

---

## The long and the short of it: California's electricity crisis

---

Lee S. Friedman

Goldman School of Public Policy  
University of California at Berkeley  
Berkeley, CA 94720–7320, USA  
E-mail: lfried@berkeley.edu

**Abstract:** Unexpected problems sometimes arise when governments attempt to introduce competition. The problem considered herein is market power and its exercise during the California electricity crisis of 2000–2001. In introducing competition, both transitional and long-run opportunities for firms to exercise market power may arise. California had transitional rules that severely limited participation of its utilities in forward markets and enhanced the market power of new generating entities. The transitional problems could have been avoided, but in the long-run a smaller market power issue should be expected to arise stochastically. This analysis suggests a new long-run institutional policy role: continual regulatory oversight of an industry that could be workably competitive most of the time. This explains why an agency like the Federal Energy Regulatory Commission should have a permanent Office of Market Monitoring. It also suggests why, in some electricity markets, stochastic market power events may arise before capacity gets strained.

**Keywords:** market power; utility regulation; electricity crisis; competition policy.

**Reference** to this paper should be made as follows: Friedman, L.S. (2009) 'The long and the short of it: California's electricity crisis', *Int. J. Public Policy*, Vol. 4, Nos. 1/2, pp.4–31.

**Biographical notes:** Lee S. Friedman is an Economist and Professor of Public Policy at the Richard & Rhoda Goldman School of Public Policy at the University of California at Berkeley. He is a graduate of Dartmouth College and received his PhD from Yale University. He joined the Berkeley faculty in 1974 to help fashion the economics curriculum of the public policy programme. He has published research on a wide variety of issues, including utility and environmental regulation, educational finance, criminal justice policies, agricultural subsidies, and consumer decision-making. He is the author of *The Microeconomics of Public Policy Analysis* (Princeton University Press), and strives to improve the effectiveness of analysis on actual public policies and practices.

## **1 Introduction**

Unexpected problems sometimes arise when governments attempt to introduce competition as a means of encouraging increased efficiency. The specific problem considered herein is market power and its exercise. If the competition introduced is too limited, then one or more of the competitors may possess substantial market power. If this market power is exercised, the resulting harm may be worse than if no competition had been introduced. That is certainly the case in the subject of this study: the California electricity crisis of 2000–2001. Borenstein *et al.* (2002) found that in the summer of 2000 immediately preceding the crisis, the exercise of market power accounted for 59% of the \$2 billion increase in wholesale electricity payments over the prior summer. The California Attorney General's Office claims that ratepayers were overcharged by \$9 billion during the crisis, of which over \$2 billion has since been recovered from the generators involved with additional sums still pending.<sup>1</sup>

This article draws lessons about the potential problem of market power of particular interest to those concerned with policy designs and regulatory reforms that involve the introduction of new market forces. The primary use of the market power concept in policy analysis has been in quite specialised agencies of government, *e.g.*, in the USA particularly the Antitrust Division of the Department of Justice (DOJ) and the Federal Trade Commission (FTC) in their considerations of whether certain industries have become overly concentrated. However, market power issues arise not just in mature, largely unregulated industries overseen by agencies like the DOJ and FTC, but in many cases where the government attempts to improve efficiency by unleashing market forces. These cases include hospital contracting, environmental allowance markets, the deregulations in many of the traditionally-regulated industries like electricity and telecommunications, the transition to markets in economies that had been relying primarily on governmental production, and others. The problem of market power in these examples may be transitional rather than long-run, which does not make it any less important to those striving for successful reforms. Appropriate policies to avoid it and to respond to its appearance may thus also be transitional.

In the case of electricity that we examine, there are reasons to expect both transitional and long-run opportunities for firms to exercise monopoly power. One particularly interesting aspect of this case concerns policies with respect to spot and forward markets. By policies, we mean any rules or restrictions on the ability of electricity buyers and sellers to participate in these markets.<sup>2</sup> California had restrictive rules that severely limited the participation of its utilities in forward markets and had the effect of enhancing the market power of some of the new generating entities in its market. To some extent, these policies were transitional ones with the intent of achieving workably competitive markets. While there were important political influences on the shape of these rules, the more analytic policy design phase could have better considered potential market power issues.

Beyond the transitional concerns, our analysis suggests a new long-term institutional policy role – one that requires continual regulatory oversight of an industry that could be workably competitive most of the time in order to prevent the abuse of market power that arises stochastically. This helps explain why agencies like the Federal Energy Regulatory Commission (FERC) and regional Independent System Operators (ISOs)

should have offices of market monitoring and clarify some of their responsibilities. It also suggests that the stochastic market power events may arise at times before system capacity is strained.

Other electricity sector policies, like reserve requirements and actions that unleash increased real-time demand responsiveness from willing consumers, are also important. However, the most general points apply to other sectors. For example, one of the most daunting global tasks ahead is to create a regulatory system to control greenhouse gases, and market-based systems like cap-and-trade have centre stage in these discussions. On the one hand, this is a long-term problem for which well-working spot and futures markets will be critical components. Both Olmstead and Stavins (2006) and Manne and Richels (2004) express concern that initial allowance distributions in a global market will leave a few countries with most of the excess permits to trade, making collusion to exercise market power a concern. On the other hand, in the absence of global agreements, individual states like California are taking action in the hope that other jurisdictions will act similarly. In the initial years, there may well be problems of market power that arise within such localised systems. California's AB1493, for example, mandates reductions from its motor vehicles and effectively allows manufacturers to trade allowances with one another – but there are only a small number of manufacturers, and market power may prevent or distort trading.

Apart from potential market power, our analysis generally assumes the markets are otherwise workably competitive. Of course they are not, *e.g.*, unregulated greenhouse gas emissions from the electricity sector are a pressing problem that must be addressed. The analytic simplification allows focus on the roles of spot and forward markets in order to clarify their functions and, in some circumstances, their dysfunctions. In part it is a misunderstanding of these time-denominated markets that caused, prolonged and exacerbated the California crisis.<sup>3</sup> Whereas the 1998–1999 (and 2002) spot prices of electricity averaged around \$30 per megawatt-hour (MW-hr) and seemed quite normal, the monthly average began to rise in 2000, exceeding \$100 in many months and reaching \$175 in August. In successive months they remained far higher than the utilities were allowed to collect from their customers (which was about \$60 per MW-hr), leading to financial crisis, rolling blackouts, and general market meltdown by early 2001.

Of course it is very difficult to take a complex situation like this one, and parse the observed results into different explanatory factors: *e.g.*, the exercise of market power, higher natural gas prices that had to be paid by many of the generators, the rise in the price of NO<sub>x</sub> emissions permits required in some parts of southern California, appropriate market scarcity pricing in (unusual peak) hours when system capacity is truly strained, *etc.* Furthermore, it is not illegal in the US for a firm to exercise market power *per se*, even if it is offensive on both efficiency and equity grounds. To be illegal the exercise must involve collusion or violate particular regulatory rules. When FERC conducted its assessment of the Western electricity markets, it reported:

“Staff concludes that prices in the California spot markets were affected by economic withholding and inflated bidding. Staff finds this violated the antigaming provisions of the Cal ISO and Cal PX tariffs and recommends proceedings to require disgorgement of profits associated with these inflated prices.”<sup>4</sup>

However, California has bitterly contested the (small) amount of disgorgement ordered by FERC in its rulings, and these controversial matters are still pending resolution.<sup>5</sup> This paper takes as its starting point only that substantial market power was exercised during

the crisis, and does not focus either on the parsing issue or the illegality issue. We focus on how and why market power can arise in creating a new market like that for California electricity, and policies to prevent it.

A few words of background will be helpful. Electricity is different from most other commodities in that, for all practical purposes, it cannot be stored – it must be used when it is produced.<sup>6</sup> Thus there can be no gradual or marginal shifting of inventories towards supply in one period or another. At any point in time, production levels are primarily a function of the capacity existing at that time and the extent to which that capacity is used. For market participants, that makes *risk management* of price and cost the main purpose of forward or futures markets that vary only in the length of time of the contracts being traded. It is not a matter of trading supply in one period in order to serve another, as would be the case with, say, oil or emissions allowances.<sup>7</sup>

California restructured its electricity markets in a way that caused almost all electricity to be bought and sold no more than one day ahead of time. Its three utility distribution companies, which continued to supply 97% of customers in the non-municipal market, initially were completely barred from any forward contracting beyond one day. Furthermore, all electricity produced by the non-divested generation assets of these companies (and all previously contracted for under ‘qualifying facility’ or QF contracts, largely with co-generators and renewable sources like wind power) was required to be passively run through the centralised auction pools. The state created two new entities to manage the restructured market: the California ISO, which had responsibility for actually running the electric grid, and the California Power Exchange, which had responsibility for running the main ‘day ahead’ market. We will treat all of these transactions as an approximate spot market, since our main interest is in understanding pros and cons of allowing contracting much further ahead – years rather than, for example, simply an extension to a ‘two-day’ ahead market.<sup>8</sup>

In Section 2, we consider a broad landscape: a variety of public policies that have been used to try and foster workable electricity markets in jurisdictions that have restructured to attempt this. These policies have been thought of as largely transitional, in that potential market power is generally thought to be at a ‘peak’ in the initial stages of moving from monopoly to competitive supply. The hope is that strong competition will emerge over time and gradually dissipate any initial market power, and obviate the need for any special transitional rules in place to control it. In the jurisdictions reviewed outside of California, none of them had the type of major problem experienced by California.

In Section 3, we explain the relations among spot and forward markets for electricity. In particular, in fully competitive markets the expected electricity spot price over time is largely unaffected by the mix of short-term and long-term contracts that may be signed in advance. To a first approximation, the forward price for any time period has present value equal to that of the expected spot price over the same period. However, the risk cost associated with future price uncertainty does depend heavily on the time-denominated mix of contract lengths, and California’s almost exclusive reliance on spot markets resulted in excessive risk cost.

