CEEPR Center for Energy and Environmental Policy Research

Did English Generators Play Cournot? Capacity withholding in the Electricity Pool.

by
Richard Green

04-010 WP March 2004

A Joint Center of the Department of Economics, Laboratory for Energy and the Environment, and Sloan School of Management

Did English Generators Play Cournot?

Capacity withholding in the Electricity Pool.

Richard Green[*]

University of Hull Business School
University of Hull
Hull HU6 7RX

Tel: 01482 465720
Fax: 01482 466216
Email: R.J.Green@hull.ac.uk

Presently visiting MIT Center for Energy and Environmental Policy Research
E40-407, MIT CEEPR
1 Amherst Street
Cambridge MA, 02139-2347

March 2004

Revised: 6 April, 2004

Abstract
Electricity generators can raise the price of power by withholding their plant from the market. We discuss two ways in which this could have affected prices in the England and Wales Pool. Withholding low-cost capacity which should be generating will raise energy prices but make the pattern of generation less efficient. This pattern improved significantly after privatisation. Withholding capacity that was not expected to generate would raise the Capacity Payments based on spare capacity. On a multi-year basis, these did not usually exceed "competitive" levels, the cost of keeping stations open. The evidence for large-scale capacity withholding is weak.

Keywords: Electricity prices, Cournot competition, capacity withholding
JEL: L94

[*] Support from the Economic and Social Research Council, under project R000 23 6828, *Developing Competition in the British Energy Markets*, and from the Leverhulme Trust, through the award of a Philip Leverhulme Prize, is gratefully acknowledged. I would like to thank the MIT Center for Energy and Environmental Policy for its hospitality, Ijaz Rasool for help with the data, and Joanne Evans and Paul Joskow for helpful comments.

1. Introduction

Few people would dispute that when an electricity industry is liberalised, market power can be a problem. Incumbent generators frequently hold a high proportion of capacity, imports are restricted by limits on the transmission lines to adjacent markets, demand is inelastic, and the product cannot be stored. The term “withholding” is often used to describe the way in which generators could exploit market power. “Economic withholding” implies that a plant would not offer its output as soon as the market price was high enough to cover its costs of doing so, but would wait until the price had risen above its costs. “Physical withholding” implies that the plant’s output is not made available to the market at any price. In both cases, the plants that are withheld from the market will generate less, and are likely to make less money, but the strategy can increase the company’s profits by raising the price received by its other units.

The strategy of withholding units from the market is a close fit to the economists’ model of Cournot competition, which is the most popular model of market power in electricity. In a Cournot model, firms decide how much output to sell in each time period, and prices are set by the intersection of the demand curve with this quantity. Cournot models can make use of the rich cost data that is usually available, and generally produce a unique equilibrium. These attractive features might be offset, however, by the fact that most electricity markets are built around auctions in which generators offer prices for tranches of output, and the fixed output assumption of the Cournot model seems empirically inappropriate.

It might be possible to rescue the Cournot approach by pointing out that generators need not offer all of their capacity to the market in every period. Kreps and Scheinkman (1983) have shown that when generators choose capacities and then compete in prices, the process can end up with prices at the Cournot level. If generators can be sure that they will collectively offer no more capacity than the market will need, then they can drive up prices without fearing that their rivals will have the spare capacity to expand output and steal their market share.

This paper asks whether British generators have been keeping capacity away from the market in order to raise prices above competitive levels. It examines two possible strategies. One is to reduce the overall level of capacity available, keeping high-cost plants from the market, which will raise the component of the British spot price designed to reward capacity when it is scarce. This strategy was described by the regulator (Offer, 1991) and studied by Newbery (1995). The second strategy is to reduce the overall level of capacity by keeping low-cost plants from the market, ensuring that a higher-cost station sets the energy component of the spot price, at a higher level than if all the capacity was available. This strategy is discussed by Wolak and Patrick (1997), who support their argument with data implying that capacity was being withheld at stations owned by the dominant British generators, which had lower availability than comparable US plants.

This paper argues that availability figures do not provide conclusive evidence of capacity withholding to raise prices. Capacity could be held back from the market at times when there was no likelihood of it being called on to generate, with no impact on prices. The stations would not be commercially available (the definition used in the British electricity market) even though, technically, they were not subject to maintenance and could be made available (a more traditional definition, and that used in the US). To look for evidence of capacity withholding related to market power, we must examine prices and generation patterns.

The next section of the paper outlines the British electricity market. The third section considers power markets and approaches to modelling them. Section 4 discusses "simple" capacity withholding, aimed to drive up the capacity payment. Section 5 discusses "sophisticated" withholding aimed at the energy price. A final section offers brief conclusions.

2. The British Electricity Market

The electricity industry in England and Wales was restructured in March 1990. The integrated Central Electricity Generating Board was divided into three generating companies and the National Grid Company (NGC), responsible for transmission. NGC also operated the daily spot market, the Pool, in which practically all electricity had to be traded. Every day, generators submitted price bids for each of their available units, and NGC calculated the least-cost operating schedule for the following day. The System Marginal Price (SMP) for each half-hour was based on the bid of the most expensive unit in normal operation during that half-hour. A second component of the spot price, the capacity payment, was based on the balance between available capacity and expected demand. NGC calculated the loss of load probability (with some smoothing) and multiplied it by the value of lost load (set at £2/kWh in 1990, and increased annually by the rate of inflation), less SMP, to give the capacity payment. SMP and the capacity payment gave the Pool Purchase Price, paid to all generators for their scheduled output. Generators that were available, but were not scheduled, also received a capacity payment, and the cost of this, and various other payments related to system support, were recovered through a charge known as Uplift.[1] The Pool Purchase Price plus Uplift gave the Pool Selling Price, paid by all buyers.

The main buyers in the Pool were the twelve Regional Electricity Companies (RECs), responsible for distribution, and with an 8-year franchise over supply (retailing) to the industry's smaller consumers. The largest consumers, however, could choose their supplier from the time of the restructuring, and the majority soon did so. Some bought from other RECs, and a few of the largest bought on their own behalf in the Pool, but the most active suppliers were the two major generators, National Power and PowerGen. These companies

[1] In 1997, some of the charges recovered through Uplift were transferred to a new charge, the Transmission Services Use of System charge. This is not formally part of the Pool Selling Price, and should be added to it in any comparison of selling prices over time, but the distinction is not relevant to this paper.

had inherited all the industry's conventional capacity (29 GW and 20 GW respectively), while the nuclear stations were given to Nuclear Electric. They had proved impossible to privatise at the start of the restructuring, although the more modern stations were eventually privatised in 1996.

