Generation Adequacy and Investment Incentives in Britain: from the Pool to NETA

Fabien Roques[1], David M. Newbery[2], William J. Nuttall[3]
September 19, 2004
CMI Electricity Project[4]
University of Cambridge, England

Abstract

There is no academic consensus on which market design provides the least distorting long-term investment incentives. Theoretical rationale and practical experience suggest that "energy-only markets" with spot prices that are allowed to reflect scarcity rents can generate sufficient income to allow capacity cost recovery by generators. However, different market designs with separate payments for capacity or reserve obligations have the advantage of not relying on infrequent price spikes to remunerate reserve capacity. Three years after the controversial change of the British market design from the compulsory Pool with capacity payments to the decentralised energy-only New Electricity and Trading Arrangements (NETA) market framework, we contrast the two market designs in terms of investment incentives. We review the biases of the Pool capacity payments design, the drought of investment following the introduction of NETA, and the reaction of the market during the first "stress-test" of NETA during the winter 2003. We recommend that NETA adopts a single marginal imbalance price as dual imbalance pricing distorts price signals in times of scarcity. The lack of long-term contracting that causes hedging and financing difficulties for power projects can be compensated by vertical and horizontal reintegration at a cost of increased market power. The case for re-introducing a capacity payment is ultimately a policy decision depending on the degree of price volatility that is considered acceptable.

Keywords: investment, electricity, market design, capacity payments

JEL-Classification: D24, D43, D92, L94

[1] PhD student, Judge institute of Management, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, England, email: f.roques@jims.cam.ac.uk; phone: +44 (0)7817541044; fax: +44 (0)1223 335299
[2] Department of Applied Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DE.
[3] Judge institute of management, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG.
[4] We acknowledge the CMI Electricity Project for supporting this research, and Heren Energy for providing the data on electricity prices.

OUTLINE

1 INTRODUCTION

The main objective of electricity market liberalisation is to improve the economic efficiency of the electricity supply industry. There is a great deal of evidence to suggest that market liberalization has, in general, improved productive efficiency, reduced operating costs of generating plants by improving labour productivity, reduced maintenance costs, and improved fuel purchasing strategies (Newbery and Pollitt, 1997). However, there is as yet insufficient experience to assess the long-term benefits from liberalising the electricity industry. The key issue is whether a liberalised electricity industry can deliver *sustainably* competitive outcomes; that is whether there will be both adequate and timely investment, built at least cost and delivering competitive electricity prices, and which market design and structure will be best to achieve these long-term objectives. The objective of this paper is to shed light upon the interactions between market design, market structure and investment incentives by reviewing the BRITISH experience and its two successive Pool and New Electricity and Trading Arrangements (NETA) market frameworks.

In the early years of liberalisation, the focus of academic research and regulatory scrutiny concentrated mainly on short-term market efficiency and competitiveness. As the first territories to liberalise – among which are included England and Wales and some U.S. states – have now reached the end of their first investment cycle, much attention is being paid to assessing the long-term dynamic performance of the liberalised electricity industry. The initial signs in England and Wales and in the U.S. were promising. In England and Wales new capacity of about one quarter of the existing (adequate) capacity was added in the first decade after liberalisation. In the U.S. 181 GW of new capacity was added between 2000 and 2003 alone. This initial enthusiasm in Britain was largely prompted by the high prices prevailing under the pre-liberalised regime, combined in some cases by market power that sustained prices at above entry levels. The sudden removal of regulatory constraints requiring approval of new build and the possibility of entering new markets may have spurred the more aggressive utility managers in the U.S. to over-invest initially. The subsequent over-capacity combined with more intense competition squeezed price-cost margins and created financial difficulties for many generating companies, pressure from their creditors, and a subsequent reluctance to invest in new plant in these markets. Elsewhere, some electricity markets are just beginning to approach their first major investment cycle as surplus capacity is absorbed (or withdrawn), so that the decisive test for liberalized electricity markets remains ahead. This is the case in the more mature European continental markets such as The Netherlands, Germany, and the Nordic countries.

The reliability of electricity supply has been one of the overriding concerns guiding the restructuring of the electric power industry. The policy imperative of "keeping the lights on" has been the principal motivation for many technical and economic constraints imposed on market designs. In Britain, there is considerable Government concern whether generating capacity will continue to be adequate to ensure security of electricity supply, and whether investment signals are timely and properly reflect the social profitability of investment, given the turbulent experience of recent investments. The market now appears considerably more risky than it did a few years ago. This raises the hurdle rate of return, and may bias investments against capital-intensive base-load plant and peaking capacity, undermining the quest for efficient and secure supply. Other market failures such as risk aversion from investors, regulatory uncertainty, and lack of long-term contracting could also distort investment incentives. Do these market failures constitute a strong enough case for regulatory intervention to support investment? Is there any 'best-practice' market design that can ensure generation adequacy in the long run at least cost while minimising regulatory interference with the market?

There is no consensus among academics on which market design provides the least distorting long-term investment incentives. Theoretical rationale and practical experience suggest that energy-only markets with spot prices that are allowed (and expected) to fully reflect scarcity rents will generate sufficient income to allow capacity cost recovery by generators. Hence as far as supply adequacy is concerned, a well-functioning energy-only market can provide the correct incentives for generation adequacy. The critical role of electricity in the economy and the political ramifications of widespread electricity shortages or price spikes have prompted many regulators around the world to take steps above and beyond reliance on market forces in order to ensure generation adequacy. Different market designs, with separate payments for capacity or reserve obligations have the advantage of not relying on infrequent price spikes (of possibly uncomfortably long duration) to remunerate reserve capacity. Three years after the controversial change of the British market design from a compulsory gross Pool with capacity payments to the decentralised pay-as-bid energy-only NETA market framework, this paper contrasts the two market designs in terms of investment incentives. It also highlights the market design changes and regulatory actions that appear necessary to maintain power generation investment and long-term generation adequacy.

The first part provides a historic perspective on the investment framework in Britain under the successive national Central Electricity Generating Board monopoly, the centralised Pool market, and the current NETA market arrangements. It reviews in particular the role that capacity payments played to facilitate entry under the Pool. The second part studies the drought of investments following the replacement by NETA, and how the market responded to the first "stress-test" of NETA during the winter 2003. The third part concentrates on potential market failures under the current form of NETA and the resulting distortions in investment incentives, focussing in particular on the issue of the imbalance pricing mechanism. The fourth part examines the impact of the lack of long-term contracting on hedging the investment decisions of power investors, showing that vertical and horizontal concentration provide an organic hedge but at the possible cost of market power. The last part reviews the regulatory actions that might reduce the investment distortions identified in the previous sections, considering in particular whether capacity payments should be reintroduced.

2 POWER INVESTMENTS BEFORE NETA

The privatisation and liberalisation of the electricity industry in Britain have profoundly modified institutional and regulatory structures, as well as the business model of the power generation companies.

2.1 Power investments before privatisation

Under the nationally owned Central Electricity Generating Board (CEGB), decisions about capacity expansions were subject to administrative planning procedures. The system was operated by a single, vertically integrated, state-owned utility comprising generation and transmission, selling wholesale power under a bulk supply tariff to twelve Area Boards each with a franchise on distribution and supply (retailing) in their respective regions. The main strength of such an integrated structure was the opportunity for balanced and consistent development of the system. Having access to information about the entire system, and owning and operating not only generation, but also transmission facilities, the CEGB was in an ideal position for comprehensive and integrated planning and realisation of capacity expansions.

However, the CEGB investment planning suffered several biases. Assessments of supply security were only based to a very limited extent on consumers' willingness to pay to cover such needs (Trade and Industry Committee, 2003a). Prices were cost-based and only adjusted to changes in underlying demand and supply conditions through the capacity element that was part of the bulk supply tariff. Consequently, there was very little information about what consumers might be willing to pay for different levels of supply security.

Second, system balancing management was based on technical criteria, such as capacity reserve levels and the likelihood of a rationing event. Based on demand forecasts, the CEGB drew up plans for new capacity requirements. The short term operation of the system required that there should be a sufficient level of generating capacity to meet demand at all times except nine winters in every century (NGT Seven Year Statement, 2002). Generation adequacy on the long term was determined by engineering targets: the reserve margin (the percentage of excess capacity as compared to peak demand) was set at no less than 24% (given the prevailing time lags in construction and forecast uncertainties). This reliance on purely engineering criteria for capacity expansion, combined with cost-based pricing, led to a high level of supply security, but also arguably to an over-expansion of capacity (International Energy Agency, 2002) and excessive costs.

Since, at least in principle, the CEGB could pass on to consumers all costs associated with capacity expansions, it was indeed very tempting to ensure a level of system capacity that minimised the risk of politically unpopular rationing. Last but not least, the engineering culture of the CEGB, combined with access to cheap public sector borrowing might have contributed towards this over-investment bias, as exemplified by the expensive British nuclear program (Helm, 2004).

2.2 Investment incentives under the Pool

One of the main arguments for privatising and unbundling the industry was that it would deliver better dynamic efficiency by making investors fully responsible for their investment choices. Being protected by its ability to pass on to captive consumers their investment costs, the CEGB had no incentive to reduce investment costs or risks. By making electricity producers bear the burden of their investment risks, this should lead to better informed and carefully assessed investment decisions. The electricity supply industry in England and Wales was restructured and privatised in March 1990. The CEGB was split into three generation companies and the National Grid Company (NGC), which owns and operates the transmission network. Twelve Regional Electricity Companies replaced the Area Electricity Boards, which had been responsible for electricity distribution and supply. Formal responsibility for generation adequacy was with the Regional Electricity Companies. They were deemed to meet it as long as they bought from the Pool (at prices that could rise as high as the value of lost load).

The Pool (meaning the Electricity Pool of England and Wales), a mandatory auction spot market, was established during 1990. The Pool operated a day-ahead market in which bids were submitted on the day before, and the least-cost unconstrained schedule then determined the system marginal price (SMP) as the most expensive generating set (genset) required to operate in each half hour, assuming that there are no transmission constraints. The Pool design facilitated entry by Independent Power Producers, and to that extent facilitated new investment. The Pool included capacity payments to encourage generators to invest and provide reserve capacity. Capacity payments were aimed at reflecting the expected cost to the user of a supply interruption, measured by the Value of Lost Load (VOLL), defined as the value that a consumer is ready to pay for the last kWh of electricity rather than being disconnected.[5] Capacity payments were made to each genset declared available to operate in each half hour, and were set equal to the Loss of Load Probability (LOLP) multiplied by the excess of the Value of Lost Load (VOLL) over the station's bid price (if not dispatched) or the SMP (if dispatched).[6] The sum of capacity payments and the SMP made up the Pool Purchase Prices, or PPP. The opportunity to sell all generation in a compulsory gross Pool at the PPP, which was therefore a highly liquid market, combined with a franchise on which Regional Electricity Companies could write long-term Power Purchase Agreements to buy from the IPPs, did much to stimulate entry and improve competition.

2.3 Strategic withholding of capacity

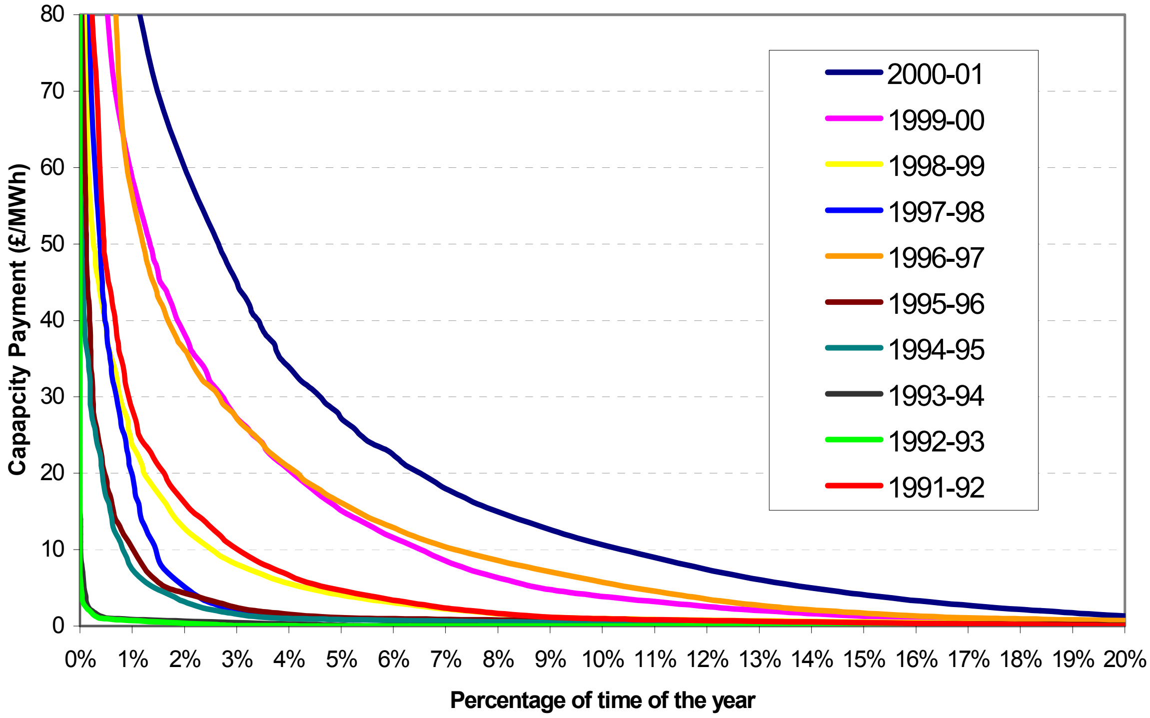
How did the capacity payments influence bidding behaviour, price levels and long-term investment incentives? The main problem of the Pool was the sustained market power of the incumbents. Sweeting (2001) and Newbery (1995) provide a review of market power leverage under the Pool. The relationship between market power and the capacity payment design is particularly interesting. The important question is whether the Pool's capacity payment design facilitated the exercise of market power and gaming. If so, would this capacity payment design have worked better if the market would have been more competitive? Newbery (1998b) asserts that a decrease in the capacity payments would not have led to a decrease in the Pool prices, as the sustained market power of the incumbents would have maintained the Pool prices at artificially high levels.

The capacity mechanism was widely criticised for not providing the right incentives to investors, allegedly because it was prone to manipulation through capacity withholding (Newbery, 1995). *Figure 1* shows the percentage of time in each year of the Pool during which capacity payments exceeded the Y-axis value. It is striking that four out of the five years during which the capacity payments were the highest were the during the last years of the Pool. Green (2004) uses a Cournot model to assess strategic withholding of capacity to raise capacity payments. He observes that such behaviours were not a significant feature of the early years of the Pool; the 1994/95 sustainably high capacity payments were subject to an Offer investigation, which found them to be the result of an unlikely combination of plant outages. During the following years, the capacity payments returned to lower values, before rising sharply again during the last three years of the Pool.

[5] VOLL was set administratively at £2,000/MWh in 1990 and was then increased annually by the RPI – in 2000, it stood at £2,816/MWh (data from National Grid)

[6] More detailed information is available on the Pool website still maintained by Elexon (http://www.elecpool.com/index.html)

Figure 1: Cost duration curve of capacity payments (£/MWh)
Data source: Elexon



This led critics to suspect that strategic behaviour had rendered the capacity payment meaningless in the last years of the Pool. The last year of the Pool recorded the highest capacity payments, while the capacity margin for 2000/01 was as high as 25.3 % (NGT, 2000). These high capacity prices were directly related to the absence of significant amounts of generation capacity during the summer 2000. Plant had been absent for a mixture of planned and unplanned outages. Some generation capacity was withdrawn for "economic reasons" and this was investigated by OFGEM under the "good behaviour clause" added to some generators' licences.

Figure 2: 2000-2001 plant availability versus total generation declared
Source: British gas trading 2001

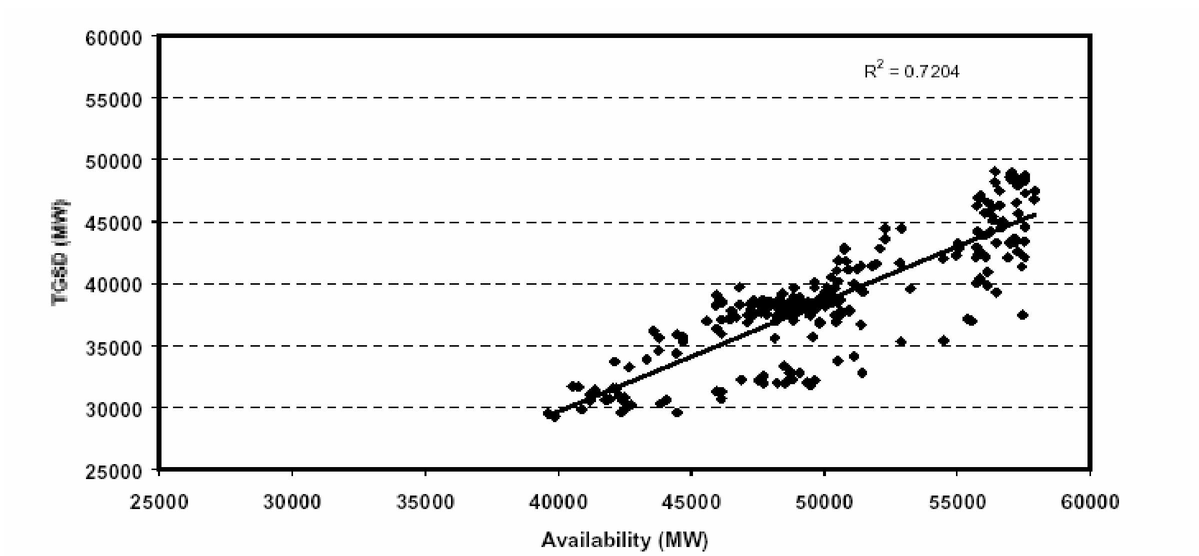


Figure 2 shows that there was a strong correlation between available generation and demand during the year 2000/2001, which is not intuitively logical. However, Green (2004) points out that these abnormal payments were not the result of strategic capacity withdrawal but rather caused by the anomalous way in which the availability factor of new plants was calculated. The next section investigates the flaws in both the LOLP and VOLL estimates that were used in the capacity payments computation.

