

DISCUSSION PAPER SERIES

No. 3010

FAILING ELECTRICITY MARKETS: SHOULD WE SHOOT THE POOLS?

Richard Green

INDUSTRIAL ORGANIZATION



Centre for **E**conomic **P**olicy **R**esearch

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP3010.asp

FAILING ELECTRICITY MARKETS: SHOULD WE SHOOT THE POOLS?

Richard Green, University of Hull and CEPR

Discussion Paper No. 3010
October 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **Industrial Organization**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Richard Green

October 2001

ABSTRACT

Failing Electricity Markets: Should we Shoot the Pools?*

This Paper discusses the electricity reforms in California and in England and Wales. In both cases, a centralized spot market played a major role, and both markets have now been abolished. This Paper argues that their disappearance is not evidence that future electricity restructuring should avoid the use of spot markets. Instead, the problems in England and Wales were largely due to market power. In California, problems arising from market power and a tightening demand-supply balance were turned into a disaster because the spot market had not been backed up by hedging contracts.

JEL Classification: L94

Keywords: electricity markets and hedging contracts

Richard Green
University of Hull
School of Economic Studies
HULL
HU6 7RX
Tel: (44 1482) 465 720
Fax: (44 1482) 466 216
Email: r.j.green@hull.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=114998

* I have gained from discussions about the Californian electricity markets with Carl Blumstein, Severin Borenstein, and Jim Bushnell, but would not like to implicate them in what follows. This Paper is produced as part of a CEPR Research Network on The Evolution of Market Structure in Network Industries, funded by the European Commission under the Training and Mobility of Researchers Programme (Contract No ERBFMRX-CT98-0203).

Submitted 16 July 2001

Failing Electricity Markets: Should we shoot the Pools?

Richard Green^{*}

University of Hull and CEPR

School of Economic Studies

University of Hull

Hull HU6 7RX

Tel: 01482 465720

Fax: 01482 466216

Email: R.J.Green@econ.hull.ac.uk

May 2001

This paper discusses the electricity reforms in California and in England and Wales. In both cases, a centralised spot market played a major role, and both markets have now been abolished. This paper argues that their disappearance is not evidence that future electricity restructuring should avoid the use of spot markets. Instead, the problems in England and Wales were largely due to market power. In California, problems arising from market power and a tightening demand-supply balance were turned into a disaster because the spot market had not been backed up by hedging contracts.

During the debate on how to deregulate the Californian electricity industry, one participant distributed campaign buttons with the message “Stop Poolco.” A vocal group was convinced that the restructuring should not be based upon the model of a single uniform-price auction market, combining electricity trading and its physical

^{*} I have gained from discussions about the Californian electricity markets with Carl Blumstein, Severin Borenstein, and Jim Bushnell, but would not like to implicate them in what follows.

dispatch, even though such a market was at the centre of the reforms in England and Wales, arguably the catalyst for the Californian reforms. When the new structure took effect, in April 1998, the California Power Exchange was only one of a number of schedule co-ordinators, reporting the results of its auction to the separate Independent System Operator. When the reform collapsed into chaos, the Power Exchange was the most prominent early casualty. It pre-deceased the Electricity Pool of England and Wales by only three months, however, for on March 27, 2001, the British government and regulator replaced the Pool with New Electricity Trading Arrangements. These were based on bilateral trading, with no place for the kind of uniform-price auction at the centre of the previous system.

Were the campaign buttons right? Are centralised electricity auctions bound to fail? This paper analyses the causes of the problems in California, and in England and Wales, to show that the auctions should not be blamed for most of them. In England and Wales, the Pool sometimes set inappropriate prices, and rule changes to improve matters were sometimes blocked by its members, but the major problem was market power. In California, market power was again a problem, turned into a disaster by regulation that required the utilities to sell at fixed prices, whatever the cost of buying power.

What do we need (and want) in an electricity market?

When designing an electricity trading system, there are things we need, and things that we should want. The division between these may depend on the commentator, but I believe that there are only two absolute needs. First, the system must remain stable, in an electrical sense - inflows must continuously match outflows, and transmission constraints must be observed. Second, all the power generated and consumed must be paid for, avoiding free riding.

My list of “wants”, characteristics for a trading system that are desirable, but not absolute requirements, is longer. Most transactions should take place at stable prices, to help participants plan their production and consumption. The system's dispatch should be as efficient as possible, minimising short-run costs. Investment decisions should also be efficient, to minimise long-run costs. In a market system, the best way to achieve efficiency is generally to have prices that signal costs. In the absence of externalities, participants who maximise profits, taking these prices as

given, should create an efficient outcome. Finally, it is best for the trading system to be stable in the sense of avoiding frequent rule changes, since these can be costly to implement, and create uncertainty for market participants, which can cost even more in the long run.

Traditionally, power stations were subject to central dispatch, in order to ensure the stability of the system. The first proto-markets, regional power pools, generally continued to operate in this way – the system operator in a “tight” power pool could give orders to any station in that pool, as if it was part of the operator’s own company. It was natural for the first true electricity markets to follow this pattern, as the industry was cautious about maintaining security of supply. Under this centralised model, all generators had to trade through the central market and follow its dispatch instructions, based on the bids they submitted. This was intended to guarantee security, and, given cost-based bidding, should ensure an efficient dispatch. All power flows were metered, and paid for at the prices set in the central market. Where price hedging was desired, it could be provided through financial contracts for differences, described below.

Once the markets based on “tight” pools were seen to be working, security of supply became less of a constraint on market design, and policy-makers were more willing to contemplate a greater degree of contractual freedom. The second wave of market designs allowed much more bilateral trading. Under these models, generators and retailers could trade bilaterally, decide which stations to schedule to perform the trades, and merely inform the system operator of their choices. Their trades would be netted off from their generation and demand when calculating how much power they were buying through the system operator. Some central trading had to continue, for some generators will inevitably produce more than they have sold (“spill” power) while others will need “top-up” supplies; similarly, retailers will not succeed in buying exactly the amount of power that their customers consume. Some mechanism had to cover these adjustments, and ensure that the power transferred through them was paid for. The system operator frequently ran this mechanism, for the adjustments were linked to the need to keep the system stable, and that remained their responsibility. In an emergency, the system operator must be able to order generators to change their output levels, paying appropriately if they do. Bilateral trading could offer participants stable prices (if they signed appropriate contracts), but efficient dispatch would depend on sufficient short-term trading to ensure that high-cost

stations that had sold their output in advance could be replaced by cheaper stations that had not contracted their output for that period.