In the attempt to privatise the nuclear stations, however, the government had only created two large conventional generating companies, and it soon became clear that they had substantial market power. The regulator[2] issued a series of reports on their behaviour, culminating with a series of Undertakings announced in 1994. The companies agreed to sell 6 GW of capacity (about 15% of their total at the time), and to keep Pool prices below specified levels in 1994/5 and 1995/6. All the capacity was sold to the same company (Eastern, one of the RECs), and market power soon resurfaced as an issue. National Power and PowerGen agreed to sell more capacity in 1998, in return for permission for vertical mergers,[3] which had been denied them in 1996. The companies have since sold more of their capacity, on a voluntary basis. Nevertheless, the regulator tried to impose a "good behaviour condition" in the regulatory licences[4] of the eight largest generators. Two of the generators resisted, and after the Competition Commission determined that the clause might be expected to act against the public interest, it was removed from the licences of those generators that had accepted it. Furthermore, the Pool was abolished and replaced by New Electricity Trading Arrangements in March 2001, which the regulator and the government believed would be less susceptible to market power.[5]

Divestitures by the incumbents were not the only way in which they lost market share over the 1990s. Most of the RECs, and many later entrants, invested in new Combined Cycle Gas Turbine stations, which have now taken nearly 20% of the market. The earlier stations were all supported by long-term contracts, which hedged the Pool price. National Power and PowerGen also sold much of their output under contract, which reduces their short-term incentive to raise the spot price. High or volatile spot prices will increase the price at which they can sell contracts in future, however, and so being heavily contracted does not eliminate the companies' incentive to exercise market power in the spot market.

National Power and PowerGen also invested heavily in new CCGT plant. In total, 18.5 GW of new CCGT capacity and 1.2 GW of nuclear capacity were added between March 1990 and March 2000. Since the system had spare capacity in 1990, and peak demand has

[2] The Director General of Electricity Supply was responsible for regulating the industry. His office was originally known as Offer (Office of Electricity Regulation) but combined with the gas regulator's office in 1999 to form Ofgem (Office of Gas and Electricity Markets), with the same person appointed to regulate both gas and electricity. In 2001, he became the first Chairman of the Gas and Electricity Markets Authority, which has replaced the individual regulators.

[3] PowerGen bought East Midlands Electricity, while National Power bought the supply business, but not the distribution network, belonging to Midlands. All of the ex-REC supply businesses are now part of vertically integrated groups.

[4] Much of the detail of the industry's regulation is governed through licences. All companies in the industry are required to hold an appropriate licence for each of their activities. These can normally only be changed with the company's agreement, or if the regulator gains the support of the Competition Commission.

[5] Not all independent commentators agree with this view (Hogan, 2000; Newbery, 1998; Wolfram, 1999). The National Audit Office, however, concluded that NETA had "facilitated" lower wholesale prices (NAO, 2003, para 6).

grown only from 48.3 GW to 51 GW, the incumbents have obviously been able to close a large amount of capacity without risking shortages. Nevertheless, the regulator has been concerned that the companies should not close an excessive amount of capacity, and the companies agreed in 1992 to allow an independent assessor to examine their closure decisions, should the regulator request this. This agreement was embodied in a new licence condition that also required the companies to give the regulator an annual statement of their plans for making each unit available to the market, a half-yearly update, and an end-of-year reconciliation, noting and explaining any significant changes. The licence condition does not lay down what the availability policies should be, but the regulator has enough information to decide whether the companies are following their stated policies, and assess their impact on the market.

The companies' price bids are placed in the public domain a few days after they are made, and the regulator has often studied them in reports on competition. In 1992, the regulator agreed that stations that were used to meet constraints were justified in bidding at a (high) level that nevertheless did not recover more than their costs over the year as a whole, if they were only occasionally called on to generate at that level (Offer, 1992). In the mid-1990s, the companies agreed to bid in such a way as to produce prices below a given level (in normal circumstances). Apart from this, the regulator did not attempt to place other restraints on the companies' bids until the introduction of the "good behaviour" licence condition. He commented on the level of prices, but tended to place less emphasis on the means used to achieve that level.

3. Power Markets and Models

Economists have been interested in the ways electricity companies can raise prices above their costs. Most electricity markets are based on some form of auction. In real time, the system operator must match demand and supply, and the price bids from an auction allow this to be done in an economically efficient manner. In the longer term, bilateral contracts are common, but most markets have some forum in which buyers and sellers can sell power on a short-term basis. In the Nordic countries, this forum is NordPool, while national regulation markets are responsible for minute-by-minute balancing. In California, the Power Exchange was one of several Scheduling Co-ordinators that match expected generation and demand in advance, while the Independent System Operator dealt with short-term adjustments and reserve. In England and Wales, the Pool was responsible for both day-ahead trading and real-time adjustment, but the New Electricity Trading Arrangements that replaced it separate the two. In most cases, however, the day-ahead market (or equivalent) provides a marker price for the industry, and one set "as if" in a first-price auction.

In a perfectly competitive market, all generators would bid their capacity into this market whenever they thought that there was a reasonable chance that it might be needed, and would do so at a price equal to their marginal cost. If no generator owns more than one unit, setting a higher price runs the risk that the unit would not be called, without affecting the price that the unit would receive, which would be set by another generator almost all of the time. A generator that does not offer its unit into the market cannot earn anything from it. Only if there is very little chance that the unit would be required should it be withheld, in order to save the costs of making it available needlessly. While most of the unit's costs are either fixed on a long-term basis (capital costs are sunk, while a basic level of staffing and age-related maintenance are needed for as long as the station is open) or directly related to output (fuel costs and maintenance based on the station's operating regime), some staff costs would be incurred if the station was made available but not required to generate.

A generator that owns many units and wishes to increase the market price has two main alternatives. The first is to bid all of its available capacity into the market, but at prices above its marginal costs. The second is to withhold some of its capacity from the market. If the capacity should have been infra-marginal (i.e. it had low costs), this would mean that a more expensive unit is now at the margin, raising the price. In England and Wales, withdrawing extra-marginal capacity increases the capacity payment, but withdrawing extra-marginal capacity might have little impact in other markets, unless other generators interpret the move as a signal to adopt higher-priced strategies.

