would increase the competitive market-clearing price within the California PX and ISO markets. If transmission became congested, then further exports would be infeasible and the quantity demanded in the California market would include only the exports up to the transmission constraint. Thus, the competitive market prices we estimate incorporate opportunities for export from California.

III. Measuring Market Power in California's Electricity Market

The fundamental measure of market power is the margin between price and the marginal cost of the highest cost unit necessary to meet demand. As discussed above, if no firm were exercising market power, then all units with marginal cost below the market price would be operating. Even in a market in which some firms exercise considerable market power, the *marginal* unit that is operating could have a marginal cost that is equal to the price. When a firm with market power reduces output from its plants or, equivalently, raises its offer price for its output, its production is usually replaced by other, more expensive generation that may be owned by nonstrategic firms.

In estimating a price-cost margin in this paper, we therefore must estimate what the system marginal cost of serving a given level of demand *would be* if all firms were behaving as price-takers. In the following subsections we describe the assumptions and data used for generating estimates of the system marginal cost of supplying electrical energy in California.

A. *Market-Clearing Prices and Quantities*

As described above, the California electricity market in fact consists of several parallel and overlapping markets. Given that generation and distribution firms, as well as other power traders, can arbitrage the expected price of energy across these commodity markets, we rely upon the unconstrained PX day-ahead energy price as our estimate of energy prices in any given hour.[15] We chose to rely upon the PX unconstrained price because the PX handled over 85 percent of the electricity transactions during our sample period and the unconstrained PX price represents the market conditions most closely replicated in our estimates of marginal costs. In particular, we do not consider the costs of transmission congestion or local reliability constraints in our estimates of the marginal cost of serving a given demand. The PX unconstrained price is also derived by matching aggregate supply with aggregate demand without considering these constraints. The resulting market-clearing price therefore reflects an outcome that would occur in the absence of transmission constraints, just as our cost calculations reflect the outcome in a market in which all producers are price-takers and there are no transmission constraints.[16]

It has been argued that the day-ahead PX price should be expected to systematically overstate the marginal cost of energy supply because sellers in the day-ahead market would include a

[15] One might be concerned that this arbitrage would not hold in light of the requirement during our sample period that the three investor-owned utilities buy all of their energy from the PX. Given the financial nature of the PX market and availability of a number of forward market products to hedge the day-ahead PX price risk, the full meaning of this requirement was ambiguous. More importantly, Borenstein et al. (2001) find that the PX and ISO prices track quite closely throughout most of our sample period.

[16] We would like to emphasize again that we use the PX price as representative of the prices in *all* California electricity markets. This is not a study of the PX market and the market power we find is not limited to the PX market. It is the amount we estimate to be present in all California electricity markets. Quantitatively similar results obtain using day-ahead or real-time zonal prices.

premium in their offer prices to account for the opportunity of earning ancillary services revenues, which require that the units not be committed to sell power in a forward market. However, if this were true, then the PX price would also be systematically higher than the ISO real-time price, which would not include such a premium because suppliers of ancillary services are also eligible to sell energy in the real-time market. Empirically this is not the case. Over our sample period, the PX average price was not significantly greater than the ISO average price.[17]

The interaction of these energy markets also requires us to use the systemwide aggregate demand as the market-clearing quantity upon which we base our marginal cost estimates. This level therefore includes consumption through the PX, other SCs, and any "imbalance energy" demand that is provided through the ISO imbalance energy market. Consumption from all of these markets is in fact metered by the ISO, which in turn allocates imbalance energy charges among SCs during an *ex post* settlement process. We are therefore able to obtain the aggregate quantity of energy supplied each hour from the ISO settlement data.

The acquisition of reserves by the ISO also requires discussion here. Since the ISO is effectively purchasing considerable extra capacity for the provision of reserves, it might seem appropriate to consider these reserve quantities as part of the market-clearing demand level. However, with the exception of regulation reserve, as described below, all other reserves are normally available to meet real-time energy needs if scheduled generation is not sufficient to supply market demand.[18] Thus, the real-time energy price is still set by the interaction of real-time energy demand—including quantities supplied by reserve capacity—and all of the generators that can provide real-time supply. Therefore, we consider the real-time energy demand in each hour to be the quantity that must be supplied, and capacity selected for reserve services to be part of the capacity that can meet that demand and, as such, to be part of our aggregate marginal cost curve.

Unlike the other forms of reserve, regulation capacity is, in a way, held out of the imbalance energy market and its capacity could therefore be considered to be unavailable for additional supply. For this reason we add the *upward* regulation reserve requirement to the market-clearing quantity for the purposes of finding the overall marginal cost of supply.[19]

B. *Marginal Cost of Fossil-Fuel Generating Units*

To estimate the marginal cost of production for an efficient market, we divide production into three economic categories: reservoir, must-take, and fossil-fuel generation. Reservoir generation includes hydroelectric and geothermal production. These facilities differ from all others in that they face a binding intertemporal constraint on total production, which implies an opportunity cost of production that generally

[17] See Borenstein et al. (2001). There is also a fundamental theoretical flaw in this argument. Though option value would cause a firm to offer power in the day-ahead market at a price above its marginal cost, arbitrage on the demand side (and by sellers that do not qualify to provide ancillary services) would still equalize the market prices. The equilibrium outcome would just have a reduced share of power sold through the day-ahead market due to the forgone option value.

[18] In other words, all reserve capacity that is economic at the market price is assumed to be used to meet energy demands in real time. Due to reliability concerns, the ISO occasionally has not utilized some types of reserve ("spinning" and "nonspinning") for the provision of imbalance energy even when the units are economic (see Wolak et al., 1998). The conditions under which this occurs are somewhat irregular and difficult to predict. For the purposes of this analysis we have assumed that these forms of reserve are utilized for the provision of imbalance energy.

[19] Regulation reserve is procured for both an upward (increasing) and downward (decreasing) range of capacity. The amount of upward regulation reserve at times reached as high as 10.8 percent of total ISO demand, although the mean percentage of upward regulation was 2.2 percent over our sample. Because the generation units that are providing *downward* regulation are, by definition, producing energy, the capacity providing downward regulation should not be considered to be held out of the energy market. Note also that by adding regulation needs to the market demand, we are implicitly assuming that all regulation requirements are met by generation units with costs below the market-clearing price. To the extent that some units providing regulation would not be economic at the market price, this assumption will tend to bias downward our estimate of the amount of market power exercised.

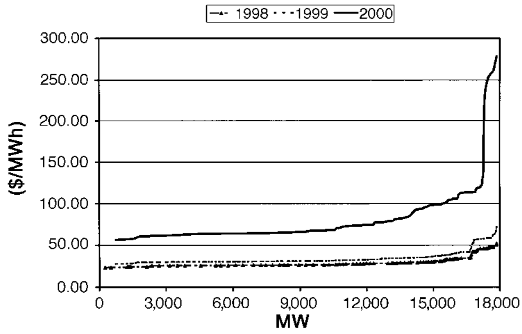

FIGURE 1. CALIFORNIA FOSSIL-FUEL PLANTS MARGINAL COST CURVES, SEPTEMBER

exceeds the direct production cost. Must-take generation operates under a regulatory side agreement and is always inframarginal to the market. Because of the incentives in the regulatory agreements, these units will always operate when they physically can. All nuclear facilities are must-take, as well as all wind and solar electricity production. We discuss below our treatment of reservoir and must-take generation.

For fossil-fuel generation, we estimate marginal cost using the fuel costs and generator efficiency ("heat rate") of each generating unit, as well as the variable operating and maintenance (O&M) cost of each unit. For units under the jurisdiction of the South Coast Air Quality Management District (SCAQMD) in southern California, we also include the cost of NOx emissions, which are regulated under a tradeable emissions permit system within SCAQMD. The cost of NOx permits was not significant in 1998 and 1999, but rose sharply during the summer of 2000. The generator cost estimates are detailed in Appendix A. Figure 1 illustrates the aggregate marginal cost curve for fossil-fuel generation plants located in the ISO control area that are not considered to be must-take generation and shows how it increased between 1998 and 2000 due to higher fuel and environmental costs.[20] Note that because the higher-cost plants tend to be the least fuel efficient and the heaviest polluters, increases in fuel costs and pollution permits not only shift the supply curve, but also increase its slope since the costs of high-cost plants increase by more than the costs of low-cost plants.[21]

The supply curves illustrated in Figure 1 do not include any adjustments for "forced outages." Generation unit forced (as opposed to scheduled) outages have traditionally been treated as random, independent events that, at any given moment, may occur according to a probability specified by that unit's forced outage factor. In our analysis, each generation unit, i, is assigned a constant marginal cost mc_i—reflecting that unit's average heat rate, fuel price, and its variable O&M cost—as well as a maximum output capacity, cap_i. Each unit also has a forced outage factor, fof_i, which represents the probability of an unplanned outage in any given hour.

Because the aggregate marginal cost curve is convex, estimating aggregate marginal cost using the expected capacity of each unit, $cap_i \cdot (1 - fof_i)$, would understate the actual expected cost at any given output level.[22] We therefore simulate the marginal cost curve that accounts for forced outages using Monte Carlo simulation methods. If the generation units $i = 1, \ldots, N$ are ordered according to increasing marginal cost, the aggregate marginal cost curve produced by the jth draw of this simulation, $C_j(q)$, is the marginal cost of the kth cheapest generating unit, where k is determined by

$$(1) \quad k = \arg\min \left\{ x \,\middle|\, \sum_{i=1}^{x} I(i) \cdot cap_i \geq q \right\}.$$

[20] Costs of generation shown in Figure 1 are based on monthly average natural gas and emissions permit prices. Here and throughout our analysis, we assume that gas prices are competitively determined and accurately reported. If gas markets were not competitive or reported prices exceeded the actual prices paid by electricity generators, as recent Federal Energy Regulatory Commission findings have suggested, then we have underestimated the full impact of market power on the wholesale price of electricity.