In the next sections, we see how the markets in England and Wales and in California applied these different models – the Pool was a centralised “gross” market, while California was built around the bilateral model.

The Pool’s market design

The people who restructured the British electricity industry liked to think of themselves as pioneers. They may not have been aware that Chile had run an electricity market for many years, or that the Norwegians were gradually expanding the scope of their inter-company trading. In any case, they believed that they were proposing something radical when the White Paper outlining the restructuring was published in 1988 (Department of Energy, 1988). The restructuring proved more complicated than most had imagined, and it was not until the autumn of 1989 that a workable model for wholesale trading had been agreed. Many in the industry had been opposed to the decision to break the Central Electricity Generating Board into competing generation companies and a transmission company (the National Grid Company (NGC)), and were determined not to take risks with system stability.

To this end, the Electricity Pool of England and Wales basically replicated the CEGB’s procedures for scheduling generation. In the past, a computer algorithm had taken vectors of cost information and operating parameters for each generating set, and calculated the least-cost schedule that would meet the demand forecast for the following day. The Pool used the same computer program, but the companies could submit five price components in place of the vector of cost information. An extra program was written to average these price components and obtain the cost of power from each station – the System Marginal Price (SMP) was based upon the bid of the most expensive station in normal operation in each half-hour.

A second component, the capacity payment, was based on the Loss of Load Probability (the risk of a power cut) multiplied by the Value of Lost Load, set administratively at £2/MWh, and up-rated annually for inflation. This is a component of the optimal electricity price (Crew and Kleindorfer, 1976) and should ensure that stations that are only required at peak times can cover their fixed costs.

SMP and the capacity payment were paid for all the generation in the day-ahead schedule. Actual generation would inevitably differ from this schedule, and

NGC's controllers had to keep the system stable. Stations would be instructed to vary their output levels to meet changes in demand, and to react to plant failures. There would be transmission constraints, where the demand in part of the system, net of the scheduled local generation, exceeded the amount of power that the network could import. Some stations in the area would have to be "constrained on", generating more than had been scheduled in order to relieve the constraint on the transmission lines, while others, outside the area, would be "constrained off" to keep total generation in line with demand. All the stations producing more than had been scheduled would have bid more than the ruling SMP (or they would have been scheduled), and so they were paid their own bid price for the additional output. All the stations producing less than scheduled were required to buy back the unwanted power, but at their own bid price, which was less than SMP (or they would not have been scheduled initially). The difference between SMP and their bid was sometimes known as a "lost profit payment". The costs of these payments, of capacity payments to stations that were available but not scheduled to generate, and of buying reserve and other ancillary services, were recovered in a price component known as Uplift.

To some extent, this approach might be portrayed as saying that the Pool dealt with transmission pricing issues by ignoring them – all electricity was sold at the same price. The combination of constrained on and constrained off payments (also known as counter-trading) ensures that generators are willing to follow the dispatcher's instructions.¹ The prospect of high constrained-on payments gives generators an incentive to locate in areas where transmission constraints create shortages of power. The lost-profit payments, however, mean that generators have no incentive to avoid areas where the constraints create power surpluses. Similarly, the Pool ignored transmission losses when deciding which plant to schedule (although the software could have taken them into account) – all stations were treated equally, and suppliers' metered demands increased pro-rata with the average losses across the system. This was despite the fact that 106 MW of power generated in the north of the country would be needed to replace 94 MW generated in the south-west, due to the marginal losses on north-south power flows.

¹ Hogan (2001) reports the failure of the first version of the Pennsylvania-New Jersey-Maryland market, which allowed bilateral trading, and dealt with congestion by allowing the system operator to constrain generators off, without giving them any lost profit payment. All the generators at risk of being constrained off quickly switched to bilateral trading, leaving the system operator in charge of insufficient plant to manage congestion.

Pool prices were expected to be volatile at times, but a hedging mechanism was also designed. Indeed, this hedging was central to the privatisation, in giving the companies involved predictable cash flows for their first few years, and ensuring that the British coal industry remained viable during the transition. Contracts for differences set payments between generators and suppliers, based on the level of the Pool Purchase Price (SMP plus the Capacity Payment). Under a one-way CfD, the generator would pay the supplier the difference between PPP and the strike price of the contract, whenever PPP exceeded the strike price, in return for an up-front payment. A contract with a strike price equal to the marginal cost of the station, and an up-front payment equal to its fixed costs (plus a profit margin), would effectively mean that the contract's buyer was buying the station's output at a price close to cost. The contract would be triggered whenever the Pool price was above the strike price, which should be the times when the station is running, if it bids at marginal cost. The generator would therefore replace the uncertain stream of Pool prices with payments equal to its costs, as long as it generates when required.

Two-way CfDs require the buyer to pay the seller whenever the Pool Purchase Price is less than the contract's strike price. They act as a straight hedge on the price, and would normally be traded with a strike price equal to the expected Pool Purchase Price (plus any margin for risk hedging) and no up-front premium. At the time of privatisation, however, a large number of three-year contracts were signed with prices significantly above the expected Pool Purchase Price. The excess allowed the two major coal-fired generators, National Power and PowerGen, to pay above the world market price for British coal, even though Pool prices were expected to relate to the marginal cost of fuel, which was cheaper imported coal. The contracts' buyers, the twelve Regional Electricity Companies, were allowed to pass the extra cost on to their smaller consumers, who were not to get a choice of supplier until 1998. Larger consumers were allowed to shop around from the start, and did not have to pay a premium for British coal – any contracts to support their purchases were based on the expected Pool price.