The first papers on market power in England and Wales looked at bid-based price increases. von der Fehr and Harbord, 1993 used an auction model to suggest that generators would follow a mixed strategy with expected bids above marginal costs, although their approach did not allow for detailed modelling of the generators' capacities and cost structures. They were able to show generators' bid curves that were consistent with their approach. Green and Newbery (1992) used the supply function approach, which can cope with more detailed information on capacities and costs, but yields multiple equilibria. The higher-priced equilibria involve supply functions that diverge from marginal cost at an increasing rate as marginal cost itself rises. Green (1994) compares generators' bids with their marginal costs, and shows that they appear to be consistent with the supply function approach, but does not use any formal statistics. Wolfram (1998) does use a statistical approach, and places her work in the auction tradition, but her findings are consistent with supply function models. Her main results are that the larger generator in the British duopoly had higher bids, relative to its marginal costs, and that all stations tended to submit higher bids when their owner had more lower-cost capacity available. Raising the price bid by one station will raise the price received by all the stations that are infra-marginal when it runs. The more of these that belong to the station's owner, the greater the incentive to bid up its price.

Most papers on other markets have followed the approach of Cournot competition, in which generators decide how much of their capacity to make available to the market in each period, and prices are set to bring demand into balance with this supply. Cournot models have been used to analyse markets in Sweden (Andersson and Bergman, 1995), the Nordic countries (Amundsen *et al,* 1999) and California (Borenstein and Bushnell, 1999), amongst others. These papers attempt to predict energy prices, and might be construed as in the "price-increase" tradition, but the strategic instrument is in fact the capacity offered to the market. Since each generator assumes that its rivals will not respond to any reduction in its own capacity, the equilibria tend to involve greater reductions in output, and higher prices, than supply function models in which generators expect their rivals' output to increase when they try to exercise market power.

Three papers clearly look at the generators' decision on whether to offer their capacity to the market. Newbery (1995) showed that the capacity payment system gives British generators an additional incentive to keep some of their capacity from the market, and quantified the size of this incentive. Anecdotal evidence suggests that generators have sometimes delayed the end of an outage if this would raise the level of capacity payments. Crampes and Creti (2001) show how generators would have an incentive to reduce the amount of capacity that they offered into the market if this would then turn the market from one with excess capacity and low bids into a capacity-constrained market with high prices in equilibrium. Wolak and Patrick (1997) introduced the idea of withholding infra-marginal capacity in order to raise prices without raising their bids above marginal cost. They show that the generators declared roughly the same amount of capacity declared available, relative to demand, in peak (November – February) and off-peak months (the remainder) between April 1991 and March 1995. Since demand was much lower in off-peak months, this meant that a large amount of capacity was not made available to the market at those times. The mean availability in peak months was 51 GW, while the mean availability in off-peak months was 43.8 GW. Furthermore, National Power and PowerGen had lower availability figures, for each type of plant that they owned, than comparable plants in North America, when new entrants to the market were achieving higher availabilities than comparable American stations.

Wolak and Patrick suggest that the generators would prefer to withhold capacity from the market, compared to raising their bid prices, because capacity withholding could be disguised as maintenance, while the regulator could easily compare each station's bid price to an estimate of its marginal costs. In practice, however, the regulator has not often referred to explicit comparisons of a station's bid with its costs. A study on constrained-on plant (Offer, 1992) gave revenue and cost figures for a number of stations that had submitted high bids to exploit their geographical position, and a report on gas turbine plant also considered station costs. There have been a number of comparisons of differing bids made by the same station at different times, with the implication that at least some of them could not have been cost-reflective. However, the overall level of prices has rarely been related to the generators' costs – a study on "costs and margins" promised in 1993 never saw the light of day, and the regulator has not referred to it since.

In contrast, however, the major generators were required to provide the regulator with an annual forecast of their capacity availability, on a set-by-set basis, a half-way update, and a reconciliation at the end of each year. The licence condition governing these statements made no attempt to lay down levels of availability, but they provided the regulator with all the evidence that he might need, were he to investigate a generator for anti-competitive behaviour. The regulator has used information on forced outage rates and demand adjusted to normal weather conditions to determine whether a temporary reduction in the plant margin, which caused high capacity payments in January 1995, was "abnormal" (Offer, 1995). The major generators also had to provide the regulator with advance notice of any plans to close capacity, and consider the option of offering it to sale to a third party. The regulator could appoint an independent assessor to decide whether the plans are reasonable, and it is unlikely that an assessor would approve closures when capacity payments are predicted to be high.[6] All in all, the generators may have had less freedom with their capacity declarations than with their price bids.

4. Capacity withdrawal and capacity payments

In some circumstances, keeping capacity away from the market is a perfectly acceptable competitive response. A peaking station is unlikely to be required to run overnight, when demand is low. The station may be technically available, in that it is not undergoing maintenance, but it would not make commercial sense to staff the station at operational levels – the staff would be wasting their time, waiting for a call from the grid operators that is not going to arrive. The station should only be made available in the sense of being staffed and ready to generate during daytime hours when it is likely to be needed. Availability figures in the Pool were based on this definition, and some mid-merit and peaking stations regularly declared themselves "commercially unavailable" at off-peak times.[7] If this is done without affecting prices, then it can be counted as a proper competitive response. This would be the case when there the chance that the station would actually be required to run is negligible.[8] If prices rise in response, however, then there is a possibility of market manipulation. The question is whether the price before the station was withdrawn was high enough to cover the costs of making it available over the relevant period. If the price was above the avoidable cost of keeping the station open, the "competitive" option would be to make it available, while if the price was below this level, a single station's owner would lose money by keeping the station in the market, and should withdraw it.

This gives us a test. If generators had perfect foresight and acted competitively, capacity payments should equal the avoidable cost of a peaking station. Higher capacity payments would imply anti-competitive behaviour. It is normal to compare costs and revenues over the course of a year, and generators do recover most of their costs at the time

[6] It is not strictly necessary to use predictions, since companies trade annual contracts to hedge capacity payments, and so actual market prices can be used to compare a station's costs with its potential revenues.