A quote from an independent generator in California, not involved in operating the divested fossil fuel generators from the utilities, illustrates the harm this can cause. Calpine Corp., now in bankruptcy, had just received unanimous regulatory approval to build a 600 MW plant in Hayward, CA, but then said it had no intention of starting

construction: “We’re not moving into new construction of plants without power purchasing contracts... Current market conditions are an obstacle to construction at this time.”<sup>9</sup> A bar on forward contracting can reduce and retard long-run expansions in system capacity because it increases the risk costs of investments in new generators (and this would in turn affect expected future spot prices).<sup>10</sup>

In Sections 4 and 5, we focus on the possibility of noncompetitive behaviour. In particular, we ask how the availability of markets for different contract lengths affects the ability of (one or more) firms to exercise market power. We show that for realistic levels of concentration among generation companies, it is much easier to exercise market power with exclusive reliance on spot markets than it is with time-varying contracts. In Section 4 we use a simple simulation model of full spot market reliance and illustrate that generator market power arises with stochastic regularity, and can be profitably exercised even by fairly small generators (*e.g.*, those controlling only 2.5% of the market). Furthermore, these stochastic opportunities may arise at times before system capacity is strained. In Section 5, however, we explain why much of that market power dissipates when it must be spread over time-varying contract lengths.

Sections 4 and 5 suggest that in many electricity markets, market power is not something that should be expected to disappear completely. Because of the unusual nature of the commodity, market power in the spot market should be expected to arise with some statistical regularity even if that market is competitive most of the time. This raises the interesting public policy issue of whether there can and should be continuing oversight agencies whose function is to enforce market rules that protect the public from the exercise of market power during these circumstances. This is a key public policy question that has not been aired sufficiently in the professional literature. The latter for the most part treats market power as a problem that is largely a structural issue, *i.e.*, one that can largely be measured by concentration ratios that are fixed in the short-run and dealt with by policies to block mergers and order the breakup of overly large firms. In electricity markets, the size of the market and the relative shares of the firms in it can change drastically due to demand or supply shifts that can occur within hours (*e.g.*, rapid temperature change, transmission line failure) as well as across seasons, recalling that there is no inventory to provide buffers from such shifts.

The final section of the paper (Section 6) discusses continuing policy issues, and then concludes with a brief summary. Important policy issues confront California and other electricity markets about how to deal with both transitional and continuing potential market power. Transitional issues involve initial divestiture policy to foster a competitive market structure, policies with regard to any mandated vesting contracts, and limits or restrictions on forward contracting. Continuing issues involve the stochastic nature by which market power can be expected to arise even with a market supply structure that functions perfectly well most of the time. Policies to deal with this might include actions to increase demand responsiveness, offices of market monitoring, and market rules like reserve requirements and certain types of price caps such as those triggered by unusual events (automatic mitigation procedures). The policy tools and remedies that are appropriate for a transitional time frame are not the same as those appropriate for continuing issues, and policies should be put in place to deal with market power in both types of settings.

## **2 Policies to prevent the exercise of market power in transitional electricity markets**

### *2.1 California*

In the years before the California restructuring, considerable public debate took place with substantial time and opportunity for stakeholders and policy analysts to offer their ideas, reactions, and criticisms to shape the reforms to come. These discussions occurred during the period 1993–1998, initiated by the issuance in February 1993 of ‘The Yellow Book’ by the California Public Utilities Commission that invited stakeholders to consider substantial regulatory reform including possible restructuring, involved numerous interim reports and proceedings, and ended with the opening of the newly-restructured market in April 1998. From the start, there was concern about the possible exercise of market power in those segments of the industry that were intended to be competitive.

This concern included both the wholesale electricity market, in which generators sell electricity to energy service providers (including the utility distribution companies), and the retail market (sales to end users of the electricity). However, it was focused virtually exclusively on the potential market power of the three existing vertically-integrated, local monopoly utility companies (PG&E, SCE, and SDG&E). It did not seem to occur to anyone that ‘new’ entrants to the California market might themselves possess market power, an analytic weakness in the planning process. The market power that the owners of utility-divested fossil-fuel generators came to possess was not foreseen because both (a) no specific plan of divestiture was ever agreed upon in advance; and (b) the market power arose not just from the nature of the actual divestiture but additionally from the unintended absence of any retail competition (the legislature decided on political grounds to mandate a rate freeze for utilities not at then-current levels as planned but at rates reduced by 10%, which created a significant entry barrier), and the peculiar nature of rules requiring that most electricity be bought and sold through the spot market operated by the Power Exchange.<sup>11</sup>

The foreseen market power issues were as follows. On the opening day of the new wholesale market, the only significant buyers of electricity would be the three Utility-Distribution Companies (UDCs) that under regulation had retail monopolies. This is because it would take time, perhaps considerable time, from the market’s opening for end-users to find and sign up for alternative energy service providers (ESPs, and no one knew how many customers would wish to do this and how many new ESPs would enter the market). Thus at least during an interim period, there must be some transition rules to prevent the UDCs from exercising their new market power. One clear action was to freeze the UDC retail rates for a period of four years (or less if certain conditions involving revenue recovery for stranded assets were met, which could not happen if wholesale market prices were high). But could not the UDCs exercise monopsony power when arranging electricity purchases from generation that they did not own (or was not governed by pre-existing contract)?

The latter concern was part of the motivation for banning the UDCs from entering into any bilateral contracts (which would primarily be forward contracts, although could involve spot contracting). Without such a ban, the UDCs could sign long-term contracts that lock up all future supplies of electricity and prevent the entry of new ESPs who wanted to compete against the UDCs. Furthermore, if most electricity was arranged for by private bilateral contracts, then how would the ‘price’ of electricity be visible to

generation entrepreneurs who are trying to determine if new generation supply would be profitable? California stakeholders wanted to ensure a healthy, vibrant, and visible spot market that would encourage competitive generation supply (and conservation efforts) from all possible sources in response to appropriate market signals. Thus it required the UDCs to arrange for delivery of all electricity through the hourly, day-ahead centralised spot market run by the Power Exchange.

## 2.2 *The UK*

The designers of the California system were much influenced by the primary market precedent, the earlier 1990 privatisation of the electricity supply in the UK. In the UK, all electricity supply was required to flow through a centralised spot market known as 'The Pool'. The Pool operated largely the way that Californians assumed their analogous Power Exchange would work. Electricity demanders (in the form of 12 regional distribution companies) did not directly schedule demands as in California, but gave information to the National Grid Company (NGC) which it used to forecast demand. The NGC then procured this amount from the Pool by arranging all bid supply from lowest to highest bid and setting price at the marginal bid that made supply equal to 'need'. However, there were three crucial aspects of the UK market that are important to note.

First, the behaviour of suppliers in this market was a matter of continual concern over its first few years of operation. Regulatory oversight was the responsibility of the Office of Electricity Regulation, headed by the Director General of Electricity Supply, then Prof. Stephen Littlechild, an economist. Most of the electricity supplies were controlled by two newly-privatised companies, Powergen and National Power. Concerns about the level and volatility of prices led to several investigations and threats of referral to the Monopolies and Mergers Commission (which had authority to break up the companies) until a 1994 agreement was reached in which price-caps were instituted and certain amounts of generation were to be divested by each of the two companies.

Second, and perhaps surprisingly, Pool prices were largely irrelevant to the revenues of companies and to the costs of consumers, because of the widespread use from the start of forward 'Contracts for Differences (CFDs)'. These are bilateral contracts that are signed to cover any period of time chosen by the contracting parties, typically a buyer and seller of electricity, in which they agree on a 'strike price'. A typical two-way contract specifies that if the Pool price for that period is below the strike price, the electricity buyer must pay the difference to the seller. If the Pool price is above the strike price, then the payment goes the other way: the seller pays the difference to the buyer. Either way, and regardless of the Pool price, it is the strike price that determines buyer costs and seller revenues for the contracted amount. Contract durations cover a spectrum of periods, from as long as 15 years to less than one year, and most buyers and sellers arrange to hold a portfolio of diverse contract lengths.

Third, the initial CFDs played the important role of 'vesting contracts', although it is not clear to what extent their intent was actually to do that. Vesting contracts are transitional contracts that are intended to ensure appropriate and reliable behaviour when ownership of an asset is transferred but its economic use is intended to remain stable for some period of time. If a UDC divests one of its generators that it had been using in a routine way to supply some of its electricity needs, it is common for such a divestiture to be arranged with a contract that binds the new owner to continue to supply most of that routine electricity at a price close to marginal cost. It may, for example, stipulate that the

divested plant is to provide 75% of its capacity to the UDC at marginal cost (or a specified price) for 3–5 years. If California had made arrangements like these involving its newly-divested generating plants, this would have removed most market power from these generators during the transitional period.<sup>12</sup>

### *2.3 The PJM market*

The PJM Market initially covered the states of Pennsylvania, New Jersey, Maryland and Delaware, and like New England, each state was responsible for its own restructuring legislation.<sup>13</sup> We will use Pennsylvania as illustrative, a state that is acknowledged as having some success in stimulating retail competition and in which the wholesale markets seem to operate largely as expected. The Electricity Generation Customer Choice and Competition Act became law on 3 December 1996. The law basically separates the generation of electricity from the services of transmitting and distributing it. The law called for a phase-in of retail choice with one-third eligible to choose by January 1998, another third by January 1999, and the remaining third by January 2000. All utilities subject to the separation requirements were required to file their restructuring plans with Pennsylvania's Public Utilities Commission (PUC) in 1997. The Competition Act encourages market participants to coordinate their plans and transactions through an ISO or functional equivalent (PJM). Electric utilities are permitted to divest themselves of facilities or to reorganise their corporate structures, but unbundling of services is required. So the utilities can retain their generating assets, and in addition bilateral contracting for any time duration is allowed. In the year 2000, 52 suppliers were licensed to sell their generation in the Commonwealth and while this has declined somewhat over time (utilities were later allowed to charge all suppliers more for the back-up power they provided, narrowing the retail price differences among traditional and new suppliers), 43 suppliers remain licensed in 2007.