Appropriate CfDs can also be combined with the capacity payments mechanism to promote efficient decisions about plant closures. Many commentators have suggested that capacity payments were intended to provide a signal for new investment, and then pointed out that it is inappropriate to base such decisions on a price that changes every half-hour. For at least twenty years, however, the level of

capacity in England and Wales has been “fine-tuned”, not by new investment, but by bringing forward or delaying the closure of old stations. CfDs exist which make payments with reference to the capacity payment alone, and the strike price should be based on the expected value of capacity payments over the year. If there is a lot of spare capacity, this expected value should be low, and less than the cost of keeping at least some of this capacity open. That can be seen as a market signal that closures would be appropriate, remembering that the capacity payment is intended to reflect the expected value of the energy that would be lost through capacity shortages. When there is less spare capacity, the expected capacity payment will be higher, raising the value of the CfD, and signalling that stations should be kept open. While the CfD can “lock in” revenues for a station that is kept available throughout the year, the station is still free to make day-to-day decisions about its availability on the basis of the capacity payment expected for the following day. It should be efficient to make the station available whenever the expected capacity payment exceeds the cost of doing so, but not when the expected capacity payment is lower. This efficient course of action should also be privately profitable, for the owner of a single station.

Experience with the Pool

The caveat at the end of the previous section is crucial for understanding the problems that have been experienced in the UK electricity market. The Pool ought to produce a reasonably efficient outcome, as long as there is no market power. In the presence of market power, any mechanism is likely to have problems, and the Pool certainly did. The government had intended to privatise the whole electricity industry, including the nuclear stations that accounted for nearly 20% of output, but had not performed well in the past – several had yet to be formally commissioned after twenty years’ construction. A large generating company, with 70% of the industry’s capacity, was created in the hope that this could absorb the risks of the nuclear stations’ future performance. A second company, with the remaining 30% of capacity, was a weak counter-balance to the first. During 1989, it became evident that the nuclear stations were felt to be too expensive and risky to be privatised, even in a very large company, and they were withdrawn from the sale. It was too late to alter the rest of the structure, and so National Power, the larger of the two companies originally planned, inherited 60% of the industry’s conventional capacity, and PowerGen inherited the other 40%. It soon became evident that they had the ability and incentive to raise

prices well above competitive levels (Green and Newbery, 1992; Wolfram 1998, 1999).

At first, Pool Prices were very low: much lower than the government had expected. The generators' contracts were at fixed prices, and so they did not worry about the level of Pool prices. They had an incentive to burn as much coal as possible, and so bid low prices. In April 1991, however, some of these contracts expired, and the generators started to raise prices, as shown in figure 1. Some large customers had been persuaded to buy at Pool prices and were exposed to the increase. Their complaints led to the regulator's first enquiry into the major generators' behaviour. There have now been more than ten reports on competition in generation.

The first reports concluded that Pool prices had been below the level of the generators' avoidable costs. This meant that it was difficult to object to an increase, although some specific bidding strategies were criticised. The regulator also criticised some features of the Pool Rules which led to price spikes. This is when the price might treble (or worse) for two or three half-hours. It can be hard to determine the cause of a particular spike, given the complexity of the scheduling program and the bids. Price-setting is a mechanistic process which occasionally seems to give absurd

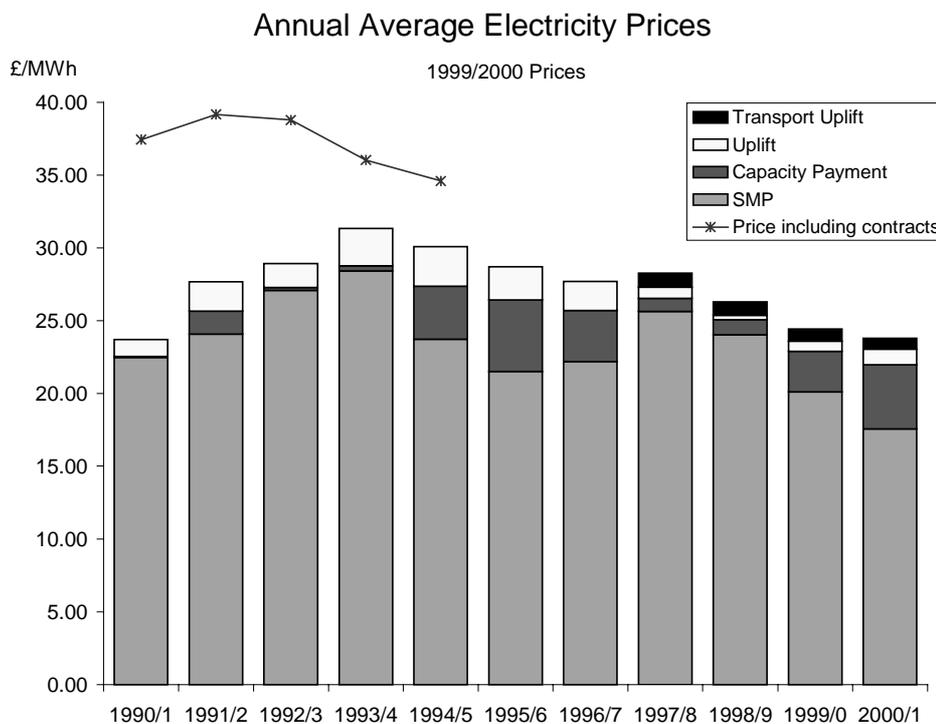


Figure 1

results. The prices have to stand, because contracts have been written around this process.

The Pool has modified some of its rules to prevent certain kinds of price spike. However, it has generally been difficult to change the Pool Rules, as a result of a deliberate policy decision taken at the time of privatisation. The government was concerned that the industry might try to make life difficult for entrants, or that generators might try to change the rules and disadvantage suppliers. Voting in the Pool was on the basis of market shares, with half the votes for generation and half for supply, so that in the early days, when some generators had established supply businesses, but suppliers had practically no generation, the generators had more than half the votes. To prevent this, the Pool Rules required super-majorities for change, and gave members who disagreed with a given proposal several opportunities to delay it, and then to appeal to the regulator. The knowledge that he might have to take on a quasi-judicial role inhibited the regulator from championing some desirable rule changes.