[7] Officially, stations in the US would still be considered available in these circumstances (Personal Communication: M. Curley of NERC, 29 August 2000), and so comparisons between US and UK availability data can be misleading – US stations should only be recorded as unavailable when shut down for, or awaiting, maintenance. In practice, Joskow and Schmalensee (1987) show that units with a low ratio of actual to available output tended to have lower availability figures. They ascribe this to the strain of frequently turning the station on and off, leading to increased maintenance, but the pattern is consistent with stations being declared unavailable simply because they do not expect to run.

[8] The calculation behind capacity payments actually uses every station's highest availability over the previous eight days, and so declaring a station unavailable overnight and at weekends should have no impact on the capacity payment.

of the system peak. That does not imply that capacity payments should be zero at all other times, however. Even in off-peak months, some plant must be made available, but not expected to generate, at the daily peak, and will incur an avoidable cost in doing this. Capacity payments must be expected to exceed zero, if only slightly, to make it worthwhile to declare this reserve capacity available.

At the annual level, the Monopolies and Mergers Commission (1996) has reported that the fixed costs of an open cycle gas turbine plant were £6/kW per year. This might be a lower bound on the expected capacity payment that a competitive generator would require in order to keep its station open. An independent assessor's report on plant closures (Offer, 1998) gave a fixed cost for a unit of £22.50/kW per year, with an extra £5/kW per year for station costs. However, the unit cost (for a 200 MW unit) included £8 million for an overhaul, spread over 4 years of operation, or £10/kW/year. For a station close to the end of its life, but capable of continued running without such an overhaul, the avoidable cost would be £17.50/kW per year. In 1992, Offer's report on constrained-on plant included output and cost figures for a number of stations – with appropriate assumptions on fuel prices, thermal efficiencies, and variable operations and maintenance costs, the implied fixed costs were around £25/kW per year. That might form an upper bound for the expected capacity payment that should trigger plant closures.

These bounds should be compared with the *expected* capacity payments at the time closure decisions would be made, which are unobservable.[9] The *actual* capacity payments will be higher than the expected level if the weather is cold (England and Wales have a winter peak) or plant failures are above average, and lower if a warm winter coincides with few plant failures. Over a number of years, however, the average *actual* capacity payment might form a good proxy for the expected level. Table 1 shows the capacity payment in each year from 1990/1 onwards. Figures are given for a complete financial year, and broken down into monthly averages for the winter, the summer, and the "shoulder" months (March, April, September and October). Competitive behaviour would imply that the annual average prices were close to the range of £6/kW to £25/kW discussed above, and that these should be concentrated in the winter months. Capacity payments in the summer and the shoulder months should be low (but not zero) if the generators are acting competitively.

The annual capacity payments are volatile, and there are a number of individual years, particularly later in the period, when they exceed our range for the competitive level. Their cumulative average over the life of the Pool, however, does remain below the upper bound of competitive values. In most years, capacity payments outside the winter months are also low enough to be compatible with a competitive outcome. In the following paragraphs, we comment on some of the payments that appear to have exceeded competitive levels.

[9] Some contracts for differences were written to hedge capacity payments alone, and would form a guide to participants' expectations, but prices for these are not in the public domain.

Table 1: Capacity payments in England and Wales, 1999-2000 prices

	Annual Capacity Payment, £/kW	Average, 1990/1 to date, £/kW	Monthly capacity payment, £/kW		
			Winter (Nov-Feb)	Shoulder	Summer (May-Aug)
1990/1	0.55	0.55	0.10	0.00	0.02
1991/2	13.73	7.14	1.59	1.24	0.26
1992/3	1.76	5.35	0.23	0.16	0.01
1993/4	2.89	4.74	0.51	0.09	0.05
1994/5	31.49	10.09	6.53	0.45	0.12
1995/6	42.59	15.50	8.21	0.76	0.64
1996/7	30.32	17.62	3.45	2.19	1.75
1997/8	7.59	16.37	1.45	0.20	0.07
1998/9	8.82	15.53	0.94	0.26	0.96
1999/00	24.10	16.39	2.18	1.90	1.94
2000/01	38.08	18.36	1.11	5.94	2.47

The table suggests that capacity withholding to increase capacity payments was not a significant feature of the early years of the Pool. The one, quite well known, exception concerns behaviour by PowerGen in the summer and early autumn of 1991. Each day, the company declared that some of its stations would not be available on the following day, raising the loss of load probability, and setting capacity payments at a higher level than if the stations had been available. The stations were then re-declared to be available, and collected the capacity payment, which had been fixed and could not be re-calculated. The company soon stopped the practice, in response to criticism by the regulator, who imposed the availability reporting requirements described above. Furthermore, the Pool Rules were changed to use every station's highest availability (declared or actual) over an eight-day period when calculating the loss of load probability. This made PowerGen's original tactics effectively impossible – a station would have had to be unavailable for a week before delaying its return to service would affect capacity payments. If a plant has been on outage for eight days or more, however, a company could gain from delaying its return, and there is plenty of anecdotal evidence that this sometimes happened. In the first few years of the Pool, however, any extended outages of this kind failed to raise capacity payments above "competitive" levels.

In the period since 1994/5, however, capacity payments were much higher, averaging more than £20/kW. The high payments in 1994/5 were the subject of an Offer investigation, which found them to be the result of an unlikely combination of plant outages, as discussed above. In 1995/6, payments were concentrated in the winter, when demand was high, and raised by exports to France (where output was reduced by strikes). In other words, they could once again be consistent with competitive availability choices.

The following year, however, capacity payments were at similar levels (on average) throughout the year. That might suggest that generators were making enough plant available at the winter peak, but withdrawing too much of it at other times of the year, raising "off peak" capacity payments. The data for 1999/00 would be consistent with this, but Ofgem (1999) provides an alternative explanation. Capacity payments depended not only on the amount of plant on the system, relative to demand, but also on its reliability. That was measured in a way which tended to make new capacity appear less reliable, and so capacity payments would be higher, the greater the proportion of new plant on the system. The plant margin was actually slightly higher in July 1999 than in July 1998, but capacity payments were five times as high. Part of this was due to an increase in the proportion of new plant: half of the increase was due to the way in which a single (lengthy) outage at a relatively new nuclear station affected its calculated reliability (Ofgem, 1999, pp 25-6). This begs the question of why the payments were calculated in this way, but also suggests that the high payments in 1999/2000 were "innocent" rather than the result of strategic capacity withdrawal.