[21] Our estimates assume that the rated capacities of the plants, cap_i, are strictly binding constraints. It has been pointed out to us that the plants can be run above rated capacity, but at the cost of increased wear and more frequent maintenance. If we incorporated this factor—about which there seems to be very little detailed information—it would shift rightward the industry supply curve and would increase our estimates of the extent to which market power was exercised.

[22] For any convex function $C(q)$, of a random variable q, we have, by Jensen's inequality, $E(C(q)) \geq C(E(q))$.

where $I(i)$ is an indicator variable that takes the value of 1 with probability of $1 - fof_i$, and 0 otherwise. For each hour, the Monte Carlo simulation of each unit's outage probability is repeated 100 times. In other words, for each iteration, the availability of each unit is based upon a random draw that is performed independently for each unit according to that unit's forced outage factor. The marginal cost at a given quantity for that iteration is then the marginal cost of the last available unit necessary to meet that quantity given the unavailability of those units that have randomly suffered forced outages in that iteration of the simulation. If, during a given iteration, the fossil-fuel demand (total demand minus hydro, must-take, and the supply of imports at the price cap) exceeded available capacity, the price was set to the maximum allowed under the ISO imbalance energy price cap during that period. Note that in nearly all cases, the PX price in that hour did not hit the price cap, so such outcomes were counted as "negative market power" outcomes in the analysis. Thus, these outcomes are not driving, and if anything are reducing, our finding of market power.

We did not adjust the output of generation units for actual outages, because the scheduling and duration of planned outages for maintenance and other activities is itself a strategic decision. Wolak and Patrick (1997) present evidence that the timing of such outages was extremely profitable for certain firms in the U.K. electricity market. It would therefore be inappropriate to treat such decisions as random events. Because we find market power in the summer months—high-demand periods in California in which the utilities have historically avoided scheduled maintenance on most generation—it is unlikely that scheduled maintenance could explain these results in any case.[23] We would expect scheduled maintenance to take place in the autumn, winter, and spring months, which is the time period over which we find little, if any, market power.

The operation of generation units entails other costs in addition to the fuel and short-run operating expenses. It is clear that sunk costs, such as capital costs, and periodic fixed capital and maintenance expenses should not be included in any estimate of short-run marginal cost. More difficult are the impacts of various unit-commitment costs and constraints, such as the cost of starting up a plant, the maximum rates at which a plant's output can be ramped up and down, and the minimum time periods for which a plant can be on or off. These constraints create nonconvexities in the production cost functions of firms. For a generating unit that is not operating, these costs are clearly not sunk. On the other hand, it is not at all clear how, or whether, a price-taking, profit-maximizing firm would incorporate such costs into its supply bid for a given hour. In fact, it is relatively easy to construct examples where it would clearly *not* be optimal to incorporate start-up costs in a supply bid.[24] We do not attempt to capture directly the impacts of these constraints on our cost estimates. Below, we discuss how nonconvexities could affect the interpretation of our results.

C. *Imports and Exports*

One of the most challenging aspects of estimating the marginal cost of meeting total demand in the ISO system is accounting for imports and exports between the ISO and other control areas. We can, however, observe the net quantity of power entering or leaving the ISO system at each intertie point, as well as the willingness of firms to import and export to and from California.

If the power market *outside* of California

[23] Scheduled maintenance on must-take resources and reservoir energy sources was accounted for under the procedures outlined in the following subsections.

[24] Consider a generator that estimates it will be "in" the market for six hours on a given day and bids into the market in each hour at a level equal to its fuel costs plus one-sixth of its start-up cost. Consider the results if the market price in one hour rises to a level sufficient to recover all start-up costs, but in all subsequent hours remains at a level above the unit's fuel costs, but below the sum of its fuel cost plus the prorated start-up costs. If this unit committed to operate in the one hour that it covers its fuel costs plus one-sixth of its start-up costs, but stayed "out" of the market in subsequent hours, it is not maximizing profits, because it could have earned an operating profit at market-clearing prices in the five remaining hours.

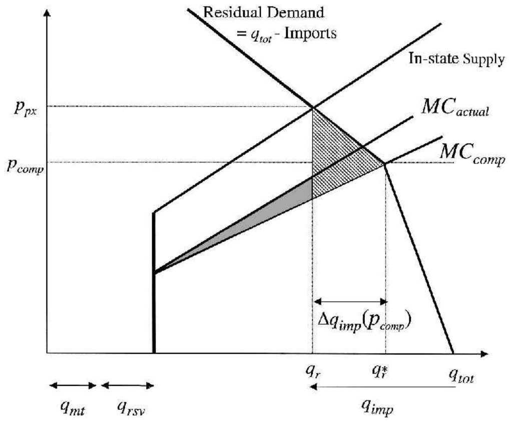

FIGURE 2. IMPORT ADJUSTMENTS AND EFFICIENCY LOSSES

were perfectly competitive, then the marginal generator that is importing into California would, absent transmission constraints, have a marginal cost about equal to the market price in California. When market power is exercised within California, this would mean that, in an effort to drive up price, some in-state generators are withdrawing (or raising the offer price on) their marginal generation and allowing more expensive imported power to be substituted for it. Thus, in the absence of market power, we would see lower imports. This means that the cost of serving the demand that remains after the competitive level of imports is netted out would be higher than the cost of serving the demand that remains after the true level of imports is adjusted for.[25]

Figure 2 illustrates a hypothetical marginal cost curve of the in-state generation, excluding must-take (q_{mt}) and reservoir energy resources (q_{rsv}). The market demand is q_{tot}, and the observed price is P_{px}. At a price of P_{px}, we see imports of q_{imp} ($=q_{tot} - q_r$) that shift the remaining demand to the left to a quantity q_r. If the price were instead set at the competitive price of P_{comp}, we would see imports at some level less than or equal to those seen at P_{px}. This would shift the residual in-state demand to a quantity q_r^*. Thus, in order to estimate the price-taking outcome in the market, we need to estimate the net import or net export supply function.

Estimating the Net Import/Export Supply Functions.—One of the primary responsibilities of the California ISO is to ensure the reliable usage of the system's transmission network. This requires that the ISO operate a market for rationing transmission capacity when its use is oversubscribed. This market is implemented through the use of schedule "adjustment" bids, which are submitted by scheduling coordinators to the ISO along with their preferred day-ahead schedules.

Scheduling coordinators submit their preferred import or export quantities and the ISO checks to see whether these flows exceed transmission capacity limits. If these proposed power flows are feasible, no further adjustments are required. In the event that the net of proposed import and export schedules does exceed transmission capacity on some intertie, the ISO initiates a process of congestion relief by adjusting schedules according to their adjustment bids. Adjustment bids establish, for each scheduling coordinator, a willingness-to-pay for transmission usage. Schedules are adjusted according to these values of transmission usage, starting at the lowest value, until the congestion along the intertie is relieved. A uniform price for transmission usage, paid by all SCs using the intertie, is set at the last, or highest, value of transmission usage bid by an SC whose usage was curtailed.

Adjustment bids reveal the willingness-to-supply imported energy of out-of-state suppliers (and exported energy of in-state suppliers) at each intertie over a wide range of quantities, not just at the observed net import/export quantity. For the vast majority of hours the aggregate net flow is into California for the relevant price range, so we refer to this as the import supply curve, but negative import supply (net export) is possible. Let the import supply curve of scheduling coordinator sc at import zone z be the net

[25] Capacity constraints on both the transmission interties into California and the production capacity of non-Californian producers complicate this intuition somewhat. If such a capacity constraint were binding at the observed California market-clearing price, then the marginal *production* cost of imports would most likely be below this market-clearing price and, thus, a perfectly competitive price within California would yield only weakly lower imports.

of its preferred import quantity and all of its incremental and decremental adjustment bids *into California* from z.

$$q_z^{sc}(p) = q_z^{sc,init} + \sum_{\hat{p}<p} q_z^{sc,inc}(\hat{p}) - \sum_{\hat{p}>p} q_z^{sc,dec}(\hat{p}). \tag{2}$$

In other words, the preferred level of imports from sc at z at a price of p, would be its scheduled imports, which are independent of price, plus the amount of *additional* supply it is willing to provide in exchange for receiving a payment less than or equal to p, minus the amount of *reduction* in supply that it would agree to in exchange for making a payment that is greater than or equal to p. The aggregate net import curve into the California ISO system for any hour can be calculated by summing the value of $q_z^{sc}(p)$ over all interties and SCs:

$$q_{imp}(p) = \sum_{sc} \sum_{z} q_z^{sc}(p). \tag{3}$$

This aggregation constitutes an upper bound on the responsiveness of net imports to changes in the California price. The ISO is in practice prevented from substituting import adjustment bids across individual scheduling coordinators or across transmission interties, so that the actual import supply curve will be a significantly steeper function of price than the curve constructed as described. The ISO will only act in the event that the initial schedules indicate that congestion will arise, even though the adjustment bids may indicate a potential Pareto-improving import adjustment. Thus, while our aggregate import supply curve assumes that all imports from all locations are perfect substitutes, and that these imports are priced at marginal cost, reality falls short of this level of import efficiency.[26]

Our analysis assumes that wholesale electricity suppliers outside of California are price-takers, so that the import supply curve represents the aggregate marginal cost curve of suppliers outside of California net of their native load obligations. Some observers have argued that the suppliers of electricity outside of California may exercise market power (i.e., offer power at above their marginal cost) when selling into the California market. If this is the case, then an import supply curve that reflected no market power from out-of-state suppliers would indicate a greater supply of imports at every price and therefore a leftward shift in the residual demand curve in Figure 6, which would lower the competitive benchmark price. Thus, our treatment of imports will tend to bias downward our estimates of the extent of market power.

D. *Hydroelectric and Geothermal Generation*

Reservoir generation units (i.e., hydro and geothermal units) present a different challenge because the concern is not over a change in aggregate output relative to observed levels but rather a *reallocation* over time of the limited energy that is available to them. Thus, the bids of hydro units do not reflect a production cost but rather the opportunity cost of using the hydro energy at some later time.