The biggest failure was perhaps the issue of transmission losses. When the Pool was established, it was impossible to reach agreement on whether the charges for losses should be differentiated on a geographical basis to reflect the marginal cost of transmission, and the issue was put in a schedule of matters to be decided later. In due course, a scheme was proposed that would effectively have reduced the net revenues of northern generators, and raised those of southern power stations, with opposite effects on suppliers. A coalition of companies who stood to lose from the changes twice appealed decisions on the issue to the regulator, and finally asked for a judicial review of the matter. The case was tied up in the legal system until it was no longer relevant, with new arrangements due to be introduced when the Pool was abolished.

The Pool was more successful in dealing with the other transmission-related charges, those recovered in Uplift. Uplift rose rapidly after privatisation, in part because NGC had to take a lot of transmission circuits out of commission in order to connect new generators, thus creating temporary constraints. NGC had no incentive to minimise the cost of these, since they were passed straight through to consumers, whereas overtime working, for example, would add to NGC's costs. Following a suggestion from the regulator, NGC and the rest of the industry negotiated an Uplift Management Incentive Scheme. If the out-turn level of Uplift was below its target,

NGC would receive a share of the savings, while if it was higher, NGC would absorb a proportion of the excess costs. Successor schemes became more sophisticated, dividing Uplift into several portions, and giving NGC greater incentives to reduce the parts that it was more able to influence, while some Uplift charges were later transferred into a Transmission Services Use of System charge. Table 1 shows that the cost of Uplift generally fell from the mid-1990s onward.²

Table 1: Uplift costs (£ million)

	Constraint costs	Total Uplift
1990/1	50	267
1991/2	145	475
1992/3	136	383
1993/4	255	592
1994/5	194	718
1995/6	73	655
1996/7	58	595
1997/8	25	412
1998/9	21	370
1999/2000	15	476

While the progress in keeping Uplift under control has been helpful, it is only one component of the Pool Selling Price. The overall level of prices is of most concern to suppliers and customers. In his first reports, the regulator did not object to the level of prices, but in 1993 he decided that recent price increases had been unjustified. He threatened to refer the companies to the Monopolies and Mergers Commission, which could have broken them up. To avoid this, they promised to keep prices down for the next two years, and to sell about 15% of their plant. They did this, and the real time-weighted Pool Selling Price fell in every year after 1993/94, with one exception, when the lower capacity payments of 1997/98 were more than offset by an increase in SMP. The major generators raised their bids to maintain their revenues, and were criticised for this by the regulator. The market was becoming

² Uplift includes the cost of Capacity Payments to stations that are available but not scheduled to generate, and so the relatively high capacity payments of 1991/2, 1994-7, and 1999/2000 will be reflected in the Uplift figures for those years.

Generation in England and Wales

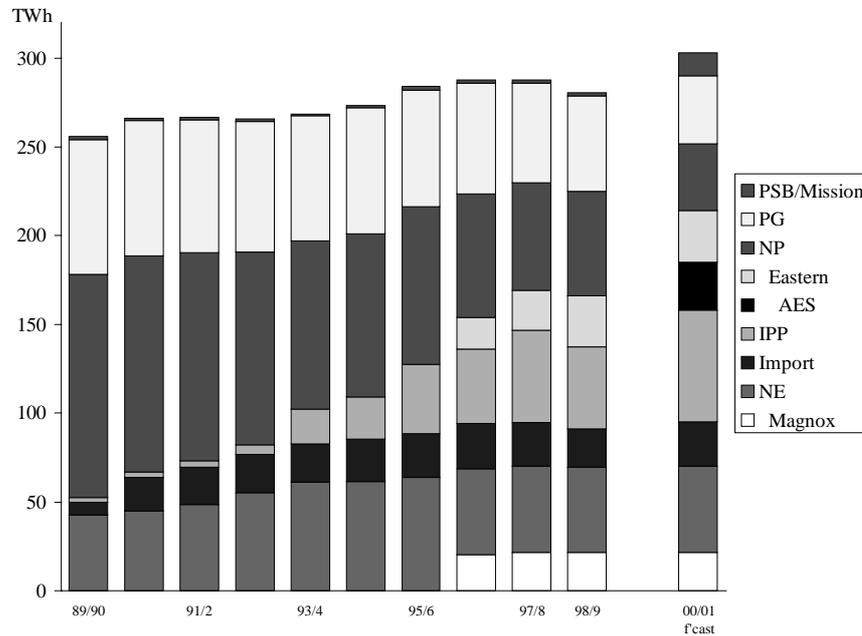


Figure 2

more competitive, however, with entry on a large scale, and further divestitures by National Power and PowerGen. Figure 2 shows this gradual fragmentation.

Despite this, in October 1999, the regulator asked the eight largest generators to accept a “good behaviour” condition in their licences. It would have allowed him to intervene if he had believed that they were acting in an anti-competitive manner. Two companies, British Energy and AES, rejected the condition, and the matter was referred to the Competition Commission. The Commission decided that the condition was not required, and the regulator withdrew it from the licences of the companies that had accepted it. One major factor in the decision was that the Pool was about to be abolished, and concerns about the generators’ ability to manipulate the Pool’s pricing mechanism would only apply for a few more months.

The New Electricity Trading Arrangements

The regulator had long been growing impatient with the Pool’s inability to change its rules and reform its procedures. In May 1997, a Labour government was elected, with historical ties to the UK’s dwindling number of coal miners. Since privatisation,

the demand for coal had nearly halved, as coal-fired stations were replaced by Combined Cycle Gas Turbines. There were accusations that this had only happened because the nature of competition in the Pool was unfair, due to its uniform-price rule. Gas-fired stations could submit a bid of zero, but still receive the Pool Purchase Price, set by coal-fired stations that had to submit higher bids, and therefore lost market share.³ The fact that all stations received the price set by the marginal bidder was also believed to make the Pool more vulnerable to the exercise of market power. In October 1997, the government announced a review of the Pool. The following year, an Energy White Paper announced that the Pool would be replaced with New Electricity Trading Arrangements (NETA), which started, after some delay, in March 2001.

NETA has replaced the Pool's centralised market with as much freedom to contract as possible. There is no central market until 3½ hours before real time, and most electricity is being traded via bilateral contracts. Several companies have set up power exchanges to facilitate trading in the last day or so before real time operation, left entirely to their own devices by the government and the regulator. In the last 3½ hours before real time operation, NGC runs a balancing mechanism, keeping the system stable by accepting bids and offers to raise or reduce generation and consumption. Each accepted trade is paid its own bid or offer.