In the summer of 2000, however, Ofgem took action when Edison Mission Energy withdrew 480 MW of capacity from the market. Edison was one of the companies that had accepted Ofgem's short-lived market abuse licence condition, and the regulator started an investigation into the company's decision. Ofgem (2000) reported that "on the basis of analysing the avoidable costs of the capacity withdrawn by the company … against spot and forward prices it appears that Edison may be in breach of the market abuse condition." Edison then agreed to return the plant to service, and Ofgem took no further action. This is the first time that the regulator has taken action against a generator for withdrawing capacity since PowerGen's blatant manipulation of 1991. The highest prices of 2000/1 came two months later, in September 2000, and Ofgem's investigation once again concluded that the shortcomings of the capacity payment's calculation, rather than deliberate action on the part of generators, was to blame.

While the capacity payments mechanism had serious flaws in its calculations, and generators did engage in "tactical" withdrawals of capacity, it is probably reasonable to conclude that for most of the 1990s, generators did not "strategically" withhold enough capacity from the market to force expected capacity payments above competitive levels.

While the level and pattern of capacity payments is generally consistent with "competitive" behaviour, the way in which the dominant generators withdrew capacity from the market does deserve a comment. In general, it is likely to be more expensive to retain two stations, closing part of the capacity at each, than to close a single station.[10] Furthermore, if a station is to be closed, making the closure permanent at once will allow the company to dispose of the site. National Power and PowerGen, however, have frequently closed stations in stages, and mothballed plant before announcing its permanent closure. Mothballing plant might have been part of an entry-deterring strategy, with the implied threat

[10] Keeping part of a station open can be justified if it is required to provide system support, perhaps by easing a transmission constraint, and can earn enough from doing so to offset the additional costs involved.

that the market could be flooded with spare capacity, should there be excessive entry. In practice, however, such threats were never carried out, and appear to have been very unsuccessful in deterring entry. Closing a station in stages will sometimes be a tactical response to the regulator's requirement that the generators should consider selling stations that they plan to close. It is practically impossible to buy part of a station, and so partial closure allows the generators to avoid selling plant that they believe is surplus to requirements. If the closure was justified on "competitive" grounds, then a stand-alone operator should not be able to make a profit from the station, but the generators may have feared that a rival would attempt to force the incumbents to withdraw other plants, accelerating their loss of market share. If the closure was not justified on competitive grounds, but was part of an anti-competitive capacity withdrawal, then the plants would be more attractive to rival generators, who could make a profit without further closures from the incumbents. The evidence above, however, and the regulator's investigations into proposed closures, suggest that the level of closures was generally justified on competitive grounds, even if their manner was in some respects dubious.

5. Capacity withdrawal to raise energy prices

The previous section has considered withdrawing high-cost capacity in order to boost capacity payments. A second kind of manipulation has been suggested by Wolak and Patrick (1997). An incumbent might withhold low-cost capacity, ensuring that higher-cost stations would set the system marginal price, thus raising it, without requiring any station to bid significantly above its costs. This strategy might be undertaken at times of low demand, and the stations made available when demand was higher, so that the system marginal price could be increased without increasing capacity payments. Wolak and Patrick suggest that it would be hard for a regulator to detect this kind of capacity withholding, since it could be disguised as maintenance.

Withholding a station in this way, however, will reduce its load factor. The stations that run instead will have higher load factors. This section of the paper uses information on load factors before and after restructuring to construct a test for this kind of capacity withholding. The CEGB regularly published the load factors of its stations, while the MMC published information on the load factors of National Power and PowerGen's coal-fired stations in its 1996 reports into their merger proposals. NGC provided load-duration curves, showing the number of hours in each year for which demand above a given level.

If all stations were available throughout the year, and there were no transmission constraints or other operating restrictions, then it would be possible to match the generators' output exactly to the shape of the load-duration curve. The most efficient stations would have a load factor of 100%, while less efficient stations would run for as long as there was demand for their output. In practice, however, it is extremely unlikely that a station would be available for the whole year, and so its load factor will be less than 100%. The "missing" output must be made up by another station, and will be outside the load-duration curve.

Operating restrictions will further reduce the output of the more efficient stations, and increase the output of those with higher costs. This will give us the pattern shown in figure 1.

The line is the load-duration curve, showing the demand for each hour in the year, ranked in decreasing order of demand. The highest demand is placed at hour 0, and the lowest at hour 8760 (8784 in a leap year, and 8640 in 2000/1, when the Pool was abolished on March 26). The rectangles show each station's output, ranked in merit order – the stations with the lowest operating costs, which ought to run the most, are at the bottom. The dark grey portions show the output that was "within" the load-duration curve, and would have been produced if generation had perfectly matched demand. The light grey portions are the output that was only required because some other stations were generating less than the system could have accepted from them – the dark grey areas do not fill the area under the load-duration curve. The total area of the light grey portions should be the same as the area of the blank spaces under the load-duration curve.

The CEGB's last published output data were for 1985/6, 1986/7 and 1987/8. Output patterns in 1984/5 were significantly disrupted by the miners' strike of that year, and earlier years might be increasingly unrepresentative of the system as it was restructured. We assume that the CEGB aimed to run its stations in the "true" merit order, and derive this from the mean load factor of each station across the three years – the coal-fired stations with the highest load factors during this period are assumed to be highest in the merit order.

Throughout the 1990s, the price of coal in the UK was sufficiently high, relative to the price of gas, that CCGT stations were always above coal-fired stations in the merit order, given their greater thermal efficiency. To obtain the merit order for the 1990s, we simply inserted the CCGT stations above the best coal-fired stations, with the newest stations at the top. This assumption would break down for 2000/1, however. In the second half of that year, the price of gas delivered to power stations was 50% higher, relative to the delivered price of coal, than over the previous eighteen months. This was the only time in the Pool's history that a CCGT would have had significantly higher fuel costs than a coal-fired station. Table 2 shows how this affected the type of plant which was setting SMP – in 1999/2000, CCGTs were only on the margin for one percent of the time. In the first half of the following year, this had risen to 4.8 percent of the time, and 13.5 percent of the time in the second half of the year. Given the change in fuel prices, this was clearly due to a change in the cost-based merit order, and not to a change in generators' strategies. Since the merit order changed in the middle of the year, it is not possible to compare 2000/1 with the years of a stable merit order. Results for each year of the Pool are reported, but the comparisons are based on the ten years from 1990/1 onwards.

