CEEPR **Center for Energy and Environmental Policy Research**

# The Wholesale Market for Electricity in England and Wales: Recent Developments and Future Reforms

by

Andrew Sweeting

00-007 WP September 2000

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

# The Wholesale Market for Electricity in England and Wales: Recent Developments and Future Reforms

Andrew Sweeting[1]

5 September 2000

COMMENTS WELCOME

***Abstract***

The England and Wales wholesale electricity market is about to undergo major reform (NETA). I describe and analyse the proposed arrangements, contrasting them with those currently in operation. I argue that while NETA will remove one or two of the Pool's problems, particularly by eliminating capacity payments, there is no reason to expect that it will significantly improve outcomes. Market power could continue to be a problem and, despite NETA's attempt to decentralise the market, the complex rules of the centralised phase operating close to real time are likely determine the level of wholesale electricity prices. Future arrangements for transmission are also considered. I argue that, if generators have local market power, these may exacerbate rather than reduce current problems.

[1] Department of Economics, Massachusetts Institute of Technology, Cambridge MA 02139. atsweet@mit.edu. I would like to thank a number of industry participants for providing helpful information. Paul Joskow provided me with funding and numerous helpful comments.

# 1. Introduction

The England and Wales (E&W) wholesale electricity market is about to undergo a radical redesign, called the New Electricity Trading Arrangements (NETA), which should go live on 21st November.[2] Although considerable effort has been spent on designing the new arrangements it is unclear how market outcomes will be improved. In particular, if market power is a problem under the current arrangements (the Pool), straightforward analysis suggests that this will remain the case under NETA even though trading and scheduling will become more decentralised.

This paper provides an integrated and fairly comprehensive analysis of the Pool and of NETA.[3] As with any electricity market the devil is very much in the details, so much space is devoted to describing the present and future arrangements. However, the analysis focuses on three main points. First, in the absence of market power, neither system would give perfectly efficient outcomes. Switching to decentralised trading does not guarantee efficiency because the final stage of trading will remain highly centralised with complicated rules. The prices that emerge at this stage are likely to determine prices in earlier markets, especially if there is market power. Second, NETA in itself provides no way of diminishing market power, as even OFGEM now seems to recognise. As at present the erosion of market power will depend on market structure becoming less concentrated and market rules becoming less complex and open to manipulation. Under NETA the rules will change, but may be even more complicated than under the Pool. Third, transmission pricing will not be integrated into NETA. This might be acceptable as transmission losses and congestion do not currently entail great costs in the UK, even though they are not priced efficiently. However, OFGEM's proposals for the future (decentralised physical rights trading) may increase rather than decrease inefficiencies associated with transmission. The failure to integrate transmission into NETA has precluded certain options that might have been cheaper and more efficient.

The paper is arranged as follows. Section 2 describes the current arrangements, while Section 3 analyses their actual and perceived problems. Sections 4 and 6 do the same for NETA. Section 5

---

[2] Financial Times, LEX Column, 14 August 2000.

[3] Two significant issues are largely ignored: governance arrangements and the incentives of the SO. Fortunately, the regulator (OFGEM, formerly OFFER) and market participants recognise that these issues are important.

provides a brief overview of the existing literature on NETA. Section 7 concludes. A glossary and list of abbreviations can be found at the back of the paper.

## 2. The Current Arrangements for the Wholesale of Electricity in England & Wales

In the Pool dispatch is centrally scheduled by the SO, based on bids made by generators and the SO's estimates of demand. These are also used to determine prices, although generators and electricity suppliers can hedge Pool price fluctuations with financial contracts. I first describe how physical supply is scheduled, and then explain how prices are determined for energy and transmission.[4]

### 2.1. Arrangements for the physical supply of electricity[5]

The centrepiece of the current arrangements is the Pool, through which nearly all electricity is traded.[6] Generators compete to supply electricity in a particular day by submitting bids for each genset by 10am the previous day. A bid consists of 5 elements: a start-up price (which is a price in £ for simply starting-up the unit), a no-load price (£ per hour, for keeping the unit warm regardless of the amount of electricity produced) and three incremental prices for power actually generated (£ per MWh) with corresponding "elbow points" which specify the ranges at which the different incremental prices apply. The stylised bid function for a particular unit is shown in Figure 1. Scottish Power, Scottish Hydro and Electricité de France (EdF) also bid to supply power over (or, much less frequently, to take power from) the interconnectors with Scotland and France. The maximum amount of power a genset can bid is its "Registered Capacity", although this can also be changed up to 10am the previous day. Although a generator can only submit one set of bid prices for each genset per day, he is able to vary his availability by half-hour, both before and after submitting bids, thereby allowing a generator with multiple plants some freedom to vary the shape of his aggregated bid function by half-hour. Generators also submit technical data concerning, *inter alia*, their ability to increase or decrease output over short timescales.