In the case of a hydro firm that is exercising market power, this opportunity cost would also include a component reflecting that firm's ability to impact prices in different hours.[27] It is important to note that even the actual observed bid prices of a small, price-taking hydroelectric firm operating in an oligopoly market would provide little information about its opportunity cost of the energy if the entire market were perfectly competitive, because the actual opportunity cost of water for these units will be influenced by the expectation of future prices,

[26] One consequence of this is that the import quantities implied by the aggregate of the adjustment bids do not exactly equal the imports that are actually observed. To realign the import supply curve implied by the adjustment bids with the observed import-price pair for each hour, we calculate the change in imports in each hour as $\Delta q_{imp}(p) = q_{imp}(p) - q_{imp}(p_{actual})$, where p_{actual} is market price during the hour under consideration. This adjustment ensures that at prices equal to the actual observed price for that hour, there would be no change from the observed level in imports when performing our counterfactual price calculation. A positive $\Delta q_{imp}(p)$ implies an increase in net imports relative to the observed price.

[27] See Bushnell (2003).

which in turn will be impacted by the ability of other firms to raise those prices.

For these reasons, we make the assumption that the actual, observed output of these resources is the output that would be produced by a price-taking firm participating in a perfectly competitive market. In practice, this assumption means that in constructing our estimate of the marginal cost of meeting load in any given hour we apply the observed production of hydro and geothermal resources for each hour and then calculate the marginal cost of satisfying the remaining demand with the state's fossil-fuel resources. We give the intuition here for why this assumption biases downward our estimates of productive inefficiency and market power. A more detailed explanation is in Appendix C.

For the purpose of calculating the impact of market power on total production cost, it is easy to see that this is a conservative assumption, one that will produce downward-biased estimates on the efficiency effects of market power. The optimal hydro schedule will, by definition, lead to weakly lower production cost than any other hydro schedule. To the extent that actual production differed from the optimal schedule, it could only raise total production cost. Thus our assumption will bias upward our estimate of perfectly competitive production cost.

For the purpose of measuring market power, we need to consider the impact of our assumption on our estimates of marginal production cost. Of concern is the possibility that the observed hydro schedule (which may include a response by hydro firms to the exercise of market power by others)—when combined with a counterfactual perfectly competitive production of fossil-fuel resources—could produce a *lower marginal cost* estimate on average than the optimal hydro schedule. However, it is straightforward to show that when system marginal production costs from nonhydro sources are convex in quantity, any reallocation of hydro energy away from the least-cost allocation will raise marginal costs more in the hours from which energy is removed than it will reduce marginal cost in the hours to which energy is added.[28] Thus our assumption of optimal hydro production can only bias our time-weighted estimates of marginal cost upwards, and therefore our estimates of price-cost margins downward.

We present results in which price-cost margins are weighted by the market volumes in each hour. To consider the effect of our hydro assumptions on these results, we need to address the possibility of a reallocation of hydro energy between off-peak and peak hours relative to the optimal schedule. A hydro firm that is attempting to exercise market power would likely allocate less hydro energy during peak hours than would be the case for a price-taking firm (see Bushnell, 2003). This strategic hydro allocation, when combined with competitive fossil-fuel production, would produce a higher weighted average of marginal cost than would the optimal schedule. To the extent the firms controlling hydro resources attempted to exercise market power with those resources, our results will therefore understate the overall level of market power.

However, the vast majority of reservoir resources were controlled by the PG&E and SCE, each of which had a fairly strong incentive to lower wholesale power costs. Therefore it is possible that these firms responded to an increase in market power with an overconcentration, relative to perfect competition, of energy during high-demand periods. As argued above, this reallocation (if allowed by the flow constraints) would raise off-peak marginal costs more than it would lower on-peak marginal costs. However, since (non-must-take) market volumes are likely to be higher on-peak, the impact on the quantity-weighted average of marginal cost is uncertain.

We examined this issue empirically by asking whether our estimates of marginal costs produce opportunities for a reallocation of hydro energy that would result in a lower weighted-average marginal cost. Such an opportunity would exist if fossil-fuel marginal costs during some high-demand period were lower (due to "overproduction" from hydro sources) than the marginal cost in some lower-demand period. If, by contrast,

[28] This is because at the least-cost allocation of hydro energy, marginal fossil-fuel costs will be equalized over all hours for which hydro flow constraints allow a discretionary use of hydro energy.

our marginal cost estimates (over a period with stable input prices) were monotonic in market demand, then no systematic opportunity for lowering the weighted average of marginal cost exists and any bias from our hydro assumption is in the direction of raising costs and lowering market power.

To examine this possibility, we estimated a kernel regression of our estimated marginal cost (i.e., competitive price) on system demand (presented in Appendix C) in order to detect whether in aggregate there are systematic deviations from a monotonically increasing relationship between demand and our estimate of system marginal cost. Such regressions for each of the three summers in our sample period show that our system marginal cost estimates were monotonically increasing in demand for each of these time periods. This leads us to conclude that it is highly unlikely that our assumption that the actual schedule of production reservoir resources was the cost-minimizing schedule creates a significant negative bias on the weighted-average estimates of system marginal costs.

E. *Calculating Cost Increase Relative to Competitive Outcome*

Utilizing the assumptions outlined in the previous sections, we estimated the perfectly competitive market price in the California energy markets for each hour of market operation from June 1998 through October 2000. The residual market demand to be met by in-state fossil-fuel units within the ISO system in hour t, q_{ff}^t, is estimated to be

$$q_{ff}^t(p) = q_{tot}^t + q_{reg}^t - q_{mt}^t - q_{rsv}^t - q_{imp_{act}}^t - \Delta q_{imp}^t(p). \tag{4}$$

Here q_{tot}^t is the actual ISO metered generation (including net imports), including generation scheduled through all energy markets associated with the ISO control area, including the PX, ISO imbalance energy market, and other SCs. q_{reg}^t represents the addition to demand due to the need for capacity dedicated to regulation reserve. The quantities q_{mt}^t and q_{rsv}^t represent the amount of energy produced by must-take generation and by reservoir generation, respectively. These quantities are all treated as price inelastic. The term $q_{imp_{act}}^t - \Delta q_{imp}^t(p)$ is the level of imported energy adjusted by the response to changes in the market-clearing price, as described above.

For each hour, we make 100 fossil-fuel generation marginal cost curve estimates, each reflecting a combination of independent Monte Carlo draws for the outage of each generation unit. For each of these draws from the system-wide fossil-fuel marginal cost curves we compute the intersection of this marginal cost curve with the residual market demand curve $q_{ff}^t(p)$. This yields an estimated marginal cost and an in-state market-clearing quantity q_{rj}^t for Monte Carlo draw j. We denote the marginal cost associated with this quantity as C_j^t. We can then compute an estimate of the expected value of the marginal cost of meeting the in-state demand that results from price-taking behavior by in-state generators as:

$$\bar{P}_{comp}^t = \frac{\sum_{j=1}^{100} (C_j^t)}{100}. \tag{5}$$

Note that there are cases in which $P_{px}^t - \bar{P}_{comp}^t$ is negative in our simulations. Absent an operational error or an attempt at predatory pricing, firms will not actually be willing to sell power at prices below their *true* economic short-run marginal costs. In other words, prices will not be below the perfectly competitive price. Nonetheless, during some hours, particularly June 1998 and during the winter and spring of 1999, PX prices were below our estimates of the perfectly competitive market price. At least three factors contribute to these outcomes.

First, our cost estimates can exceed the actual marginal cost because we do not consider the dynamic effects of unit commitment constraints, such as start-up costs, ramping rates, and minimum down times. These constraints can create opportunity costs of shutting down units that, in essence, lower the true marginal cost of operating that plant. Of course these same constraints also can create opportunity costs that, at other times, raise the true marginal cost. This is one reason why we include the negative markups in our results; we did not want to exclude the off-peak impact of these constraints on our cost estimates, since there is an opposite effect on our estimates during peak hours.

Second, cost information for generating units are not exact data on which all parties agree. For the most part, we used values submitted to state and federal regulatory agencies under the former regulated regime. For this reason, our estimate of a unit's marginal cost may be slightly higher than the cost level at which it is capable of operating in a market environment. Therefore we include negative price-cost differences in order to prevent truncating the effect of data uncertainty on our cost estimates.

Third, and probably most important, our calculations do not control separately for the output levels of *reliability must-run* (RMR) generation. Some fossil-fuel generation units have been declared must-run for *local* grid reliability under certain system conditions. These generators get separate nonmarket payments when they are called under the RMR contracts they have signed with the ISO. RMR units are not dispatched through the price-setting process. Because they are held out and paid a different price, the resulting price in the PX can be below the marginal cost to the system if the power provided by RMR units were instead provided as part of the full dispatch of the system. In fact, due to the level of RMR calls by the ISO during some periods during our sample, particularly the spring of each year, it is possible that no other fossil-fuel generation was economic during these time periods. Under these circumstances, the highest (opportunity) cost units selling in the PX could be hydro or out-of-state coal plants, either of which have lower marginal cost than any of the fossil-fuel plants we examine. However, these periods are likely to occur when the PX price is extremely low, not extremely high. In such cases, import energy with costs below those of in-state fossil-fuel generation could be the marginal generation, and the actual PX price could be lower than the marginal costs of any of the fossil-fuel units we have examined. Because we don't account for the RMR units, our estimates could still indicate that a fossil-fuel unit is marginal and its cost is the system marginal cost, so our estimated system marginal cost would be above the actual PX price due to unaccounted for RMR calls.[29]

If the estimated MC is above the PX price for either the first or second reason, then it seems that the most accurate estimate of market power would come from including the "negative market power" outcomes in our calculations. However, total start-up costs for the fossil-fuel units in California are about \$39 million during our sample period, less than 1 percent of total fossil-fuel generation production costs during the period and less than 1 percent of the market power rents we find.[30] In addition, there are other reasons to think that start-up costs explain only a minor part of the deviations from marginal cost pricing. First, the units that turn on to meet peak demand during the summer have little start-up costs (fuel oil or jet fuel units) or none at all (hydroelectric units), so the impact at the times we find the greatest market power is likely to be low.[31] Second, our estimates of market power are substantially greater in summer 2000 than in summer 1999, but the amount of electricity produced per start-up is 5.7 percent *lower* in summer 1999, implying that start-up costs would likely be a greater factor in 1999 than in 2000. Similarly, the ratio of start-up costs to our estimated fossil-fuel production costs was

[29] This implies that neglecting RMR calls could underestimate market power. In addition, it appears that the initial RMR agreements exacerbated some of the *local* market power problems that they were designed to mitigate. See Bushnell and Wolak (1999).