Table 2: Plants setting System Marginal Price, by fuel type (number of half-hours)

	1999/2000		April – Sept 2000		Oct 2000 – Mar 2001	
Coal	14,260	81.2%	6,861	78.1%	5,416	63.7%
Interconnector	2,336	13.3%	843	9.6%	1,592	18.7%
PSB	757	4.3%	652	7.4%	339	4.0%
CCGT	170	1.0%	422	4.8%	1,147	13.5%
OCGT	43	0.2%	0	0.0%	0	0.0%
CCGT/CHP	2	0.0%	5	0.1%	0	0.0%
No SMP was set	0	0.0%	1	0.0%	2	0.0%
Totals	17,568	100%	8,784	100.0%	8,496	100.0%

Returning to figure 1, we could simply take the ratio of the blank area under the load-duration curve to total generation as an indicator of the "mis-match" between generation and demand, and compare this across years. This would ignore the role of forced outages, however. The CEGB (along with most other system planners) assumed that a proportion of plant would be unavailable at any time, for technical reasons. This means that we should never expect to fill the area under the load-duration curve with generation. We can get a more realistic "ideal" pattern if we shrink the load-duration curve along the horizontal axis, while stretching it along the vertical axis to preserve the total area. Using a 10% forced outage rate (the CEGB norm) the maximum length of running expected would shrink to 90% of the year (7884 hours), while a peak demand of 45 GW would be raised to 50 GW. Figure 2 shows the dotted original load-duration curve, and the revised curve. A greater proportion of the generator's output is under this revised curve, reflecting the fact that the industry needs enough capacity to meet the peak demand plus a margin for unavailable stations.

We will make a second adjustment, to take account of new stations, and the nuclear stations. Once they have been commissioned, most new stations (and all built during our period) would be at the top of the merit order, running nearly continuously. While they are commissioning, however, new stations will only produce intermittently, to a timetable dictated by their testing requirements, rather than by market conditions. The amount of plant being commissioned varied over our period, and it would be unreasonable to count output that was "lost" during commissioning as evidence of capacity withholding. Similarly, while nuclear stations attempted to run at high load factors, they all too often failed to achieve these, for reasons that had nothing to do with National Power and PowerGen (which did not own any nuclear stations). The striped area in figure 2 represents the output lost due to plant commissioning or poor nuclear performance, compared to a target load factor of 85%. Note that this still allows for some "under-performance" compared to the vertical section of the adjusted load-duration curve, which is at a load factor of 90%.

Table 3 shows how these calculations work in practice, for four stations in 1990/91. Hinkley Point B was a nuclear station which achieved a load factor of 76.5%, and output of 7,510 GWh. At the top of the merit order, we assume that a "perfect" performance would have been a load factor of 90%, and an output of 8,830 GWh. The difference between 8,830 GWh and 7,510 GWh is a measure of the lost output due to its failure to match the load-duration curve – some 1,320 GWh. Since Hinkley Point B was a nuclear station, however, we discount shortfalls below a load factor of 85%, and so only 490 GWh (8,830 – 8,340) are treated as truly lost.

Table 3 – illustrative calculations for selected stations, 1990/91

Station	Hinkley Point B	Drax	Cottam	Drakelow C
Capacity	1120	3750	1920	448
Actual Load Factor	76.5	79.1	76.1	25.6
LDC load factor	90	90	84.3	15.1
Output at this LF	8,830	29,565	14,179	593
Output at 85% LF	8,340			
Actual Output	7,510	25,994	12,805	1,006
Lost Output	1,320	3,571	1,374	0
Truly Lost Output	490	3,571	1,374	0

The large coal-fired station at Drax was also high enough up the merit order to have a potential load factor of 90%. Its actual load factor was 79.1%, and the shortfall was 3,571 GWh (29,565 – 25,994), all of which was counted as "truly lost". Another large coal-fired station, Cottam was slightly lower down the merit order, and had a potential load factor, had all stations above it been operating at their maximum load factors, of 84.3%. Instead, its load factor was 76.1%, and so it accounted for 1,374 GWh of lost (and "truly lost") output. Finally, Drakelow C was a low merit station with small, less efficient, units. Its potential load factor would only have been 15.1%, had all the stations above it been operating at their maximum load factors, but their shortfalls actually allowed it a load factor of 25.6%. With an actual output greater than the amount under the load-duration curve, it was not responsible for any lost output.

In general, the adjustment for new plant is made for the first financial year in which a new station generated. Some stations started their commissioning late in the year, however, and did not have time to produce much output before the year end. It would be inappropriate to discount a large amount of lost output on account of a plant that was only just starting the commissioning process, however. Therefore, new stations with load factors of less than 25% were moved down the merit order to be treated as peaking plant, where they would not affect the amount of lost output. The adjustment for output lost in commissioning was made in the following year. Taking the two years together, this will reduce the amount of lost output that is discounted due to the plant's commissioning.

Many new plants had several units, but they were generally commissioned simultaneously, or nearly so, and therefore treated together. The main exception is the Drax coal-fired station, which commissioned its last two units (out of six) in 1985 and 1986. Drax's reported output is not broken down by unit, and any output produced by the new units before they were formally commissioned was not assigned to the station, but reported as a total for all such units, (which included several nuclear stations at this time). Since 2.5 GW (3.1 GW) of capacity had already been commissioned before 1985/6 (1986/7), it would be inappropriate to apply the adjustment for new plant to the entire station, when only one 625 MW unit was being commissioned in each year. We therefore assumed that the commissioning unit achieved a load factor of only 25%, and deduct 3.3 TWh (625 MW × 60% × 8,760 hours) of output from the calculated amount of lost output in 1985/6 and 1986/7. This almost certainly over-estimates the amount of output to be discounted on behalf of these units, improving the performance of the CEGB. Since the post-privatisation industry turns out to be better than the CEGB at fitting its output to the load-duration curve, any assumption that improves the CEGB's performance will narrow the gap between the two. This adjustment also moves Drax to the top of the merit order for coal-fired stations, based on the implied performance of its "established" units.[11] Two post-privatisation stations, South Humber Bank and Seabank, were also commissioned in two stages, and each part is treated as a separate station in this analysis.