2.4 The flawed design of the capacity payment

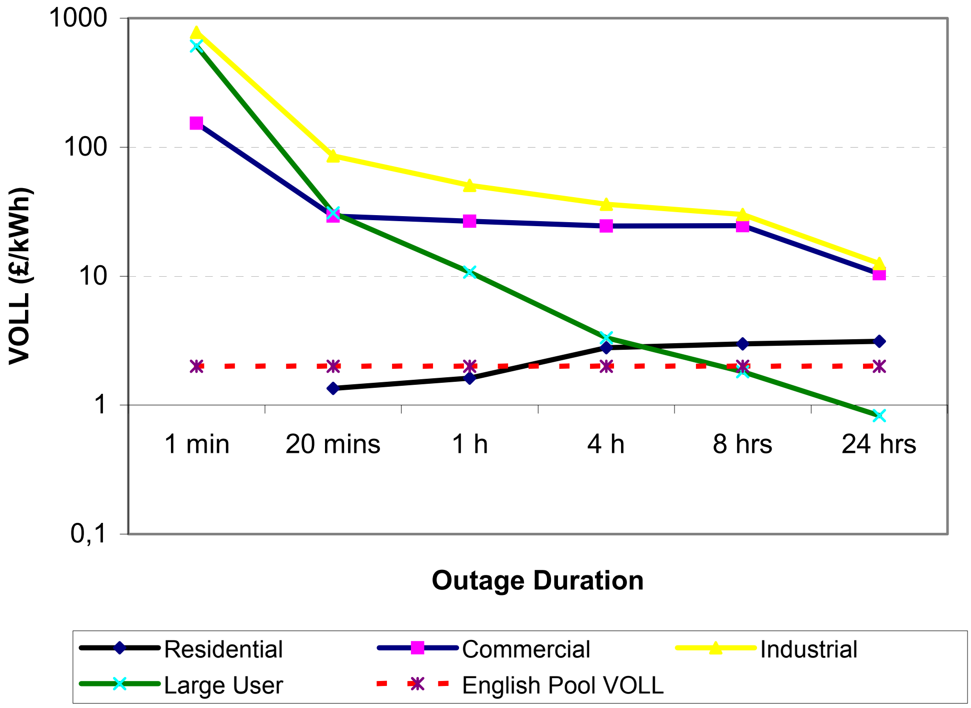
Newbery (1998b) argued that the computer program used to calculated the LOLP almost certainly overestimated the chance of a power failure, estimating it corresponded to a probability of a failure on one of the 10 peak days greater than 99.88%. The LOLP calculation indeed suffered several flaws that led to overestimating the probability of losing load. LOLP calculation used the standard error in the demand forecast, and probabilities of "disappearance" of each generating unit between the date from which availability was deduced and the time of the forecast (Green, 1997).

The way the disappearance ratios were calculated systematically underestimated the amount of capacity actually available and despatched at peak times, and made it easy to game for generators. For pre-1990 plant, these disappearance ratios were set equal to their historical pre-1990 values, even though subsequent capacity payments had provided strong incentives to improve reliability and availability. The reliability of post-1990 plant was based on operating performance in the previous year. The main issue was that the calculation used average availability, ignoring the fact that some plants are fully available at peak times but less available off peak. The disappearance ratio represented the probability of a plant not being available on any random day of the year, given that it was available in the previous week: it made no allowance for the various reasons why plant was not available. Moreover, an unplanned outage at a relatively new plant in one month could significantly increase the LOLP and hence capacity payments in the following month, despite high reliability of the plant on a day-to-day basis.

The second main flaw that can be identified in the LOLP calculation is that the software looked only at the absolute difference between generation and demand, not the relative difference. Thus a generation margin of 5000 MW would create the same price signal in the summer as on a winter peak, although the absolute demand level might be different by 200%. Lastly, demand reduction blocks offered by some large customers received capacity payments, but this demand reduction was anomalously not included within the LOLP calculation, thus increasing capacity payments.

The second element of the capacity payment, the LOLP which was set administratively, was also subject to criticism. Newbery (1998b) found indirect evidence using price elasticities that the VOLL figure was probably underestimated. VOLL estimates are rendered difficult by the lack of direct measurement of the value consumers put on electricity security of supply, as power outages are exceptional phenomena. Current estimates in use in electricity markets around the world are often arbitrary determined, and a better approach would consist in deriving them from consumer willingness to reduce load at times of scarcity.

Figure 3: VOLL for different interuption durations
Source: UMIST 1995



As exemplified by *figure 3* above, the problem is that there is no such thing as a generic VOLL for all consumers. VOLL varies greatly among consumers categories, large industrial and commercial consumers having a much higher VOLL which can justify the installation of back-up systems, while individual consumers have a VOLL close to the £2,000/MWh chosen administratively in 1990. Moreover, VOLL varies with the duration of the power shortage: while VOLL decreases with time for large industrial users, which have the possibility to start back-up systems or to adapt their activity, VOLL increases with time for domestic consumers whose activity crucially depends on electricity (e.g. freezer loss after a certain number of hours).

The level of the capacity payment was determined by the product of LOLP and (VOLL-SMP), and therefore it might be that the overestimation of LOLP was partly compensated by the underestimation of VOLL. Newbery (1998b) pointed out that although each calculation of VOLL and LOLP might seem somewhat arbitrary, it is their combination that matters to determine the level of the capacity payment and thus the investment incentives. Refining the analysis, the proportion of the two components actually has an important impact on the riskiness of investing in reserve generation capacity, and therefore on generation investment incentives. If VOLL were increased, and LOLP reduced so that their products remains unchanged, the capacity payment would have been higher but paid less often to generators. In other words, reserve generation plants would be run less often but earn the same profit on average as they would be paid higher capacity payments. As a consequence, one can expect that the generators would have adapted their bidding strategies and investment strategies to the new risk profile of the market. Both the optimal portfolio of plants for a generator and its optimal bidding strategies would have been modified by the increased risk in the market for reserve capacity. In the light of this example, one realises that the over-estimated LOLP combined with a relatively low

VOLL might have been a 'politically' strategic choice: by providing a relatively constant flow of revenues, the so defined capacity payment made investment in power plant easier than if generators would have had to rely on higher but rarer and thus more uncertain cash flows. The choice of the 'low risk' design option for the capacity payment appears designed to encourage entry and make the market more competitive, while guaranteeing that the lights would not go out and take the gleam off privatisation.

It is interesting to compare the British and Australian experience of determining capacity payments. In Australia VOLL is paid whenever the system has inadequate capacity to maintain supply, so that the LOLP is the actual *ex post* probability of loss of load. The VOLL was set at the British level, but several actual black-outs lead to a reconsideration of the VOLL, which had to be substantially increased to make peaking capacity commercial.

To conclude, despite the flawed calculation of the capacity payment, one should not forget the main advantages of the Pool design: it facilitated entry by Independent Power Producers, and to that extent contributed to decrease the market concentration. The opportunity to sell all generation in a compulsory gross Pool, which was therefore highly liquid market, provided a transparent and simple price signal on which to develop contracts for differences and other hedging financial instruments. Lastly, the franchise on which Regional Electricity Companies could write long-term Power Purchase Agreements did much to stimulate entry and improve competition. It is interesting to notice that the Pool design principles, a compulsory gross pool and a capacity mechanism to encourage investment, is close to the Stand Market Design (SMD) recommendations of the FERC, which Hunt (2002) considers as the "clear market design winner".

The Council of Australian Governments issued a report on the workings of the Australian National Electricity Market after studying a number of electricity markets (Nordpool, PJM, NETA). It concluded that there were good reasons for retaining the single price gross pool design. The report observed that NETA's incentives for individual balancing created "a significant inefficiency that adds cost to the system" (CoAG, 2002, p103).

3 HAS NETA UNDERMINED INVESTMENT?

3.1 The current market arrangements

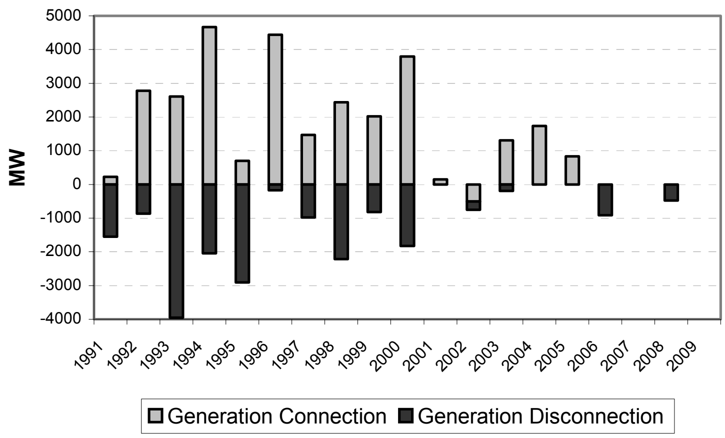
In October 1997, the Minister for Science, Energy and Technology asked the Director General of Electricity Supply (DGES) to review the electricity trading arrangements and to report results by July 1998. The recommendations of the resulting *Pool Review* were accepted and NETA went live on 27 March 2001. The process that led to the eventual ending of the Pool and its replacement by NETA has been extensively described and criticised (e.g. Newbery, 1998a,b). Under NETA, electricity is now traded in four voluntary, overlapping and interdependent markets operating over different time scales. The most obvious difference between NETA and the Pool is that under the Pool all generation was centrally dispatched while under NETA plant is self-dispatched. The obligation to balance output with contracted demand is now placed on each generator (and suppliers are similarly required to match contracted with actual demand), with the System Operator (SO)'s task confined to ensuring system stability.

The Pool, that acted as both a wholesale market for all electricity and allowed NGC as SO to balance the system, was replaced by a Balancing Mechanism (also operated by NGT – the successor company to NGC - as SO) for the residual imbalances of parties that fail to self-balance. Whereas the Pool operated as a uniform single-price auction for buying and selling all power (including that needed for system balance), the Balancing Mechanism is run as a discriminatory (pay-as-bid) auction. Lastly, capacity payments have been abandoned and there is no separated remuneration of reserve capacity: NETA is an 'energy-only' market, as opposed to the Pool and other markets in the world (Spain, PJM), which remunerate both energy and capacity through price- (capacity payments) or quantity-based (capacity obligations) arrangements.

3.2 Investments indicators evolution since privatisation

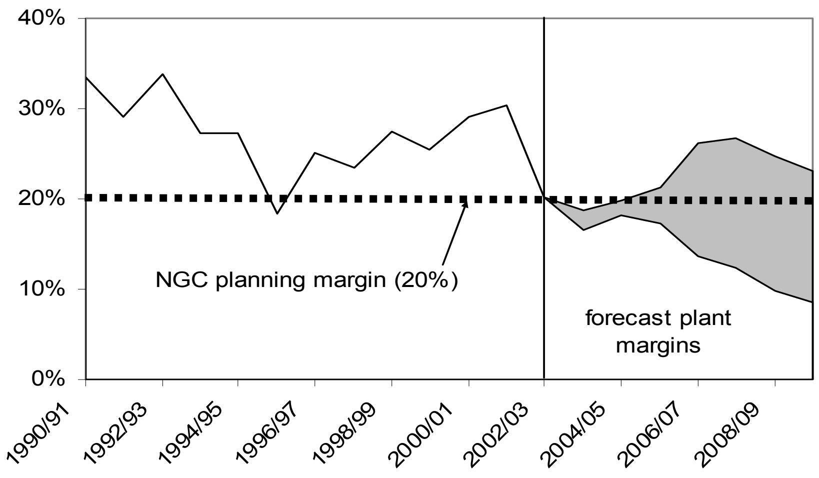
Figure 4 below shows that since privatisation in 1990, generation connections and disconnections have remained at high levels during the Pool but decreased dramatically since the implementation of NETA. It is noticeable that very few power stations have been commissioned since 2001, or are under development.

Figure 4: Past and declared future Generation Capacity Changes (source: NGC SYS 2003)



Similarly, the capacity margin, defined as the percentage of installed capacity in excess of peak demand over a given period, has fallen dramatically after 2001 (see *figure 4)*. Under the Pool, capacity payments and relatively high wholesale electricity prices resulted in large investments in generation during the second half of the 1990s, characterised by the so-called "dash for gas". The temporary drop of the capacity margin in the middle of the 1990s can be attributed to anticipations of generation companies of the "dash for gas", which resulted in the decommissioning of many uncompetitive coal power stations as a result of the announcement of the construction of more cost efficient combined cycle gas turbines (CCGT). However, since the creation of NETA in 2001, falling wholesale prices have contributed to more 'mothballing' of electricity generating plant, to the postponement of construction on a number of new power stations which had received planning permission, and to reductions in levels of capacity margins (JESS, 2003a).

Figure 5: UK electricity system capacity margin
Source: NGT, April 2003



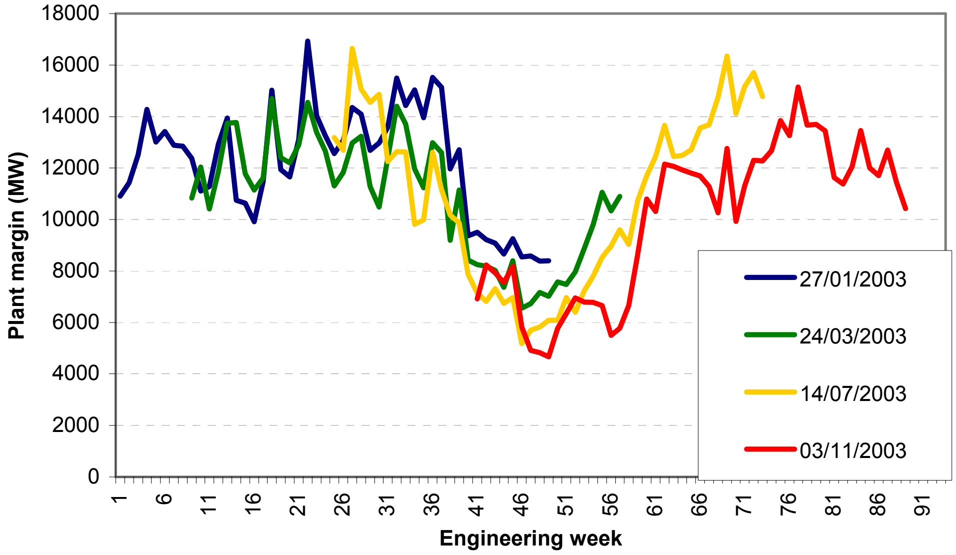
There is no single figure that defines an optimal reserve margin, because the reserve requirement depends greatly on the characteristics of the electricity system, namely on transmission and distribution capacities, pumped storage capacity, the size of the system, and the portfolio of generating plants. According to the International Energy Agency (IEA, 2002), "reserve levels in the range of 18 to 25% of total generating capacity are often considered appropriate". The IEA (2002) mentions also that "reliability criteria may appropriately be relaxed, as the flexibility of electricity systems to respond to a surge in demand increases". The IEA details several factors that might further increase flexibility in the future, such as demand-side measures intended to increase the responsiveness of consumers to supply conditions, the gradual development of bilateral electricity trade, the deployment of distributed generation, and the integration of markets. The former national-owned CEGB operated a 24 % planning capacity margin (in the years when the time to complete a large base-load station was 7-10 years), while the current operating target of NGT is a 20 % capacity margin (CCGT can be built in 2-3 years). *Figure 5* shows that under some future scenarios the capacity margin might fall below the 20 % operating target.

3.3 Winter 2003: the first stress test of NETA

3.3.1 A forecast capacity shortage

Under the Pool, the surplus of capacity inherited form the CEGB, combined with favourable conditions for new entry maintained the capacity margin at consistently high levels. During the year 2003, the NETA arrangements were confronted with their first "stress test", as British capacity margin forecasts for the winter 2003/2004 dramatically fall during the spring and summer of 2003. This was the result of the closure or mothballing of a number of plants in 2002 and early 2003, following the historically low levels of wholesale prices observed during 2002. On 1 April 2003, Powergen closed their High Marnham (756MW) and Drakelow (333MW) coal-fired plants. Powergen also took the decision to mothball from 1 April 2003 Killingholme (450MW, CCGT) and Grain (1350MW, oil-fired) from 1 April 2003 (OFGEM, 2004a).

Figure 6: Evolution of Capacity margin Forecasts 52 weaks ahead
Data source: NGC SYS 2003

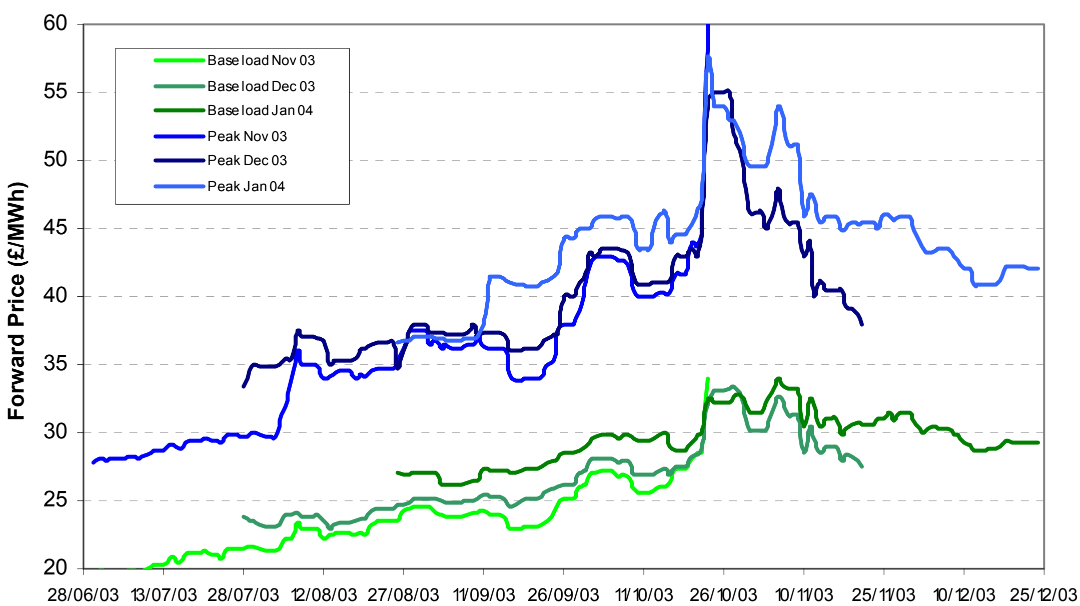


This withdrawal of plant resulted in NGT capacity margin forecasts falling to a low point of 16.2 per cent for the winter 2003/2004 in May 2003. *Figure 6* shows the evolution during the year 2003 of the capacity margin forecast for the coming winter 2003/2004. The capacity margin for December 2003 (engineering weeks 43 to 47) was forecast to be 8200 MW in January 2003, but only 5200 MW in May 2003. At this point NGT was concerned that such a low capacity margin forecast might not allow it to operate the system in extreme conditions during the winter 2003/2004.