In order to schedule production and set prices for each half-hour, the SO, based on weather information and historic usage, uses a price-inelastic estimate of demand. A computer algorithm, SUPERGOAL, sorts the generators' bids into a "merit order" to produce an unconstrained

---

[4] The structure of the current arrangements is the same as that introduced in 1990/91 when the CEGB was first broken up.

[5] Description based on OFFER (1998a) and NGC (2000).

schedule to meet the forecast demand and reserve requirements at minimum cost. The schedule is unconstrained in the sense that no account is taken of probable transmission constraints, although consideration is given to technical parameters and intra-temporal optimization across half-hour periods.[7] Pool Purchase Prices (described below) and the expected operational dispatch schedule (now adjusted to take into account transmission constraints) are released by 4pm the day before real time. In the sense that the SO selects the generators who are to produce, the Pool is a "centrally dispatched system", a characteristic which will change under NETA.

Although the SO uses an inelastic demand forecast in the Pool, it is not quite true to say that the demand side is completely passive in the price-setting process. In 1993 the demand-side bidding scheme was introduced which allows a small number of large customers (now 39) to place bids to reduce their demand into the Pool in the same way as generators. For the purpose of determining prices, these bids are treated as additional generation i.e., they are included in the merit order. This is the only source of demand elasticity that can affect day-ahead pool prices. However, the demand reductions associated with the bids below the market price are not included in the *operational* schedule determined by the SO, which only includes actual generation. Of course, for physical balance it is the actual price responsiveness of total demand that matters (i.e. how does total demand react to higher pool prices), rather than the elasticity implied by these bids or the inelasticity assumed for the rest of demand that matters. As many more than 39 customers have contracts related to Pool prices, demand is likely to be more elastic than is suggested by the way the Pool operates.

On the day itself, changes in plant availability, outages and demand shocks require the SO to act to balance the schedule in real time. The SO can do this either by calling on extra production (or reduced production) from available generators who submitted bids into the Pool or by calling on various generators who have contracted with the SO to provide ancillary services.[8] The proportion met by contracts depends on the type of service concerned: for example, reactive power and black start capability are primarily provided via contracts, while reserve is primarily provided through the Pool. Transmission constraints are dealt with almost entirely through the

---

[6] Exempt generators are typically very small, and are connected to the distribution networks of the RECs rather than NGC's transmission network. For ease of description, this generation is ignored.

[7] For example, it may be cheaper not to switch plants off- and on- between periods, but instead to run several plants at levels significantly below full capacity during short periods of reduced demand.

Pool (less than £1m is spent per year on constraint contracts, in contrast to around £35m through the Pool).[9] Contracts are sold by the SO either through competitive tender, or by bilateral negotiation. Large customers also take part in reserve auctions, indicating their willingness to have their supply curtailed at short notice. In contrast to the demand-bidding scheme, these customers not only affect prices in the auctions, but are also operationally scheduled.

## 2.2. Financial arrangements for the wholesale of electricity

### 2.2.1. Payments to generators for energy[10]

The two components of the Pool Purchase Price (PPP), at which the Pool buys power from generators, are determined on the day before real time. The System Marginal Price is given by the bid of the marginal unit in the unconstrained schedule.[11] The other part of the PPP is a capacity payment for making capacity available. This payment equals LOLP * (VOLL - max (SMP, bid price)) where LOLP is the "loss of load probability" and VOLL is the "value of lost load". LOLP is intended to be an estimate of the probability that the system will fail due to lack of generating capacity and is based on the amount of capacity bid into the Pool relative to forecast system requirements.[12] VOLL is an administratively-set estimate of the loss final consumers would suffer from being cut-off.[13] PPP is the sum of SMP and the capacity payment, and this is the £ per MWh price received by generators included in the unconstrained schedule who actually supply. The capacity payment, termed in this case the Unscheduled Availability payment (USAV), is also received by generators who bid, but are not called to supply. This is supposed to act as an incentive for generators to make capacity available and to help to maintain a margin of spare capacity. As discussed in Section 3, there has been criticism of how the capacity payment

---

[8] Under the Pool, the SO does not contract for energy in advance, except through purchasing "ancillary services". As described below, this could change under NETA.