This analysis assumes that the shape of the load-duration curve is the only constraint on generators. In practice, some stations have to reduce their output because of constraints on the transmission system. Output from low-cost stations has to be replaced by output from higher-cost stations that are closer to the load centres, in order to avoid overloading the transmission lines from the more distant stations. Transmission constraints rose significantly in the early 1990s, in part because of the amount of new plant being connected to the transmission system and old plant being disconnected. If this meant that high-merit stations had to generate less, it might be expected to worsen the fit between the load-duration curve and actual generation. This would provide an "innocent" explanation for any deterioration in performance in the early 1990s. In practice, we will see that this fit improved after privatisation, despite any impact of constraints.

Our test, therefore, is to measure the blank area under the adjusted load duration curve, as a proportion of total output. The shaded area above the load-duration curve can be denoted "excess output", the striped area under the load-duration curve is the part of this that is due to low output levels from nuclear plant or new entrants, and the blank area under the load-duration curve is "truly excess output". If generators are withholding capacity that

[11] Drax was also the first station to have Flue Gas Desulphurisation equipment fitted, and while this slightly reduced its thermal efficiency, the Environment Agency required National Power to run Drax in preference to other stations without FGD, reinforcing its position at the top of the coal-fired merit order.

should have been in merit, in order to raise the system marginal price, the amount of truly excess output should rise after restructuring. Table 4 gives the figures.

Table 4: Output and the load-duration curve

Year	Total Output (TWh)	Within load-duration curve	Lost new and nuclear output	Truly excess output (TWh)	Truly excess %
1985/6	231,983	200,269	12,801	18,913	0.082
1986/7	237,785	192,715	21,359	23,711	0.100
1987/8	243,912	206,643	13,940	23,330	0.096
1990/1	261,537	224,515	16,730	20,291	0.078
1991/2	263,259	232,226	13,392	17,641	0.067
1992/3	262,655	229,947	14,538	18,171	0.069
1993/4	266,523	229,584	14,049	22,890	0.086
1994/5	268,953	227,025	21,192	20,736	0.077
1995/6	283,779	245,138	23,756	14,885	0.052
1996/7	286,851	249,705	12,214	24,931	0.087
1997/8	288,193	252,083	10,870	25,239	0.088
1998/9	293,368	254,962	10,773	27,633	0.094
1999/00	296,741	252,269	23,723	20,749	0.070
2000/1	306,083	245,925	23,596	36,562	0.119

On this definition, 9.2% of output was "truly excess" in the three years of CEGB operation. In almost every year after the restructuring, the proportion of truly excess output was lower. The one glaring exception is 2000/1, the last year of the Pool, and the way in which the changing fuel prices during the year make the comparison inappropriate is discussed above.

We can test the hypothesis that the mean percentage of truly excess output was the same in the first three years and the next ten years of data, using Excel's built-in t-test. If we assume that the two samples have the same variance, the p-value of a two-tailed test for equal means is 0.075, while if we assume unequal variances, it is 0.079.[12] In either case, we reject the hypothesis of equality at the ten percent level. After privatisation, the generators were significantly better at fitting their output to the load-duration curve. If they had been withholding output in order to raise prices, this would have worsened their fit to the load-duration curve.

[12] The variances are 9.11E-05 and 1.55 E-04, which are hardly equal, but an F-test does not reject the hypothesis of equality at the 10% level.

Does this result depend upon the values chosen for forced outages and the target load factor for new and nuclear plant? Table 5 shows how the p-value for the hypothesis of equal means varies with these parameters. The table always reports the greater of the two values for equal and unequal variances.

Table 5: p-values for testing the hypothesis that the proportion of truly excess output did not change between 1985/6-7/8 and 1990/1-99/00

Target load factor for new and nuclear plant	Availability after forced outages				
	1.00	0.95	0.90	0.85	0.80
1.00	0.015				
0.95	0.023	0.016			
0.90	0.037	0.028	0.044		
0.85	0.056	0.046	0.079	0.174	
0.80	0.067	0.055	0.094	0.153	0.632
0.75	0.059	0.048	0.074	0.097	0.334
0.70	0.049	0.041	0.058	0.074	0.147
0.65	0.040	0.034	0.046	0.058	0.089

If the rate of forced outages is 0.1 or lower, then we continue to get a significant difference in the proportion of truly excess output, at the ten percent level. If the rate of forced outages is greater than this, and the target load factor for new and nuclear plant is close to that expected for established conventional plant, the difference is not significant at this level. None of the tested parameter values implied that the CEGB had a better performance than the privatised industry, however. Furthermore, the qualitative results do not depend upon the treatment of new plant. If we simply look at the proportion of output under the load duration curve, with no adjustment for nuclear outages or new plant (apart from continuing to ignore those with load factors under 25% in their first year), it rises from 0.840 under the CEGB to 0.864 in the first ten years of the Pool. The improvement is significant at the five percent level if we assume equal variances, but not significant (p = 0.252) if we assume that the variances are unequal.[13]

6. Conclusion

We have studied two ways in which generators could play "Cournot" strategies, raising prices by withholding capacity from the market. From the evidence available, neither strategy seems to have been significant in England and Wales. After an initial abuse of the capacity payment, the regulator has required the major generators to provide regular information on their capacity and its availability. It appears that the companies have

[13] The variances are 7.33 E-04 and 1.40 E-04 respectively. An F-test for their equality is not rejected at the 5% level, but is rejected at the 10% level.

generally made capacity available in a competitive manner, declaring plant whenever it is technically available and can expect to earn more than its avoidable costs. Anecdotal evidence suggests that an outage will sometimes be prolonged unnecessarily with the intention of keeping capacity payments high, and there were times near the end of the Pool's life when the regulator found that individual companies had acted in an anti-competitive manner. For much of the period, however, the overall level of payments implies that such practices were not used to raise prices by a significant amount. The industry's overall performance at fitting its output to the load-duration curve improved significantly after privatisation, when systematic capacity withholding would have worsened it.

Cournot models of the electricity industry do have attractive features. They can support detailed cost modelling, and do not usually suffer from the multiple equilibria common with other modelling approaches. Against this, however, they do not give a good representation of the way in which prices are set in electricity markets, and their price predictions have generally been too high.[14] This paper suggests that models in which firms set prices, rather than quantities, are likely to give better results.