### *2.4 New England*

The New England ISO received conditional FERC approval on 25 June 1997. It was created by the utilities of six New England states: Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island and Vermont. Each state determined its own restructuring legislation, and we will use Massachusetts as illustrative. On 27 November 1997, HB 5117, the Electric Utility Restructuring Act, was signed by Governor Paul Cellucci to restructure the industry in Massachusetts. The Restructuring Act mainly affects the Commonwealth's eight investor-owned distribution companies, which supply 87% of the electricity in Massachusetts. Retail access was required by March 1998, and a simultaneous rate cut of 10% to be followed 18 months later by an additional 5% cut was made law. The law encouraged divestiture of generation, and the generation portion of the electric industry is now virtually all owned by independent power producers.<sup>14</sup>

In fact, except for Vermont all generation throughout the New England ISO area has been divested from the utilities that had owned them. By early 2000 about 25.2 GW of capacity had been sold. According to Bushnell and Saravia (2002), in 2001 the only supplier with more than 10% of system capacity in the New England ISO was a non-utility, and it had a vesting contract to sell most of its power back to the divesting utility at pre-specified prices.



## 2.5 Summary

In sum, this section shows that restructured electricity markets emphasised differing design strategies in their initial years of operation to try and avoid serious market power problems. Only California failed in this task, although one cannot say this is due to an inadequate planning period. The UK managed the transition in good part through extensive use of forward ‘contracts for differences’, which included important elements of vesting contracts. Pennsylvania in the PJM market continued to allow bilateral contracting for any length and vigorously promoted retail competition. In the New England market, divestiture of generators from utilities to independent companies was aggressively pursued to keep market shares small.

It is no easy task to restructure such a complex system as electricity. Because one subset of arrangements, the wholesale mix of short-term and long-term or forward electricity purchases, is so important to the functioning of electricity markets and to understanding the California experience, we turn now to a more detailed examination of this aspect.

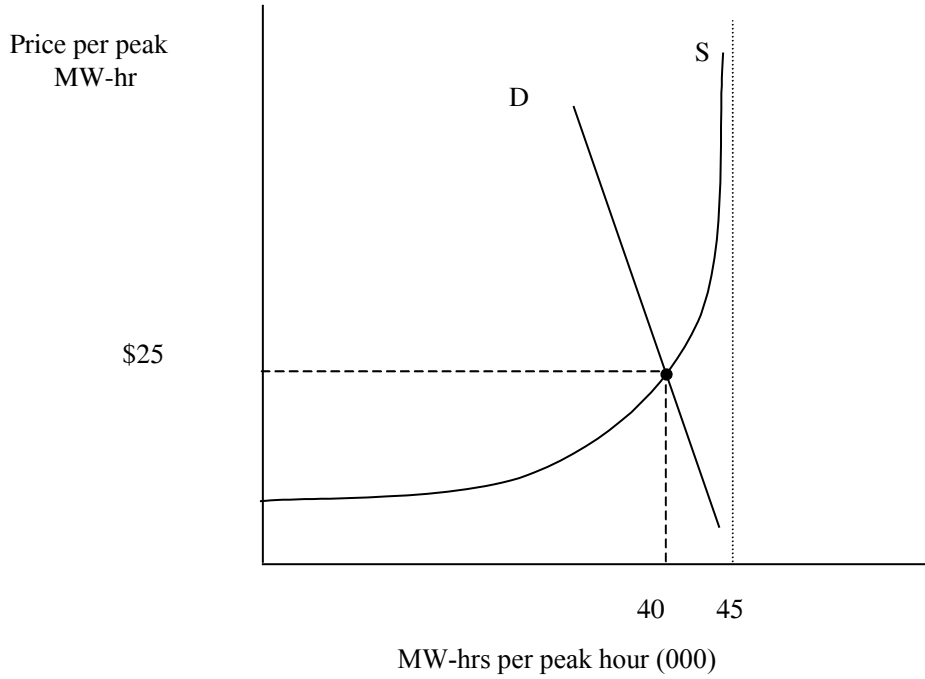
## 3 Forward contracting in competitive electricity markets

Imagine that all electricity is provided through a spot market, and Figure 1a shows the competitive supply curve with the aggregate spot demand. The spot demand curve is highly inelastic (very few consumers are aware of changes in the spot price at the time that they are consuming).

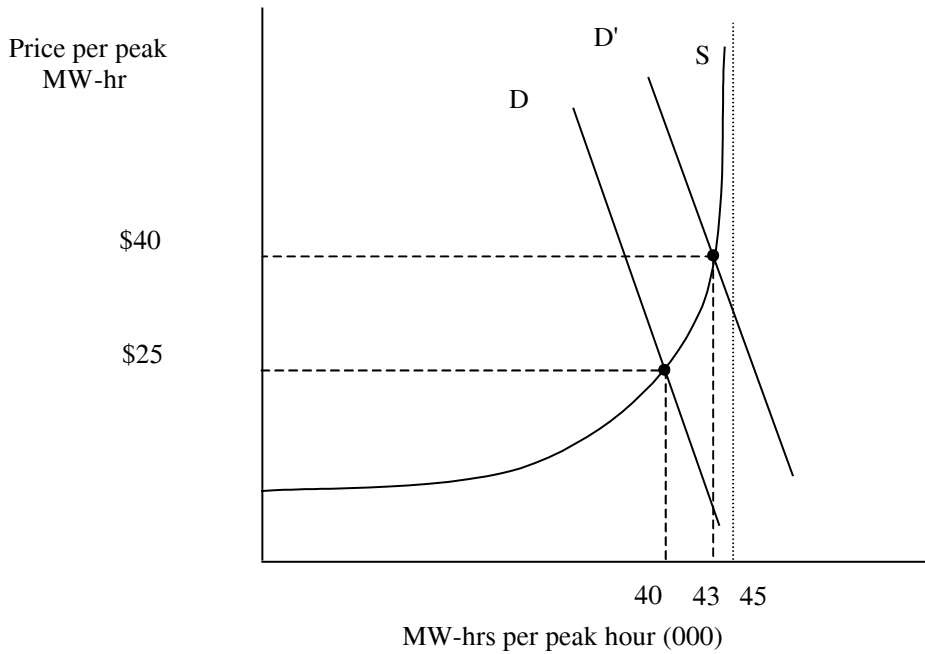
The equilibrium price during a normal peak-period hour is shown at \$25 per megawatt-hour (MW-hr), and the equilibrium quantity at 40 Gigawatt-hours (GW-hrs, 1000 MW = 1 GW). Note that the supply curve is shown as rising steeply beyond 40 GW-hrs. The only capacity left is from the most expensive generators – *e.g.*, peaking units that are designed to operate only during times when electricity prices justify their higher short-run marginal costs, old inefficient generators on their last legs, those that are heavy polluters in urban areas, those that need repairs because they burn far more fuel than when in good repair – and it is very expensive to produce more than 45 GW-hrs because all capacity is essentially fully utilised.<sup>15</sup> Should the day turn out to be an unusually hot one, in which air conditioners must work harder and demand for electricity shifts from **D** to **D'** in Figure 1b, then these more expensive generators would be called into service and the equilibrium is shown at \$40 and with quantity of 43 GW-hrs.

Now permit forward contracts for some of this demand and supply, such that some portion of the market participants in Figure 1 ‘moves’ to the forward market. We will describe shortly how this is motivated by a reduction in risk cost. However, our first point is that this movement does not affect the equilibrium in the residual spot market. For simplicity, assume that these contracts are signed two years ahead of the actual period we have depicted. We will also abstract away from whether the forward market is a centralised exchange or simply the result of bilateral agreements, and assume that the average contracted price is \$30 and that these contracts are for 30 GW-hrs.

Figure 1 Equilibrium with no forward contracting



(a) No forward contracting: equilibrium is at  $Q = 40, P = \$25$



(b) No forward contracting: unusual Demand ( $D'$ ) shifts equilibrium to  $Q = 43, P = \$40$

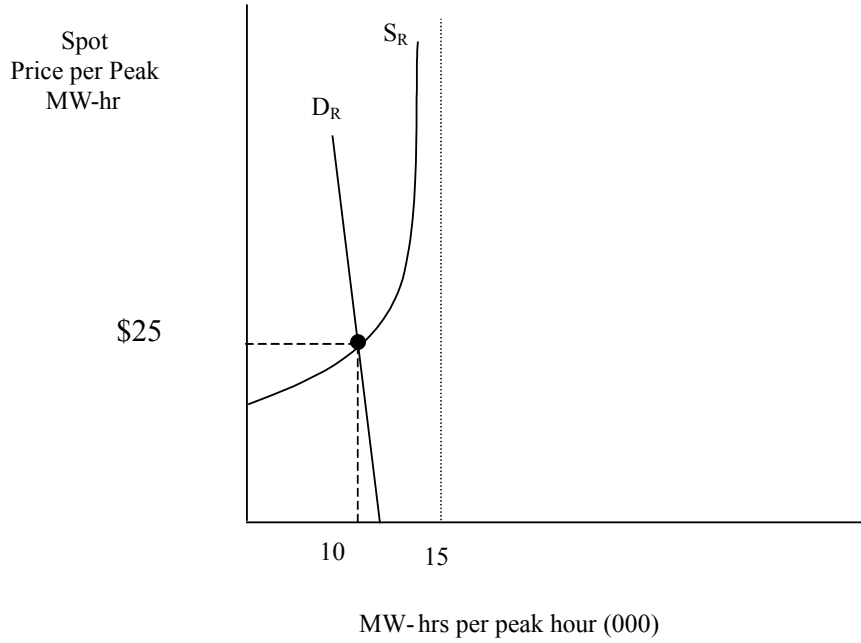
At the time these contracts are being signed, no one knows for sure what conditions will be like two years in the future. The average contracted price could end up higher or lower than the spot price when the contracts become due. If the only two possible price outcomes were those shown in Figures 1a and 1b and the probability of the unusual day was .2, then the expected spot price would be  $\$28 = .2*40 + .8*25$ . If the expected spot price is \$28, then risk-averse consumers will pay more than \$28 to insure themselves of future supply at a reasonable cost. It may also be true that generators or suppliers are risk-averse, and they may be willing to accept less than the expected spot market price in order to guarantee some sales. So it is not obvious whether the forward price will be higher or lower than the expected spot price; it depends on the distribution of risk preferences among buyers and sellers. Apart from the risk adjustment, one would expect the forward price to remain close to the present value of the expected spot price as any large deviation would create a profitable arbitrage opportunity.