3.3.2 The market reaction

During the summer 2003 winter 2003/04 forward prices increased substantially. *Figure 7* below shows the evolution of the forward base-load and peak-load electricity prices for the November, December 2003, and January 2004.

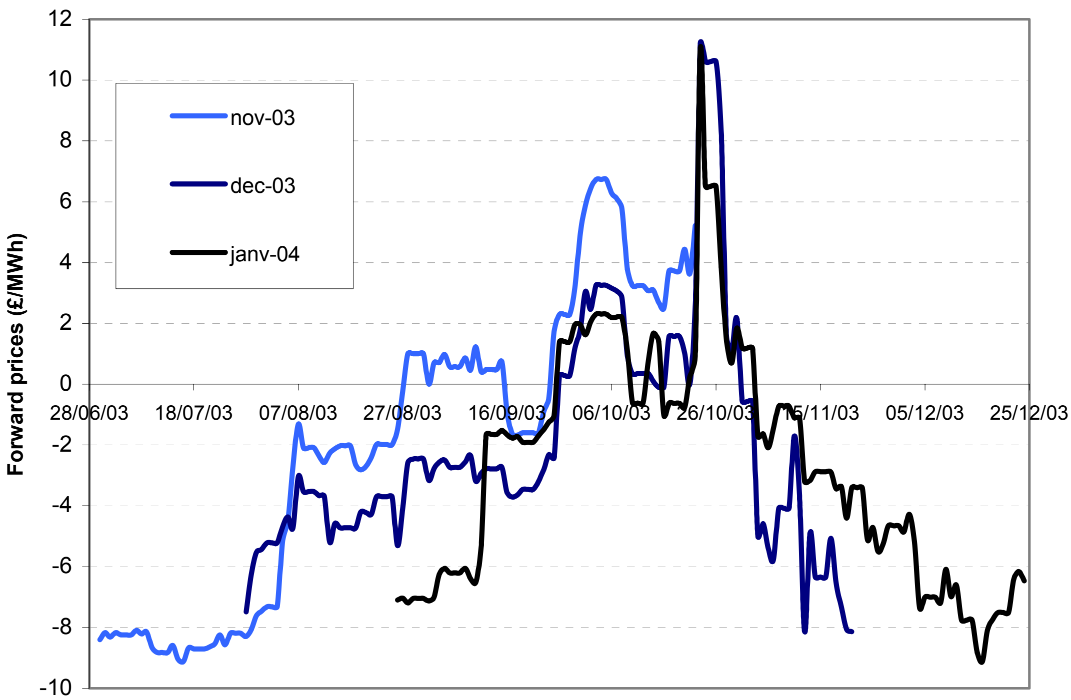
Figure 7: Winter 2003/04 forward electricity prices (£/MWh)
Data source: Heren



Base load prices increased from 23 £/MWh to 33 £/MWh between July and October 2003 (43% rise). The rise of peak load prices is even more impressive, from 35 £/MWh in July 2003 to 55 £/MWh in October 2003 (57 % rise).

In interpreting these electricity prices trends as a market response to the forecast capacity shortage, one needs however to asses the impact of the underlying determinants of forward electricity prices. In particular, forward gas prices have an important impact on forward electricity prices, as the forecast 'spark spread' (the spread between the price of electricity and the price of the gas required to generate that electricity) determines the forecast profitability of gas power stations. In other words, was the observed increase in forward electricity prices a healthy reaction of the market to the scarcity forecast, or simply caused by the rise of the forecast gas prices? Forward gas prices for delivery during the winter increased slightly during 2003, so that the increase in forward prices was not completely translated into increased profitability. *Figure 8* shows the forecast peak-load winter 2003/2004 spark spread evolution during 2003.[7] This can be considered as a better indicator of market expectations of the capacity tightness during the winter. The graph shows that the forecast spark-spread for December 2003 rose from –4 £/MWh to 11 £/MWh, thus confirming that most of the forward electricity prices increase was due to perceived scarcity in the market. Similarly to forward prices, wholesale electricity prices increased considerably during the second half of 2003. For instance, the average spot price of £20.4/MWh was 34 per cent higher than during the second half of 2002; for 2003 as a whole the increase on 2002 was 20 per cent in money terms.

Figure 8: Forecasted Winter 2003/04 Peak-load Spark Spread
Data source: Heren



This price increase during the summer 2003 motivated some generators to bring back on stream some mothballed plants for the winter 2003 (BBC, 2003). On 19 August 2003 it was announced that

[7] Data sources: Electricity and NBP gas forward prices from HEREN, assuming a plant heat rate of 7000 BTU/KWh and a thermal efficiency of 46.8% corresponding to the 2002 UK average efficiency of gas turbines according to NGT.

unit one of Grain would be returned to service, which increased available capacity by 675 MW. On 29 September 2003 Edison announced that one unit of the pumped storage plant at Dinorwig (288MW) would be returned to the system. Further mothballed plants have returned during November, and the January 2004 *Seven-year Statement* update published by NGT reported a capacity margin at 21.6 per cent, as compared to the forecast 16.2% in May 2003. *Figure 8* shows that base load and peak-load prices settled down respectively around 28 £/MWh and 42 £/MWh during the winter. OFGEM concluded in its retroactive winter 2003 report that the actions taken by firms during the second half of 2003 is to be expected in a competitive market whereby the availability of capacity responds to prices, which reflect underlying supply and demand conditions (OFGEM, 2004a).

3.3.3 What role did NGT play in signalling scarcity?

While there were doubts on the capacity of market prices to respond to this scarcity forecast, the actions taken by generators in the second half of 2003 demonstrated that the NETA arrangements had passed their first "stress test" with success. But if the market responded quickly to the forecast capacity shortage during 2003, which tools did NGT use to induce this response? Did NGT just made its concern known or did it proactively contract for reserve? During the summer 2003, NGT warned on a number of occasions during industry meetings that under some cold weather scenarios the Operational Planning Margin Requirement (OPMR) would not be met in a number of weeks during the coming winter. But most importantly, following analysis of its likely reserve requirements for winter 2003/04, NGT initiated a Supplemental Standing Reserve Tender (SSRT) on 14 October 2003. The tender closed on 27 October 2003, and NGT received 22 tenders in total. NGT accepted 20 of these tenders and procured a total of 852MW of Supplemental Standing Reserve (SSR) at a total cost of £18.87 million (NGT, 2004a). The majority of this volume was provided from plant that had previously been mothballed and a significant amount came from the demand side.

The role that this supplementary tender for reserve played in signalling NGT concerns about the forecast capacity margin is demonstrated by the two previous figures, which reveal that the period of NGT tender between the 14 and 27 October 2003 coincided with the forward electricity and spark-spread price spikes. Lastly, NGT also developed a Maximum Generation Service (MGS) for winter 2003/04. MGS applies to non-firm generation that is not commercially viable for a generator to offer to the market all the time. It allows generators to produce – and be paid for – more than their registered plant capacity. MGS would only be used in emergency situations, and its use was not required in 2003/04. Discussions are ongoing within the industry about the development of an enduring MGS (NGT, 2004a).

3.4 NGT re-defined role as regard to short term security of supply

Throughout the process of introducing NETA, there was extensive consultation regarding the role of NGT versus the role of the market in ensuring electricity balancing and short-term security of supply, which was defined as the period from day ahead to real time (OFGEM, 1999a, b)). Under NETA, NGT is not required to contract in advance to ensure that there is sufficient generation capacity to meet peak demand, nor is NGT the provider/buyer of last resort. Market participants are responsible for ensuring that generation capacity is sufficient to meet peak demand. The trading arrangements provide commercial incentives on market participants to balance their contracted and physical positions and therefore ensure that the market as a whole matches generation and demand for each half-hour long balancing period. In its role as residual balancer NGT is responsible for ensuring that demand and supply are balanced on a moment by moment basis; managing the physical consequences of any plant failures that occur on the network for the short period (e.g. 3-4 hours) until the market is able to respond to such a failure; and managing the physical consequences of any unexpected increases in demand for a short period until the market is able to respond to such an increase (OFGEM 2004c).

Prior to winter 2003/04, NGT had concerns in relation to forecast capacity margin for the winter period and asked OFGEM for further clarification of OFGEM's interpretation of its obligations and how they related to the way that it procured short-term reserve. Following clarification of OFGEM's interpretation, there has been a crucial change in the methodology used by NGT when procuring reserve. The modification deals with *when* NGT procures reserve, i.e. how much in advance of real time. This is as much, if not more, important for security of supply as *how much* reserve NGT procures. NGT has indeed the commercial flexibility to procure its short-term reserve requirements through forward tenders/contracts or options and also via the Balancing Mechanism. By contracting to procure its short time reserve requirement, NGT sends signals to the market. In particularly, how much in advance of real time NGT contracts for reserve is crucially important to allow the other market players to anticipate an impending shortage, and to adjust their positions accordingly.

In considering when to procure its short-term reserve requirement (i.e. in advance vs. on-the-day), NGT considers the cost of procuring short-term reserve ahead of time versus the expected cost of procurement close to real time. There is a trade-off between the degree of certainty that NGT achieves in respect of securing its short-term reserve requirements in view of its wider licence obligations and the balancing costs that it incurs. Under the pre-winter 2003/04 approach, NGT failed to comprehend this reserve procurement impact on other market players: it was procuring short-term reserve based purely on narrow economic trade-offs, without giving explicit consideration to its wider obligations to ensure short term security of supply. Under this approach when procuring short-term reserve via the standing reserve tenders, NGT's assessment was based on consideration of the relationship between what NGT terms the 'assessment price'[8] and the 'equivalent price'[9]. Based on this assessment, NGT entered into forward contracts for its reserve requirements up until the point where the equivalent price and the assessment price equalled each other, without giving explicit consideration to its wider obligations to balance the system in real time.

Following OFGEM's clarification, NGT's approach to reserve procurement during the winter 2003/04 was different. In response to its concerns in relation to capacity margin levels, and based on an assessment which included consideration of its wider short term security of supply obligation, NGT procured short-term reserve that it would not have procured solely on the basis of the narrow economic trade-offs described above. According to OFGEM's clarification (OFGEM, 2004c), NGT's current approach to procuring reserve "gives explicit consideration to the trade-off between the degree of certainty that it achieves in respect of securing its short-term reserve requirements in view of its wider licence obligations and the balancing costs that it incurs. For example, if NGT forecasts that there is a significant risk of there being insufficient plant available on the day, it can enter into forward

---

[8] The assessment price represents the forecast value of the service to NGT and is the avoided cost of alternative reserve services. It is based on historic price curves with appropriate adjustments for market drivers.

[9] The equivalent price represents NGT's estimate of the actual cost of the service. It is based on the total forecast cost of the tender (which is the sum of the availability cost and the forecast utilisation cost) divided by the capacity.

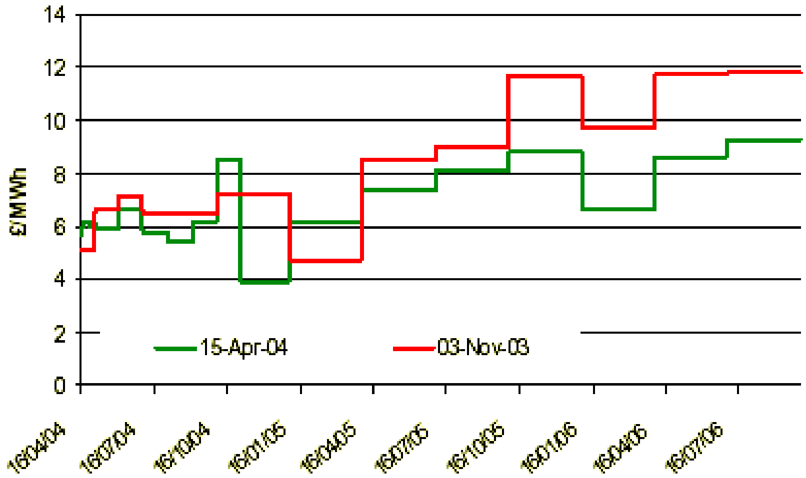
contracts that might not otherwise appear to be economic, based on a narrow assessment such as that undertaken previously, in order to reduce the risk that it would not have sufficient short-term reserve available on the day."

While OFGEM regards "the subtle differences between the two approaches" as a modification that should let "NGT's procurement of short-term reserve largely unaffected" (OFGEM, 2004c), such clarification on the timing of reserve procurement by NGT, and on the extension of NGT responsibility as regard to short-term security of supply should greatly contribute to improving the price signals quality in the market by helping other players to anticipate capacity shortages, as exemplified by the impact that the Supplementary Reserve Tender (SST) that NGT ran in October 2003 had on electricity prices. It is also interesting to note that NGT was not charged for the cost of procuring this supplementary reserve. OFGEM indeed approved the NGT request that the costs of the SSRT should be considered as an 'Income Adjustment Element' (IAE), i.e. that an adjustment (set at £5.54 million) should be made to its Incentivised Balancing Costs (IBC) to remove the impact of the SSRT from its performance under the SO incentive scheme (OFGEM, 2004d).[10]

3.5 The outlook for future winters

One might wonder if the cyclical seasonal pattern in prices and mothballing-de-mothballing of plants observed in 2003 will be reproduced in the coming years. Will there be sufficient investment to increase the reserve margin to more comfortable levels? The spark spread is an important indicator for potential investors, since it indicates whether the investment is likely to be profitable.[11] Where the spark spread widens one would expect companies first to respond by de-mothballing existing gas- (or coal-) fired capacity and then by building new gas-fired plant.[12]

Figure 9: Forecast spark spread 2005-06
Source: JESS 4th report



---

[10] The forecast costs associated with the eight tenders which would have been accepted under 'approach 1' was £0.87 million. The forecast cost of the remaining twelve contracts that NGT accepted, due to its consideration of wider issues under 'approach 2', was £18 million.

[11] The gas spark spread assumes a standard efficiency (Platts assumes 55%) while the coal spark spread assumes 34%.

[12] In Europe, given projected gas and carbon prices, CCGT is likely to remain the least-cost choice, although under some price scenarios new coal might be competitive.

Figure 9 shows that the gas spark spread is currently expected to remain around £6/MWh, before rising to a maximum of £9/MWh during late 2006. Given that the short run operating costs ('O&M' costs, such as staff, maintenance, etc.) for a combined cycle gas turbine can be estimated between £6 and £8/MWh, these spark spread forecasts make it profitable to de-mothball a plant whose fixed costs have been amortised, but suggest that there will be little investment in new gas turbines. One interesting point to note is that the April 2004 forecast spread over summer 2005 and 2006 is significantly lower than that six months ago. This change is largely due to a fall in expected electricity prices, possibly reflecting the fall in emission allowances prices from the beginning of the year to around £5/tCO_2e from £9/tCO_2e.[13]

If prices are still insufficient to allow new entrants to recover fixed costs, the availability of mothballed plants ready to be returned to generation within short time scales will be crucial to maintain system security in the event of a future capacity shortage such as the one of winter 2003. Prior to the winter 2003, a total of 4.2 GW of mothballed plant was identified, of which 1.6 GW had the physical potential to be returned within six months. In the event a total of 3 GW actually returned prior to the winter (NGT, 2003a). This shows that the return of plant to service can be accelerated beyond initial forecasts if operators judge the market conditions to be favourable. However, by 2004 there was much less mothballed plant capacity available in the system than there was in 2003, and that might undermine the ability of the system to respond as quickly as in 2003, would another capacity shortage be forecast.

The availability of information on mothballed plants, as well as the flexibility of the re-connection to the system, are crucial elements for short-term security of supply. Another potential impediment to returning plant for short periods at the peak is that the charge for using the grid is levied per MW of capacity for the entire year. To facilitate the return of mothballed plant within year, NGT has proposed an amendment to the Connection and Use of System Code (CUSC), which would allow generators to apply for capacity on the NGT transmission network in 4-week blocks, rather than the current minimum of 12 months. First Hydro has recently raised an alternative option to this draft amendment proposal, which would allow generators to apply for short notice (2 weeks), short-term transmission entry capacity, available in 6 week blocks. The proposal is being considered by an industry working group with a view to its implementation prior to the 2004/05 winter. Moreover, National Grid has been requested by OFGEM to consider changes to the Grid Code in response to a number of information gaps identified by the DTI/OFGEM Joint Energy Security of Supply working group (JESS) about the estimated return to service times of mothballed Generating Units, and the capability of gas-fired Generating Units to operate using alternative fuels, such as oil distillate, or alternative gas supply (NGT, 2003b). These changes would aim at facilitating a mechanism for Generators to provide to National Grid information relating to these issues.

---

[13] Carbon prices are added to fuel prices and included in this spark spread calculation.