Traders have to notify Elexon (formerly described as the Balancing and Settlement Company) of their electricity contracts, which are compared with their metered demand or output. Companies with imbalances between their contracted positions and their metered quantities have to cash them out at unattractive prices. Companies needing to buy power (generators who have under-generated, and suppliers with consumers who have bought more than the supplier contracted for) have to pay the System Buy Price (SBP), the average of the prices that NGC paid to buy power in the Balancing Mechanism. Companies needing to sell power get the System Sell Price (SSP), the average of the prices that NGC received for selling power in the Balancing Mechanism. The intention is to give companies a strong incentive to contract accurately in advance, for the SBP is normally expected to be greater than the cost of power bought in advance, while the SSP is expected to be lower. In the first month of NETA's operation, the median difference between the

³ In other words, the incumbents kept prices high, encouraging entry, and therefore losing market share.

SBP and SSP was £30/MWh, and the maximum was £2,354/MWh. The SBP exceeded £150/MWh in 9% of half-hours, while the SSP was negative in nearly a quarter.

It is very early to assess NETA's performance. Volatile balancing market prices were to some extent a design objective, to encourage early contracting. Generators who cannot predict their output in advance, such as wind generators, are likely to make losses because of this volatility (Bathurst and Strbac, 2001), and the regulator has announced an enquiry into their position.

For consumers, however, the overall price of power will be a more important measure of NETA's success or failure. The regulator has complained that Pool prices were as high at the end of the 1990s as at the start of the decade, despite significant reductions in generating costs. Figure 1 shows that this is true, and that it is misleading. The effective price for most electricity trades are set by contracts, and average contract prices were well above Pool prices in 1990/1. Over the next few years, however, the differential between contract prices and Pool prices fell.⁴ This means that the effective price of electricity *has* declined significantly over the decade, even though it may still be above the level of a new entrant's costs.

The government and the regulator have argued that moving from uniform pricing to bilateral trading will reduce the scope for generators to exercise their market power. Studies in auction markets for government securities (quoted by Newbery, 1999) imply that moving from uniform pricing to a discriminatory auction will make prices more competitive for a given number of traders, but may discourage entry, reducing the number of traders, which could make prices less competitive overall. Bower and Bunn (2000) used a simulation model to predict that moving from the Pool to a discriminatory pricing auction would increase prices. The regulator and the government have pointed to the reduction in Pool prices at the end of the 1990s as evidence that the market was anticipating NETA.

It is hard to see how an expected reduction in the future price of a commodity can affect its current price, when the commodity in question is non-storable. The normal arbitrage conditions via storage fail to apply. That means that the price

⁴ The line in the figure takes actual revenues per kWh generated for National Power, PowerGen and Nuclear Electric, and assumes all other sales were made at the Pool Purchase Price. From 1992/3 onwards, entrants linked to the RECs generated an increasing proportion of electricity. Their contract prices were well above Pool prices, and calculations based on the major generators' selling prices would have given an underestimate of average prices from the mid-90s onwards.

reductions preceding NETA have more to do with a reduction in market power, as National Power and PowerGen sold a number of their stations to other companies, than the future change in market rules. The companies' incentive to use their market power may also have been reduced by a fear that the new regulator would become less tolerant in the run-up to NETA.

To find the true impact of NETA, we must await a study of pricing behaviour, linked to market structure, once the market has been operating for a reasonable period. At present, we can only say that although the Pool's governance structure proved inadequate, it might have been reformed for a lot less than the £500 million that NETA has cost to implement. NETA will need effective arbitrage between a number of separate markets if it is to produce an efficient dispatch. California chose a market design with several markets, and it is to that design that we now turn.

California's Market Design

California started to deregulate its electricity industry in the mid-1990s. The California Public Utilities Commission created an outline plan known as the Blue book in 1994, but the restructuring was the formal creation of the State Legislature, in the form of Assembly Bill 1890, passed unanimously in 1996.

The open nature of American law-making ensured that a vigorous debate preceded AB1890. As already mentioned, the vociferous lobby which opposed the idea of a compulsory centralised market won the argument. Some people were arguing from conviction, in the belief that a monopoly market-maker was likely to be inflexible and inefficient, compared to organisations forced to compete for customers. Stoft (1997), however, points out that some participants in the debate could have had other motives. Energy traders, or power marketers, make most of their profits from exploiting arbitrage opportunities – selling something for more than it cost to buy. In markets where arbitrageurs are successful, this normally implies that they are performing a useful service, often bearing risk, and ensuring that prices converge to efficient levels. Stoft's argument, however, is that some participants in the California debate had an incentive to deliberately design an inefficient market, maximising the arbitrage opportunities that they could profit from.

Arbitrage depends on the same product being traded in more than one market, and the Californian reform certainly achieved this pre-condition. An independent system operator was established to control the network, running a real-time market to

keep generation and demand in balance. The ISO also ran five markets for different kinds of reserve, ranging from regulation reserve (part-loaded plant able to respond instantly to a shortage of power) to half-hour reserve.

While the ISO was responsible for keeping the system stable, it was not responsible for scheduling plant. Any scheduling co-ordinator (SC) could submit a balanced schedule, of matched generation and demand, to the ISO. The SC could include ‘incs’ and ‘decs’, prices for which the generators it controlled would be willing to increase or decrease output. The ISO then had to assess whether the schedules were feasible, or whether they would create congestion on the lines (“Path 15”) linking the north and the south of the state. If the initial schedules were infeasible, the ISO would use the information in the inc and dec bids to find a feasible alternative dispatch, and would suggest this to the scheduling co-ordinators. Only after the scheduling co-ordinators had returned to the ISO with their revised schedules could the ISO take over the problem of congestion management, buying and selling itself to ensure a feasible dispatch. If path 15 was congested, the state would be split into two zones, and the ISO’s real time market would produce a different price for each.