[30] For 65 of the 92 units in our fossil-fuel cost curve, the unit-specific formulae for determining the cost per start-up (as a function of both input fuel costs and the price of electricity) were submitted to the ISO as part of the RMR contract renegotiation process. For the remaining 27 units, we estimated the unit-specific start-up formula by using parameters from a similar unit that did have an RMR contract. We used the daily price for the input fuel used by that unit and daily average annual retail price to industrial customers for the electricity price in the start-up cost formula.

[31] Additionally, if marginal cost functions turn upward smoothly around the rated capacity, rather than having a strict L-shape, the typical argument that a competitive plant would bid its start-up costs for the "single hour" it would run are incorrect. In that case, even the last plant turned on would run for many hours because it would be replacing higher-cost output from other plants that would otherwise be producing along the steepest parts of their MC curves.

higher in summer 1999 than in summer 2000, 0.91 percent versus 0.37 percent.

Likewise, it is unlikely that much of the negative market power outcomes could be the result of cost-data errors. Many PX prices in June 1998, for instance, were well below the costs that anyone has claimed for operation of fossil-fuel generating units.[32] Thus, it is most likely that the cost estimates that exceed the PX price occur because there were no fossil-fuel generating units that were economic to run at the time. Only fossil-fuel units running under RMR contracts were active. In that case, the marginal cost of the system, and thus the market price, is being set by much cheaper out-of-state coal plants, by nuclear plants, or by hydro or geothermal plants. If this is the case, then the proper treatment would be to truncate the results, resetting any finding of "negative market power" to set marginal cost equal to price. Still, in order to avoid biasing the results in favor of finding market power, we do not truncate the negative outcomes in the primary results we report.

IV. Results

We computed the expected perfectly competitive price each hour for the months of June 1998 through October 2000 using the algorithm described above. From the import adjustment bids, the median hourly reduction in imports from the observed level at the PX price versus the level at our estimated competitive price was 2.4 percent. For each hour, we can calculate an arc elasticity implied by the adjustment bids for the import response from the change between the competitive and actual price and the resulting change in imports. The median arc elasticity of import supply for these hours is 0.63.[33]

The added wholesale cost of energy due to departures from a competitive market, ΔTC, is calculated by taking the difference between the PX price and our estimate of competitive benchmark price and multiplying it by the total ISO metered generation less the must-take energy for that hour.[34] That is, for hour t,

$$(6)\quad \Delta TC^t = [P^t_{px} - \bar{P}^t_{comp}] \cdot [q^t_{tot} - q^t_{mt}],$$

where $\bar{P}^t_{comp}$ is the expected competitive price in period t. This expectation is taken with respect to the distribution of generating unit outages, as shown in (5).

For any set of hours $\mathcal{S}$, our measure of market performance is

$$(7)\qquad MP(\mathcal{S}) = \frac{\sum_{t \in \mathcal{S}} \Delta TC^t}{\sum_{t \in \mathcal{S}} TC^t}.$$

$MP(\mathcal{S})$ is the proportional increased wholesale cost of electricity during all hours in $\mathcal{S}$. Defining $MP(\mathcal{S})$ in this manner is consistent with the view, reflected in our competitive benchmark Monte Carlo simulation, that the observed market price is conditional on a realization from the joint distribution of generating unit outages. To reflect this fact, let $\hat{P}^t_{px}$ denote the observed PX price for hour t and $E(\hat{P}^t_{px})$ the expectation of this magnitude with respect to the joint distribution of generating unit outages. Unlike the counterfactual case of price-taking behavior, we cannot draw from the distribution of generating unit outages and compute a distribution of market prices that reflect the current level of market power. This would require a model for the strategic interaction among players in the California market. However, by defining $MP(\mathcal{S})$ as shown in equation (7), we can take advantage of the law of large numbers to prove that our measure is a consistent estimate of the proportional cost increase. To show this, rewrite the index as:

$$(7')\quad MP(\mathcal{S})$$

$$= \frac{1/Card(\mathcal{S}) \sum_{t \in \mathcal{S}} [\hat{P}^t_{px} - \bar{P}^t_{comp}] \cdot [q^t_{tot} - q^t_{mt}]}{1/Card(\mathcal{S}) \sum_{t \in \mathcal{S}} \hat{P}^t_{px} \cdot [q^t_{tot} - q^t_{mt}]},$$

[32] If we were to ignore any "negative market power" outcomes for prices below, say, \$18/MWh, virtually all of the "negative market power" effects would be eliminated.

[33] We calculate the arc elasticity as $\dfrac{(P_1 + P_2)/2}{(Q_1 + Q_2)/2}\,\dfrac{(Q_2 - Q_1)}{(P_2 - P_1)}$.

[34] By taking the observed quantity as the market demand, we are, for the reasons discussed earlier, implicitly assuming that demand is price inelastic.

TABLE 2—ACTUAL PRICE AND ESTIMATED MARGINAL COST

Month	Year	Mean of Actual Production Per Hour (MWh)	Mean of PX Price ($/MWh)	Mean of Marginal Cost ($/MWh)	Sum of ΔTC ($ million)	Aggregate $\Delta TC/TC$ (percent)
June	1998	24,134	12.09	22.55	−44	−51
July	1998	28,503	32.41	27.33	103	28
August	1998	31,256	39.53	27.71	220	39
September	1998	28,209	34.01	26.28	134	33
October	1998	25,043	26.65	26.21	13	5
November	1998	24,107	25.74	27.53	−4	−2
December	1998	24,953	29.13	25.40	45	17
January	1999	24,480	20.96	22.41	−5	−2
February	1999	24,079	19.03	21.20	−12	−7
March	1999	24,734	18.83	20.80	−12	−7
April	1999	24,763	24.05	24.50	4	2
May	1999	24,625	23.61	25.34	0	0
June	1999	27,081	23.52	25.89	13	5
July	1999	29,524	28.92	27.12	63	17
August	1999	29,813	32.31	30.64	56	14
September	1999	28,573	33.91	30.25	63	16
October	1999	27,558	47.63	34.38	186	31
November	1999	26,046	36.91	28.87	105	26
December	1999	26,647	29.66	27.73	30	9
January	2000	26,377	31.18	27.66	48	13
February	2000	25,961	30.04	29.52	10	3
March	2000	25,618	28.80	31.38	−17	−6
April	2000	25,728	26.60	32.43	−43	−16
May	2000	27,038	47.22	40.43	150	25
June	2000	30,644	120.20	53.59	1,152	63
July	2000	30,343	105.72	59.37	801	50
August	2000	32,310	166.24	76.19	1,475	56
September	2000	29,981	114.87	76.86	577	36
October	2000	27,422	101.51	68.06	443	34

where $Card(\mathcal{S})$ is the cardinality or number of elements (hours) in the set $\mathcal{S}$. For sets $\mathcal{S}$ with a large number of elements, the index is approximately equal to

$$(7'') \quad MP(\mathcal{S}) = \frac{\sum_{t \in \mathcal{S}} [E(\hat{P}^t_{px}) - \bar{P}^t_{comp}] \cdot [q^t_{tot} - q^t_{mt}]}{\sum_{t \in \mathcal{S}} E(\hat{P}^t_{px}) \cdot [q^t_{tot} - q^t_{mt}]}$$

which is equal to the ratio of the expected cost increase relative to the perfectly competitive benchmark, due to the current level of market power and market imperfections, divided by the expected cost of purchasing electricity under current market conditions.

Table 2 reports the PX price, estimated marginal cost, and the added cost of power due to prices that exceeded marginal cost for each month in the sample period. As is evident from Table 2, June 1998 produced very idiosyncratic results, with an average PX price considerably below our estimate of marginal cost. The market was only in its third full month of operation at this time and a number of fossil-fuel generation units were going through ownership transfer and regulatory approval of these transfers. The CTC mechanism provided the three investor-owned utilities with an incentive to induce low energy prices and the utilities were still operating and bidding many of these units through June 1998. As described above, the use of reliability must-run contracts, which paid some fossil-fuel units to run in exchange for payments that were above the market price, was also widespread during June. For these reasons, we believe the market results from June 1998 do not provide much meaningful information on the state of competition in the California market. Nevertheless, for completeness, we have included these results. For the set of all hours over the entire 29-month period from June 1998 to October 2000, the $MP(\mathcal{S})$ is equal to 33 percent, amounting to total payments in excess of competitive levels equal to \$5.55 billion with a standard error of \$1.19 billion.[35]

Having generated estimates of price-cost margins for each hour of the 29-month sample period, we can examine subsets of the data to gain insight into the underlying dynamics of the market. One test of the credibility of our results is whether our estimates of market power vary in the way that economics would predict. We would expect market power to be quite low during the off-peak months, December through April. Electricity demand is low in these months and supply is relatively large due to the resurgence in hydro production from winter rains. In December 1998–April 1999, we find an average $MP(\mathcal{S})$ of 1.9 percent, and in December 1999–April 2000, we find an average $MP(\mathcal{S})$ of 1.8 percent, neither of which is significantly different from zero. Thus, we find that there was essentially no margin between prices and marginal cost during the period in which supply was most abundant compared to demand and sellers had the least ability to exercise market power. In addition, these results provide evidence that significant short-run operating costs are not missing from our cost estimates, because negative or zero margins would not be observed over such an extended period of time.