4 IMPROVING PRICE SIGNALS

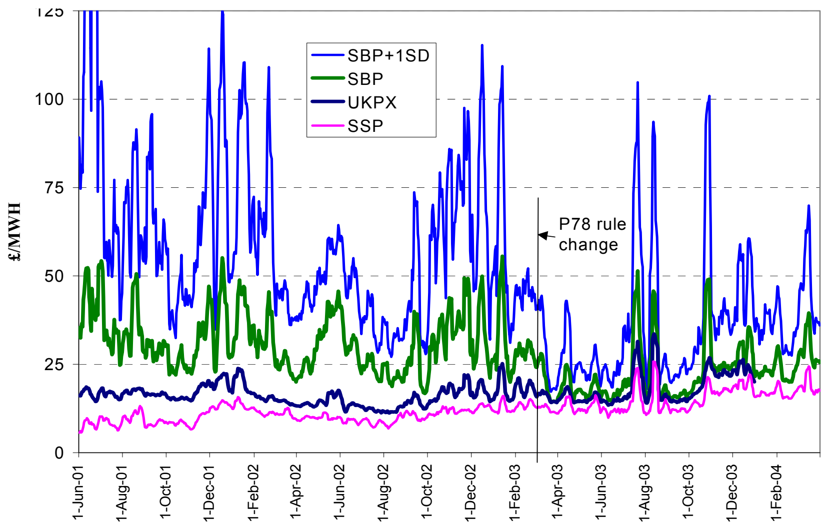
4.1 The Balancing Mechanism design flaws

Whereas the Pool operated as a uniform single-price auction for buying and selling all power (including that needed for system balance), the Balancing Mechanism (BM) is run as a discriminatory (pay-as-bid) auction. Elexon determines two cash-out prices: the weighted average of accepted offers (to increase output or reduce demand) determines the System Buy Price (SBP) and that of bids (to reduce output or increase demand) the System Sell Price (SSP). The imbalance pricing scheme has been subject to much criticism since the introduction of NETA (Cornwall, n.d.; Newbery and McDaniel, 2003), as it is based on innovative principles, namely decentralised despatch, dual-imbalance pricing and a volume-weighted average calculation of the imbalance charge. These characteristics of the Balancing Mechanism could be changed independently of the overall NETA design concept. The following sections detail successively the drawbacks of decentralised despatch ("nominations"), of having two different prices for being short and long, and of using average instead of marginal imbalance pricing. Moving to a single marginal imbalance price would provide better price signals, but this idea has so far been rejected by OFGEM in the first step of the cash-out mechanisms launched in its review of March 2004 (OFGEM, 2004b).

4.1.1 The Balancing Mechanism increased the cost of imbalances

The balancing prices are considerably more volatile and unpredictable than the Pool prices that served as a more liquid balancing market. *Figure 10* below shows 7-day moving averages of the buy and sell prices, and, to give a sense of the risk in the SBP, gives one standard deviation of the 7-day half-hourly buy prices, as well as the underlying spot price.

Figure 10: Spot and cash-out weekly moving average prices June 2001-April 2004
Data source: Elexon



As a result of the initially extreme volatility of the balancing prices a considerable number of modifications were made. One of the more important (P78, shown on *figure 10*) made the reverse balancing price (i.e. the price facing parties who were in the opposite position to the overall market, e.g. long when the market was short, and hence aiding balance) would revert to the spot price, and hence not penalise those helping balance the system relative to their selling in the spot market (Elexon, 2002).

The critical feature of having two different prices for being long or short is that these prices are normally different (SBP ≥ SSP),[14] and penalise each party's imbalances, whether or not they amplify or reduce the system imbalance as a whole. Combining two different imbalance prices with an average imbalance calculation makes it more risky for a generator to offer balancing services. If a generator has an accepted offer to increase output, and then suffers a loss of output, he is likely to have to pay more than he is paid. He may therefore prefer to retain the spinning reserve for his own insurance. This has led to claims that the Balancing Mechanism increases the costs of balancing to the detriment of non-portfolio generators (i.e. new entrants and British Energy) and intermittent suppliers like wind (Henney, 2002).[15]

Newbery (2004) estimates the risk of having to pay the buy price (SBP) after a generator suffers a forced outage, in order to meet an assumed contract position. If a generator fails at a random moment and stays off-line for 24 hours and is unable or unwilling to re-contract before gate-closure), the cost will be the 24-hour average of the SBP from that moment.[16] In the year before the P78 rule change indicated above, the expected cost of such an outage (relative to an assumed variable cost of £12/MWh) was £17/MWh or £0.4/kW/event compared to £13/MWh or £0.32/kW/event under the Pool for 1997-8. The variance was, however, twice as high as under the Pool. In the year following P78, the average cost had fallen to £11/MWh or £0.3/kW/event and the variance had also fallen to 150% that of the Pool.

Figure 11 illustrates the cost duration curve for balancing under NETA from April 1 2003-31 Mar 2004 compared to the Pool in 1997-98. Thus 5% of the time the cost would be £30/MWh for the following 24 hours in both the Pool and the recent BM and 1% of the time it would be £70/MWh in the BM compared with £44/MWh under the Pool. The risks in the early days of NETA were very much higher and led to claims that plant was inefficiently part-loaded to avoid penal imbalance costs, at considerably higher cost.

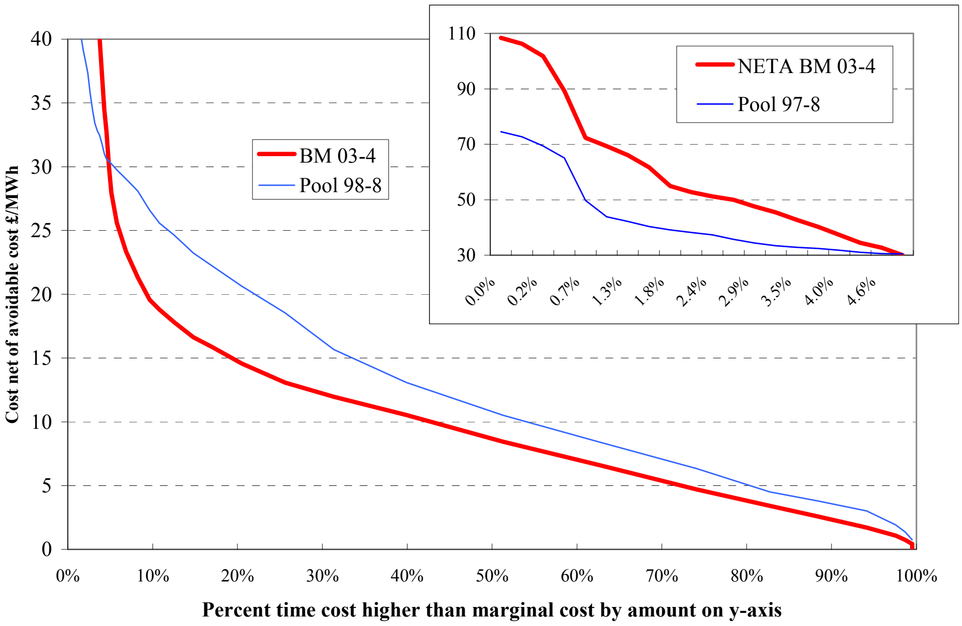
One should interpret this finding with some care, as the Pool required bids to remain valid for 24 hours while bids and offers to the BM can be changed on a short time scale and in response to a perceived tightening of the market when a large unit goes off-line, making it more risky for generators to handle outages. Even if we ignore such responses, if a large plant were to go down, the demand in the BM might be such as to considerably increase the short-run cost, but without knowing the shape of the bids and offers it is hard to estimate by how much.

[14] The prices were equal by about 25% of the time, and SSP exceeded SBP very occasionally (0.1% of the time) in the first 18 months.

[15] Although the net surplus of the Balancing Mechanism is recycled, there are transfers between different types of participants, while there are extra real costs in requiring all participants to replicate the SO balancing function, especially in maintaining spinning reserve.

[16] The assumption of not being able to re-contract is conservative, tending to exaggerate the risk involved.

Figure 11: Cost of 24-hour failure under the Pool and NETA
Source: Newbery 2004



4.1.2 Decentralised despatch increases system balancing costs

An argument that is often levelled against NETA is the inefficient decentralisation of despatch ("nominations") from a system-wide perspective. While under the Pool NGT, as System operator, centrally coordinated the despatch of the generators, under NETA generators are individually responsible to self despatch. Hunt (2002) argues that the reliance on self despatch actually increases total system balancing costs, because of the loss of system multiplexing, whereby some imbalances cancel each other, and because the system-wide forecasting accuracy is better than the one of individual generators. The fundamental issue is whether the burden of providing – and paying – for reserve capacity should be decentralised and be borne individually by the players of the market, or centrally optimised by the system operator, and paid collectively by the market players through the balancing mechanism costs. A consensual starting point is that the costs of balancing the system and providing for reserve should be minimised. However, this is not equivalent to NGT minimising the costs of the balancing mechanism, as part of the system security costs are borne by the generators through their individual hedging strategies.

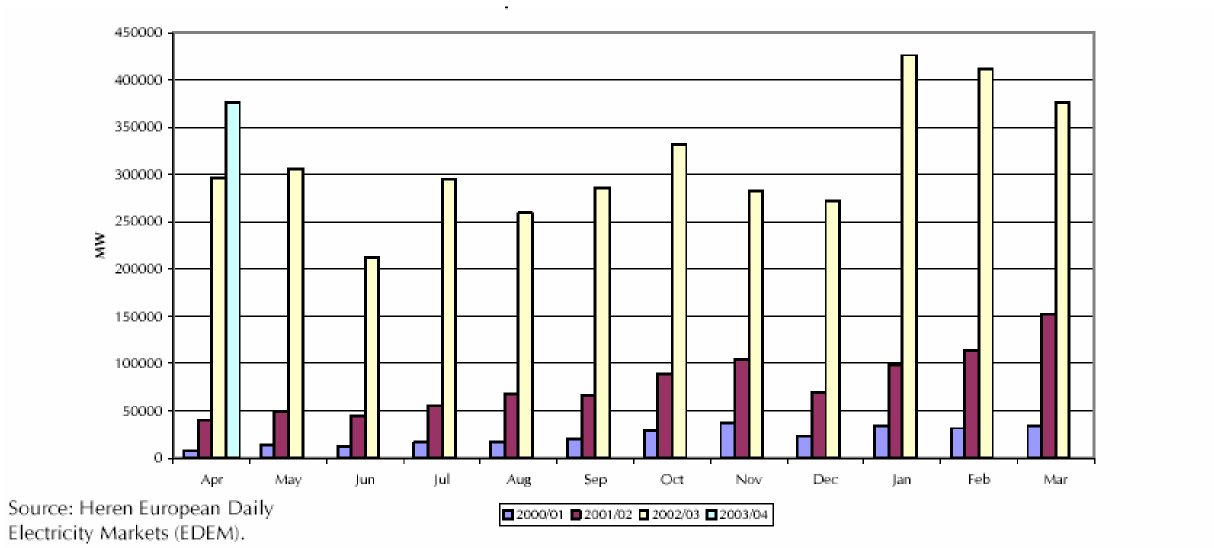
An optimal design of the market arrangements would thus minimise the sum of the costs of NGT balancing operations and the players' hedging costs. Let us examine two opposite hypothetical designs. At one extreme, encouraging the players to rely on self-insurance results in low balancing costs for NGT, as generators hedge in order not to face the imbalance charge, and because the costs of reserve capacity is depressed as all players tend to be long. But NGT's low balancing costs might be more than outweighed by the high private hedging costs of the generators and suppliers. In particular, the loss of the system multiplexing, which allows some positions to offset each other, as well the inefficiency of decentralised demand forecasting (individual demand forecasting accuracy is lower than system-wide demand forecasting accuracy) represent supplementary costs.

At the other extreme, reserve capacity is centrally despatched by the system operator, who procures it through a tendering procedure. This would result in higher balancing costs for NGT, but would encourage liquidity in the balancing mechanism, which would thus better reflect the true cost of reserves. By collectivising the risk and hedging, one can theoretically reduce the costs of hedging as the tendering procedure and central coordination result in a more cost efficient despatch of reserve. The ideal balancing mechanism design lies somewhere in-between these two extreme cases, and would optimise incentives for market players to balance their positions, and the collective efficiency gains of a centralised despatch. It is thus more likely to be based on a single cash-out price to take advantage of system multiplexing, with some reasonable penalty charge for being in imbalance if the system marginal price is considered not to give sufficient incentives to contract.

4.1.3 The lack of liquidity of the Balancing Mechanism

One of the arguments for implementing dual imbalance pricing was that making the balancing market a poor guide to the system balancing price would encourage contracting. Liquidity would increase through bilateral trading. As expected, ‘over-the-counter’ power trades have strongly increased under NETA, as demonstrated by *Figure 12*, which shows that the volume of the volume of the reported OTC power trades (‘over-the-counter’) on the UK Power Exchange (UK PX) have strongly increased during the second year of NETA (Brown, 2003).

Figure 12: Reported UK OTC Power Trades Volumes Evolution
Source: Ofgem 2003



The diversification of contract types confirms this positive evolution, as demonstrated by the following *table 1*. Several price reporters have entered the market since 2001 (i.e. Platts, Heren, Energy Argus and Anderson Spectron Power Index) and the number of registered contracts types has tripled.

Table 1: Diversity and volume of forward contracts

Source: Heren European Daily Electricity Markets (EDEM)

Financial Year	Number of different contract types	Number of reported trades
2000/2001	137	8,351
2001/2002	322	26,538
2002/2003	336	81,794

On the other hand, the lack liquidity of balancing mechanism is a real source of concern under NETA. One of the advantages of the compulsory participation into the old English Pool is that it created *de facto* liquidity. While the forward contracts under NETA appear to be quite liquid, the volume of NGT's balancing actions is comprised between 4 and 6% of the total demand out-turn. Henney (2002) argues that this lack of liquidity undermines the quality of the price signals of the balancing mechanism, which do not reflect the underlying fundamentals. Stoft (2003) argues that in a multi-market framework such as NETA, investment signals crucially depend on the ability of the price signals to feed in without distortion in the successive market layers characterised by different time scales. Under NETA this price feed-in mechanism between the balancing and contract markets is distorted by the lack of liquidity of the balancing mechanism, and the reliance on average imbalance pricing.

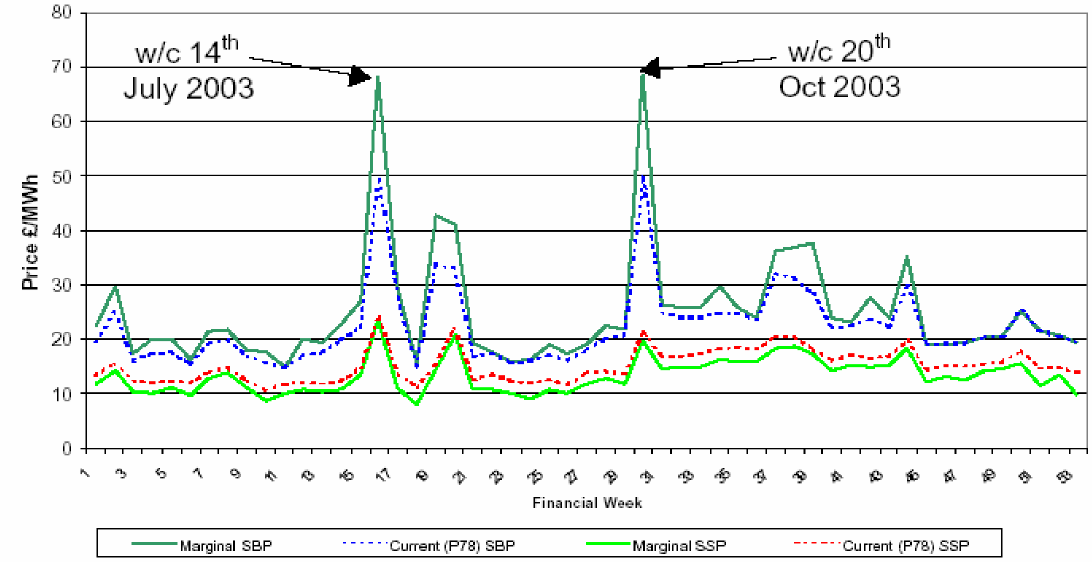
The events of 10 December 2002 illustrate these problems. On that day the system demand was the highest then recorded, and it exceeded the level forecast by National Grid. Whilst the price in the day-ahead market showed only a slight increase over the system peak to around £30/MWh, similarly to the average System Buy Price in the Balancing Mechanism which increased only to £71.6/MWh, the System Operator accepted offers in the Balancing Mechanism at £9,999/MWh for the marginal System Buy Price (Ball, 2003). The insulation of the energy market from the costs of short term balancing, which is both due to the average pricing formula and to the lack of liquidity of the balancing mechanism, is most significant at times of scarcity, but creates the risk that the market will fail to deliver appropriate price signals for long-term investments. A price that might warn of impending shortage may indeed not materialise until the market is under severe stress, and the delay in the price signals might undermine timely investment decisions.

4.1.4 Marginal pricing would provide better scarcity signals

The second source of distortion for the price feed-in between the Balancing Mechanism (BM) and the contract markets lies in the weighted average formula used to calculate imbalance charges. The BM mutes scarcity signals by paying generators their bid price and not the marginal price (in order to mitigate market power and possibly reduce volatility). Market fundamentals dictate that during times of shortage, electricity prices should rise to the marginal cost of generation required to meet demand. The current arrangements result in imbalance prices which can be significantly lower than the marginal energy balancing price, particularly at times of shortage. The main concern in this respect is that the energy price prevailing in the forward market will be artificially capped at the cost of being in imbalance, calculated as the *volume weighted average* price of providing balancing energy, rather than the price at the margin. This could lead to inappropriate contracting decisions as energy from plant with a marginal cost greater than this may not be purchased in the forward markets as it becomes economic not to buy sufficient energy in the forward market and face paying potentially lower imbalance prices. In the longer term this could lead to closure of generation with higher marginal costs, and in the absence of a specific mechanism to ensure sufficient capacity is available to meet demand, may lead to an erosion in the levels of security of supply enjoyed to date.

National Grid and Barclays Capital have proposed in respectively the P136 and P137 modifications to the Balancing and Settlement Code that imbalance prices should be calculated using a marginal methodology. These three Modification Proposals have been rejected by OFGEM (NGT, 2003b).[17] NGT argues that this could undermine security of supply, and that better signals would be provided by marginal pricing, which would in the longer term allow forward prices (driven by supply-demand fundamentals) to better reflect the value of capacity, promoting new build and long-term security of supply delivered by the market. One should realise however that the use of average rather than marginal prices gives equally incorrect imbalance signals, as the imbalance charge depends on the relative contracting position of the generator – short or long – as compared to the market global imbalance. But from a security of supply perspective, the crucial issue is that the signals are right at times of capacity shortage. Marginal pricing can be expected to deliver better reliability incentives in times of scarcity than average pricing as in times of scarcity the whole system can be expected to be short, and thus all generators being short in the same direction as the system would face the 'real cost' of their imbalance.