[14] A rare exception is Bushnell (2003), which shows that a Cournot model, updating the assumptions of Borenstein and Bushnell (1999), gives a good fit to the prices experienced in California in the summer of 2000. While this is before the final crisis, the market was hardly performing well at this stage.

References

Amundsen, E., L.Bergman and B.Andersson (1999) *Trade and Market Power on the Northern European Market for Electricity*, mimeo, Stockholm School of Economics

Andersson, B., and L.Bergman (1995) "Market Structure and the Price of Electricity: An *Ex Ante* Analysis of the Deregulated Swedish Electricity Market" *The Energy Journal* vol 16 no 2, pp 97-109

Borenstein, S. and J.Bushnell (1999) "An Empirical Analysis of the Potential for Market Power in California's Electricity Market." *Journal of Industrial Economics*, Vol 47, No. 3, pp 285-323

Bushnell, J. (2003) *Looking for Trouble: Competition Policy in the U.S. Electricity Industry* CSEM WP-109, University of California Energy Institute

Crampes, C. and A.Creti (2001) *Price Bids and Capacity Choice in Electricity Markets*, mimeo, Institute d'Economie Industrielle, University of Toulouse

von der Fehr, N-H M., and D. Harbord (1993) "Spot Market Competition in the UK Electricity Industry", *Economic Journal*, vol 103 no 418 pp 531-46

Green, R.J. (1994) "Britain's Unregulated Electricity Pool" in (ed) M.Einhorn *From Regulation to Competition: New Frontiers in Electricity Markets*, Boston, Kluwer

Green, R.J, and D.M.Newbery (1992) "Competition in the British Electricity Spot Market" *Journal of Political Economy* vol 100 no 5, pp 929-53

Hogan, W.W. (2000) *Making Markets in Power* Cantor Lecture to the Royal Society for the Encouragement of Arts, Manufactures and Commerce, London

Joskow, P.J. and R.Schmalensee (1987) "The Performance of Coal-Burning Electric Generating Units in the United States: 1960-1980", *Journal of Applied Econometrics*, vol. 2, no. 2, pp 85-109

Kreps, D. and J.Scheinkman (1983) "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes", *Bell Journal of Economics*, vol 14 pp 326-337

Monopolies and Mergers Commission (1996) *National Power PLC and Southern Electric plc: A report on the proposed merger,* Cm 3230, London, HMSO

National Audit Office (2003) *The New Electricity Trading Arrangements in England and Wales: Report by the Comptroller and Auditor General*, HC 624 of Session 2002-2003, London, The Stationery Office

Newbery, D.M. (1995) "Power Markets and Market Power", *The Energy Journal*, Vol 16, No 3, pp 39-66

Newbery, D.M. (1998) "The Regulator's Review of the English Electricity Pool" *Utilities Policy*, vol 7 no 3, pp.129-142

Offer (1991) *Pool Price Enquiry* Birmingham, Office of Electricity Regulation

Offer (1992) *Report on Constrained-On Plant*, Birmingham, Office of Electricity Regulation

Offer (1995) *Generators' Pool Price Undertaking 1994/95* Birmingham, Office of Electricity Regulation

Offer (1998), *Independent Assessor's report on plant closure. 17 April 1998* Birmingham, Office of Electricity Regulation

Ofgem (1999) *Rises in Pool prices in July: a decision document. October 1999* London, Office of Gas and Electricity Markets

Ofgem (2000) *Ofgem's investigation of Edison First Power under the market abuse licence condition: Initial findings*, London, Office of Gas and Electricity Markets

Wolak, F. and R.H.Patrick (1997) *The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market* Power Working Paper PWP-047, University of California Energy Institute

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) "Electricity Markets: Should the Rest of the World Adopt the United Kingdom's Reforms?" *Regulation*, Vol. 22 (4): pp.48-53

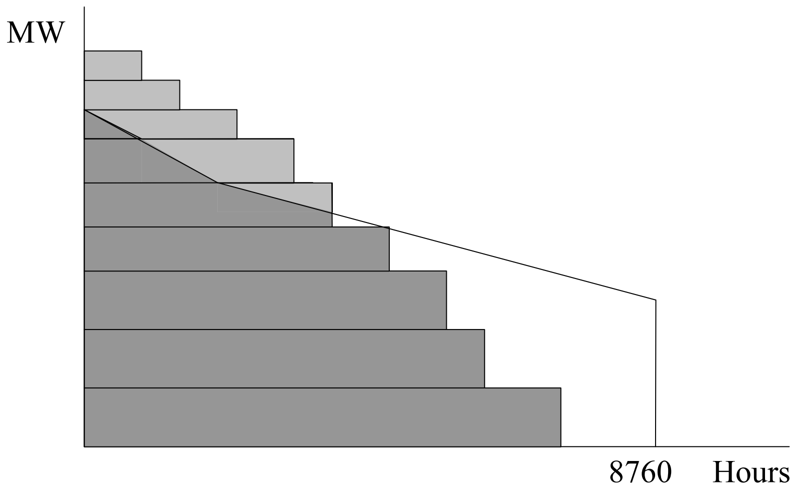

Figure 1: The Raw Load-Duration Curve

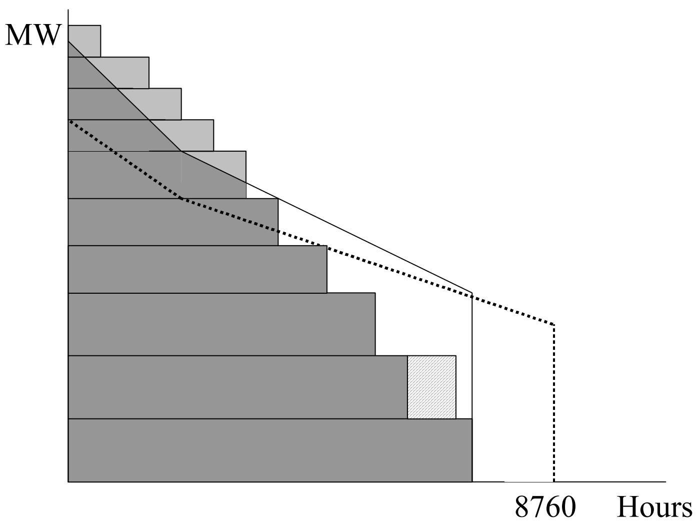

Figure 2: The Adjusted Load-Duration Curve