[9] OFGEM (2000c), Table 3.1, pg. 26

[10] Based on OFGEM (1998a), pgs. 9-12

[11] The SO combines the different elements of bids into a single price per genset. The methodology used varies on whether the half-hour in question is designated a Table A or a Table B period, with Table A periods being those with expected higher demand. Table A period prices include no-load and start-up prices, which are intended to represent generators' fixed costs. For more information see NGC (2000), Appendix E.

[12] In fact LOLP is based on the highest amount of capacity available from a genset in the previous seven days (OFGEM (1999b), pg. 26). This means, for example, that lack of availability for a few days due maintenance for example will not affect LOLP.

[13] The value of lost load in 1997/8 was £2599/MWh (OFFER (1998a), pg. 11).

is set and may be manipulated. Under NETA, there will be no distinct payment similar to the capacity payment.

Gensets in the unconstrained schedule but not included in the actual schedule (because of lower demand than expected or transmission constraints) receive PPP less their bid price.[14] Gensets that, in contrast, were not in the unconstrained schedule but are called to produce receive their bid price (plus the capacity payment). The following table summarises payments to generators.

**Table 1 : Payments to Generators (£/MWh)**[15]

| | | *Actual Schedule* | |
|---|---|---|---|
| | | **In** | **Out** |
| ***Unconstrained schedule*** | **In** | SMP + LOLP(VOLL-SMP) | SMP - BID + LOLP(VOLL-SMP) |
| | **Out** | BID + LOLP(VOLL-BID) | LOLP(VOLL-BID) |

*Source: Based on Table 1, OFGEM 1998, pg. 12*

A capacity payment, determined in the same way as for generators, is also made to those customers in the demand-side bidding scheme whose bids are above the SMP (and so do not affect the SMP). However there is no payment to participants whose bids are below the SMP. These participants benefit from receiving slightly lower electricity prices when their bids help to reduce SMP.

Finally, generators and customers also receive payment based on any ancillary contracts they hold with the SO.

[14] The rationale for this is that the bid price should represent the avoidable cost of generation. Therefore when the genset is asked not to generate, PPP less the bid price should represent the profit which would otherwise be forgone by not generating. Of course, if the bid price is greater than cost, the genset will be undercompensated for not being called.

[15] Gensets with only part of their capacity scheduled receive the appropriate payments for their scheduled and unscheduled capacity.

### 2.2.2. Charges to electricity supply companies and customers purchasing directly from the Pool for electricity[16]

Electricity is purchased from the Pool at the Pool Selling Price (PSP). This is composed of two elements : the PPP described above, and Pool Uplift which is only determined after real time. Subject to the SO's incentive scheme, Pool Uplift recovers the cost of USAV payments for plant not included in the unconstrained schedule and additional energy costs resulting from differences between forecast and actual demand, and planned availability and actual availability of generators (for example, if plant is not available NGC will have to purchase more power at higher bid prices).[17] The metered supply to customers is also scaled up on a uniform-geographical basis in order to reflect losses: this is done on an average, ex-post basis, so in the number of MWh paid for by the Pool will always exactly equal the number of MWh charged to those receiving power.[18]

### 2.2.3. Transmission Charges

An alternative to this system of charging for losses was approved by a majority of Pool members in 1995: both suppliers and generators would have been charged based on (scaled) zonal marginal transmission loss factors. However, this scheme was challenged by two Northern generators, and has been in judicial review ever since.