Figure 13: Comparison of Average and Marginal Pricing Methodologies, Financial Year 2003/04
Source: NGT 2004



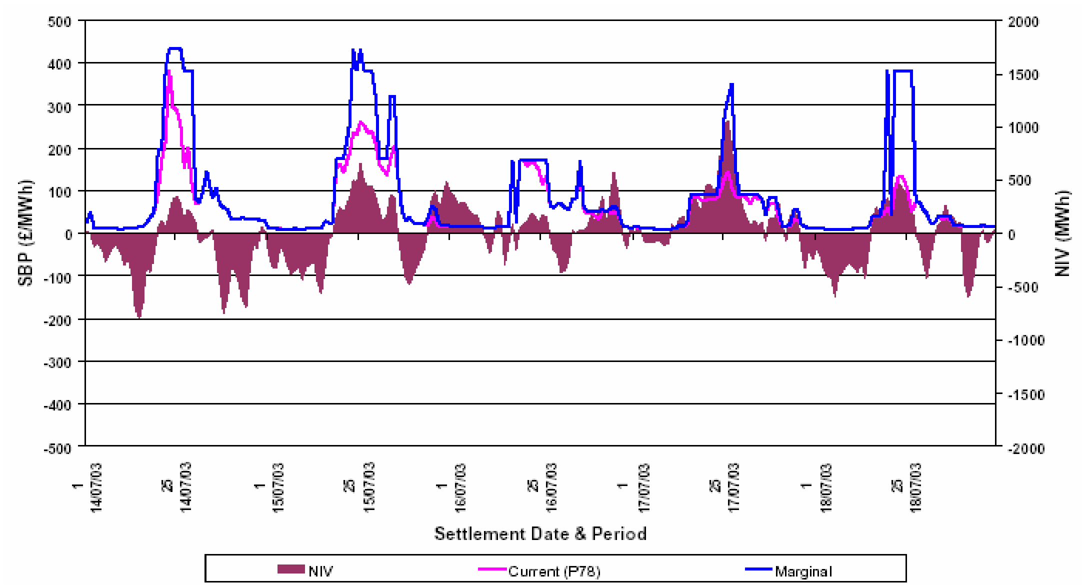
In response to the cashout review, NGT has performed a simulation to compare the imbalance price calculated using the current methodology with an imbalance price resulting from a marginal methodology for every settlement period in the financial year 2003/04 (NGT, 2004b).[18] *Figure 13*

[17] The current version of the Balancing and Settlement Code (the 'Code') can be found at http://www.elexon.co.uk/ta/bscrel_docs/bsc_code.html
[18] NGT analysis replicates the current tagging methodologies, but contains the following assumptions:

shows a comparison of prices calculated by the current and marginal methodologies, aggregated into weekly averages across the financial year 2003/04. This simulation shows that calculating 2003/04 imbalance prices using a marginal pricing regime as compared with the existing arrangements would have increased mean SBP by £2.78/MWh and decreased SSP by £1.80/MWh, thereby increasing the spread between the prices by £4.58/MWh on average. The standard deviation of SBP would have increased by £14.39/MWh indicating an increase in volatility, whilst the increase in standard deviation of SSP was much smaller at £1.23/MWh.

Figure 14 provides a detailed assessment of the week beginning on the 14th July 2003, which is one of the two weeks where the difference between the weekly average of SBP using a marginal methodology and SBP using the current methodology is greatest. The week commencing 14th July 2003 saw the system under a considerable amount of stress predominantly caused by high demands (primarily due to air conditioning load) and lack of plant capacity (low overall capacity margins and plant on outage). The contractual position of the market was short for many periods in this week. NGT believe that this is a example of when the current methodology understates SBP (NGT 2004b). The graph shows that the marginal methodology would have provided an SBP between twice and three times higher at times where the system was short of capacity (reflected by a negative National Imbalance Volume (NIV) on the Figure 14).

Figure 14: Comparison of current Average and simulated Marginal Pricing Methodologies, 14th - 18th July 2003
Source: NGT 2004



In rejecting BSC Modification Proposal, OFGEM expressed two specific concerns regarding the proposal (OFGEM, 2004, Decision letter in relation to BSC Modification Proposals P136 and P137, 30 March 2004). The first concern was that a full marginal main cash-out pricing methodology would increase the incentives on participants to game the market; the second concern was that marginal cash-out prices could create distortions because they could be set based on a very small volume of energy

· No Arbitrage actions were taken
· BPA has been added to marginal prices for the purposes of easier comparison with the current methodology

accepted by the SO. As the market appears today both on the generation and supply-side much more competitive as it was when the review of the Pool was launched, OFGEM's concerns about market gaming could be relaxed. Besides, NGT's simulation shows that marginal prices are not likely to be set by very small volumes of actions (NGT 2004b). OFGEM suggested using the volume weighted average of a pre-defined volume of actions (a marginal "chunk"). This could represent a way of ensuring that a marginal methodology does not derive a price that is unrepresentative of the price of energy at the margin due to a small acceptance volume. NGT's analysis suggests that there is very little difference between the frequency of prices generated under the current, marginal or chunky marginal methodologies. On the occasions when SBP is in excess of £80/MWh, the marginal "chunk" methodology has a dampening effect on the 'pure' marginal price (unsurprisingly, proportional to the size of the chunk).

What would be the impact of the move towards a marginal imbalance calculation for the two imbalance prices on bidding and contracting behaviour? There has been reluctance among the generators to change to marginal imbalance prices, as non-portfolio concerned that if they suffer a breakdown at times of plant shortage, being charged a marginally calculated imbalance price could be financially crippling. NGT concludes from its analysis that "marginal prices would not put undue risks on Market Participants" (NGT 2004b). However, one could expect that the £4.58/MWh average spread increase between SBP and SSP and the increased volatility would lead generators and suppliers to adjust their contract cover. This might worsen some of the above-mentioned problems of the current balancing mechanism, namely the lack of liquidity and the tendency of generators to over-contract to reduce their imbalance exposure.

To conclude, a marginal methodology for the calculation of imbalance prices would provide more appropriate signals to the forward markets than the current methodology, especially at times of scarcity. However, such a modification is likely to increase volatility and the spread between the SBP and SSP to levels comparable to those experienced at the beginning of NETA, prior the P78 modification, and would thus worsen the well-known problems of NETA, namely over-contracting and low liquidity of the balancing mechanism. It is therefore questionable whether the proposed change would have the desired positive outcome. Rather than going into this direction, OFGEM should consider implementing marginal pricing with a single-imbalance price. Given the change of market structure since the creation of NETA, the concerns about market-gaming that led OFGEM to advocate pay-as-bid dual imbalance pricing could be appropriately relaxed. Therefore, OFGEM should consider a more profound reform of the Balancing Mechanism as part of the second stage of the cashout review, which would come back to the sound economic principle of a single marginal imbalance price.

4.2 Demand-side participation should be improved

Demand flexibility and participation to the balancing mechanism are key to a secure and cost effective operation of the system in real time. One issue that the current England and Wales electricity system shares with all other electricity systems is that the balancing of supply and demand relies mainly on supply side flexibility, because of the lack of demand participation in the balancing mechanism. Greater demand response in electricity markets is needed to help ensure that electricity markets are always able to clear, *i.e.* by rationing electricity supply according to price rather than through brownouts or blackouts. A stronger demand response will help mitigate market power in electricity markets and provide potential investors with more predictable energy (and ancillary service) prices and therefore decrease investment risks.

The introduction of the new electricity trading arrangements was meant to tackle the demand side deficiencies of the Pool (Henney, 2002). OFGEM declared in July 1999 that "*increasing the role of the demand side in the new trading arrangements is seen as a major development and has always been a key objective".* The change from a centralised pool to a two-sided market was designed to increase the scope for demand to play a meaningful role in price setting. However there seems so far to be little improvement in demand side participation. Table 2 below shows little evolution since the introduction of NETA in demand side share of frequency response, and a slight increase for fast reserve and standing reserve provision .

Table 2 - Demand participation in reserve services provision (Source: NGT)

Service	2000/2001	2001/2002	2002/2003
Fast reserve	0%	5%	5%
Standing reserve	23%	29%	29%
Frequency response	29%	29%	28%

Cornwall (2002) notices that "although there is interest from the demand side in ancillary services there has been, as yet, scarcely any participation in the Balancing Mechanism from the demand side. On top of that commercial load management has virtually disappeared". Cornwall (2002) argues that this is largely due to the Balancing Mechanism participation requirements, which have raised practical and legal barriers to demand side participation. In 2002 a group was set up within the industry under OFGEM's direction, the Demand Side Working Group (DSWG), to look at the issue of demand side participation, and why the hoped for response had largely failed to materialise. This resulted in further enhancements in NGT's procurement strategy for balancing services, though there remain some important inequalities in the treatment of demand compared to generators.

One might question whether there is something pervasive in the perception of electricity that undermines a proper demand response. Power cuts are very infrequent and most customers cannot signal the value they place on reliability. This creates a bias in the perception of the commercial value of electricity. Electricity provision and reliability are today perceived as a right in Britain, which is reasonable as the value of providing the reliability is clearly high relative to current prices, but there is a lack of connection between this value and the price paid. Moreover, electricity is not purchased as such, but rather for the service it brings to people, whether it be lighting, heating, or playing music. As people actually buy a service and value each service differently, it is hard to associate a single value with the electricity required to deliver the associated service. But if access to electricity supply were to be regarded as a right, security of electricity supply need not be. Rochlin (2002) notes, however, that as long as security of supply continues to be treated as a right, "customers and politicians should know it involves significant stranded costs in both generation and transmission". Improved demand participation allowing for load-shedding in times of scarcity would be much less costly than a gold plated electricity system.

The low price elasticity of electricity demand, especially for small customers, is at least partly due to the inability to reward consumers for adjusting their consumption when prices are high. Only about half of total demand has real time metering equipment and most customers buy on contract. Patrick and Wolak (1997) showed in their econometric study of the inter-temporal substitution of electricity consumption that even a small increase in demand elasticity can have a big impact on power prices. Providing each individual consumer with half-hourly meters does not currently seem cost effective, although ENEL is investing heavily in south Italy to install individual consumer meters, arguing that the specific poor payments recovery rates of the area make it worth the investment. Borenstein (2004) concludes from a simulation exercise that the efficiency gains of real time pricing (RTP) for the largest customers are likely to far outweigh the costs, but that the incremental benefits of putting more customers on RTP are likely to decline as the share of demand on RTP grows, so that the gains from putting smaller customers on RTP might not justify the costs. However, it may be cost-effective to replace existing meters with RTP meters whenever they need replacement for ageing reasons, and then to offer such customers additional services that such meters make possible, much as installing water meters was voluntary in much of Britain and gradually increased penetration.

Even if the extreme solution of RTP meters proves too expensive, a mix of technological improvements and financial contracts for delivery insurance can potentially greatly improve the demand responsiveness. Doorman (2003) describes a capacity subscription scheme, based on the theoretical concepts of priority service and self-rationing, whereby demand is limited to a pre-subscribed capacity through a Load Limiting Fuse activated in times of scarcity by the System Operator. Such technological device would be much less expensive to install than individual meters. Another improvement would be the development by suppliers of standardised insurance contracts which would compensate for loss of load on the basis of the reliability level chosen and paid for by the customer. These contracts would simply be a systematisation of the current case by case compensation of power cuts (NGT, 2001).[19] Their time-varying and consumer-type depending attributes provide a basis on which to elaborate. Should insurance contracts for loss of load be implemented, the customer could chose to pay for a cheap or more expensive insurance contract on the basis of the reliability level s/he would like to be compensated for in case a shortage occurs. These insurance contracts would not require any technological change and would provide valuable signals on consumers' valuation of security of supply.

5 IMPROVMENTS TO HEDGING AND FINANCING

5.1 The current difficulties of financing power projects

The long-run nature of investments in the electricity industry makes them heavily dependent on the ability to raise capital or finance. Very few power stations have been commissioned since 2001, or are under development. This contrasts with the large number of projects that have been given building permission, but are not at present proceeding. Raising finance appears to be currently the major obstacle to building new plants currently. Bankers and lenders appear extremely reluctant to engage in

[19] Under the Guaranteed Standard on Supply Restoration, domestic customers are entitled to £50 compensation (business customers £100) if their electricity is not restored within 18 hours. Customers are entitled to an additional payment of £25 for each subsequent period of 12 hours that supply is not restored.

merchant power projects, and not only in Britain. De Luze (2003) explains bankers' current lack of confidence in the power sector around the world by the confluence of many seemingly unrelated events, including the California crisis, the fraud and bankruptcy of ENRON in late 2001, the questioning of the credibility of deregulation and of the new power sector business models that rely heavily on financial trading.

In the British market, additional factors have also played a role, particularly the collapse of power prices after 2000, the near total US withdrawal from the market, and the financial problems of TXU Europe and British Energy. Rating agencies have lowered their credit rating of most power companies in the past years, and Monnier (2003) from Fitch Rating recognises that "investment-grade rating is currently extremely difficult to achieve for power projects" and that 40% of market capacity controlled by operators in financial distress, either insolvent or in restructuring. OFGEM does not appear to be worried by potential bankruptcies. It says that "the capital market does not appear to be a barrier to the market delivering secure supplies" (OFGEM, 2004a). Bower (2003) has an opposite position and asserts that OFGEM and the Financing Services Authority (FSA) have undermined security of supply by failing to request generation and supply companies to have sufficient capital to withstand losses arising from financial and credit risk exposures. One can however expect that if a company goes into liquidation, the receivers will continue to operate and sell power, as happened with TXU.

De Luze (2003) believes that it will take time for the confidence in the power industry to rebuild. PriceWaterHouse Coopers (2002b) recognises that power companies need to question their practices and business model profoundly to regain investors' confidence. There is no question that there will be profitable investment opportunities in the future, but changes in corporate governance, improved risk management strategies, more integrity and transparency in financing and in financial statements will be required. A survey of the major world utilities from Market and Opinion Research International confirms that utility companies do not seem to be on the same wavelength as investors (PriceWaterHouse Coopers, 2002a). There is indeed a huge gap between the rating of the importance of key indicators by companies and the ratings given to these indicators by investors.

It will take some time before both investors and lenders understand the new allocation of investment risks in the electricity industry. One of the theoretical major benefits of liberalisation lies indeed in the redistribution of risk among the different stakeholders of the electricity industry. By spreading the risk over the whole chain, instead of having all the risk burden borne by end consumers, one gives all stakeholders in the investment process an incentive to consider carefully the best investment strategy. Ideally end consumers, retailers, investors, and even lenders should all bear some of the investment risk. Both the market design and the market structure contribute to define the risks that a generator will face when investing in a new power plant. One crucial issue lies in understanding how the market design and market structure interact as regard to long-term investment hedging and financing. In particular, generators and suppliers can hedge each other risks by either contracts or organic integration. For the regulator, there are important trade-offs to be made as regard to market concentration allowed and contract duration. A too stringent position might indeed lead to an appropriate allocation of risk among the different stakeholders of the investment.

5.2 The lack of forward contracting makes investment risky

5.2.1 Forward contracting under NETA

Contracts hedges between generators and consumers provide a basis for hedging the risks associated with developing new power plants. However, there appears to be relatively limited interest on the part of end-consumers to sign up for long-term contracts. Larger consumers are more likely to be interested in long-term arrangements in order to stabilize costs of inputs. In many markets, the existence of surplus capacity and relatively low price since liberalization have discouraged long-term contracts. Littlechild (2002) points out that large consumers continue to rely mainly on one to three-year contracts rather than contracts of longer term. One can expect that there will be little interest from end-consumers for entering into long-term contracts before they experience periods of high prices spikes.

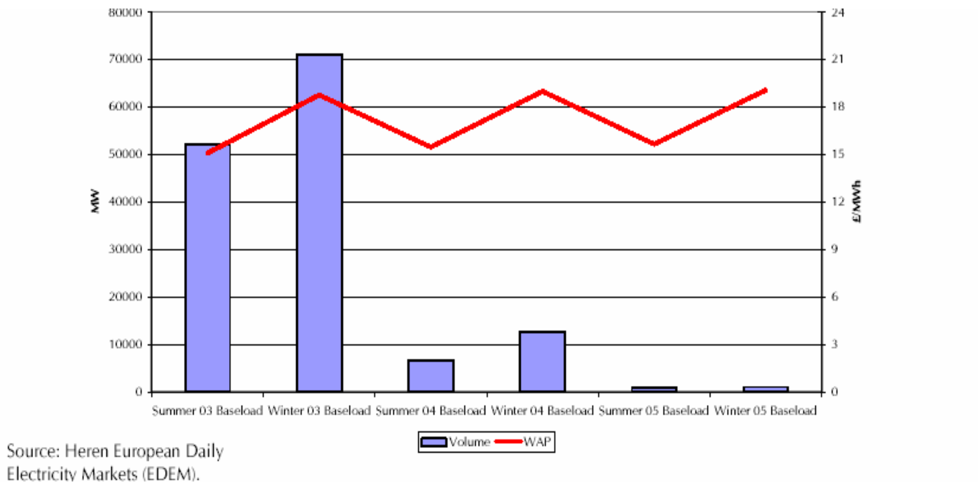
In most instances, however, generators do not sell their electricity directly to final consumers but to supply (or retail) companies which act as intermediaries. In this perspective, long-term contracting by retailers might be the missing link between the investors' desire for a long-term contract and some consumers' tendency to rely on short-term markets and spot prices. The possibility for long-term contracting with an electricity supplier with a large and relatively stable base of customers could be a way forward for investors searching for a long-term hedge. British supplier Centrica has in the past years entered into long-term agreements with Intergen for a UK gas-fired power plant in 2002, and with British Energy for a four-year contract in 2003 (British Energy, 2003). Somerset (2002) details the example of a new co-generation power plant in the Netherlands financed by a 15-year power purchase agreement between the project developer (Intergen) and the supplier (Nuon). The prices in the first five years of the agreement are fixed by the contract (in turn backed by a gas supply agreement), with prices and quantities in years 6 to 15 shared between market and fixed price agreement.