The series of events that led to the California electricity crisis in 2000–2001 began with dramatic price increases during the summer of 2000. Many policy makers and regulators have argued that the competitive performance of the market fundamentally changed during summer 2000, thereby initiating the crises. In order to make such comparisons, however, one must account for differences in the relative levels of demand during these periods. As described above, we would expect the estimated market power to increase as the demand faced by in-state nonutility sellers rises relative to the capacity of these players. Figure 3 shows a kernel regression of this relationship for late summer of 1998, 1999, and 2000.[36] The horizontal axis of Figure 3 is the demand faced by these firms after accounting for actual imports, must-take, and reservoir production. The vertical axis is the ratio $\Delta TC^t/TC^t$, which is equal to the Lerner index for that hour.[37]

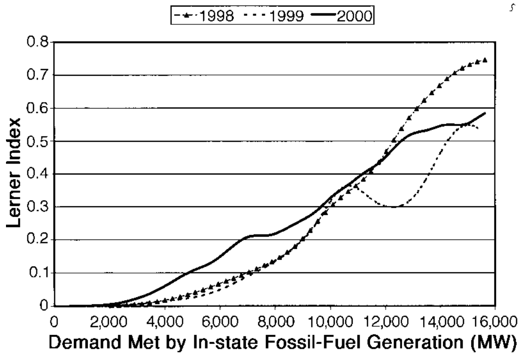

FIGURE 3. KERNEL REGRESSIONS OF LERNER INDEX, AUGUST 7–SEPTEMBER 30

The results summarized by this figure show that market power steadily increased with the demand faced by the nonutility in-state suppliers consistent with the earlier discussion of the nature of competition in the electricity industry. During lower demand hours and months, as well as springtime months when significant hydro energy is available, no single firm can affect prices significantly. During higher demand hours, however, competitive sources of energy begin to reach their capacity limits and the pool

[35] Appendix B outlines our procedure for computing this standard error, which accounts for the error associated with the randomness of forced plant outages.

[36] To be precise, the data are for August 7 through September 30 of each year. We focus on this period because the ISO energy price cap varied during our sample period, but it was set at the same level, \$250/MWh, for August 7 through September 30 of all three years. August and September are historically two of the highest demand months of the year in California, and they can exhibit the lowest supply availability due to declining hydro resources late in the summer.

[37] Because the Lerner index is not symmetric around zero, negative values of the ratio are set to zero in estimating the kernel regressions.

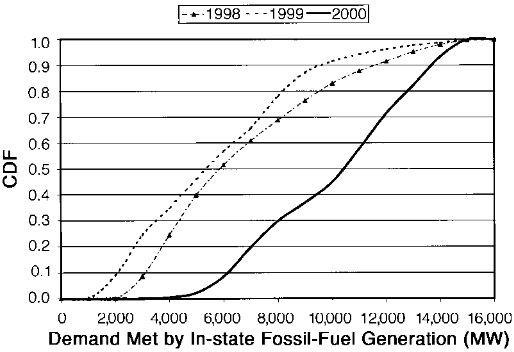

FIGURE 4. CUMULATIVE DISTRIBUTION FUNCTIONS (CDFs) OF DEMAND MET BY IN-STATE FOSSIL-FUEL PLANTS, AUGUST 7–SEPTEMBER 30

of potential competitors for additional supply dwindles. Because of the lack of significant storage capacity and the inelasticity of demand, firms can take advantage of the capacity limits of their competitors during these high-demand hours. This is consistent with the effects detected from the oligopoly equilibrium simulations in Borenstein and Bushnell (1999). This sequence of events does not imply a shortage of generating capacity to serve the energy or ancillary services needs of the California ISO control area. However, the combination of the concentration of ownership of generating assets and the level of demand did combine to create circumstances where one or more market participants recognized that their capacity was needed to meet the ISO's energy and ancillary services needs regardless of the actions of other market participants. Under these circumstances, firms find it in their unilateral interest to bid to raise prices even though there is sufficient capacity available to meet the California ISO's total energy and ancillary services requirements.

Our results also indicate that, given the supply and demand conditions during that period, the performance of the market was not dramatically different in 2000 from that in 1998 and 1999. Figure 4 shows the cumulative distribution functions for the demand met by in-state fossil-fuel generation for the late-summer period during 1998–2000. Although total market demand was only 6 percent higher during late summer 2000 than in late summer of 1999 and 5 percent higher during that period in 1998, the demand met by in-state fossil-fuel plants, nearly all of which were unregulated by 1999, increased from an average of 6,639 MWh during 1998 and 5,690 MWh during 1999 to 10,007 MWh during 2000. This is largely due to a substantial decline in imports from an average of 5,069 MWh in 1998 and 6,764 MWh in 1999 to 3,627 MWh in 2000. Thus, although the performance of the market controlling for the demand faced by in-state fossil-fuel generation did not change significantly during 2000, the distribution of this demand did change. Far more hours spent at higher residual demand levels created larger average margins during 2000. This combined with the fact that marginal costs also nearly tripled between 1999 and 2000, which meant that similar Lerner indices reflected much larger absolute dollar margins, producing extremely large wealth transfers.

V. Deadweight Loss and Rent Division

Even without a market power analysis, it is clear that the extraordinary prices that began in the summer of 2000 created large transfers of wealth. The analysis we have carried out, however, allows us to parse the changes in wholesale payments for electricity between three mutually exclusive and exhaustive categories: changes in the competitive cost of generating electricity, changes in the level of competitive inframarginal rents (which would have occurred without any market power), and changes in seller rents due to the exercise of market power. Some of the rents due to market power became profits of electricity producers or marketers, but some were dissipated in production efficiency losses: efficiency losses resulting from the operation of higher-cost production units when a firm with lower-cost production exercises market power and restricts output.

A. *Deadweight Loss*

We begin by estimating the loss in economic efficiency due to the imperfections in the market. Because the demand for electricity in the California market was effectively perfectly inelastic with respect to the wholesale market price, efficiency losses would stem primarily from the inefficient allocation of production.[38] With asymmetric producers, there will be efficiency losses from the substitution of higher-cost production from price-taking firms, or even smaller strategic firms, for the lower-cost production of the larger firms that are exercising market power.

This would be a fairly straightforward calculation if there were no imports into the state. With no imports and perfectly inelastic demand, we could simply compare the efficient production costs of a given quantity of power, using the approach described in the previous section, with the actual production cost of that quantity of power. Due to imports that vary in quantity with the exercise of market power in state, however, we also need to account for the substitution of higher-cost imports for lower-cost in-state generation.

Thus, we divide the efficiency loss into these two components: the loss due to misallocation of a given production quantity of output among the fossil-fuel plants inside the ISO system, and the loss due to misallocation of production between fossil-fuel plants within the ISO and plants outside the ISO (imports). The vast majority of power imported into California originates from regulated or publicly owned firms, and most of these firms have substantial native demand obligations. We have therefore assumed that the adjustment bids from these firms reflect the actual opportunity cost of their production (i.e., that they are price-takers). When calculating wealth transfers, this is a conservative assumption. However, when calculating the impact of market power on efficiency losses, it is not. By assuming that import bids reflect the marginal cost of the supplier, we assume that increased production from these imports due to market power exercised by firms within California creates an increase in total production cost. If the adjustment bids from firms outside of California contain margins over their own marginal cost, this margin will be counted as an efficiency loss, when it is in fact a transfer from consumers to those producers.

TABLE 3—PRODUCTION COSTS AND RENT DISTRIBUTION ($ MILLION) JUNE–OCTOBER

	1998	1999	2000
Total actual payments	1,672	2,041	8,977
Total competitive payments	1,247	1,659	4,529
Production costs—actual	759	1,006	2,774
Production costs—competitive	715	950	2,428
Competitive rents	532	708	2,101
Oligopoly rents	425	382	4,448
Oligopoly inefficiency—in state	31	31	126
Oligopoly inefficiency—imports	13	24	221

The two components of deadweight loss are illustrated in Figure 2. The inefficiency from the reallocation of the actual production quantity among fossil-fuel resources inside the ISO is illustrated by the solid gray area between the competitive marginal cost curve and the "actual" marginal cost curve just above it.[39] Our estimates of the total in-state fossil-fuel production inefficiency for June through October of 1998, 1999, and 2000 are shown in Table 3. The expected cost from additional imports is illustrated in the striped area of Figure 2 and reflects the difference between producing the quantity $\Delta q^t_{imp}(p_{comp})$ from imported production and producing that same quantity from in-state production along the marginal cost curve MC_{comp}. Again, we have assumed that the adjustment bids of importing firms reflect their actual marginal production costs. Our estimates of the total production inefficiency due to higher-than-optimal imports for the summer months of our sample are shown in Table 3.

In Figure 5, we illustrate the relationship between our estimated in-state productive inefficiency and aggregate demand faced by California fossil-fuel plants using a kernel density regression. Given our findings in the previous section, it is not surprising that we observe low levels of

38 This is not true if the exercise of market power caused some "interruptible customers"—customers that have agreed to curtail consumption upon request from the utility in return for lower electricity rates overall—to significantly reduce their demand. This happened on 26 occasions during our sample (5 in 1998, 1 in 1999 and 20 in January–October 2000, but we have no way of estimating the deadweight loss from these demand reductions.

39 This is a rough representation since the cost difference need not rise with the quantity of in-state fossil-fuel production, as we discuss in what follows.

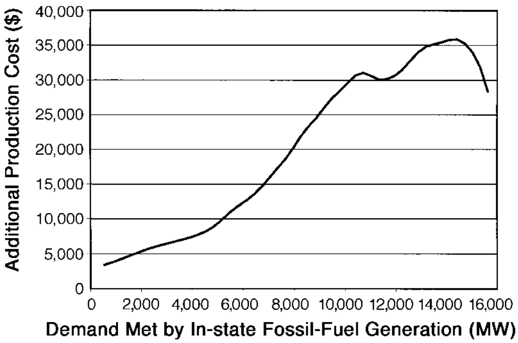

FIGURE 5. EFFICIENCY LOSSES

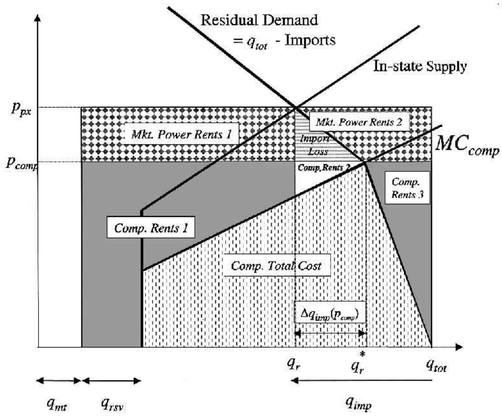

FIGURE 6. CALCULATION OF DIVISION OF RENTS

production inefficiency at low levels of system demand, when there are low levels of market power. What may be surprising at first is that productive inefficiency declines as demand nears system capacity even though our estimates of market power continue to increase. When the system is near capacity, however, very small output reductions can yield enormous price increases. Thus, while exercise of market power at those times may cause large wealth transfers, the resulting productive inefficiency is small because nearly all resources are running in any case.