Scheduling co-ordinators who sent power across the constrained boundary would have to pay (or receive) the difference between the two zonal prices as a transmission charge. The ISO also had to resolve intra-zonal congestion, which it handled in the same way as the English system, buying from and selling back to appropriately located stations, and spreading the cost across all consumers. The ISO attempted to hedge the cost of some constraints by signing “reliability must run” contracts with strategically located plant.

The largest scheduling co-ordinator was the California Power Exchange, set up by the state as a not-for-profit organisation. The PX ran a day-ahead auction in which generators and retailers submitted supply and demand schedules for each hour, and the PX found the hourly market-clearing price and quantity at the intersection of the schedules. The PX had devised an elaborate set of rules for adjusting bids if the auction did not at first produce an equilibrium, designed to ensure rapid convergence to an efficient price, but it turned out that these were not required and the PX continued as a single-iteration auction. The PX later added “day of” auctions to allow participants to adjust their positions after the first auction had closed, but before the ISO’s real-time market opened.

The three large investor-owned utilities in California, Pacific Gas and Electric, San Diego Gas and Electric, and Southern California Edison, were required to buy all of their requirements from the PX, and to schedule their generation through it. They were also required to divest at least half of their fossil-fuelled generating capacity, to reduce the potential for market power. In return for this, the reform included a Competitive Transition Charge (CTC), designed to enable the companies to recover their stranded costs – the difference between the market value of their generation assets after deregulation, and the amount that the former system of regulation would have allowed the companies to earn for them.

The level of the CTC in any hour was not set in advance, although each company was given a maximum amount that it could recover before December 31, 2001, when the CTC would expire. AB1890 had included a 10% rate reduction for residential customers, to take effect at liberalisation, followed by a four-year rate freeze. In each hour, the CTC for each utility was the difference between its frozen rate and the sum of the PX price and its (still regulated) charge for transmission and distribution. This provided the utility with an apparently perfect hedge in the short term – if it bought in the PX, the three cost components (transmission and distribution, electricity purchase, and CTC) would exactly equal its revenue. If the PX price was less than expected, the CTC would be increased, and the company's stranded costs would be paid off more quickly. The total amount to be recovered would not change, however, leaving the company insulated from the impact of the PX price. Once each utility had recovered all of its stranded costs, the price it was allowed to charge would equal the PX price, plus its charge for transmission and distribution. As before, the utility would be hedged against fluctuations in the PX price, which would be passed on to consumers.

The first exception to this apparently perfect hedge was that it did not fully protect the utilities against high PX prices. As long as the CTC was going to be fully paid off before December 2001, the utilities' revenues would be equal to their transmission and distribution charges, plus the PX price, plus their stranded costs – in other words, equal to their costs. If PX prices became too high, however, the CTC in each period would be lower, and it might not be possible to pay off all the stranded costs before December 2001. In that case, the utilities' revenues over the first four years would be fixed by the rate freeze, but their costs would depend upon the level of the PX price, leaving them potentially vulnerable to increases.

The second exception would occur if it turned out not to be possible to pass on increases in the PX price to customers, even after the CTC had expired. Most discussions of retail competition in the US have been based on the idea that the incumbent utility would offer some kind of “default service”, passing through the spot cost of buying electricity, and that customers who wanted the added value of a hedging contract would have to obtain it from a new entrant. In some states, customers who switched away from the incumbent utility were given a generous “shopping credit”, well above the expected wholesale price, which allowed competitive retailers to cover the costs of buying power, make a profit, and still undercut the incumbent. Retail competition was relatively “successful” in these states, in that a significant number of consumers chose to buy the cheaper power.⁵ The competing retailers had the option of making arrangements to hedge their power purchase costs in order to offer fixed prices. In California, however, the shopping credit was set equal to the PX price, leaving no margin for a retailer to cover its own costs. Very few customers switched, and most of those did so in order to buy “green power” from renewable generators. This meant that in California, practically all customers would be exposed to changes in the PX price, once the CTC had expired. Should PX prices ever be seen as excessive, this could lead to political problems.

California’s Experience

At first, the Californian electricity markets appeared to have got off to a good start. Some problems appeared in the ISO’s reserve markets, where one of the market rules required the ISO to buy fixed quantities of each kind of reserve, whatever the price, even if a higher quality reserve (i.e., one available at shorter notice) was available at a lower price. The ISO was allowed to remove the restrictions, and things improved.

At first, most companies bidding into the ISO’s markets were required to base their bids on their costs, but the Federal Energy Regulatory Commission gradually gave an increasing number of participants the authority to set “market-based rates”. Because of the fear of market power, there was a bid cap of \$250/MWh in the ISO’s real-time market. This bid cap naturally affected the shape of the demand curves in

⁵ The UK adopted an equivalent system, in that the incumbents’ charges were split into generation, transmission, distribution and supply, and entrants only had to pay the incumbents for transmission and distribution, leaving a margin to cover their own supply costs.

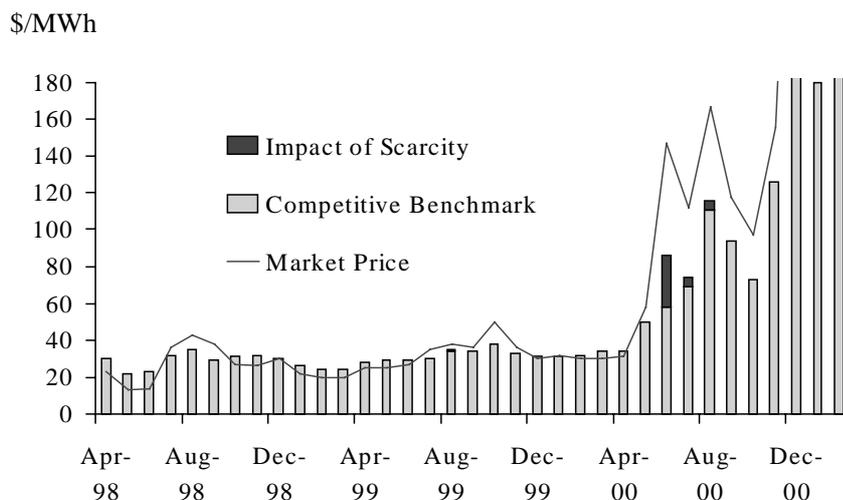
the PX's day-ahead market – no buyer would wish to pay more than \$250 in the PX, when they knew the price in the real-time market would be capped at this level. More generally, we should expect prices in the real-time market and the PX to track each other closely, in that the PX price should be an unbiased predictor of the real-time price, in the absence of risk-aversion and transaction costs (Borenstein *et al*, 2000). Borenstein *et al* found that there were significant differences between the two prices during 1998, but that the prices converged during the first half of 1999. They took this to imply that market participants were learning over time, and that the interactions between the markets were gradually becoming more efficient, although some inefficiencies persisted. They are careful to point out that this is not a test for the efficiency of the individual markets. Nevertheless, their evidence does not support the argument that a centralised pool is essential because separating electricity trading across several markets will inevitably be inefficient.