However, such long-term arrangements are to date unusual (at least in liberalised markets).[20] The British NETA market arrangements, as other electricity markets, is characterised by the lack of longer term forward contracting. The third JESS report points indeed out the short time horizon of market players by showing that the volumes of electricity traded in forwards contracts is very poor, and that the time horizon of traders is roughly limited to two years, as exemplified by *figure 15* (JESS, 2003b). One should be aware that this lack of liquidity of long-term contracts is a general feature of electricity markets. Woo et al. (2003) notice that trading of forward and futures contracts is thin in liberalized markets. MacKerron and Shuttleworth (2002) suggest that removing some of the industry restrictions on the use of long-term contracts could benefit investment. Such regulatory restrictions to the use of long-term contracts, such as the adoption of annual tariffs for transmission access and retail sales, were indeed motivated by preventing incumbent players from tying up their customer base and using "foreclosure" to stop other companies from competing with them. The incumbents' loss of market share makes such restrictions less necessary today, and possibly damaging to investment prospects. In this perspective, the US Federal Energy Regulatory Commission (FERC) had proposed that retailers in wholesale electricity markets in the United States be able to make arrangements in supply for up to three years in advance of real time. By including a forward contracting requirement in

[20] In many developing and transition countries, private generator investment invariably requires a long-term power purchase agreement, often with sovereign risk guarantees.

its proposals, the FERC hoped to encourage greater reliance on forward contracts rather than on the more volatile spot markets.

Figure 15: Seasonal forward Baseload volumes traded and weighted average prices
(Source: OFGEM)



The important issue is whether this lack of forward contracting more than a year ahead does have an adverse impact for investment in an industry in which construction lead times are much longer, usually between three (Combined Cycle Gas Turbine) and six years (Nuclear). De Vries and Hakvoort (2002) take the view that this represents an important market failure, because of the impossibility for investors to secure long-term returns on investments through long-term contracts. However, when looking at other commodity industries with a comparable lack of liquid longer term forward markets, such as oil for instance, one notices that investments are made on the basis of expectations and not thanks to binding forward contracts. Investors and traders can use forward curves representing expectations of the near-term future prices to guide their investment decisions, but have to consider longer-run market fundamentals to predict future prices and returns. Thus the lack of forward contracting as such should not represents an important impediment for long-term generation investment, provided that capitalist entrepreneurs develop an appropriate understanding of the risks and returns at stake. Indeed, a survey of the major world utilities from Market and Opinion Research International reveals that although utilities recognise that market short-termism is a problem, the majority of the utilities surveyed claim that it does not inhibit investment (Price WaterHouseCoopers, 2002a).

5.2.2 The role of forward markets in price discovery

If the lack of liquidity of forward electricity markets as such might not undermine long-term investment, it has several negative side effects. First, this lack of liquidity has an indirect adverse impact on the ability of market players to make correct forecasts of future electricity prices. This is due to the non applicability of the standard forecasting techniques used to build forward curves. Indeed, most of the approaches used in other commodity markets to build forward curves are based on arbitrage theories between spot and futures contracts, and not between spot and forward prices. These in turn are strongly influenced by the ability to store most commodities, which provides the necessary intertemporal price links.

Second, forwards and futures commodities markets are often used by risk managers to hedge their risk (particularly for storage), with liquid forward prices helping the price discovery mechanism to determine the fair value for future delivery. Therefore the illiquidity of forward markets in electricity might undermine the revelation of the price expectations of the different market players, and represents a major threat to market transparency. Third, prices from these markets are also the key inputs to many derivative-pricing models – energy derivative prices can be evaluated given the forward curve and the associated volatilities of the forward prices. Thus, the lack of liquidity of forward electricity prices might undermine the ability of market players to hedge against risks.

More fundamentally, electricity markets are incomplete markets and cannot provide sufficient hedging products to cover the long-term risks associated with an investment in power generation. A complete market would provide for a full set of forward and spot markets and risk-management tools, for each specific product/time/place. As already detailed before, forward and future markets for electricity are traded on a much shorter time horizon (a couple of years at best) than the period over which a power plant has to be amortised (several decades). Financial markets thus cannot provide hedging tools of sufficient length as compared to the investment cycle. It is indeed unlikely that there will ever be sufficient demand for a long-term electricity futures contract to ever become a standard financial product.

5.2.3 Do investors actually rely on forward prices?

The real issue underlying the lack of liquidity of forward electricity market lies in identifying to what extend forward market price signals are taken into account by investors. Do merchant generators consider forward electricity prices as accurate forecasts of future power prices? Many economic models use such an argument to demonstrate the importance of forward markets (Stoft, 2003). If one indeed assumes that electricity markets are perfect, then arbitrage justifies the argument that forward price are equal to the expected value of spot price at the time of the forward contract delivery.

However, one can doubt that inventors in the power business rely on forward prices to assess their investment opportunities. Indeed, forward price do not necessarily represent a market consensus on future supply and demand balances of the physical underlying gas or electricity. This is primarily because forward contracts are pieces of paper and cannot be directly arbitraged in advance with actual future physical deliveries, particularly as regard commodities such as electricity that cannot be stored. What the forward price does reflect is the supply and demand for forward contracts; which in turn can hinge on the supply need for hedges; which is a function of contracting patterns in the market including speculative positions, the relative risk aversion of market players and the perceived cost of risk, as well as a reflection of fundamental supply and demand issues as regards the physical commodity. Therefore, one needs to be cautious in interpreting forward prices. It would be a major and contentious theoretical leap to assume that forward prices, which can vary from place to place and from day to day, are the 'best forecasts of future spot prices'.

Investors are thus more likely to use their own estimates and forecasts of the future evolution of prices according to fundamentals. In this perspective, one might wonder to what extent investment decisions are taken independently of price signals. The extreme case in which investors rely on their own forecasts and business strategies to decide long-term investments and would take market signals into account only in a short to medium perspective for their hedging and operational strategies. An example is given by the de-mothballing of power plants during the autumn of 2003 in Britain in response to the increasing forward electricity prices for the winter 2003/04. Such disconnection between market signals and investment decisions would constitute a rational explanation of *herd behaviour* and hysteresis cycles in electricity market investments, as modelled by Ford (1999). Ford's model of investment dynamics in electricity markets reveals that the electricity market are likely to be characterised by boom and bust construction cycles, as in other capital-intensive industries with long building times and inelastic demand.

5.2.4 Does retail competition hamper long-term contracts?

A related issue is the impact of retail liberalization on the incentives for retailers to enter into long-term power purchase agreements with generators. It can indeed be expected that retailers will be less interested in signing long-term power purchase agreements if consumers can switch suppliers. Neuhoff and de Vries (2004) argue that retail liberalization undermines long-term contracting between generators and retailers, because of the free riding opportunities it creates. They argue that generating companies would only sign a limited volume of long-term contracts with retail companies in an environment of strong retail competition. The issue stems from the fact that retail companies may lose their customers to new retail companies in times when their long-term contract price exceed the short-term price. Indeed, in periods with average wholesale prices and retail prices above long-term contract prices, retail companies benefit and generators lose from their long-term contracts. In exchange, generators would expect to win from long-term contracts in periods with low wholesale prices. But in such periods, new retail companies may enter the market and offer cheap retail electricity. Thus retail companies with existing long-term contracts would incur losses, and some would eventually go bankrupt and would not honour their contracts.

Littlechild (2002) argues against this belief that retail competition precludes the signing of long-term contracts. He notices that "if the contract is really worth signing, the utility could match any price reductions to customers and still come out ahead. A consequence of retail competition is that suppliers who wish to sign long-term contracts have to back their own judgement rather than pass the risks to customers; this is likely to improve the quality of decision-making". Perhaps more to the point, building societies make fixed-interest mortgages available while short-run rates vary, exposing them to similar risks which are typically handled by demanding a pre-agreed exit or renegotiation fee for ending a contract that is out-of-the-money for the borrower.

Neuhoff and de Vries (2004) propose to solve the problem by institutional change, which would depart from retail competition and create a credible counterparty for generators to sign long-term contracts. If, for example, retail companies held regional monopolies, consumers would not have the option to switch. The most direct way to maintain generation adequacy would therefore be to retain the consumer franchise (Newbery, 2002). Littlechild (2002) notices however that a disadvantage of retail monopoly is that utilities and regulators, who do not have to test their judgements in the market, are typically not well placed to judge the costs and risks of long-term contracts. They can nevertheless force customers to enter such contracts and to bear the resulting costs and risks.

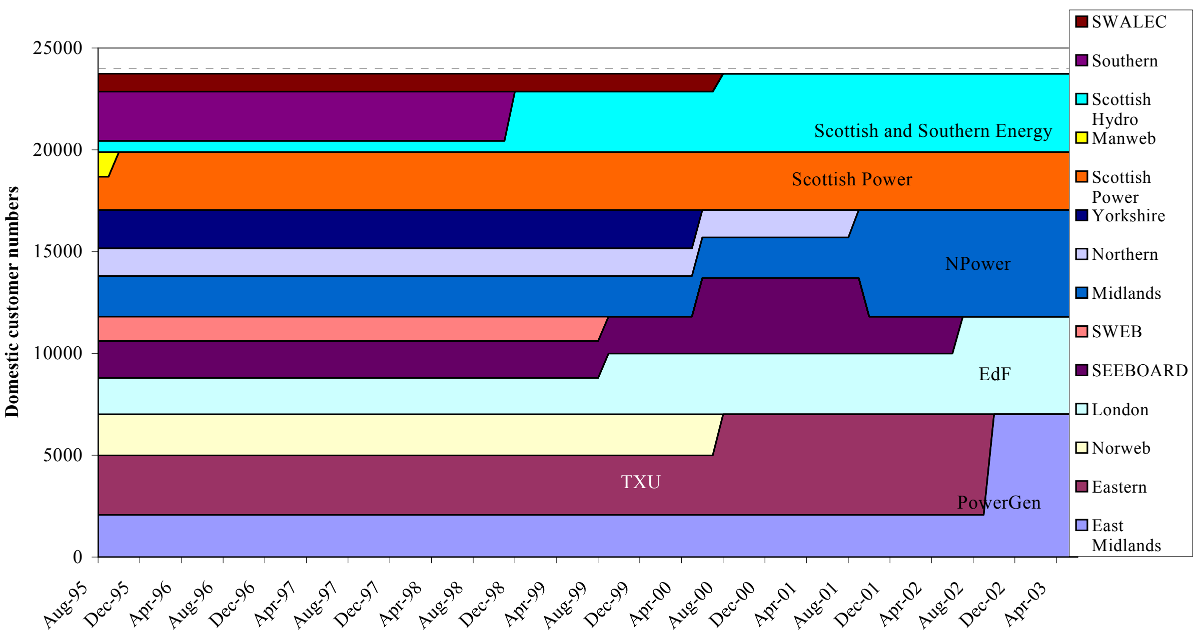
This debate over the detrimental impact of retail competition on long-term contracting needs to nuanced in several ways. First, retail supplier do actually sign long-term contracts as well as shorter term ones in Britain and other countries (Norway, Sweden, Germany, New-Zealand, and several States in the US), even if the volumes are small. In liberalised electricity markets, suppliers and generators hedging strategies are integrated with their speculation strategies: power company trading departments adjust their contract portfolios constantly, in order to both optimise the hedging position of the company and to take advantage of arbitrage opportunities. Long-term as well as short-term contracts are naturally part of the hedging portfolio of power companies, and are optimised together. Most long-term contracts are in turn hedged against and balanced by shorter positions. Thus, the fact that power companies no longer have a fixed consumer base is not the determinant factor for power companies to decide which proportion of long-term contracts to sign.

5.3 Alternatives to long-term contracts

5.3.1 Organizational hedges: vertical and horizontal re-integration

Vertical integration between suppliers and genrators constitutes an organisational hedge to improve the prospects of stable cash flows as a source of finance for larger, capital intensive investments. *Figure 16* shows the consolidation of British supply market over the last seven years; five generation companies now provide power to most of the domestic UK consumers.

Figure 16: Ownership of supply businesses in the UK
Source: Data supplied by J Bower



Horizontal integration between gas and electricity businesses is another hedging strategy for companies making a significant investment in gas-fired generation. The risk of volatile gas prices can be hedged against by acquiring companies with upstream natural gas assets in order to hedge fuel cost risks associated with gas-fired power generation. The last decade has seen several gas-electricity cross-sectoral mergers and acquisitions in the USA and in Europe. Meritet (2001) analyses the convergence of natural gas and electric power industries in the United States, and the consequences of mergers and acquisitions between electric power firms and natural gas (principally gas distributors). Gas-electricity convergence mergers may be attractive to investors for several reasons. One motivating factor is the ability of the merged company to hedge the risk between natural gas production and gas-fired power generation. Besides, the ability to generate electricity from gas, or stop generating and sell gas, if that is more profitable, helps the merged company manage the price volatility in natural gas and electricity, in the absence of long-term contracts or liquid forward markets. On top of these organisational hedging motivations, mergers can be an effective way of gaining economic efficiency by increasing the efficiencies of the firms. Higher equity financing is a way to cover more risky investments, but comes, however, at the price of higher financing costs. Indeed, the growth in power firms through mergers is, for this reason, not surprising. Yet, the emergence of the size and scope of these firms has also raised concerns about concentration of the industry.

5.3.2 Consolidation raises market power concerns

Liberalization of the European Union gas and electricity sectors has triggered an ample consolidation movement of utilities, which have so far taken full advantage of the slowly evolving regulatory environment. Mergers and acquisitions have led to the emergence, in Europe, of seven large power companies (EDF, E.On, RWE, ENEL, Vattenfall, Endesa, and Electrabel) that already hold a significant market share of existing power assets, and are expected to contribute a significant portion of new investment from internal sources. However, the prospect of utility assets being concentrated into the hands of a select group of companies may undermine EU plans for an open European energy market. Stronger regulatory intervention could have a profound effect upon the ambitions of the key utilities players. Waltenspuel (2003) compares the ongoing consolidation in the European electricity market to a poker game, where the "wildcards are the regulatory regimes under which the utilities operate and the risks of political intervention".

Mergers between electricity and gas companies are indeed under close scrutiny of the competition regulatory authorities, as they could have a detrimental impact on market competition in the electricity and gas markets. Vertical reintegration or horizontal concentration can create market power that can be abused to reduce competition. Convergence mergers of gas suppliers with electric utilities raise competitive concerns if it results in market power over the supply of fuel to a supplier. European and national Competition authorities intervened in several instances to prevent mergers and acquisitions that might lead to excessive market power. For instance, in Spain, government intervention on competition grounds put a halt to the merger of Union Fenosa and Hidrocantabrico, while the scale of remedies required for a tie-up between Endesa and Iberdrola caused the parties themselves to call it off. For the regulator, there are important trade-offs to be made as regard to horizontal and vertical integration and investment incentives. On the one hand, financing long-lived investments such as power plants in a liberalised industry may require companies having a critical size. On the other hand, regulators must be careful that these companies do not acquire too much market power. These conflicting objectives require coordination between the electricity market regulator and the competition authority and a fuller understanding of the nature of market power in electricity markets, that differ in important ways from normal product markets.

5.3.3 Developing innovative financing and hedging techniques

Electricity markets are still in their infant stage, and innovative financing and hedging techniques are expected to develop further. Other well-established commodity markets, such as oil markets, have developed numerous financial tools to hedge against risks spread over various time scales. As electricity market traders gain confidence, one can expect that comparable financial innovations will develop and greatly facilitate hedging (novel insurance products for risk sharing). Commonly used electricity derivatives traded in OTC markets include classic financial products such as forward price contracts, swaps, and options. Products more specific to the electricity industry are developing rapidly. For instance, s*park spreads* are cross-commodity options designed to minimize differences between the price of electricity sold by generators and the price of the fuels used to generate it. Innovative products that do not focus on price risk *per se* have had so far mixed success in the electricity industry. These include emissions trading, weather derivatives, and insurance contracts. In order to cover the risk from such low-probability events such as a plant outage, multiple-trigger derivatives and specialty insurance contracts are used to complement normal derivative products.

Besides these new hedging products, longer-term innovative financing contracts should also provide new ways of sharing risks among the different shareholders of the industry. The technique of *mezzanine financing*, which is a method of financing allowing companies to raise capital by increasing debt without giving up large equity positions, should for instance facilitate investment for independent power producers. Customised financing projects adapted to the risk/return profile of the stakeholder of the power plant investment are also a promising avenue. An example of such innovative financing is given by the planned development of a new nuclear plant in Finland by the firm TVO. The long business cycle of the paper industry has led these large electricity users to group in an organisation of consumers to bear all the investment risks of building a power plant, in exchange of a fixed-price long-term contract for electricity. In fact, TVO is a purchasing co-operative organising upstream integration as a joint venture. TVO owns two nuclear power plants already in operation, and sells electricity at cost to its investors in proportion to their contribution to the investment. With a guaranteed customer base who agrees to cover all costs, TVO expects to be able to finance future investment in a power plant at a very reasonable rate, making a nuclear plant its most cost-effective option (Tarjanne, 2000). This kind of alternative financing is likely to be well suited for the development of on-site distributed generation. Opportunities for economic combined heat and power (CHP) generation and a need for higher reliability should increase the attractiveness and feasibility of such distributed options.

6 REGULATORY ACTIONS TO SUPPLEMENT MARKET SIGNALS

6.1 Reducing regulatory uncertainty

Regulatory uncertainty is one of the major sources of risk for investors in a fast transforming regulatory framework (de Vries and Hakvoort, 2002). One major source of uncertainty potentially discouraging new investment is future Government policy to renewables and low-carbon generation options. The design of the ETS is still evolving and with it the incentives to invest in various types of generation. In the face of uncertainty about the future carbon price and the future incentives for particular forms of generation, that will be influenced by the exact rules on European emissions allocations for both old and new plant, the natural response will be to delay investment, and to prefer cheaper and quicker-build plant such as CCGT. One straightforward recommendation is for the government to adopt sustainable market rules, and carefully review the necessity of any change. At the political level, there is a major asymmetry in the consequences of forecasting errors. Therefore it is far from certain that any Government would be able to resist the political pressure to intervene during periods of high prices and enforced supply reduction. In Britain, many parliamentarians question the government reluctance to interfere with energy markets. They assert that because electricity is vital to the economy, the Government should not restrict its role simply to monitoring the electricity market (Trade and industry Committee, 2002).