The more important point for this part of our story is not the forward prices, but the irrelevance of these transactions for allocation decisions when we get to the day of reckoning. When the time comes to actually supply the electricity, an obligated supplier (one who agreed to supply electricity in the forward market) with higher marginal costs than unobligated suppliers would take advantage of the 'room for a deal'. That is, rather than generating it on its own, the obligated supplier would arrange for the lower cost generator to actually produce the power (for example, simply by buying it on the spot market). Similarly the obligated demander, should it turn out to have purchased forward more than it wishes based on the current spot price, will simply offer to sell the excess in the spot market (the aluminium company in footnote 15 is an extreme example). In other words, the actual demanders and suppliers on the day of reckoning will turn out to be exactly those depicted in Figure 1a or 1b, regardless of which of them signed the forward contracts. The contracts only determine financial gains and losses to the different participants, but do not effect actual supply and consumption decisions.<sup>16</sup>

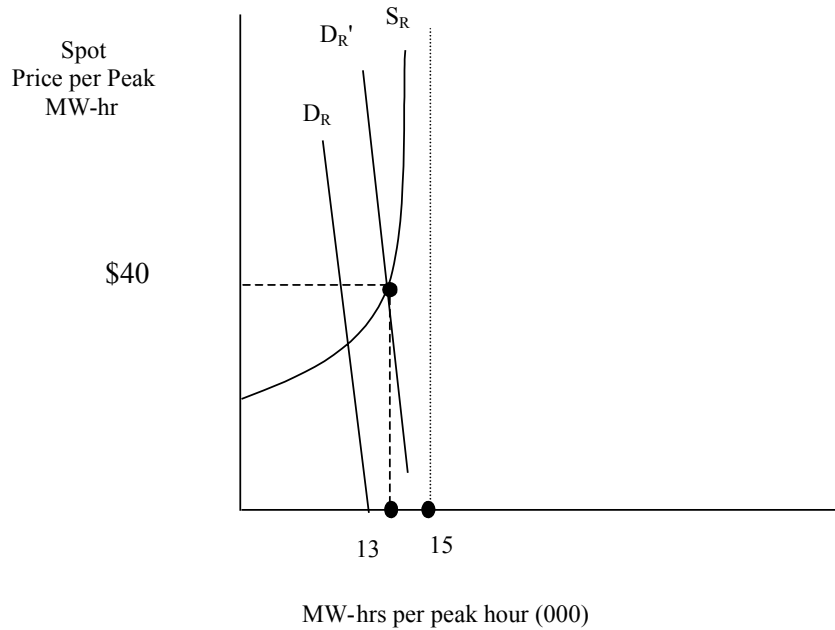
Figure 2a shows the unobligated residual demand and residual supply in the spot market for that peak-period hour two years hence. The residual or spot demand curve  $D_R$  is simply the original demand curve  $D$  with the first 30 GW-hrs demanded removed, since that demand will be met by pre-arrangement. That is, no matter which demanders signed the forward contracts, if only 30 GW-hrs were available on the market at the time of reckoning, there would be transference of the contracts among the demanders until those with the highest willingness to pay at delivery time held them. Similarly, the residual or spot supply curve  $S_R$  is simply the original supply curve  $S$  with the first 30 GW-hrs of supply removed, since that quantity of supply has been pre-committed. Again, no matter which suppliers made the commitments, the ones that would actually deliver them would be those with the least marginal costs due to mutually beneficial transfers of the contract obligations. The equilibrium in the spot market for unobligated transactions occurs at exactly the same \$25 price as before, and ten previously unobligated GW-hrs are purchased.<sup>17</sup>

Figure 2b shows how the spot market equilibrium is affected if the demand turns out to be unusually heavy (analogous to Figure 1b). As before, the market-clearing spot price is \$40, but with 13 unobligated GW-hrs being purchased in the spot market. In other words, the presence of the forward market has no effect on the allocation of electricity on the day of reckoning. Its only value is in reducing the risk cost that market participants would otherwise bear.

Figure 2 Spot market equilibrium with some forward contracting



(a) Forward contracting for 30 GW-hrs: spot market equilibrium is at  $Q = 10$ ,  $P = \$25$



(b) Forward contracting for 30 GW-hrs: unusual Demand ( $D'$ ) shifts spot market equilibrium to  $Q = 13$ ,  $P = \$40$

#### 4 Market power in electricity markets

At this point we return to a situation somewhat similar to our competitive markets of Figure 1 (with no forward markets), but introduce the presence of some firms with market power in the market in order to illustrate how they could create a situation like that of the California crisis. One main point of this section is to illustrate how variable the degree of market power is even when the capacity of the system and its ownership remain fixed throughout. Again, there are many other factors that contributed to the real crisis that are not being studied here. For example, the price of natural gas used as fuel by many generators in California more than doubled, which by itself would have approximately doubled electricity costs. The costs used in this model reflect the lower, pre-crisis price of natural gas and are held constant.

##### 4.1 *Why market power in an electricity market may vary dramatically within a short-run period*

The models of interest to us are short-run models in which the constellation of electric generating plants is fixed for the period in question, and the decisions about whether and how much to operate them depend upon their short-run marginal costs in relation to output prices. We consider models in which the Short-Run Marginal Cost (SRMC) curve looks like that of Figure 1 or 3. The characteristics of this curve that are important are that the supply is highly elastic at relatively low levels of output, but becomes more inelastic as output grows, and extremely inelastic as output levels strain the capacity of the system. Thus even in a perfectly competitive market, price would be expected to be much higher whenever demand conditions are such that they begin to strain capacity, even if most of the time demand is such that it crosses the relatively flat portion of SRMC where prices are low.

Within the above framework, we wish to consider the class of models in which there is an oligopoly segment of the market owning generators with moderate short-run marginal costs. Most firms in the market will still be considered competitive price-takers, including those owning the generators where short-run marginal costs begin to rise rapidly (the ‘peaker’ segment). Thus we will use a competitive fringe model of oligopoly behaviour. We wish to consider the question of when it might be profitable for the oligopolists to withhold some of their output from the market in order to raise price and increase their total profit on the non-withheld output that they will still sell. The market power of an oligopolist in a competitive fringe model may be written as follows, using Lerner’s market power measure:<sup>18</sup>

$$\frac{P - MC}{P} = \frac{s}{\varepsilon_D + (1 - s)\varepsilon_S}$$

where:

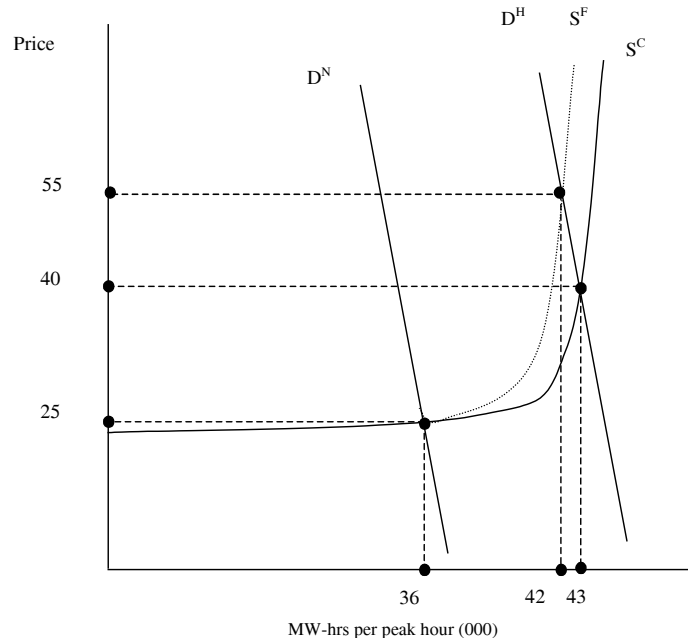
- $s$  = the market share of the oligopolist
- $\varepsilon_D$  = the (absolute) market elasticity of demand
- $\varepsilon_S$  = the supply elasticity of the competitive fringe.

The more elastic are demand and fringe supply, the less the market power of an oligopolist of given size. With highly inelastic demand and fringe supply, which often characterises electricity, even a relatively small oligopolist can have significant market power.

In Figure 3 we draw a supply curve like that of Figure 1 and two different demand curves, one 'high demand' intersecting the inelastic portion of supply and the other the 'normal' demand intersecting the elastic portion. Assume that demand elasticity is about .1. Supply elasticity on the elastic portion approaches infinity, and let's assume it is about .2 on the inelastic portion. Imagine an oligopolist that controls 4 GW of power located on the flat portion of the supply. When demand is normal, this oligopolist with 1/9 of the competitive output level has no market power because of the high supply elasticity of the competitive fringe. If it withholds some output and shifts the supply curve from  $S^C$  to  $S^F$ , there is no apparent shift on the horizontal segment. But with high demand, even though the oligopolist's share of the competitive output level drops to 1/11, market power is .32 by the Lerner measure. The same withholding produces a noticeable shift of the vertical supply segment and raises price from \$40 to \$55. Thus even a relatively small oligopolist can sometimes have substantial market power in an electricity market.

Note that we have not yet provided any explicit analysis of the oligopolist's behaviour. In this class of models, any withholding by an oligopolist is offset to some extent by replacement supplies from the competitive fringe. In Figure 3 a new after-withholding equilibrium is shown at 42 GW-hrs, but the diagram does not show how much the oligopolist withheld and how much extra the competitive fringe provided. In order to try and make California-type outcomes a bit clearer, the next section is based on a more complete empirical illustration.

**Figure 3** The oligopolist can raise price to \$55 with high demand, but cannot raise price at all with normal demand



#### 4.2 *An illustrative model with California parameters*

Details of the model for this section are explained in the Appendix. Its purpose is to clarify the nature and effects of market power exercise in a system like California's, with varying assumptions about the size of the market available for withholding. It deliberately strips away almost all of the real-world complexity that could lead to substantial price variation in the absence of market power exercise; this is not a model that helps one to distinguish the exercise of market power from other factors that could cause price increases. Indeed, the idea is to make the exercise of market power very transparent. Given a demand and supply structure that is like California's (with factor prices held constant), it illustrates plausible levels of price increases that could be achieved by an entity controlling varying amounts of the system capacity and exercising its market power.