The specific inefficiency that Borenstein *et al* mention is market power. Puller (2001), for example, found direct evidence of market power during peak hours in 1998 and 1999 – companies had unused capacity with marginal costs below the market price. On average, however, the ISO's Department of Market Analysis found that prices were below a competitive benchmark (based on marginal cost bidding) in many months during the first two years of the market's operation (figure 3).⁶ Prices were 20% above the benchmark during the summer of 1998, however, and also rose above the benchmark level in the summer and autumn of 1999.

This was a warning sign – demand had been growing rapidly in California, and there had been little new investment for a decade. Some commentators blame tight environmental restrictions; others point out that no plant has actually been refused on environmental grounds, although some potential investors may just have seen no point in making an application they felt sure would be refused. In the middle of the decade, uncertainty over new market rules is likely to have discouraged investment.

⁶ Hildebrandt's figures for July – August 1998 seem inconsistent with his price graph for the same period, and so my figure has been drawn to match a graph by Sheffrin (2001), using a (possibly updated version of) the same analysis.

Wholesale electricity prices in California



Source: Hildebrandt, 2001, table 2-1; Sheffrin, 2001

Figure 3

In May 2000, as demand rose for the summer, the competitive baseline cost rose. Prices rose by more, however, as the narrowing gap between demand and supply increased the scope for the exercise of market power. The problem became far worse in June, when the reserve margin fell below 10% of demand in a significant number of hours. The mark-ups achieved in those hours have been added to the competitive benchmark prices in the figure, since a competitive market would require a “rationing price” when demand comes close to exceeding supply. Prices were even higher in August, and then fell slightly at the start of the autumn.

Note that the competitive price was more than double the level of the previous year. Part of this was due to the narrowing gap between demand and capacity, part was due to an increase in the price of gas. Natural gas spot prices at the Henry Hub in Louisiana rose by 50% between the spring and autumn of 2000. The price in California was almost identical to the Henry Hub price in April 2000, but 20% higher for a few weeks at the end of the summer. This differential, linked to a shortage of capacity on the pipeline into California, reappeared in November and worsened – at one point in early December, the Californian price exceeded \$50/MMBTU, while the Henry Hub price was below \$10/MMBTU. The capacity shortage may well have been affected by market power in the gas market – its consequences in the electricity market were disastrous.

Both the “competitive benchmark” (based on the California spot gas price) and the actual average market price went “off the graph” in December 2000 – the ISO calculated the average price paid as \$395/MWh, compared to a competitive benchmark of \$285/MWh. The utilities could not cope with prices at this level. They were buying most of their power on the spot market, but selling it at prices fixed under the rate freeze. In San Diego, prices had been unfrozen, but the California Public Utilities Commission capped prices again in response to consumer protests. The utilities had received authority to hedge a small volume of purchases outside the PX, but had taken little advantage of this, apparently because the prices available at the time had seemed excessive (Faruqui *et al*, 2001). The utilities stopped paying for power, leaving some generators unable to buy fuel, and forcing them off-line. Other generators seem to have stopped supplying because they feared they would not be paid in future. Supplies from outside the state fell for a similar reason, and because low rainfall had reduced the amount of hydro-electric energy available. Demand exceeded the available supply (although the winter is *not* the peak season for demand in California), and the ISO had to commence a programme of rolling blackouts at peak times of day.

The ISO had to resort to physical rationing, because the mechanisms for rationing by price were missing. Most consumers were still paying capped retail rates. Relatively few customers had time-of-use metering, that would allow their suppliers to charge prices that varied within the day. Some of those customers were on time-of-day pricing schemes, paying more at peak times, but those schemes were not linked to actual wholesale market prices. By contrast, 50% of energy consumption in England and Wales goes through meters that record consumption by the half-hour, and a significant proportion of these consumers were on “pool-related” tariffs. Those tariffs passed through the actual spot price of power, giving consumers real incentives to reduce demand when prices are high.⁷

The Federal Energy Regulatory Commission imposed a “soft cap” on prices in the PX. Any generator bidding more than \$250/MWh had to be prepared to justify its bid, and while those generators would receive their own bid, the market clearing price for generators bidding less would be capped at \$250/MWh. The PX was unable to re-

⁷ The Pool did not have an adequate mechanism to ensure that these demand reductions fed back to the equilibrium price, although there was a rudimentary scheme for “demand side bidding” in the daily auction, which sometimes held prices down at peak times.

design its software to implement the new rules, and declared itself bankrupt, unable to run its market, in January 2001. Such caps were not always effective in keeping prices down, in any case – some generators reduced the amount of power they were selling through the formal markets, forcing the ISO to buy emergency supplies at uncapped prices. The California Department of Water Resources started to buy electricity on behalf of the utilities through bilateral contracts, paying fearfully high prices. At the time of writing, large-scale power shortages during the summer of 2001 seem inevitable.

What went wrong?

California's reform was derailed by an unexpected rise in gas prices and the predictable narrowing of the gap between capacity and demand, which gave (also predictable) opportunities for generators to exploit their market power. Once the system had begun to break down, the complexity of the markets' organisation made things worse, since many generators were able to switch between markets to obtain the best prices in the face of different regulatory limits. The markets' complexity was not the primary cause of the disaster, however. Nor was the dominance of the PX's uniform price auction, despite the arguments used against the Pool in England and Wales – the State is still paying extremely high prices for its bilateral purchases.