The crucial issue is the definition of the *extreme circumstances* in which the government might have to intervene in the market. An example appeared during the summer of 2003 with the near bankruptcy of British Energy. The government participated in a £5 billion rescue plan for the embattled nuclear utility (Guardian, 2003). Indeed, in the event that British Energy had gone bankrupt *and* closed its power stations, the impact on capacity margins would have inevitably lead to power shortages in short term. Laughton and Watkiss (2003) showed that immediate closure would have reduced the capacity margin to around 10%, even assuming all interconnector capacity could be added to generation capacity. A worrying consequence of this Government intervention is that irrespective of whether government intervention was necessary, it has created a dangerous precedent: there is indeed a risk in the future that market players expect that Governments will come to the rescue in serious conditions, and thus under-invest in security themselves. However, it is implausible that any sensible system of administration of bankrupt power companies would cease plant operations, as this would foreclose a source of cash flow required by creditors. The main problems lie with bankrupt supply companies who would not be supplied with power and who would therefore have to cut off their customers (as happened in California).

There is a delicate trade-off to be found with the necessity of governance adaptability and flexibility in a relatively new electricity markets and the need for predictability. Indeed, as the market matures, adaptability and reactivity of the regulatory framework is critical for maintaining security of supply. The lack of flexibility of the governance arrangements was one of the criticisms levied at the Pool. NETA's governance structure seems to have enabled frequent modifications to be made to rules. During the first two years of NETA, 146 modifications have been done to the Balancing and Settlement Code, and 64 to the CUSC (Brown, 2003). Government agencies sometimes claim that investors should have anticipated changes in the rules and that therefore any financial problems reflect mistakes or inefficiencies of the investors. These agencies imply that regulatory risk can hedged against. However, Wright et al. (2003) demonstrate in a recent study that regulatory uncertainty is impossible to hedge against, and thus has an impact on the cost of capital. It is important the Government and the regulator recognise that their decisions to change market and regulatory institutions may create stranded costs, i.e. costs that a company incurred reasonably under the prior arrangement, but which it is not possible to recover under future arrangements.

Parliamentary sovereignty in Britain prevents any government from limiting the decision-making powers of future governments. Thus the required assurance must emerge from the accumulation of accepted customs and practices, backed up by law or statutory instrument where possible. Licences are the standard form of providing such reassurance, as these are commercially enforceable contracts that can only be changed by consent or due process, and go a considerable way towards providing regulatory stability. To further reduce regulatory uncertainty, MacKerron and Shuttleworth (2002) propose that British Governments recognise and apply at all time the *revenue standard*, which would prevent or compensate any shortfall of cost recovery resulting from the introduction of a new regulation. They suggest that some legally binding clarification of regulators' statutory duties with regard to the revenue standard might help to reinforce investment incentives.

6.2 Reintroducing a capacity mechanism?

6.2.1 Energy-only markets are risky and volatile

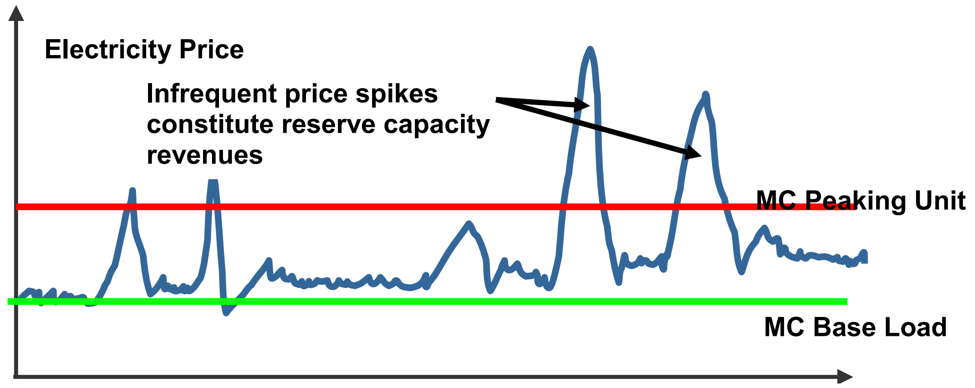
Energy only electricity markets have been adopted in the original (defunct) California design, in Nordpool, the Australian Victoria pool (although with an *ex post* VOLL) and in the British NETA design. In 'energy only' markets, there is no separate payment for energy and capacity; the primary income sources for recovery of capacity cost is the difference between the market clearing price, or the contract price in a pay-as-bid system such as NETA, and the generators' marginal costs. When ancillary services are procured separately by the system operator, generators can earn additional revenue by selling ancillary services, such as regulation and spinning reserve capacity.

The theory of spot pricing claims that electricity spot markets can provide efficient outcomes both in the short and in the long term (see Caramanis et al., 1982 and Caramanis, 1982). The spot prices should result in an efficient dispatch and allocation of available resources and should also signal the need for additional generating capacity. In a long-term equilibrium of energy only markets, the optimal capacity stock is such that scarcity payments to the marginal generators when demand exceed supply will exactly cover the capacity cost of these generators. A shortage of capacity will increase scarcity rents producing profits in excess of what is needed to cover the amortized capacity cost. Such profits will attract generation expansion. On the other hand, excess generation capacity will eliminate scarcity rents driving prices to marginal cost. When this occurs, generators on the margin will not be able to cover their investment cost. Unless such generators receive extra revenues through some form of capacity payments this will results in early retirement or mothballing of plants which will reduce capacity and drive prices back to their long term equilibrium level.

In practice, however, reliance on energy prices to cover capacity costs through scarcity rents raises many legitimate concerns. The non-storability of electricity, demand and supply uncertainty, inelastic demand and the steepness of the supply curve at its high end all contribute to high price volatility when reserve margins are low. In turn, this volatility might distort investment signals, by biasing investment against risky peak-load plants, and capital intensive technologies. Reserve capacity is provided by peaking units running only at times of high prices. As there is no separate payment for capacity, reserve generation is paid for by infrequent price spikes in periods of scarcity, as shown by *figure 17*. In theory, the scarcity rent that these peaking units are able to earn at times when generation is scarce should allow them to recover both their fixed and operating costs.

The reliance upon periodical price spikes to signal the need for peaking capacity has some significant weaknesses. Strbac and Kirschen (2000) that the uncertainty of the returns leads risk-averse investors to under-invest in peaking units. Neuhoff and de Vries (2004) model the impact of risk aversion on investor investment choices and consumer contracting decisions. They show that risk-averse investors who fund their generation investment with spot market returns will provide less generating capacity than risk-neutral investors. The logical solution is for generators to offer one-sided contracts for differences with a high strike price, for a fixed sum, This allows consumers to hedge against the price spikes (as their maximum payment is the strike price) while the revenue from the sale of these contracts provides the capacity payment needed to provide power in the peak periods.

Figure 17: Energy only markets rely on infrequent price spikes to



1.1

Further, Grimston (2004) claims that electricity markets discriminate against capital-intensive technologies such as nuclear and renewables. Indeed, while the various risks facing an investor in liberalised electricity markets affect all generation technologies, technologies which have a higher specific investment for capacity (nuclear, wind) might appear to be more exposed to electricity price risk. A firm relying on a capital-intensive technology may be competitive in terms of short run marginal cost, but will be more exposed to electricity price risk in the long term to cover the cost of capital employed. [21]

The most convincing argument against energy only markets lies in the difficulties for regulatory authorities to distinguish between the exercise of market power and legitimate scarcity rent (IEA, 2003). While some temporary high prices reflect legitimate economic signals that are needed in order to attract investment and reduce demand, they might be politically unacceptable if it is impossible to differentiate between legitimate scarcity rents and high prices resulting from market power abuse, or from strategies such as "*Hockey Stick*" bidding that exploit the inelastic demand and flawed market rules. In response, regulators frequently suppress high energy prices through interventions such as price caps and market mitigation, or by the choice of market design. This can create a revenue deficiency for the generator that may cause insufficient investment in generation capacity. Often the perceived threat of regulatory interference to curb scarcity rents is sufficient to inhibit capital

---

[21] Note that the mean profit will be higher the lower is the variable cost, and the variance of profits will be lower as profits will be bounded below by zero more often the higher is variable cost, so the risk is lower for running such technologies. The problem is that the capital at risk is higher.

formation and raise the capital cost for investment in generation capacity, leading to higher average prices but no greater security. Such interference is due to misperceptions and difficulties in distinguishing between market power abuse and legitimate scarcity rents. Thus, capacity payments or capacity obligations that stimulate capacity markets can be viewed as remedial measures needed to offset the suppression of energy prices and reassure the market in order to ensure generation adequacy.

6.2.2 Alternative market designs rely on capacity payments or obligations

Many regulators have been concerned that energy prices occurring in the various restructured systems are not sufficiently high to cover generators' capacity costs and to prompt adequate investment. Arriaga et. al. (2002) review the different capacity provision mechanisms available to regulators to encourage investments in generation. Capacity mechanisms pay generators in exchange for the undertaking to supply electricity if required. In one version of capacity mechanisms the regulator sets a price for capacity and lets the market determine the amount of capacity available. In the other version, the regulator sets the amount of capacity that has to be available and lets the market determine its price. These are known, respectively, as capacity payments and capacity requirements. Ensuring generation adequacy through capacity payments has been implemented in Britain under the Pool, Spain and several Latin American countries. The eastern pools in the US including PJM, NYPP, and New England ensure generation adequacy by imposing an installed capacity obligation (ICAP) on load serving entities (LSEs).

The concept of capacity payment is rooted in the theory of peak load pricing whose application in the context of electric power was pioneered by Boiteux (1949). According to this theory generation of electricity requires two factors of production, capacity and energy where the amount of energy that can be produced in any given time period is constrained by the available capacity. According to the basic theory, energy is priced at marginal cost and a capacity payment that would recover the fixed capacity cost is imposed on the peak-period energy users. Subsequent developments of peak load pricing theory focused on two important aspects of electricity supply: uncertainty and technology mix (Chao, 1983). The effect of uncertainty leads to redefining the basic ingredient of electricity service as energy and reliability where reliability is manifested by the LOLP calculation as a function of available capacity relative to load. The distinction between peak and off-peak then becomes a matter of degree. This perspective rationalizes levying a time-varying capacity charge on all consumption and its payment to available generation capacity whether or not dispatched on the ground that such capacity provides added reliability. The capacity payments employed in Britain under the Pool to augment energy prices and compensate available non-dispatched capacity were based on this perspective.

Under optimal capacity planning the marginal cost of incremental capacity equals the marginal cost of unserved load, which can be approximated by the marginal value of unserved load (VOLL) times the probability or fraction of time that load must be curtailed due to insufficient capacity. Hence, two alternative methods for capacity payment calculation (which are, in theory, equivalent under optimal capacity configuration) are to base the payment on the cost of peaking technology (e.g. a combustion turbine, CT) or to use the expected value of unserved load estimated by the product VOLL x LOLP.

The basic motivation for the ICAP requirements is similar to the argument in favour of capacity payments. The capacity markets prompted by the obligation provide generators with the opportunity to collect extra revenue for their unutilised reserve generation capacity and provide incentives for the building of reserves beyond the reserves that meet the short term needs for ancillary services. If one considers generation capacity as a separate product that is needed in order to provide reliable electricity service, then the supply of that product can be controlled either through prices in the form of capacity payments or through quantity control in the form of capacity obligation; then the case for quantity control can be supported by the classic prices vs. quantities argument. The basic argument is that the demand function for capacity is nearly vertical while the supply function is flat. Thus a small error in price will result in a large error in quantity so that direct quantity control is superior (Oren, 2003).

6.2.3 The drawbacks of the existing capacity mechanisms

The reliance of capacity mechanisms on engineering-based calculations has been one major source of criticism. Both the price (VOLL) used in capacity payments and the quantity (the suppliers capacity allocation) used in capacity obligations are administratively set by the system operator. In the case of capacity payments, Arriaga et al. (2002) point out that the VOLL chosen by the system operator has no market base, as exemplified by the VOLL used in the former English Pool (see section one). Chuang and Wu (2000) suggest that using VOLL figures based on demand side bidding would provide better VOLL estimates. However, the fundamental issue lies in the fact VOLL varies among consumer categories, and with the duration of the supply interruption as well as whether the interruption is scheduled and consumers forewarned or not. Therefore an efficient capacity payment design based on a realistic VOLL estimate would have to be time and consumer-type varying, which would lead to a very complex calculation. Choosing a second-best average VOLL is the most that is likely to be feasible, coupled with demand side bidding for large consumers.

Capacity obligations imposed on suppliers and ICAP markets suffer a similar problem: the allocation of capacity among the different suppliers is done *ex ante* by the system operator according to engineering simulations. However, many electricity dispatch constraints only manifest themselves in real time, and thus an *ex post* revision of the capacity obligation of each supplier would be required. Besides, Oren (2003) argues that one of the fundamental problems with capacity markets is their disconnectedness from energy markets. In the long run, the expected social cost of unserved energy as reflected by the energy-only market prices should equal the marginal cost of incremental capacity. However, the separate capacity markets created for trading reserve capacity requirement set through engineering based methods may produce prices that are not in equilibrium with the energy market prices. For instance, overestimating the expected cost of lost load would create artificially inflated demand for capacity and result in high capacity prices which in turn will lead to overcapacity that results in suppressed energy prices and socially inefficient production and consumption. Similarly, capacity payments based on such calculations would tend to suppress energy prices to or below marginal cost resulting in excess consumption and excess generation capacity.

Another problem that capacity payments and obligations share lies in the LOLP calculation. Graves et al. (1998) and Hirst et al. (2000) point to the fact that the LOLP calculations often employ simplistic models of probabilistic failure (e.g. Poison arrivals) and do not account for more complex phenomena such as the incentives of operators to keep plants running during peak price periods. Both the arbitrariness in the VOLL and the approximate nature of the LOLP calculation are likely to result in a mismatch between energy market prices and capacity values set directly or via a capacity market induced by capacity obligations. Furthermore, as British Pool reviewed in section one revealed, the predictability of calculated capacity payments can lead to gaming and manipulation of the payments in concentrated markets.

Most importantly, capacity payments or mechanisms that attempt to mitigate price volatility can be thought of as targeting the symptoms rather than the cause of electricity mechanism deficiencies. Fraser (2003) argues, that the lack of demand response in electricity markets is the heart of the problem, and that FERC's proposed Resource Adequacy Requirement (RAR) in the U.S. Standard Market Design (SMD) does not address the roots of the problem. Any regulatory intervention requires caution, since measures taken to ensure generation adequacy may have the effect of suppressing energy prices due to excess capacity or perverse incentives so that the necessity of such measures becomes self-perpetuating. This has clearly been the case in Argentina, where large capacity payments that were paid on the basis of generated energy induce generators to bid below marginal cost so as to increase production and capacity payment revenues.

Borenstein and Holland (2002) show that attempts to correct the level of investment through taxes or subsidies on electricity or capacity are unlikely to improve matters, because these interventions create new inefficiencies. They assert that a subsidy to capacity ownership financed by a tax on retail electricity is particularly problematic. Similarly, Shuttleworth (1997) recommends that regulators concentrate on creating an appropriate investment climate by eliminating all extraneous sources of risk, such as regulatory risk, and other obstacles to investment. If some form of transitional capacity obligation might be desirable in markets that are not already workably competitive, these mitigation measures should be designed with a strictly limited lifetime (a sunset clause), and/or to end under predetermined conditions (most importantly, once there is adequate demand response).

6.2.4 The British case

Should a capacity payment or obligation be reintroduced in Britain? Capacity mechanisms were used in the England & Wales Pool market until 2001, and section one studied the flaws of the Pool capacity payment design. The standard argument in favour of capacity payments is that regulatory intervention is needed to compensate for regulatory interference in the energy market. Provided that the UK government sticks to its commitment to "interfere the minimum with market forces", and that regulatory uncertainty is reduced in the eyes of investors and their bankers, this argument does not appear to make a strong case. A capacity mechanism would provide steadier revenues for generators and thus better investment incentives, but the important issue lies in how the costs of such mechanism do compare to its benefits.

The UK Government commissioned a study by NERA in 2002 to review the case for a capacity payment to be reintroduced (NERA, 2002). NERA estimated that introducing a capacity payment instrument could increase costs to consumers by some £151 million per year and therefore recommended the UK Government not to adopt such a strategy. On the basis of this study, the 2003 Energy White Paper concluded that "the case has not been made for such an instrument in the UK" (DTI, p86 at 6.43). NERA's calculation appears, however, to be highly debatable and greatly overestimates the cost of such a capacity instrument. First, the chosen methodology consisted of doing a basic calculation of the increase in customers' electricity bills if the system reserve is to be maintained artificially at higher levels than the assumed market reserve level. Such a methodology does not take into account the benefits to customers of maintaining a higher capacity margin, namely a lower Loss of Load Probability, i.e. less frequent power shortages. A serious cost-benefit analysis would have included the value consumers placed on better reliability (this could be done by multiplying their VOLL by the difference of LOLP in the two scenarios). Second, the assumptions on which NERA based its calculation greatly over-estimated the reserve difference in the two scenarios. If maintaining a 20% capacity margin is a realistic estimate of the level a system operator would choose, the assumed 8% market-determinated capacity margin appears very low compared to the capacity margins observed in the market. In 2003 British capacity margin was expected to be at an historical low point at 16.2 %, but market pressures subsequently increased it above 20%. Re-introducing a capacity payment instrument would therefore probably increase costs to consumers by much less than the £151 million NERA estimate, and could provide a comfortable 'safety-belt' to the government.

However, as far as investment incentives are concerned, the danger is to jump conclusions and confuse the source of the problem and its symptoms. The roots of the problem under NETA lie in the flawed balancing mechanism design, and the reforms should concentrate first on improving price signals in times of scarcity (see section three). If OFGEM sticks to the average imbalance pricing methodology, the price biases introduced in times of scarcity could lead to inadequate investment decisions, and a capacity mechanism will ultimately be required to maintain generation adequacy. However, reforming the Balancing Mechanism to come back to a single marginal imbalance price would provide better price signals, and the case for reintroducing a capacity payment in a well working 'energy-only' market then becomes a policy one.

Ultimately, the choice between a well working energy-only system and a market supplemented by capacity payments or obligations is a policy choice as regard to the level of volatility that policy makers are willing to see in the market. One could indeed consider the introduction of capacity payments as a proactive measure in the form of a mandatory hedge or insurance that will assure that prices stay within a socially acceptable range, which indeed bears a cost. Besser et. al. (2002) make such an argument to defend the requirement of a capacity obligation in FERC's SMD, by saying that the volatility of energy-only markets has a socially and politically unacceptable cost. Thus, while electricity markets may be delivering adequate levels of investment, price spikes are testing government commitment to allow markets to sort things out.