California's electricity supply is represented by three segments totalling 45 GW of capacity. The entire supply structure approximates the actual marginal cost of California generation over the range of consumption observed during the 1998–2001 restructured period.<sup>19</sup> Ten gigawatts of low marginal cost baseload resources are treated as competitive (representing primarily hydro and nuclear sources that were largely on a 'must-run' basis with remuneration covered by regulatory side agreements rather than PX prices). Twenty gigawatts of moderate marginal cost fossil-fuel generation are assumed to be owned by firms large enough to try and exercise market power (representing the generators that were divested primarily to five out-of-state companies). The rest consists of 15 GW of relatively-small sources that have increasingly higher marginal cost. The latter 'peaker' segment is treated as competitive, with 10 GW that can be supplied with moderate marginal cost increases (the shoulder where the supply curve is rising) but the last 5 GW characterised by rapidly increasing marginal costs (the most vertical segment of the supply curve, reflecting older inefficient units including those in the Los Angeles area subject to very expensive NO<sub>x</sub> emission charges).

The model uses several assumptions about the nature of the middle 'oligopoly' segment, but is not intended to be a complete model of oligopoly behavior.<sup>20</sup> Rather, we examine how much of this middle segment must be under the control of a profit-maximising firm (or multiple colluding firms) in order for it to exercise market power, and what prices result from a firm's control of various amounts of this segment. This depends on demand conditions, and we examine how different demand conditions affect the profitability of trying to exercise market power under each assumption about the oligopolist's size.

One assumption is that the middle-segment firms collude and act like they are controlled by one large profit-maximising firm. An alternative assumption is that there is no collusion and one oligopolist with 4000 MW of capacity tries to exercise market power (while the rest behave competitively). A final alternative assumption is that one firm with only 1000 MW of capacity tries to exercise market power (while the rest behave competitively). In all of these cases, the oligopolists consider withholding some of their capacity in order to raise price in the market and increase the total profit that they receive on the remaining capacity that they sell. They figure out the profit-maximising capacity to withhold. The results from these assumptions are compared with a benchmark case in which all firms behave competitively. While the market-power models are not

equilibrium models, they do represent the range of outcomes possible from any such models (since we include the two extremes of full monopoly and competitive behaviour within the segment).

The model also simulates the effects of varying demand conditions. In particular, we examine how three different demand conditions affect the results: demand on a normal day, a 'corner' day, and a hot day. The normal day is one in which the demand curve crosses the competitive supply curve on the flat portion not yet near the corner; the 'corner' day has greater demand in which the demand curve crosses the competitive supply near the corner, but before capacity is strained; and the hot day is one in which the demand curve crosses the competitive supply beyond the corner on the steeply rising segment.

Table 1 contains the simulation results made under the different demand conditions and different assumptions about the size of the oligopoly that is attempting to exercise market power. The entries under 20 GW are those where the oligopolists are assumed to act like a single monopoly. Those under 4 GW are for one oligopolist controlling this amount of capacity and acting alone; and those with 1 GW are for one smaller oligopolist wondering if it too might have market power. The profit column shows the amount accruing to the withholding oligopolist. The next two columns show respectively the quantity supplied by the withholding oligopolist and the total quantity supplied in the market from all sources. The extra rent column shows the increase in consumer expenditures compared to the same quantity competitively priced, and the final column shows the percent by which the oligopolist has raised price.

**Table 1** Results of simulations of oligopoly behaviour with competitive fringe

<i>Outcomes</i> →		<i>Price</i>					
<i>Situation</i> ↓		<i>(\$)</i>	<i>Profit (\$)</i>	<i>Q olig</i>	<i>Q mrkt</i>	<i>Extra rent (\$)</i>	<i>% Price increase</i>
Comptv	Normal	23		20.00	36.00		
	Corner	25		20.00	40.00		
Olig ↓	Hot	103		20.00	41.74		
20 GW	Normal	212	1,548,384	8.09	32.23	6,075,355	822
20 GW	Corner	259	2,418,685	10.11	35.31	8,275,360	936
20 GW	Hot	298	3,272,994	11.76	37.83	7,385,142	189
4 GW	Normal	23	12,000	4.00	36.00	0	0
4 GW	Corner	70	105,000	2.11	39.10	1,754,473	180
4 GW	Hot	109	334,047	3.76	41.62	238,795	6
1 GW	Normal	23	3,000	1.00	36.00	0	0
1 GW	Corner	34	20,796	0.61	39.81	371,939	36
1 GW	Hot	103	83,000	1.00	41.74	0	0

There are several interesting features to note about these results. Of course other things being equal, the larger oligopoly has greater market power that results in higher prices, less output, and more profit. In normal times, neither the 1 GW nor 4 GW oligopolist has any market power. It would take withholding by multiple 4 GW



oligopolists to raise price in this situation. On the other hand, even the quite small 1 GW oligopolist has market power when demand meets supply at the corner, and can raise price from \$25 to \$34 by withholding as little as 390 MW-hrs of power. In this situation, the cost of withholding is the small profit foregone at the low competitive price, while the benefit is a very substantial increase in price caused by the scarcity of inexpensive replacement resources.

The hot day is an interesting case because at first the results may not be intuitive: one might think that it is easy to exercise market power when system capacity is so strained, but the results suggest otherwise. When the competitive price is already quite high, the cost of withholding is therefore very high for these relatively low marginal cost generators. Still, because demand and supply are both quite inelastic, withholding can increase profit if the residual supply is sufficiently ample. That is, to achieve a given price increase a fixed amount of power must be withheld, but the profitability to the withholder depends on how much power it has left to sell at the high price. The 1 GW generator does not have ample enough supply to do this profitably; and the 4 GW oligopolist has just enough to make a small withholding profitable. Only the fully collusive oligopoly has the means to substantially raise price and increase profit at this level.

If we compare these outcomes to the actual prices observed in California, adjusting them downward by about \$40 for the lower fuel costs assumed here, it is hard to argue that reality looks like the full collusion version of this model (actual prices did not tend towards the high 20 GW levels shown here, although in a small number of instances it did reach them). On the other hand, the prices achieved by one 4 GW oligopolist acting alone seem to be somewhat of an underestimate of what happened, as in August 2000 when the actual average price hit \$175. This is of course merely suggestive, but to the extent that the exercise of market power did explain what happened, then it looks like a situation in which more than one 4 GW oligopolist engaged in some withholding (somewhere in between the one independently acting firm situation and that of full collusion).<sup>21</sup> Furthermore, the results here may help explain the motivation for these outages on days that would normally not strain capacity.

## **5 The effect of forward markets on market power**

The key insight of this section is that the existence of forward markets, or equivalently long-term contracting, will serve to dissipate many of the opportunities to exercise market power. If it can be shown that the oligopolists will sell a substantial portion of their power in the forward markets, then it follows readily enough that they will have dissipated opportunities to exercise market power in the spot market. This is because the forward contracts obligate them to supply most of their power for a *predetermined* revenue. To achieve a given price increase in the spot market, the same quantity must be withheld as with no forward market. But the benefit of this withholding can only be accrued on the much smaller unobligated residual quantity that in almost all cases will not be enough to make the withholding profitable. For example, if the 4 GW oligopolist sells 3 GW forward, then it effectively becomes a 1 GW oligopolist in the spot market (its maximisation problem is numerically equivalent to that of the 1 GW oligopolist).

Cannot the oligopolists refuse to sell in the forward market, in order to retain their opportunities to exercise market power in the spot market?<sup>22</sup> The key reasoning here is the free-riding incentive that makes it difficult for any single oligopolist to withhold even in a simple spot market: the profit from letting another oligopolist do it is far greater than doing it yourself. To illustrate how that same incentive affects participation in a forward market, let us begin with the observation that the forward price will, apart from adjustment due to the distribution of risk preferences, equal the present value of the expected spot price. Any significant deviation from this creates an arbitrage opportunity to increase expected value. If the forward price is too high, for example, then generators will increase expected profits by selling additional quantities forward until the equilibrium between forward and spot prices is restored.

What is the expected spot price? Recall the earlier example from Section 3 where we assumed that normal hours occur 80% of the time and hot ones 20%, and assume no uncertainty about generation costs. Also assume that buyers and sellers are equally risk-averse, so that the risk-reduction of forward contracting does not cause any deviation between the forward and expected spot prices. If we also assume settlement of the forward contracts is at delivery time, then there is no need for present value calculations. The expected competitive spot market price, using the models from the prior section (and simplifying to ignore the corner), is  $.8 (\$23) + .2 (\$103) = \$39$ . If one believed, however, that one 4 GW oligopolist will withhold power as in Table 1, then the expected spot price becomes  $.8 (\$23) + .2 (\$109) = \$40$ .

Every oligopolist believing that some oligopolist will withhold as above can enjoy the advantages of free-riding in advance by selling forward contracts at \$40 (to risk-averse buyers with the same belief). That is, the \$20 certain profit from a forward sale equals the expected profit from selling the same unit in the spot market  $.8 (\$3) + .2 (\$89) = \$20$ , but it is better because there is no risk. Clearly zero forward selling cannot be an equilibrium, as both contracting parties are better off by agreeing to the \$40 forward contract. But \$40 is too high to be the equilibrium price. None of the oligopolists wants to resist these contracts and be stuck at the end as the one that actually has to do the withholding. As every oligopolist becomes obligated to some extent by the future contracts, they each have less unobligated capacity left for the spot market. Everyone realises that the initial assumption about the expected amount of withholding is wrong; it will be lower, the equilibrium spot price will be lower, and therefore the equilibrium price for the future contract must be less than \$40. But for any positive amount of withholding still expected, one can still free-ride in advance by selling additional forward contracts at the revised expected spot price. This further reduces the unobligated capacity, and forward selling continues until the expected spot price approaches the competitive level. That is, forward transactions will grow until the unobligated capacity is small enough to essentially dissipate the market power that might otherwise have arisen in the spot market.<sup>23</sup>

This argument becomes stronger once one accounts for the diversity of different time-dimensions for forward contracts that will be demanded. That is, appropriate hedging given the uncertainties and volatility of actual electricity spot prices implies that both generators and electricity buyers will want to hold a portfolio of contracts that vary in length from 'day-ahead' to 'many years-ahead'. The better the forward market in terms of diversity of lengths offered, the more attractive it is to commit to forward obligations and of course to maintaining a balance across holdings of different durations. Put

somewhat differently, the greater the time diversity of contract opportunities, the more elastic is the demand and supply to any one of them, and the less will be any market power.