The key problem was not the dominance of a uniform price auction, but the dominance of a spot market with no hedging. The apparent hedge provided by the CTC reduced the risk that was taken here, but it did not eliminate it. The disaster could probably have been averted if the utilities had signed appropriate hedging contracts. Assuming that the PX would have continued as the main spot market, these would have needed to be contracts for differences, as in England and Wales. Alternatively, physical bilateral contracts could have been combined with a spot market for the remaining trades. These contracts could have insulated the utilities from most of the impact of the rising spot prices. Furthermore, contracts reduce generators' incentives to abuse their market power (Green, 1999; Newbery, 1998), helping to keep spot prices down. If the generators had hedged their output prices but not their fuel costs, of course, the utilities' problems would simply have been transferred upstream when gas prices rose. In the UK, however, the generators who hedged their output prices also hedged most of their fuel purchases. Several electricity contracts indexed the electricity price to the fuel price, while at least one plant had a

fuel contract indexed to the electricity price, implying that its gas supplier was taking some of the risk.

Such a package of contracts would have insulated most of the market from the gas price rises in the short term. Since CfDs are for fixed volumes, however, the price signals would remain relevant at the margin, giving incentives for conservation, could they be passed on to consumers. Would the long-term impact have been any different, however, or would contracts merely have postponed the reckoning?

To the extent that spot prices are high because of a fundamental mis-match between demand and capacity, and an increase in the competitive price of gas, contracts would (and should) only slow the adjustment of average prices to the level in spot trades. To the extent that prices have risen because capacity has been withdrawn because of the fear of default, contracts would have produced lower prices, if they had ensured that the utilities could pay their bills. Contracts can also have an enduring impact on market power. While many discussions of contracts as a market power mitigation tool focus on vesting contracts imposed on generators at the time of restructuring, the experience in the UK shows that generators voluntarily sold additional, annual, contracts to hedge most of their sales, even though this reduced their spot market power. While the British generators' incentive to abuse their market power was probably reduced by an effective regulator, those who argue that only involuntary contracting can have an impact on market power are probably overly pessimistic. Similarly, long-term contracts in the gas market might have reduced the differential between California prices and those at the Henry Hub.

Conclusions

California's electricity reform has been a disaster that will affect the future path of electricity restructuring. The aspect of the reform that made the disaster possible was not the fact that most power was traded through the PX's auction, but the fact that practically none of the trades were hedged. Contracts for differences allow generators and retailers to hedge their price risks, while still using a central spot market as the best way of obtaining an efficient dispatch. This model can also accommodate a reasonable proportion of physical bilateral trading, as long as enough generators are willing to adjust their plans in the light of signals (and payments) from the central spot market. A market that decided that California's problems stemmed from the use

of a central auction, and decided to rely entirely on bilateral trading, would be learning the wrong lessons from the disaster.

In England and Wales, the Pool that inspired so many restructuring efforts around the world has been abolished. As a pioneer, it had a design that others improved on elsewhere. If its governance had been better, it might have been able to incorporate some of those improvements itself. Instead, it was unable to evolve, and generators with market power were able to abuse its rules. Those generators could well have been able to raise prices, given any set of market rules, such was their dominance of the industry. That dominance declined over time, however, and after ten years, the industry was probably workably competitive. Prices declined as competition grew. Prices under NETA will probably be lower than they were under the Pool. It would be a mistake to link this to the abolition of the Pool, rather than the growth in competition.

The true lessons from California, and from England and Wales, are that centralised spot markets can work, but that they need to be combined with suitable price hedges, and regulators must be vigilant, and take action, against market power.

Bibliography

- Bathurst and Strbac (2001) *The value of intermittent renewable sources in the first week of NETA*, Tyndall Briefing Note No 2, Tyndall Centre for Climate Change Research
- Borenstein, S, J.Bushnell, C. Knittel and C Wolfram (2000) *Price Convergence in California's Deregulated Wholesale Electricity Markets*, mimeo, University of California Energy Institute
- Bower, J. and D.W.Bunn (2000) "Model-based comparisons of Pool and Bilateral Markets for Electricity" *The Energy Journal*, vol 21 no 3 pp 1-29
- Crew, A.M. and P.R.Kleindorfer, (1976) 'Peak Load Pricing with a Diverse Technology', *The Bell Journal of Economics*, Vol.7, No.1, Spring, pp.207-31
- Department of Energy, (1988) *Privatising Electricity*, Cm 322, London, HMSO
- Faruqui, A., H-P Chao, V.Niemeyer, J.Platt and K.Stahlkopf (2001) *The California Syndrome*, mimeo, Electric Power Research Institute
- Green, R.J. (1999) "The Electricity Contract Market in England and Wales" *Journal of Industrial Economics*, Vol XLVII, No 1, pp.107-124

- Green R.J., and D.M.Newbery (1992) 'Competition in the British Electricity Spot Market' *Journal of Political Economy* 100 (5), October, pp 929-53.
- Hildebrandt, E. (2001) *Further Analyses of the Exercise and Cost Impacts of Market Power in California's Wholesale Energy Market*, mimeo, California Independent System Operator
- Hogan, W.W. (2001) *Electricity Market Restructuring: Reforms of Reforms* mimeo, Harvard University
- Newbery, D.M. (1998) "Competition, contracts and entry in the electricity spot market", *RAND Journal of Economics*, vol 29 no 4 pp 726-749
- Newbery, D.M. (1999) *Evidence on Discriminatory Auctions*, mimeo, Department of Applied Economics, University of Cambridge
- Puller, S. L. (2001) Pricing and Firm Conduct in California's Deregulated Electricity Market Power Working Paper PWP-080, University of California Energy Institute
- Sheffrin, A. (2001) *California Power Crisis: Viewpoint of the System Operator*, presentation to the World Bank Group Infrastructure Forum, May 9 2001, California ISO
- Stoft, S. (1997) "What Should a Power Marketer Want?" *The Electricity Journal*, June
- Wolfram, C.D. (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
- Wolfram, C.D. (1999) "Measuring Duopoly Power in the British Electricity Spot Market", *American Economic Review*, vol 89 no 4, pp 805-826