B. *Rent Division*

With the calculation of deadweight loss due to productive inefficiency, we are now in a position to parse the total wholesale market payments into costs, competitive rents, and rents due to the exercise of market power. In Figure 6, the quantity $q_{tot} - q_{mt}$ represents the amount of power traded in the wholesale market in a given hour, q_{mt} being must-take power that is not compensated at the market price.

Total wholesale market payments are the sum of all the shaded areas. When the areas labeled *Mkt. Power Rents* and *Import Loss* (together the area above P_{comp}) are removed from the total, the result is the total wholesale payments that would have resulted if the market were perfectly competitive. The quantity q_{rsv} represents hydro and geothermal production during the hour. For purpose of calculating the change in rents during summer 2000 and how those rents were divided, we assume that hydro and geothermal power have zero marginal cost, though this assumption has no effect on the calculation of the change in rents.

Under competition, the quantity q_r^* is produced by in-state generation units and the quantity $q_{tot} - q_r^*$ is imported. The area labeled *Comp. Total Cost* is the variable production costs of in-state units (other than must-take production) and of imported power for their respective shares of production. Competition generates inframarginal rents equal to the sum of the areas labeled *Comp. Rents 1* and *Comp. Rents 2* for in-state fossil-fuel and reservoir generators, and the area *Comp. Rents 3* for imports.[40] Together, these areas—*Comp. Rents* and *Comp. Total Cost*—account for all wholesale market payments under perfect competition.

With market power, the quantity q_r is produced by in-state generation units and the quantity $q_{tot} - q_r$ is imported. The areas labeled *Comp. Rents 2* and *Import Loss* are the additional variable production costs of the imported power under the assumption that imports are bid competitively. In addition to the inframarginal

[40] We have not calculated the import elasticity for prices below P_{comp} so we assume that marginal costs decrease in a linear fashion and that the marginal costs of imports are zero when the import quantity is zero. The qualitative results do not change if we use the actual import bids down to zero and assume that all imports that made no (or negative price) adjustment bids had marginal costs equal to zero.

rents represented by the area *Comp. Rents 1,* in-state producers receive the area *Mkt. Power Rents 1,* but some of these rents are dissipated through inefficient production as described previously. Imports receive inframarginal rents *Comp. Rents 3* as well as the area labeled *Mkt. Power Rents 2.* Together, these areas account for all wholesale market payments with market power.

Between the summers of 1998 and 2000, the wholesale market cost of power rose from $1.67 billion to $8.98 billion. Efficient production costs more than tripled between these periods and with the marginal unit having higher costs, competitive rents for lower cost units quadrupled. Oligopoly rents, however, increased by an order of magnitude, from about $425 million to $4.44 billion between these summers. Thus, while a substantial portion of the increased market cost of power was due to rising input costs and reduced imports, these factors also increased the dollar magnitude of the market power that was exercised by suppliers. As the results in the previous section indicate, the underlying competitive structure of the market does not appear to have changed substantially between 1998 and 2000. Rather the higher demand and lower import levels in 2000 created more frequent opportunities for in-state fossil-fuel producers to collect large margins on increased costs, leading to the tenfold increase in oligopoly rents to suppliers.

The inefficiencies that resulted from the reallocation of production within California were much more modest, remaining at about 3–5 percent of total production costs through all three summers. The inefficiencies due to increased imports in power did grow substantially during our study, rising from 2 percent to 8 percent of total production costs by the summer of 2000. To the extent that prices from importing firms did not reflect their actual production costs, but their own market power, this figure will include oligopoly rents earned by producers outside of California as well as actual productive inefficiencies.

VI. Conclusions

Restructuring of electricity industries has been predicated on the belief that workably competitive wholesale electricity markets can be attained. The debate over whether that assumption is correct and what must be done to ensure competition in electricity generation is ongoing. We have attempted here to reliably estimate the degree to which California's wholesale electricity market has deviated from the competitive ideal.

Though a great deal of cost data are available for electricity generation units, we still had to make a number of assumptions in order to reach an estimate of the extent of market power in California. In most, though not all, cases, we have made assumptions that, if anything, are likely to produce results indicating less market power than actually exists.

The results indicate that market power in California's wholesale market was a significant factor during the summers of 1998, 1999, and 2000, though somewhat less so in 1999. These estimates should serve as a reminder that the problem of producer market power that was addressed in a purely regulatory framework for most of the twentieth century has not completely disappeared with the recent restructuring. Our results demonstrate that market power is most commonly exercised during peak demand periods, which is not at all surprising given the current inability of wholesale demand to respond to high hourly spot prices. This underscores the importance of designing wholesale electricity markets that maximize the likelihood that wholesale price signals will be reflected in retail electricity rates.

These estimates demonstrate the degree to which prices exceed system marginal costs, the price level that would occur if all firms behaved as competitive price-takers. We have not attempted to assess the profitability of any generation firms selling in California, because such profits are not necessarily an indication of market power, just as the absence of profits is not an indicator of competitive behavior. In all markets with durable assets, such as is the case in this industry, there are likely to be periods of high and low (or negative) profits regardless of the competitiveness of the market.[41]

[41] It is also worth noting that we have analyzed only the energy markets in California. Most generation units were

Finally, we want to emphasize again that this study is intended to develop an index of the extent and severity of market power. This is separable from the important debate over what index levels indicate a need for some form of market intervention. Years of electricity regulation confirmed the belief that government intervention can be costly and can result in very inefficient production and pricing. The balancing of the costs and benefits of such intervention will require a great deal more study in this industry as restructuring proceeds.

Appendix A: Data Sources

Fossil-Fuel Generation Data

Heat rates for fossil-fuel generation units that are not must-take and are located within the ISO control area are primarily taken from the California Energy Commission's data set on generation within the Western States Coordinating Council for use with General Electric's multi-area production cost model. This is the data set used in Borenstein and Bushnell (1999). Some unit heat rates were taken from the data set used by Southern California Gas Company in its 1995 performance-based rate-making simulation studies (Luis Pando, 1995). This data set was also used by Kahn et al. (1997) in their simulation analysis of the WSCC. Capacities of all units are the California ISO's "available capacity" figures. For NO*x* emission credit costs in the SCAQMD air basin, we used the quantity-weighted monthly average price paid for emissions credit trades registered with SCAQMD. Emissions rates of generation units are taken from the Environmental Protection Agency's Continuous Emissions Monitoring data.

An overwhelming share of California fossil-fuel generation uses natural gas as the energy source. For the time period studied, we used daily average natural gas spot prices reported by Natural Gas Intelligence at PG&E citygate and the California–Arizona border. The former were used for generation units north of path 15 while the latter were used for generation units in the south. Both sets of prices were adjusted by the distribution rates of the gas utility serving each generator.

A small number of California generators use either fuel oil or jet fuel as their primary fuel. We use the Energy Information Administration's reported monthly average Los Angeles spot price for jet fuel and no. 2 fuel oil.

Unit forced outage factors are taken from the National Electricity Reliability Council's (NERC) 1993–1997 Generating Unit Statistical Brochure, which reports aggregate generation unit performance data by fuel type and nameplate capacity. The forced outage factor that we used for our Monte Carlo simulations were derived from the NERC-reported unit Equivalent Availability Factors (EAF) and unit Scheduled Outage Factors (SOF). The former gives the fraction of total hours in which a generation unit was available, including an adjustment for partial outages, while the latter gives the fraction of hours in which each unit was unavailable due to scheduled maintenance procedures. Our derived forced outage factor (FOF), which reflects the fraction of time a unit was not available for production for unplanned reasons, was

$$FOF = 1 - \frac{EAF}{1 - SOF}. \tag{A1}$$

Demand and Generation Output Data

Total ISO quantity for every hour is based upon the ISO's real-time metered generation and is taken from ISO settlement data. The output of must-take, hydro, and geothermal generation for each hour is also taken from these data. Imports are calculated from the settlement data as metered imports minus exports aggregated over all transmission interties connecting the ISO's control area with neighboring control areas. The Mohave generation plant, although located outside of California, appears in metered data as a must-take generating facility and not as an import. Production from all

eligible to earn additional revenues under reliability must-run contracts and from the sale of ancillary services. Total ancillary services plus RMR revenues were approximately \$977 million in 1998, \$879 million in 1999, and \$1.859 billion in 2000.

other generation units owned by SCE, but located outside of California, appear as imports in the settlement data.

Appendix B: Calculation of Standard Error for $\Delta TC(\mathcal{S})$

The observed PX price in hour i and day d, $\hat{P}^{id}_{px}$, depends on a single realization of the joint distribution of generation unit-level outages during that hour. A different realization of unit-level outages for that hour could result in a very different observed PX price for that hour. Because we do not know the precise nature of the competitive interaction among market participants, specifically the bidding strategies of both generators and loads that gave rise to the observed market-clearing PX price, we are unable to replicate the actual price-setting process for a large number of draws from the joint distribution of unit-level outages in order to compute the expected value of the PX price, $E(\hat{P}^{t}_{px})$. In contrast, we can compute the perfectly competitive market counterfactual equilibrium because it entails the assumption of marginal cost bidding by all in-state fossil generation unit owners. This allows us to compute the realized value of the marginal cost of the highest cost unit operating in that hour for a large number of draws from the joint distribution of unit-level outages which we can use to compute an estimate of the expected value of this marginal cost to an arbitrary degree of precision for each hour. We found that the 100 realizations from the joint distribution of unit-level outages led to a very precise estimates of this expected marginal cost for each hour. Consequently, the remaining source of uncertainty in $\Delta TC(\mathcal{S})$, the total cost difference due to deviations from competitive prices over time period $\mathcal{S}$, that our standard error estimate accounts for is the uncertainty in PX prices caused by actual forced outages.