## **6 Continuing policy considerations**

Not all horizontal market power problems in a wholesale generation market can be resolved by healthy forward markets, no matter what degree of concentration there is in the industry. These forward contracts are based on expectations, and reality at the time of reckoning may turn out to be quite different from expectations. In healthy markets, perhaps 75%–80% of a generator's electricity would be contracted forward, with the remaining 20%–25% sold only days-ahead or less.<sup>24</sup> If this is correct, then in a market of California's size and concentration even our 4 GW companies will have market power some of the time (similar to that of the 1 GW company in Table 1). Larger entities would have greater market power that would arise more frequently even with ample forward markets. Further investigation is necessary to begin to understand the magnitude of this type of market power problem in systems with substantial use of forward contracting.

Are there policies that can mitigate this type of market power without causing other problems? In the FERC Notice of Proposed Rulemaking (NOPR) on Standard Market Design, the mitigation of market power was a substantial part of its aim.<sup>25</sup> It proposed three mandatory market rules and a fourth voluntary one as part of the Standard Design, although due primarily to resistance by states with more traditional regulation it backed off imposing these and withdrew the proposal in 2005. The first measure was aimed at an important problem that we, for purposes of simplification, assumed away. Our model assumes a free-flowing market, and does not consider any localised transmission constraints or bottlenecks that might create market power within a region. Of course these can be substantial – they occur every time a grid operator determines that a particular generator 'must run' for grid reliability, and FERC proposed that such generators have contracts with the grid that cap their prices at these 'must run' times. PJM has received approval to use a variant of this rule, in which a cap for local market power is imposed only if in real-time there are insufficient alternative resources available (its three pivotal supplier test).<sup>26</sup>

The second mandatory FERC measure is a safety-net bid cap, like the \$1000/MW-hr cap currently used in the PJM and Texas markets. Its purpose is to account for the lack of demand responsiveness on the 'spot' because most electricity users have no idea what these prices are at the time they are consuming. That is, the fact that people might consume electricity at \$10,000 per MW-hr does not mean that their informed choice would be to do so, and the FERC measure is intended to mitigate the high unintended costs of this information failure. More aggressive movement within the states to facilitate the use of 'smart' meters with more dynamic rate structures, like real-time or 'critical peak' pricing, would be more effective and efficient than the safety-net cap.<sup>27</sup> California spent \$35 million installing advanced real-time meters for its 20 000 largest electricity consumers in 2001, and its utilities offer voluntary dynamic pricing programmes intended to enrol 5% of system peak demand by 2007. Its utilities are also proceeding with plans for dynamic pricing for its residential and commercial customers, *e.g.*, PG&E has been installing advanced meters for all customers and is introducing voluntary dynamic rate plans for these customers in 2007.

It is worth thinking more about whether behavioural economics might offer further guidance on how to improve welfare in this 'uninformed consumer' situation, but it seems clear that the usual welfare calculations that assume a rationally-determined demand curve are completely inappropriate for the spot market.<sup>28</sup> Indeed, that is why these calculations were not included when reporting results in Table 1. But it also seems clear that this general safety-net bid cap is not designed to mitigate the type of stochastically reoccurring market power problems identified here. The cap of \$1000 is too high to mitigate most of the problems, and even the much lower western region cap of \$400 would not do so.

The third mandatory mitigation measure is a resource adequacy requirement. The rough idea is that all load-serving entities would be required to maintain reserves of 12%–18% above their expected peak loads, similar to the reserve margins that are applied to most vertically-integrated utilities. As of 2006, California requires a 15% reserve margin for its load-serving entities. The NOPR correctly points out that, because of the way that shortage costs are shared by most participants on a grid, individual participants do not have sufficient incentive to maintain reserves on their own (grid reliability is a public good to all grid users).<sup>29</sup> A very important aspect of this proposed measure is that the reserve requirement may be met not only by securing generation resources, but also by securing demand reduction resources. In other words, the requirement promotes conservation on an equal basis with additional generation resources. To the extent that this reserve requirement causes an increase in system capacity above what it would be otherwise, then of course it would reduce the number of stochastic instances in which market power would arise. This is a very promising mitigation measure. Further study might clarify the value of this measure in terms of reduced instances of market power.

The fourth voluntary measure is for unusual noncompetitive conditions that are thought to be temporary but for a persistent period of time, perhaps like the California crisis situation in the summer months of 2000. Then it became apparent that prices were unusually high during this entire period. The NOPR suggests that, should such circumstances arise, it might be appropriate to examine and possibly limit the spot market bids (day-ahead and real-time) from individual suppliers. It envisions that such mitigation would be automatically triggered by predetermined conditions (*e.g.*, a drought, prices above competitive levels for some well-defined time period).

The NOPR mentions that this fourth voluntary measure is similar to the Automatic Mitigation Procedures (AMP) used by the New York, New England and California ISOs. CAISO, for example, uses the AMP on a real-time, ongoing basis. That is, every bid for supplying generation in real-time is screened (by computer) to see if it deviates significantly from a 'reference curve' for that generator based on its previous three months bids for the same period, adjusting for several possible factors like fuel cost changes. Bids that fail the AMP screening test are automatically mitigated to the reference bid and the company immediately notified. If AMP limits scarcity rents even if they are competitive, this could have negative consequences for attracting needed supplies. The test does not identify any actual withholding, and a 'high' bidder could be trying to take advantage of another's withholding. But it may work to place limits on price-raising ability due to market power. This mechanism deserves further study as a means to mitigate stochastic market power.

The analysis offered in this paper helps to explain how California's overreliance on the spot market contributed greatly to its crisis. But it does ignore a key motivation for the initial bar on utilities from forward contracting, which was to promote retail competition by ensuring that an ample supply of electricity would be available to new energy service providers who would enter and compete with utilities for customers. That is, if the utilities signed forward contracts for all of the electricity, how would one expect a new service provider to find a reliable supply that it could offer to woo customers? If one is to allow for the growth of retail competition in California, then some limit on the amount of UDC forward contracting may be required. Similarly, one might require the UDCs to make some of their retained generation (almost exclusively hydropower and nuclear) available to other ESPs on a competitive basis. Since the electricity crisis, customers have been barred from any new 'direct access' to independent (non-utility) suppliers, but the California Public Utilities Commission in 2007 has opened new hearings to decide if this bar can and should be lifted.

The horizontal market power that was an important cause of California's crisis can be substantially mitigated by increased reliance on forward markets, by ample reserve requirements, and by market structure policies that limit the size of firms operating within one market. Even with these, stochastically there will still arise situations in which market power exists temporarily, but more aggressive use of dynamic rate structures combined with smart technology for consumers, market monitoring institutions and perhaps AMPs can contribute to the mitigation of these events.

It should be clear that restructuring in electricity is a complex, difficult process that fully challenges the capabilities of policy designers as well as those who work within the system and those who rely upon it. The question arises: are there gains large enough to justify all of the imperfect efforts and striving for improvement? Why not just leave the conventional systems alone – 'if it ain't broke, don't fix it'. In the evaluations that have been done of the restructured systems to date, there have been some modest gains. For example, Domah and Pollitt (2001) evaluate the UK distribution side, and conclude that there have been significant benefits since 1990, but that consumers have only begun to gain since 2000. For the USA, Fabrizio *et al.* (forthcoming) find that generating plant operators in restructured markets have achieved labour and nonfuel cost savings of 15%–20% relative to government and cooperatively-owned plants. The question is whether gains like these will continue so that they become widespread, broadly distributed gains over the long-run and outweigh any errors and missteps that also occur. While the answer to that is uncertain, hopefully the analysis and prescriptions in this paper will contribute towards better use of market-based regulatory mechanisms for the future, both in electricity sectors as well as others.

## References

- Allaz, B. and Vila, J-L. (1993) 'Cournot competition, forward markets and efficiency', *Journal of Economic Theory*, Vol. 59, pp.1–16.
- Bessembinder, H. and Lemmon, M. (2002) 'Equilibrium pricing and optimal hedging in electricity forward markets', *The Journal of Finance*, Vol. 62, No. 3, pp.1347–1382.
- Blumstein, C., Friedman, L. and Green, R. (2002) 'The history of electricity restructuring in California', *Journal of Industry, Competition and Trade*, Vol. 2, Nos. 1–2, pp.9–38.
- Borenstein, S. (2002) 'The trouble with electricity markets: understanding California's restructuring disaster', *The Journal of Economic Perspectives*, Vol. 16, No. 1, pp.191–211.