In Britain, however, there are two factors that should help give the government greater confidence to rely on market mechanisms to resolve any capacity crisis. First, there has been a long history of open electricity markets, which have led to a better utilisation of generating capacity and lower electricity prices over several years. The fact that customers have already enjoyed several years of benefits should increase confidence that the electricity markets do create benefits for end consumers. Second, both the industry and the regulator have gained considerable experience on the operation of electricity markets: the maturity of the market players, combined with relative flexibility of the NETA markets arrangements, which have proven to be much easier to amend than the Pool, should help going through sustained price spikes periods without the need for regulatory intervention.

7 CONCLUSIONS

The old vertically integrated franchise monopoly model under state ownership or cost-of-service regulation was normally able to finance any required capacity in generation or transmission to deliver defined security standards and to meet energy policy requirements of energy independence or fuel diversity. That model occasionally experienced financing difficulties if governments restrained final prices (although that was more of a problem in developing countries) and certainly provided poor incentives for delivering investment in a timely and cost-effective way.

In liberalised markets investment must be profit motivated, and under current EU Directives capacity choices are left to the market, except if there is a potential threat of shortage (at which point tenders can be accepted for new plant). The reliability of electricity supply has been the principal motivation for many technical and economic constraints imposed on market designs. There is as yet no clear academic consensus on which market design provides the least distorting long-term investment incentives. The Pool design had many advantages in this regard: its compulsory character created liquidity and a reference marginal system price to build hedging contracts on. Besides, capacity payments provided an additional source of revenue which facilitated entry by Independent Power Producers. However, the flawed calculation of the VOLL and LOLP and excessive market concentration lent itself to gaming by generators.

Three years after the controversial change NETA for a decentralised pay-as-bid energy-only market framework, there has been much concern recently that the drought of investment might endanger generation adequacy. It is worth noting that the ending of the Pool coincided with a massive decrease in generation concentration (Newbery, 2004), so that comparisons before and after risk conflating changes in market design and market structure. Theoretical rationale and practical experience suggest that energy-only markets with spot prices that are allowed to reflect scarcity rents will generate sufficient income to allow capacity cost recovery by generators. NETA went through its first stress-test during 2003 with success, as forward prices increased in response to a forecast capacity shortage for the winter, triggering the reconnection or mothballed plants. The clarification of NGT role in procuring reserve capacity showed that OGEM recognised the essential role of the system operator in energy only markets to ensure security of supply.

However, several market design changes appear necessary if NETA is to work well as an energy only market. The current decentralised procurement of reserve capacity increases the costs of system imbalance, and dual imbalance pricing combined with a weighted average calculation mutes scarcity signal that constitute peak generator revenues in an energy only market such as NETA. Calculating two marginal imbalance prices does not appear as a satisfactory solution, as it would raise the spread between the two prices and thus worsen the current tendency of market players to over-contract. We thus recommend that NETA adopts a single marginal imbalance price. Further, NETA has failed to improve demand participation in the balancing mechanism, especially for small consumers. If individual metering equipment does not appear cost effective, simple fuse-activated load reduction devices and financial quality differentiated contracts should be encouraged by the regulator.

Turning to the current hedging and financing difficulties of power projects, we show that the absence of liquid forward markets and corresponding contracts for more than a few years makes merchant investment in electricity more risky, and the wave of IPP projects financed by long-term

PPAs ended with the ending of the franchise in Britain. Ending the domestic franchise in 1999 removed the ability of the RECs to write long-term contracts with generators that could be passed through to the price-regulated consumers. Financing capital intensive investments will therefore require alternative forms of finance.

On the other hand, the trend to vertical integration in most markets, and horizontal integration on the Continent, suggests that larger companies are actively reducing risk exposures. Horizontal integration may confer sufficient market power to increase the price-cost margin to the point that investment becomes attractive, either to themselves if they wish to maintain market share, or to entrants attracted by the prospect of high prices sustained by market power. Size itself and the ability to operate in different markets also reduces the risk of individual investments. The regulator and competition authority should however carefully monitor what could be defined as a *competition versus generation adequacy* trade-off: a critical size is required for generators to be able to withstand the risks of investment in a liberalised market, but horizontal reintegration raise market power concerns. Vertical integration in the presence of market concentration also raises concerns and tends to further reduce market liquidity and raise entry barriers.

Different market designs, with separate payments for capacity or reserve obligations have the advantage of not relying on infrequent price spikes to remunerate reserve capacity (or not relying on the market to devise suitable risk-sharing contracts). Re-introducing a capacity payment could lower market volatility, but if poorly designed could entail substantial costs and negative side-effects. If a reform of the Balancing Mechanism to limit price signal distortions is not undertaken, and if the market is to remain as unconcentrated as seems desirable, such a mechanism to support investment will eventually be required to maintain generation adequacy. However, provided that the price distortions of the balancing mechanism are reduced, the demand response improved, and that the government is willing to withstand the political pressure during sustained periods of price spikes, such a change in market design may not be necessary.

BIBLIOGRAPHY

Abbott, M. (2001), "Is the Security of Supply a Public Good?", *The Electricity Journal*, August/September 2001.

Arriaga, I., Vasquez, C., and Rivier, M. (2002), "A market approach to long term security of supply", *IEEE Transactions on power systems*, Vol 17, no 2, p. 349-357, May 2002.

Ball, R. (2003), "A day in the Life… - An operators view of Tuesday 10th December 2002", National Grid Company.

BBC (2003), "Anger over power meltdown" (29/08/03), "Power cut causes chaos" (28/08/03), "Power Units to be restored" (16/10/2003) available on the website archive: http://news.bbc.co.uk/.

Besser, J., Farr, J. and Tierney, S. (2002) "The Political Economy of Long-Term Generation Adequacy: Why an ICAP Mechanism is Needed as Part of Standard Market Design", *The Electricity Journal* 15 (7): 53-62.

Boiteux, M.(1949), "La tarification des demandes en pointe: Application de la théorie de la vente au coût marginal", *Revue générale de l'électricité*, 1949.

Borenstein S. and Holland S. (2002), "Investment Efficiency in Competitive Electricity Markets With and Without Time-Varying Retail Prices", *Center for the Study of Energy Markets (CSEM) Working Paper 106*, University of California Energy Institute, November 2002.

Borenstein S. (2004), "The Long-Run Effects of Real-Time Electricity Pricing", *Center for the Study of Energy Markets Working paper 133*, University of California Energy Institute, June 2004.

Bower, J. (2003), "Why OFGEM?", Oxford Institute for Energy Studies, September 2003.

British Energy, 2003, "British Energy signs major supply deal with Centrica", British Energy Press Release, 7 February.

British Gas Trading, 2000, Paper to Pool Members, "Anomalous Pool Prices during 2000, Problems and Remedies", 24th October 2000.

Brown, S.(2003), "England & Wales wholesale market – 2 years on", available on OFGEM website.

Caramanis, M.C., Bohn, R.E. & Schweppe, F.C. (1982) "Optimal Spot Pricing: Practice and Theory", *IEEE, Transactions on Power Apparatus and Systems* PAS-101 (9): 3234-3245.

Caramanis, M.C. (1982) "Investment decisions and long-term planning under electricity spot pricing", *IEEE, Transactions on Power Apparatus and Systems* 101 (12): 4640-4648.

Chao, H. (1983), "Peak load pricing and capacity planning with demand and supply uncertainty", *Bell Journal of Economics*, Vol 14, (1983) pp. 179-190.

Chuang A. and Wu F. (2000), "Capacity payments and pricing of reliability in competitive generation markets", *Proceedings of the 33rd Hawaii International System Science Conference*, January 2000.

COAG (2002) Council of Australian Governments, "Towards a Truly National and Efficient Energy Market", (The Parer Report), Canberra, 20 December.

Cornwall, N. (2002), "NETA, always the bridesmaid, never the bride", *Power UK,* 100, June 2002

Cornwall, N. (n.d.), "The Need for Balance in NETA Imbalance Pricing", Cornwall Consulting website.

Cousins, J. (2003), Interim Energy Management Limited (IEML) – "Report to the JESS Working Group", 28 April 2003.

Department of Trade and Industry (2003) – Energy White Paper, "Our energy future: Creating a low carbon economy", February 2003.

Doorman G. (2003), "Capacity subscription and security of supply in deregulated electricity markets", *SINTEF Energy Research*, August 2003.

DTI, (2003b), Energy Markets Unit, "Proposals for a special administrator regime for energy network companies", Consultation document, 16 April 2003.

Electricity Act (1989), available upon request from HMSO and NGT website.

Elexon (2003), "Modification report for Modification Proposal P137", Prepared by ELEXON on behalf of the BSC Panel, 12 December 2003, Document reference: P137MR.

Elexon (2002), "Modification report, Modification proposal P78" – Revised Definition of System Buy Price and System Sell Price, prepared by ELEXON on behalf of the Balancing and Settlement Code Panel, 16 August 2002, Document Reference P078RR.

The Energy Review (2002) "A Performance and Innovation Unit Report" (PIU report) – February 2002 – Part 4: Security in the energy system.

von der Fehr, N. (2002), "Supply Security", Working Paper, *Inter-American Development Bank*, Washington, May 2002.

Ford, A. (1999), "Cycles in competitive electricity markets: a simulation study of the western United States", *Energy Policy* (27), 637-658.

Fraser, H. (2003), "A Critique of the Resource Adequacy Requirement Proposed in FERC's SMD NOPR", *The Electricity Journal*, April 2003, p.23.

Graves F., Hanser P., and Earl R. (1998), "The death and resurrection of electric capacity pricing", *Working Paper, The Brattle Group*, October 1998.

Green, R.J. (1997), "The political economy of the Pool", in M.D. Ilic, F. Galiana and L. Fink (eds.) Restructuring, power Systems Engineering and Economics, Norwell, Ma: Kluwer.

Green, R.J. (2004), "Did English generators play Cournot? Capacity withholding in the Electricity Pool", *CMI Working Paper 41*, March 2004.

Grimston M. (2004), Imperial College, "Generating Profits? Experience of liberalised electricity markets".

The Guardian (2003), "Outages hit British Energy's coffers", 04 November 2003.

Helm, D., *Energy, the State, and the Marke:, British Energy Policy since 1979*, Oxford University Press, 2004 .

Henney, A. (2002), "An independent review of NETA", 21 November 2002.

Hirst, E. and Hadley, S. (1999), "Generation Adequacy: Who decides?", *The Electricity Journal*, October 1999, vol 12, No 8.

Hirst E., Kirby B., and Hadley S. (1999), "Generation and transmission adequacy in a restructured US electricity industry", *Report to the EEI*, March 1999.

Hunt, S. (2002), "Making Competition Work in Electricity" (Wiley 2002).

International Energy Agency (2002) – "Security of Supply in Electricity Markets, Evidence and Policy issues".

JESS (2003a), Joint Energy Security of Supply Working Group DTI/OFGEM, Second report, February 2003.

JESS (2003b), Joint Energy Security of Supply Working Group DTI/OFGEM, Third report, November 2003.

Laughton, M., and Watkiss, P. (2003), "Narrow margins revisited, the impact of recent developments on reserve margins in the UK", Sept. 2003.

Littlechild, S. (2002) "Competition in retail electricity supply", *CMI Working paper 09*, 2002.

de Luze, G. (2003), "Investment in power generation: a banker's perspective", presentation at the 25th March 2003 IEA/NEA conference in Paris.

MacKerron, G. and Shuttleworth, G., (2002), NERA, A report for Powergen, "Guidance and Commitment: Persuading the private sector to meet the aims of energy policy" – 17 December 2002.

Meritet S. (2001) , "The Convergence Of Electric Power And Natural Gas Industries In The United States And In Europe", *CGEMP Research Paper,* Université Paris Dauphine.

Monnier, L. (2003), Fitch Rating, "The battle for Drax – Are UK generators back in power?", October 2003.

National Grid Transco (2004a), Grid Code consultation document, "Development of Maximum Generation Service", Consultation Ref F/04, Issue 1, 28 May 2004.

National Grid Transco (2004b), Response to OFGEM Consultation, Electricity and Gas Cashout Review, Executive Summary.

National Grid Transco (2003a), "Winter operation report 2003/04", 10 October 2003.

National Grid Transco (2003b), "Proposed revisions to the BSAD Methodology Statement required to facilitate Marginal Imbalance Pricing as proposed by BSC Modification Proposals P136 and P137", 25 November 2003.

National Grid Transco (2003c), Grid Code Consultation Document, Proposed Grid Code Modifications to Incorporate New Provisions Relating to the Flow of Information on Mothballed Plant and Alternative Fuels, 16 December 2003.

National Grid Transco, Seven Year Statement (2002) – Section on generation, available in the 'archive' section of NGT online library (http://www.nationalgrid.com/uk/library)

National Grid Transco (2001), Regulation 5 in Statutory Instrument 2001 No. 3265: 'The Electricity (Standards of Performance) Regulations'.

Neuhoff K. and de Vries L. (2004), "Insufficient Incentives for Investment in Generation", CMI Working Paper 42, March 2004.

NERA (2002), "Electricity Markets and Capacity Obligations", A report for the DTI, 13 December 2002.

Newbery, D. M. (1995), "Power Markets and Market Power", *Energy Journal* 16(3), p. 41-66

Newbery, D. M. (1998a), “The Regulator's Review of the English Electricity Pool”, *Utilities Policy*, 7(3), 129-41.

Newbery, D. M., (1998b) `Pool Reform and Competition in Electricity', ch.5, pp117-166 in M. Beesley (ed.) *Regulating Utilities: Understanding the Issues*, London: Institute of Economic Affairs.

Newbery, D.M. and Pollitt, M.G. (1997), `The Restructuring and Privatisation of the CEGB- Was it worth it', *Journal of Industrial Economics*, XLV (3), 269-303.

Newbery, D.M. and McDaniel, T. (2003) ‘Auctions and Trading In Energy Markets - An Economic Analysis’, Ch 10, pp195-234 in *CRI Regulatory Review2002/ 2003* P. Vass (ed.) (ISBN 1 85790 1142) Bath: Centre for the Study of Regulated Industries.

Newbery, D.M. (2004), “Electricity liberalisation in Britain: the quest for a satisfactory wholesale market design”, Energy Journal, to be published.

OFGEM (2004a), “Security of supply April to September 2003, Six month retrospective report”, February 2004.

OFGEM (2004b), “Decision document on modifications P136 and P137”, 30 March 2004.

OFGEM(2004c), NGT System Operator incentive scheme from April 2004, Proposals and statutory licence consultation Office of Gas and Electricity Markets, 3 February 2004.

OFGEM (2004d), Income adjusting event under National Grid Company’s 2003/04 System Operator incentive scheme, A decision document, July 2004.

OFGEM, (2003) “Asset Risk Management Survey: Composite Industry Report”, January 2003.

OFGEM (1999a), ‘The new electricity trading arrangements: Volume 1: Consultation Document’, July 1999.

OFGEM (1999b), ‘NGT System Operator incentives, Transmission Access and Losses under NETA: Consultation Document’, December 1999.

Oren S. (2003), "Ensuring Generation Adequacy in Competitive Electricity Markets", University of California Energy Institute's (UCEI) Energy Policy and Economics Working Paper Series, Energy Policy and Economics 007, June 2003.

Patrick, R. and Wolak, F. (1997), “Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices”, August 1997.

Price WaterHouseCoopers (2002a), “Competing for capital: Recognising utilities’ worth”, 2002, available at www.pwcglobal.com/energy/trading.

PriceWaterHouseCoopers (2002b), “Energy trading: Re-establishing sound foundations”, 2002, available at www.pwcglobal.com/energy/trading.

Rochlin C. (2002), “Is Electricity a right?” – *The Electricity journal, March* 2002, Vol. 15, no 2.

Shuttleworth, G. (1997) “Getting Markets to Clear”, *The Electricity Journal* 10 (3): 2.

Somerset, M. (2002), “Lessons Learned from an IPP”, *Converging Gas and Power Markets,* Amsterdam, December 2002.

Stoft S. (2002), “Power Systems Economics - Designing Markets for Electricity”, IEEE Press.

Stoft S. (2003), “The Demand for Operating Reserves: Key to Price Spikes and Investment,” *IEEE Transactions on Power Systems*, May 2003, Vol. 18 Number 2 - Special Section on Tools for Managing Restructured Energy Systems), January 9, 2003, www.stoft.com..

Strbac, G. and Kirschen, D. (2000), “Who should pay for reserve?” - *The Electricity journal*, October 2000, Vol. 13.

Stirling, A. (1994), “Diversity and Ignorance in electricity supply investment. Addressing the solution rather than the problem”, *Energy Policy*, March 1994.

Sweeting, A. (2001), “The effect of falling market concentration on prices, generator behaviour and productive efficiency in the England and Wales Electricity Market”, *MIT Working Paper*, May 2001.

Tarjanne R. and Rissanen S. (2000), “Nuclear power: least cost option for base-load electricity in Finland. The Uranium Institute 25th Annual Symposium, 30 August-1 September 2000, London. http://www.world-nuclear.org/sym/2000/pdfs/tarjanne.pdf.

Trade and Industry Committee (2002), *House of Commons, Fourteenth Report*, “Energy Supply: how secure are we?”, 12 February 2002.

Trade and Industry Committee (2003a), *House of Commons, Fifth Report*, “The British electricity trading and transmission arrangements”, 8 April 2003.

Utilities Act (2000), HMSO (Her Majesty’s Stationary Office) or OFGEM website.

de Vries, L. and Hakvoort, R. (2002), “The question of generation adequacy in liberalised electricity markets, *Delft University working paper.*

Waltenspuel C. (2003), High stakes in the European utilities poker game, Ernst and Young publication.

Wilson, J., (2000), “Scarcity, market power, and price caps in wholesale electric power markets”– *The Electricity Journa*l, November 2000, Vol. 13.

Woo, C., Lloyd, D. and Tishler, A., (2003). “Electricity market reform failures: UK, Norway, Alberta and California”. *Energy Policy* 31, 1103-1115.

Wright, S., Mason, R., and Miles, D., (2003), “A study into certain aspects of the cost of capital for regulated utilities in the UK”, on behalf of Smithers & Co Ltd., for the UK economic regulators and Office of Fair Trading, February 2003.