In constructing the standard error estimate for ΔTC, we account for arbitrary correlation in $[\hat{P}^{id}_{px} - \bar{P}^{id}_{comp}]$ across the 24 hours of the day and general forms of autocorrelation in these 24 prices across days. We can write the total cost difference due to deviations from competitive prices over time period $\mathcal{S}$ as

(B1)

$$\Delta TC(\mathcal{S}) = \sum_{d \in \mathcal{S}} \sum_{i=1}^{24} [\hat{P}^{id}_{px} - \bar{P}^{id}_{comp}][q^{id}_{tot} - q^{id}_{mt}]$$

$$= \sum_{d \in \mathcal{S}} \sum_{i=1}^{24} mkup^{id} qn^{id},$$

where $qn^{id} = [q^{id}_{tot} - q^{id}_{mt}]$ and $mkup^{id} = [\hat{P}^{id}_{px} - \bar{P}^{id}_{comp}]$. This expression can be rewritten as:

$$\text{(B2)} \quad \Delta TC(\mathcal{S}) = \sum_{d \in \mathcal{S}} \sum_{i=1}^{24} E(mkup^{id}) qn^{id} + \sum_{d \in \mathcal{S}} \sum_{i=1}^{24} \varepsilon^{id} qn^{id},$$

where $\varepsilon^{id} = mkup^{id} - E(mkup^{id})$ and $E(\cdot)$ denotes the expectation taken with respect to the joint distribution of unit-level forced outages. Therefore, the variance of $\Delta TC(\mathcal{S})$ can be written as:

$$\text{(B3)} \quad \text{Var}(\Delta TC(\mathcal{S})) = \text{Var}\left(\sum_{d \in \mathcal{S}} \sum_{i=1}^{24} \varepsilon^{id} qn^{id} \right) = \text{Var}\left(\sum_{d \in \mathcal{S}} Z_d \right),$$

where $Z_d = \sum_{i=1}^{24} \varepsilon^{id} qn^{id}$. Under suitable regularity conditions on the sequence Z_d, for example those assumed in Whitney K. Newey and Kenneth D. West (1987) or Donald W. K. Andrews (1991), we can show that $(DAY(\mathcal{S}))^{-1/2}(\sum_{d \in \mathcal{S}} Z_d)$ converges in distribution to a $\mathcal{N}(0, V)$ random variable, as $DAY(\mathcal{S})$ tends to infinity, where $DAY(\mathcal{S})$ equals the number of days in time period $\mathcal{S}$. A consistent estimate of V can be constructed as follows:

$$\text{(B4)} \quad \hat{V} = g_Z(0) + 2 \sum_{\tau=1}^{q} k(\tau/(q+1)) g_Z(\tau),$$

where $g_Z(0) = 1/DAY(\mathcal{S}) \sum_{d=1}^{DAY(\mathcal{S})} (Z_d)^2$, $g_Z(\tau) = 1/DAY(\mathcal{S}) \sum_{d=\tau+1}^{DAY(\mathcal{S})} (Z_d Z_{d-\tau})$, and $k(t)$ is a weight function satisfying restrictions given in Andrews (1991). Using this asymptotic distribution result, an estimate of the variance of $\Delta TC(\mathcal{S})$ is $(DAY(\mathcal{S})\hat{V})$, which implies a standard error of $(DAY(\mathcal{S})\hat{V})^{1/2}$.

To operationalize this procedure we need to construct an estimate of the unobservable, ε^{id}, the difference between $mkup^{id}$, and its expectation. As discussed above, in order to compute the exact expectation of $mkup^{id}$ we would need to know how all market participants bid into the PX (and other markets) as a function of current conditions in the market (system load and local reliability energy levels) and the realization of unit-level outages. Because we do not know the bidding strategies of even a single market participant and we do not observe actual generation unit-level outages during our sample period, a reduced-form approach to construct the estimate of $E(mkup^{id})$ is necessary to compute an estimate of ε^{id}. We use a linear predictor of $mkup^{id}$ constructed using hour-of-day, day-of-the-week, and month-of-sample period dummies, along with the level and square of both the forecast ISO load and day-ahead total RMR requirements for that hour of the day. Our estimate of ε^{id} is the residual from the regression of $mkup^{id}$ on these variables for all hours in our sample period. For the same reason that the squared difference between a random variable and its conditional expectation is always less than the squared difference between that random variable and its best linear predictor using those same conditioning variables, the variance of our estimate of ε^{id} should be larger than the true variance of ε^{id}. Consequently, we view our standard error as a very conservative estimate of the uncertainty in $\Delta TC(\mathcal{S})$ due to unobservable forced outages and their impact on realizations of the PX price.

Applying this procedure with the Barlett kernel, $k(s/(q+1)) = s/(q+1)$, for a value of q equal to 10 yields a standard error for $\Delta TC(\mathcal{S})$ for our entire sample period of \$1.19 billion on the estimated $\Delta TC(\mathcal{S})$ of \$5.55 billion. For the June–October periods of 1998, 1999, and 2000, the estimates of $\Delta TC(\mathcal{S})$ and their standard errors (all in \$ million) are \$425 (\$46), \$382 (\$41), and \$4,448 (\$430), respectively.

Appendix C: Cost Implications of Reservoir Energy Assumption

In this Appendix we discuss the implications of our assumption that the observed production from reservoir resources (hydroelectric and geothermal) is equal to the optimal schedule that would be produced by rational price-taking firms in a perfectly competitive market. We argue that, although this assumption is surely not true, any bias it creates in our estimates of the competitive benchmark price will be in an upward direction, leading us to understate the amount of market power. We do this by first characterizing an optimal hydro schedule, and then discussing the cost impacts of deviations from that optimal schedule.

Characterizing the Optimal Hydro Schedule.—Assume that there are n producers that control both hydro and thermal generation resources, where thermal resources include nuclear and fossil-fuel generation. Let $q_{it} = q_{it}^{Th} + q_{it}^{h}$ represent the total output of firm i in time t, where q_{it}^{Th} is the thermal output and q_{it}^{h} the hydro output of firm i. The thermal output of a firm is required to be nonnegative and also no more than its total thermal capacity, $q_{i,\max}^{Th}$.

Each producer $i = 1, \ldots, n$ has a portfolio of thermal generation technologies with an associated aggregate production cost of $C_i(q_i^{Th})$ and marginal cost of $c_i(q_i^{Th})$. We assume that $c(\cdot)$ is a strictly monotone increasing function of q^{Th}.

We can characterize the hydro systems of the suppliers as having a reservoir of $\bar{q}_i^{h}$ units of available water (measured in units of energy), $q_{i,\min}^{h}$ units of required minimum flow in each period, and a maximum flow of $q_{i,\max}^{h}$ per period. We assume that any inflows that occur during the time periods modeled (say a week or a month) do not disrupt the aggregate hydro output decisions of each firm. In other words, the limits on the total reservoir capacity are not binding during this relatively short-term planning horizon, so that any unexpected inflows are added to storage. We also assume that demand, although responsive to price and varying with time, is deterministic.

Let $p_t(Q_t)$ represent the inverse demand function for the market at time t. Given the output of the other firms, a price-taking firm i has an optimal production problem defined as

$$\text{(C1)} \qquad \text{Max}_{q_{it}^h, q_{it}^{Th}} \sum_t p_t q_{it} - C_i(q_{it}^{Th})$$

subject to the constraints

$$q_{i,\min}^h \le q_{it}^h \le q_{i,\max}^h \quad \forall t$$

$$q_{it}^{Th} \le q_{i,\max}^{Th} \quad \forall t$$

$$q_{it}^h,\ q_{it}^{Th} \ge 0\ \forall t$$

$$\sum_t q_{it}^h = \bar{q}_i^h$$

where $q_{it} = q_{it}^{Th} + q_{it}^h$, and $Q_t = \Sigma_i\, q_{it}$, the total market output in that period. Note that this problem would be separable in t, except for the last constraint, which limits the total hydro production over the T periods. We assume that each firm's single-period profit is concave in its own output. For decreasing price functions, p_t, it can be shown that this problem has a concave objective function so that the problem as a whole is convex.

In the above formulation, we do not allow for the "spilling," or free disposal, of hydro energy. Under extreme circumstances it may be profitable for a firm to withhold energy by spilling water, but this would only occur if the firm had enough reservoir quantity that it had driven marginal revenue to zero for all periods in which it was not at a maximum flow constraint.

To characterize the optimal solutions, we assign Lagrange multipliers to each of these constraints. The multipliers of interest are ψ_{it} for the thermal output limits, γ_{it} and δ_{it} for the hydro production limits, and σ_i on the total available water to the strategic hydro producer. The term σ_i is therefore this firm's *marginal value of water* in this model. This value represents the additional profit to the firm that would arise if an additional unit of water could be used for generation during the time frame of the optimization. The optimal solution is characterized by the following conditions:

$$\text{(C2)} \qquad \frac{\partial \mathcal{L}}{\partial q_{ft}^{Th}} = p_t - c_i(q_{ft}^{Th}) - \psi_{ft}$$

$$\le 0 \perp q_{ft}^{Th} \ \ge\ 0\ \forall t;$$

$$\text{(C3)} \qquad p_t = -\gamma_{ft} + \delta_{ft} + \sigma_f;$$

$$\text{(C4)} \qquad \psi_{it}(q_{it}^{Th} - q_{i,\max}^{Th}) = 0\ \forall i,t;$$

$$\text{(C5)} \qquad \gamma_{it}(q_{i,\min}^h - q_{it}^h) = 0\ \forall i,t;$$

$$\text{(C6)} \qquad \delta_{it}(q_{it}^h - q_{i,\max}^h) = 0\ \forall i,t;$$

$$\text{(C7)} \qquad \sigma_i\left(\sum_t q_{it}^h - \bar{q}_i^h\right) = 0\ \forall i;$$

$$\text{(C8)} \quad q_{it}^{Th} \le q_{i,\max}^{Th},\ q_{i,\min}^h \le q_{it}^h,$$

$$q_{it}^h \le q_{i,\max}^h, \qquad \sum_t q_{it}^h \le \bar{q}_i^h\ \forall t;$$

$$\text{(C9)} \qquad q_{it}^{Th},\ \psi_{it},\ \gamma_{it},\ \delta_{it},\ \sigma_i \ge 0\ \forall i,t$$

where the symbol $\perp$ indicates complementarity. Combining (C2) and (C3) shows that $c_i(q_{it}^{Th}) + \psi_{it} \ge -\gamma_{it} + \delta_{it} + \sigma_i$ for all t. When met with equality, conditions (C2) and (C3) represent the condition that price, p_t, equals marginal cost.