- Borenstein, S. (2005) 'The long-run efficiency of real-time electricity pricing', *The Energy Journal*, Vol. 26, No. 3, pp.93–116.
- Borenstein, S., Bushnell, J. and Wolak, F. (2002) 'Measuring market inefficiencies in California's restructured wholesale electricity market', *The American Economic Review*, Vol. 92, No. 5, pp.1376–1405.
- Bower, J. and Bunn, D.W. (2000) 'Model-based comparisons of pool and bilateral markets for electricity', *The Energy Journal*, Vol. 21, No. 3, pp.1–29.
- Brandts, J., Pezanis-Christou, P. and Schram, A. (forthcoming) 'Competition with forward contracts: a laboratory analysis motivated by electricity market design', *Economic Journal*.
- Brennan, T.J. (2001) 'The California electricity experience, 2000–01: education or diversion', *Resources for the Future Report*.
- Brennan, T.J. (2007) 'Checking for market power in electricity: the perils of price-cost margins', in A.N. Kleit (Ed.) *Electric Choices: Deregulation and the Future of Electric Power*, Chap. 8, Lanham, MD: Rowman & Littlefield.
- Bushnell, J. (2003) 'A mixed complementarity model of hydrothermal electricity competition in the Western United States', *Operations Research*, Vol. 51, No. 1, pp.80–93.
- Bushnell, J. (2007) 'Oligopoly equilibria in electricity contract markets', *Journal of Regulatory Economics*, Vol. 32, No. 3, pp.225–245.
- Bushnell, J. and Saravia, C. (2002) 'An empirical assessment of the competitiveness of the New England electricity market', *Center for the Study of Energy Markets WP-101*, University of California Energy Institute.
- Domah, P. and Pollitt, M. (2001) 'The restructuring and privatisation of the electricity distribution and supply businesses in England and Wales: a social cost-benefit analysis', *Fiscal Studies*, Vol. 22, No. 1, pp.107–146.
- Fabrizio, K., Rose, N. and Wolfram, C. (forthcoming) 'Do markets reduce costs? Assessing the impact of regulatory restructuring on U.S. electric generation efficiency', *The American Economic Review*.
- Friedman, L.S. (2002) 'Bounded rationality versus utility maximization: a test of energy price responsiveness', in R. Gowda and J. Fox (Eds.) *Judgments, Decisions and Public Policy*, Cambridge, UK: Cambridge University Press, pp.138–173.
- Friedman, L.S. and Hausker, K. (1988) 'Residential energy consumption: consumer behavior and rate design', *The Journal of Consumer Policy*, Vol. 11, No. 3, pp.287–313.
- Joskow, P. (2001) 'California's electricity crisis', *Oxford Economic Policy*, Vol. 17, No. 3, pp.365–388.
- Joskow, P. and Kahn, E. (2002) 'A quantitative analysis of pricing behavior in California's wholesale electricity market during summer 2000', *The Energy Journal*, Vol. 23, No. 4, pp.1–35.
- Landes, W. and Posner, R. (1981) 'Market power in antitrust cases', *Harvard Law Review*, Vol. 94, pp.937–996.
- Liski, M. and Montero, J-P. (2006) 'Forward trading and collusion in oligopoly', *Journal of Economic Theory*, Vol. 131, pp.212–230.
- Lowrey, C. (1997) 'The pool and forward contracts in the UK electricity supply industry', *Energy Policy*, Vol. 25, No. 4, pp.413–423.
- Manne, A. and Richels, R. (2004) 'US rejection of the Kyoto Protocol: the impact on compliance costs and CO<sub>2</sub> emissions', *Energy Policy*, Vol. 32, No. 4, pp.447–454.
- Olmstead, S. and Stavins, R. (2006) 'An international policy architecture for the post-Kyoto era', *The American Economic Review*, Vol. 96, No. 2, pp.35–38.
- PJM Interconnection (2003) 'State of the market 2003', <http://www.pjm.com/markets/market-monitor/som.html>.
- Powell, A. (1993) 'Trading forward in an imperfect market: the case of electricity in Britain', *The Economic Journal*, Vol. 103, No. 417, pp.444–453.

- US Department of Energy (2000) *The Changing Structure of the Electric Power Industry 2000: An Update*, Chap. 8.
- US Federal Energy Regulatory Commission (2002) 'Remedying undue discrimination through open access transmission service and standard electricity market design', *Notice of Proposed Rulemaking*, Docket No. RM01-12-000.
- US Federal Energy Regulatory Commission (2003) *Final Report on Price Manipulation in Western Markets*, Docket No. PA02-2-000.
- Wilson, R. (2002) 'Architecture of power markets', *Econometrica*, Vol. 70, No. 4, pp.1299–1340.
- Wolak, F.A. (2000) 'An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market', *International Economic Journal*, Vol. 14, No. 2, pp.1–39.
- Wolfram, C. (1998) 'Strategic bidding in a multiunit auction: an empirical analysis of bids to supply electricity in England and Wales', *RAND Journal of Economics*, Vol. 29, No. 4, pp.703–725.

## Notes

- 1 Recovered payments include \$417 million from Williams, \$281.5 million from Dynegy/NRG/West Coast Power, \$207.5 million from Duke Energy, \$749.7 million from Mirant & Affiliates, and \$460 million from Reliant. See the Energy section of the website for California's Office of the Attorney General (<http://www.ag.ca.gov/antitrust/energy/>).
- 2 We do not focus on certain design aspects of the markets themselves, like auction methods for running a spot market or whether or not to have centralised and standardised future contracts as opposed to bilateral contracts that are custom-tailored for each transaction. These of course are also important questions, although not central to the points of this analysis. For further information about design aspects like these, see for example Bower and Bunn (2000) and Wolfram (1998).
- 3 There were multiple causes of the crisis, including a lag in expanding supply capacity, high natural gas prices (which might themselves have been partially caused by the exercise of market power in the natural gas market), rising NO<sub>x</sub> emission permit prices, the lack of appropriate price incentives for consumers, and jurisdictional issues among regulatory agencies. For more detailed information about the crisis, see Blumstein *et al.* (2002). Valuable analyses of the multiple causes of this crisis include those by Borenstein (2002), Brennan (2001) and Joskow (2001).
- 4 P. ES-2 of US FERC (2003).
- 5 In June 2007 the US Supreme Court ruled in favour of California by upholding the September 2004 finding of its Ninth Circuit Court of Appeals in Opinion No. 02-73093 that (p.13136): "The California energy market was subjected to artificial manipulation on a massive scale. With FERC abdicating its regulatory responsibility..."
- 6 Hydropower ('hydro') has characteristics that make it seem like an exception to this, although it is not. River water is captured in reservoirs and directed to flow down onto turbines that generate electricity. By controlling the flow, one can effectively 'store' the electricity that would have been generated 'offpeak' and create a stronger flow during the most valuable 'peak' times. This applies both within a daily cycle and across a yearly cycle as well (*e.g.*, saving spring runoff water for late summer generating). Of course it is the water that is being stored, not electricity itself. This ability does create two concerns: environmental ones about the effects of altering river flows, and market power concerns about just when the power is generated (*e.g.*, a company in a market environment holding both hydro and nonhydro generators could deliberately withhold some peak hydro generation if doing so will raise sufficiently the price of its other peak generation). We will not treat the hydro problems specifically in this paper. For more thought on hydro, see Bushnell (2003).
- 7 There are other important intertemporal considerations in electricity generation aside from the pure risk management function enabled by forward markets, although we do not focus upon them. One, for example, is due to the strict yearly emissions limits under which some

generators operate. For them, operation during one part of the year reduces the amount that they can operate during the rest of the year. Another is due to maintenance: the more a generator runs without shutting down for maintenance, the more the risk of mechanical failure (which threatens future production). Taking account of these can be important for empirical studies that attempt to explain observed prices and output levels.

- 8 The ISO actually operated a real-time imbalance market in order to make up any differences between real-time demand and the amount contracted before on the day-ahead market. Technically it is the ISO market that is the true spot market and the PX day-ahead market is a very short forward market. However, in the first two years of operation only 5% of electricity was purchased in the ISO market, while 85% was purchased in the PX (and the remainder bilateral trades from those entities allowed to make them). Furthermore, there were no significant differences in prices in the two markets during this period, which is what one would expect given the ease with which bidders could switch from one to the other. See Borenstein *et al.* (2002).
- 9 Reported in 'Hayward: Panel Gives Nod to Energy Plant in Hayward', San Francisco Chronicle, Sept. 12, 2002.
- 10 Others have written about long-run supply issues and have noted the importance of short-run prices being enough above operating costs to recoup capital costs if entry is to occur, *e.g.*, Brennan (2007).
- 11 For more detail see Blumstein *et al.* (2002).
- 12 In the case of the UK, the politics of coal also helps to explain the initial vesting contracts. British Coal wanted to be sure that generation plants would continue to demand their coal as they had been before the privatisation, and secured vesting CFDs that ensured coal would be bought at an agreed upon strike price for three years following privatisation. The generators that had to sign these vesting contracts demanded, in turn, that the regional distribution companies sign vesting CFD contracts with them to buy the electricity bought with this coal. Initially, about 95 percent of generation was covered by these contracts. For further information about the UK market, see Powell (1993) and Lowrey (1997).
- 13 This market has been expanded and now includes coverage for thirteen states and the District of Columbia. See PJM Interconnection (2003) for additional information.
- 14 See US Department of Energy (2000).
- 15 If the price is high enough, additional electricity may be found from somewhat unusual sources. During the California crisis when spot prices were in the hundreds of dollars per megawatt hour, Kaiser Aluminum Company in the Pacific Northwest received its electricity at a low contracted price from Bonneville Power. It decided to shut down production during much of 2001 in order to resell its electricity in California at a reported profit of \$460 million, far more than the company profit in 1998 of \$28 million.
- 16 We abstract away from certain practical problems like the time necessary to start or stop different types of generators. The point of the model is that forward contract commitments made years in advance will be carefully reviewed in light of conditions in the months, weeks, days, and hours ahead as the time of reckoning approaches, and there is much opportunity to take advantage of any room for deals as suggested.
- 17 The 10 GW-hrs represent the quantity traded that is not obligated by any forward contract agreed to earlier. There could be additional trading in the spot market to transfer some of the forward obligations, but most of this would be done in advance of the actual spot market. Our diagram is not meant to exclude such trading from the spot market, but is simply intended to show the residual, unobligated demands and supplies after accounting for the settling of obligations. If we included any spot market trades to settle obligations, it would not affect the equilibrium but would make less clear the size of the unobligated market.
- 18 A derivation of this is provided in the appendix of Landes and Posner (1981).
- 19 The marginal cost curve roughly approximates the actual curves shown in Borenstein (2002) and Joskow and Kahn (2002) with natural gas at about \$2.50/Mcf and NO<sub>x</sub> allowances at \$25/lb.