Each price-taking firm will schedule its hydro releases so as to equate its marginal costs across all periods in the time horizon, where marginal costs include a component for the shadow price of generation capacity when the capacity constraint is binding. In the hours in which there are no binding flow constraints on hydro production, prices will be set equal to the marginal value of water, σ_i, which is constant across the time periods of the planning horizon. Let $\Upsilon \in \{0, \ldots, T\}$ denote the subset of hours when neither flow constraint binds. Then,

$$\text{(C10)} \qquad p_t = \sigma = c_i(q_t^{Th}) \quad \forall t \in \Upsilon.$$

Deviations from the Optimal Hydro Schedule.—In order to evaluate the potential impact of a suboptimal hydro allocation on the equilibrium conditions, we need to consider possible deviations from the optimal hydro allocation described above. These fall into four potential categories:

1. A reallocation from a period when the maximum flow constraint was binding to a period when the minimum flow was binding;

2. A reallocation from a period when the maximum flow constraint was binding to a period in which no flow constraints were binding;
3. A reallocation from a period when no flow constraints were binding to a period in which the minimum flow constraint was binding;
4. A reallocation between periods in which no flow constraints were binding.

A reallocation of energy away from hours when the maximum flow constraint was binding under the optimal allocation can only raise marginal cost since by condition (C3), hours in which this constraint was not binding would have lower prices (marginal cost) than hours in which it was binding. The same is true for a reallocation from an hour in which no flow constraints bind to an hour in which the minimum constraint was binding. Last, by condition (C3), marginal costs in hours in which no flow constraints bind must be equal. A reallocation from one such hour to another would necessarily raise the average marginal cost if the cost curve were convex. Therefore any hydro schedule that deviates from the optimum schedule can only raise the unweighted average marginal cost of production.

To address how a suboptimal hydro schedule might impact the volume-weighted average of marginal cost over a given time period, we utilize the following result.

RESULT: *Any feasible allocation of energy that is not equal to the optimal allocation and that produces a monotonic relationship between marginal cost,* $c(q_t^{Th})$, *and total demand,* q_t, *will produce a higher demand-weighted average marginal cost than the optimal schedule.*

To prove this result, consider again the four possible types of reallocations of energy away from the optimal schedule that are listed above. Any reallocation of energy away from a higher-demand period to a lower-demand period, cases 1, 2, and 3, will clearly raise the demand-weighted average marginal cost. Similarly a reallocation away from an unconstrained hour with higher demand to an unconstrained hour with lower demand would also raise the demand-weighted average of marginal costs. However, a

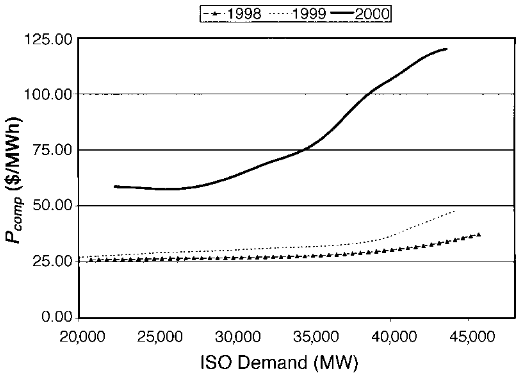

FIGURE C1. KERNEL REGRESSIONS OF COMPETITIVE PRICE VS. ISO DEMAND (JUNE–SEPTEMBER)

reallocation from an unconstrained hour with lower demand to one with higher demand would necessarily raise marginal cost in the lower demand hour to a level that is higher than the marginal cost in the next highest demand level since the marginal costs in the two hours were previously equal to each other. Such a reallocation would therefore create a nonmonotonic relationship between marginal cost and demand.

In light of the above result, we feel confident that the observed hydro schedule that we utilize in our calculations will produce a higher demand-weighted marginal cost than the unobserved optimal hydro schedule as long as we observe that our calculations of marginal cost are monotonically increasing in aggregate demand. Figure C1 illustrates this with graphs that are the result of kernel density regressions of this relationship for August–September of 1998, 1999, and 2000, respectively. In each case we do in fact observe a monotonic relationship between our estimates of marginal cost and demand, leading us to conclude that our assumption about the allocation of energy from reservoir resources is a conservative one that will understate the degree of market power that we find.

REFERENCES

Andrews, Donald W. K. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix

Estimation." *Econometrica*, May 1991, *59*(3), pp. 817–58.

Bohn, Roger; Klevorick, Alvin and Stalon, Charles. "Second Report on Market Issues in the California Power Exchange Energy Markets." Prepared for the Federal Energy Regulatory Commission by the Market Monitoring Committee of the California Power Exchange, March 1999. Available at ⟨http://www.ucei.org/Restructuring_Archive/pxMMC_report99.pdf⟩.

Borenstein, Severin. "Frequently Asked Questions about Implementing Real-Time Electricity Pricing in California for Summer 2001." Mimeo, University of California Energy Institute, March 2001. Available at ⟨http://www.ucei.org/PDF/faq.pdf⟩.

______. "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster." *Journal of Economic Perspectives*, Winter 2002, *16*(1), pp. 191–212.

Borenstein, Severin and Bushnell, James B. "An Empirical Analysis of the Potential for Market Power in California's Electricity Market." *Journal of Industrial Economics*, September 1999, *47*(3), pp. 285–323.

Borenstein, Severin; Bushnell, James B. and Knittel, Christopher R. "Market Power in Electricity Markets: Beyond Concentration Measures." *Energy Journal*, October 1999, *20*(4), pp. 65–88.

Borenstein, Severin; Bushnell, James B.; Knittel, Christopher R. and Wolfram, Catherine D. "Trading Inefficiencies in California's Electricity Markets." POWER Working Paper PWP-086, University of California Energy Institute, October 2001. Available at ⟨http://www.ucei.org/PDF/pwp086.pdf⟩.

Borenstein, Severin; Bushnell, James B. and Stoft, Steven. "The Competitive Effects of Transmission Capacity in a Deregulated Electricity Market." *RAND Journal of Economics*, Summer 2000, *31*(2), pp. 294–325.

Bushnell, James B. "Transmission Rights and Market Power." *Electricity Journal*, October 1999, *12*(8), pp. 77–85.

______. "A Mixed Complementarity Model of Hydro-Thermal Electricity Competition in the Western U.S." *Operations Research*, January/February 2003, *51*(1).

Bushnell, James B. and Wolak, Frank A. "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, September 1999. Available at ⟨http://www.ucei.org/PDF/pwp070.pdf⟩.

Cardell, Judith B.; Hitt, Carrie C. and Hogan, William W. "Market Power and Strategic Interaction in Electricity Networks." *Resources and Energy Economics*, March 1997, *19*(1–2), pp. 109–38.

Joskow, Paul L. and Kahn, Edward. "A Quantitative Analysis of Pricing Behavior In California's Wholesale Electricity Market During Summer 2000." *Energy Journal*, 2002, *23*(4), pp. 1–35.

Joskow, Paul L. and Tirole, Jean. "Transmission Rights and Market Power on Electric Power Networks." *RAND Journal of Economics*, Autumn 2000, *31*(3), pp. 450–87.

Kahn, Edward; Bailey, Shawn and Pando, Luis. "Simulating Electricity Restructuring in California: Interactions with the Regional Market." *Resources and Energy Economics*, March 1997, *19*(1–2), pp. 3–28.

Mansur, Erin T. "Environmental Regulation in Oligopoly Markets: A Study of Electricity Restructuring." POWER Working Paper PWP-088, University of California Energy Institute, November 2001. Available at ⟨http://www.ucei.org/PDF/pwp088.pdf⟩.

Newey, Whitney K. and West, Kenneth D. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, May 1987, *55*(3), pp. 703–08.

Pando, Luis. Testimony and Workpapers in Southern California Gas Company, California Public Utilities Commission Application No. 95-06-002, 1995.

Patrick, Robert H. and Wolak, Frank A. "Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices." Mimeo, Stanford University, 1997. Available at ⟨http://www.stanford.edu/~wolak⟩.

Puller, Steven L. "Pricing and Firm Conduct in California's Deregulated Electricity Market." POWER Working Paper PWP-080, University of California Energy Institute, July 2001. Available at ⟨http://www.ucei.org/PDF/pwp080.pdf⟩.

Wolak, Frank A. "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market." *International Economic Journal*, Summer 2000, *14*(2), pp. 1–40.

Wolak, Frank A.; Nordhaus, Robert and Shapiro, Carl. "Preliminary Report on the Operation of the Ancillary Services Markets of the California Independent System Operator (ISO)." Prepared for the Federal Energy Regulatory Commission by the Market Surveillance Committee of the California ISO, August 1998. Available at ⟨http://www.ucei.org/PDF/isomsc.pdf⟩.

Wolak, Frank A. and Patrick, Robert H. "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market." POWER Working Paper PWP-047, University of California Energy Institute, April 1997. Available at ⟨http://www.ucei.org/PDF/pwp047.pdf⟩.

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*, Winter 1998, *29*(4), pp. 703–25.

______. "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review*, September 1999, *89*(4), pp. 805–26.