- 20 A complete model would have to demonstrate each outcome as the equilibrium of a specific oligopoly constellation. For example, Bushnell (2007) provides such a model based on Cournot behaviour of equal-sized oligopolists.
- 21 Joskow and Kahn (2002) have shown that outage rates from the AES/Williams, Dynegy, and Reliant generators during the summer of 2000 were far above their historical norms, concluding that much of this was due to strategic withholding.
- 22 The argument that follows is similar to the Cournot duopoly analysis originally offered in Allaz and Vila (1993). This model has recently been extended by Bushnell (2007), who found it empirically consistent with behaviour in the New England and PJM markets. Similarly the empirical findings of Wolak (2000) for the Australian electricity market, and the experimental evidence of Brandts *et al.* (forthcoming), also support the Allaz-Vila model result that forward contracting can substantially reduce market power. However, one must be careful not to overgeneralise these results. Other markets besides restructured electricity ones may have features that make behaviour in them quite different. The theoretical literature has other model variants – for example, repeated game models rather than a single production period like Allaz-Vila – including Liski and Montero (2006) with the opposing prediction that collusion can increase through forward markets. It is an open question if there are actual markets that operate like the Liski and Montero model, but they do not include the restructured electricity markets that have been studied to date. For other literature on electricity forward markets, see: Powell (1993), Lowrey (1997), and Bessembinder and Lemmon (2002).
- 23 Bushnell (2007) uses a Cournot model in which the equilibrium does not quite collapse to the competitive solution, although a similar dissipation of market power results. In his model, market power of  $1/n\epsilon$  possessed by one of  $n$  equal sized firms in a pure spot market with demand elasticity of  $\epsilon$  becomes reduced to  $1/n^2\epsilon$  by the addition of one round of forward contracting (*i.e.*, adding one round of forward contracting is equivalent to squaring the number of competitors). The Allaz-Vila model shows that as the number of contracting rounds increase, the equilibrium approaches the competitive level.
- 24 According to Wilson (2002, p.1326): “Mature systems show a pattern of up to 80% contracted long term, 20% day-ahead, and less than 10% spot.”
- 25 See US Federal Energy Regulatory Commission (2002).
- 26 This test is described in Appendix J of the 2006 PJM *State of the Market Report*, available online at <http://www.pjm.com/markets/market-monitor/som.html>.
- 27 See Borenstein (2005) on the efficiency of real-time pricing. Critical peak pricing is a rate structure that triggers a fixed but much higher than usual price, with advance customer notice of up to one day, in response to prespecified adverse system conditions (*e.g.*, unusually high wholesale spot prices, or aggregate demand in an area near system capacity for that area). For example, the ‘Good Cents Select’ plan of Gulf Power in Florida is a critical peak plan. Such plans are usually combined with technology like ‘smart’ thermostats that are pre-programmed by customers to do things like allow air-conditioned temperatures to rise several degrees or shut-off swimming pool heaters.
- 28 The behavioural approach has been used for another electricity problem, that of understanding tiered rates. For an exposition of the problem, see Friedman and Hausker (1988). For an econometric investigation of it, see Friedman (2002).
- 29 Advances in public goods mechanisms designed for honest revelation of preferences could, in the future, make it possible for these grid users to make an efficient determination of the reliability level.
- 30 According to California Energy Commission data on the fossil fuel plants, PG&E divested 2881 MW to Duke Energy and 3166 MW to Mirant (formerly Southern). SCE divested 4706 MW to AES, and 4019 MW to Reliant (formerly Houston Industries). SDG&E divested 1347 MW to Dynegy and NRG Energy, and SCE also divested 1583 MW to NRG Energy and Destec. This data is available on the web: [www.energy.ca.gov/electricity/divestiture.html](http://www.energy.ca.gov/electricity/divestiture.html).

## Appendix

### *A competitive fringe model with California parameters*

In this model, we assume that 10 GW of power with the lowest short-run marginal operating costs of \$15 per MW-hr are 'competitive' baseload resources (these represent primarily hydro and nuclear sources that were operated largely on a 'must-run' basis). These generators will always operate at full capacity in the model. We assume that the next 20 GW of power are owned by firms that can try and exercise market power, and will use several alternative assumptions about the nature of the firms and their behaviour. This 'middle' 20-GW segment of the market is intended to represent the fossil-fuel generators that were divested primarily to five out-of-state companies, and for simplicity they are assumed to operate at short-run marginal cost of \$20 per MW-hr.<sup>30</sup>

Finally, the third segment of the model is the more expensive part of the competitive fringe consisting of about 15 GW of capacity, characterised by short-run marginal costs that rise gradually from \$20 to \$25 for the least expensive units up to 10 GW, but then rise very steeply for the next 5 GW of capacity such that the marginal cost of the most expensive unit in this segment is \$250 per MW-hr (there is no actual capacity limit, but further supplies are so expensive that they will not be utilised). Because the third segment is intended to include peaker plants, we will simply refer to it as the peaker segment. The use of this 'piecewise linear' competitive fringe allows for simple calculation of profit-maximising strategies for the oligopolist, and the entire supply structure assumed approximates the actual marginal cost of California generation over the range of consumption observed during the 1998–2001 restructured period.

This model is designed to yield competitive solutions with prices and quantities similar to outcomes that sometimes occurred in California. We compare these solutions to those achieved when the oligopoly firms attempt to exercise market power. They have to figure out the profit-maximising capacity to withhold. This is made complicated by the competitive peaker segment, which will increase its supply in response to any oligopoly withholding. The market clearing equilibrium condition in this segment is that the peaker marginal cost equal the market price. The oligopolist must take this response into account in order to choose the profit-maximising quantity to withhold.

Market demands in this model are given by the following equations (with all quantities in Gigawatt-hours and prices in Megawatt-hours):

$$\begin{aligned} \text{Normal demand:} & \quad Q_D^N = 36.46 - .02P \\ \text{Corner demand:} & \quad Q_D^C = 40.50 - .02P \\ \text{Hot day demand:} & \quad Q_D^H = 43.80 - .02P. \end{aligned}$$

The competitive supply curve is identical to the assumed marginal cost structure:

$$Q_s = \begin{cases} 0 & P < 15 \\ 10 & 15 \leq P < 20 \\ 30 & P = 20 \\ 2P - 10 & 20 < P < 25 \\ 39.444 + .0222P & P \geq 25. \end{cases}$$

The competitive solution for a normal day is at  $P = \$23/\text{MW-hr}$  and  $Q^N = 36 \text{ GW-hrs}$ . This means that the last 6 GW-hrs come from the peaker segment. At the corner, the price is only slightly higher at  $P = \$25$  even though the equilibrium quantity has grown to  $Q^C = 40$ . For the hot day, the competitive solution is at  $P = \$103.21$  and  $Q^H = 41.7$ ; price has risen very sharply while quantity has only increased by a relatively small amount. This is a crude approximation of competitive conditions when  $\text{NO}_x$  pollution permit prices in southern California rose dramatically, raising substantially the marginal cost of several peaking sources in that area. In our model, this implies that the competitive peaker fringe is providing 11.7 GW-hrs, thus utilising much more expensive marginal sources than on a normal day.

Now let us turn to the oligopoly behaviour: will it be in the interests of an oligopolist to withhold some production in order to raise market price and increase profit on the quantity sold?

The oligopolist will calculate as follows. Beginning with the market demand equation, subtract from it the competitive baseload supply, the peaker segment competitive supply, and any portion of the oligopoly segment that is assumed not under its control and to behave competitively. This will give the residual demand quantity as a function of market price that the oligopolist of a given size faces. To find the profit-maximising quantity, calculate the marginal revenue function from the residual demand and identify the quantity at which marginal revenue equals the marginal cost of \$20. This quantity determines market price, the oligopolist's profit, competitive fringe supply, and consumer electricity expenditures. The optimal withholding is identified simply as the oligopolist's capacity less the profit-maximising quantity.

Let us illustrate this calculation for the hot day, and assume that one of the oligopolists controlling 4 GW attempts independently to exercise market power. The oligopolists controlling the other 16 GW of capacity will be assumed to behave competitively and supply it at any price greater than or equal to \$20.

The residual demand may be written as follows:

$$Q_R^H = \text{Hot day demand} - \text{Baseload} - \text{Competitive part of oligopoly supply} \\ - \text{Competitive peaker fringe.}$$

Since the price will end up being at least \$100 (recall the competitive outcome is  $P = \$103.21$ ), we know that the baseload supply will be the full 10 GW-hrs of capacity, and similarly for the 16 GW-hrs of competitive oligopoly capacity. We also know the competitive fringe supply  $Q_F$  will be the last segment of the competitive market supply curve (after removing from it the 30 GW-hrs of baseload and oligopoly supply that are being treated separately here):

$$Q_F = 9.444 + .0222P.$$

Therefore we can write the oligopolist's residual demand equation:

$$Q_R^H = (43.8 - .02P) - 10 - 16 - (9.444 + .0222P)$$

or

$$Q_R^H = 8.3556 - .04222P.$$

Solving for  $P$  as a function of  $Q_R^H$  we get:

$$P = 197.895 - 23.6842 Q_R^H.$$

Since this is linear, its associated marginal revenue function is simply:

$$MR = 197.895 - 47.3684 Q_R^H.$$

Finally, equating  $MR = MC = 20$  to find the profit-maximising quantity, we find:

$$20 = 197.895 - 47.3684 Q_R^H$$

$$Q_R^H = 3.765, P = \$108.95, \Pi = 334,047, Q_F = 11.865, Q = 41.621.$$

In other words, during one peak hour of the hot day the 4 GW oligopolist will find it worthwhile to withhold a small amount of power, 244 MW-hrs (= 4.000 – 3.756 GW-hrs). Because the competitive price would be so high (\$103.21), the profit foregone from this withholding is substantial: \$20,303. But the extra profit of \$5.74 per MW-hr on the 3756 MW-hrs sold is slightly greater, \$21,559, making the withholding profitable by \$1256. Note that all of the other generators in the market are free-riders who benefit from the withholding oligopolist. In particular, any other oligopolist that also controls 4 GW of power but supplies all of it enjoys a much greater increase in hourly profit of \$22,960. Absent collusion, there may be many missed opportunities to exercise market power as each chooses to deliver its full supply and hopes to free ride on another's withholding.

Table 1 in the text contains the results of analogous calculations made under different demand conditions and different assumptions about the size of the oligopoly that is attempting to exercise market power.