Transmission Services Use of System charges (TSUoS) are charged to electricity supply companies and others taking power from the transmission system. Again subject to a SO incentive scheme, these recover costs resulting from transmission constraints (i.e. the cost of additional payments to constrained-on and constrained-off plant) and ancillary services, including reactive power.[19]

Both generators and purchasers of wholesale electricity face connection charges and Transmission Network Use of System (TNUoS) charges. Connection charges recover the costs of providing the assets to connect the participant to the transmission network. However, apart from in the case of spurs to generating plants, these are calculated on a "shallow" basis so costs

[16] The word "customers" is used to refer to those purchasing power from the Pool, and so in effect refers only to electricity supply companies and certain large customers. The term "final consumers" is used to refer to end users.
[17] The SO's incentive scheme means that in a range around a target level of Uplift the SO keeps some of the gain when Uplift payments are reduced, and pays some of the costs when Uplift payments increase.
[18] OFFER(1998a), pg. 16. Losses are subject to a similar SO incentive scheme.
[19] OFFER (1998a), pg. 16

resulting from having to strengthen the network generally as a result of the connection are not included. TNUoS charges (£ per MW of registered capacity for generators and per MWh of peak demand for purchasers) are calculated on a zonal basis to recover the long-run costs associated with constructing and maintaining the transmission network.[20]

### 2.2.4. Financial contracts

Although all physical dispatch is arranged through the Pool, generators and suppliers use financial contracts to hedge against Pool price changes fluctuations. The most common instruments are Contracts for Difference (CfDs). A two-way CfD specifies a particular value of a Pool price (the SMP, PPP or the capacity charge) and a particular volume of electricity. When the relevant Pool price rises above the specified level the generator compensates the supplier by the difference in prices on the volume specified, and vice-versa. A one-way CfD is effectively an option contract where the supplier (typically) can seek compensation when the spot price rises above the specified level. In return the generator receives a fixed-fee. Since Vesting a large proportion of total output has been covered by CfDs of different lengths.[21] Initially contracts were drawn up by the government to make privatisation more attractive and to pass on the costs of the coal contracts which the generators were made to sign. Since the expiry of the initial contracts, the level of contract cover has fallen although it remains substantial (above 50%). New entrants into generation have signed very long-term CfD contracts with RECs in order to guarantee their revenues, so that they could raise debt finance.[22]

Another type of financial instrument is an Electricity Forward Agreement (EFA). These standardised two-way CfDs for small volumes of power are traded by brokers and allow parties to fine-tune their contractual positions. The volume of EfAs traded has increased substantially in the last few years, and particularly in the last few months, as shown by Figure 2.

[20] For more details see Green (1997b) and the discussion in Section 3 below.
[21] However the freedom of the large generators and RECs to sign CfDs was potentially limited by the regulator, which maintained the right to investigate contracts, especially those with margin sharing arrangements, above a certain volume. This reflected concerns about how vertical relationships might weaken the development of competition in generation and supply.
[22] For details of estimated CfD volumes see OFGEM (1998), pg., 36-37. New entrants into generation also signed long-term contracts with gas suppliers, so they were able to accurately predict their costs as well.

## 3. Analysis of the existing electricity trading arrangements

The current arrangements have three main problems. First, in the absence of market power, prices may tend to be too high due to capacity payments and the way demand is modeled for setting prices. One consequence may have been excess entry. Second, the current arrangements do not prevent market power from being exercised. There is evidence that market power has been and continues to be a problem. Third, transmission pricing is inefficient but the costs resulting from this are not necessarily huge in absolute terms. The discussion follows these three arguments.

Before detailing the imperfections of the Pool, it is worth noting that in many ways it has been reasonably successful. OFGEM (1999b) records a list of achievements including the fact that security and quality of supply have been maintained, entry by new generators has been facilitated and competition in supply has been introduced, even for residential customers.[23] Of course, some of these successes can be qualified: for example, excess entry could lead to inefficient levels of system security.

### 3.1. The current energy trading arrangements in the absence of market power

While market power is a potential major problem in the Pool, many imperfections in Pool pricing would exist without market power, even though the simple method for setting SMP appears to be close to optimal in this case.

Consider how SMP is set assuming price-taking generators. The Pool effectively operates a uniform first-price sealed bid auction, so as long as each generator attaches zero probability to being the price setting marginal generator, it will be a dominant strategy for each generator to bid his avoidable cost of generation. If this happens, the SO is able to use the true merit order in dispatch, so the cheapest generation will be scheduled. In addition, if the SO uses the best forecast of the actual demand curve, then SMP will also be set efficiently (as so far is possible a day-ahead) at the point where marginal valuation of power equals the avoidable cost of the marginal generating unit.[24] Setting SMP the day ahead should not affect the efficiency of

[23] OFGEM (1999b), pg. 22

[24] Of course, this will only be exactly true to the extent that the SO's method of combining the various elements of the bid price leads to an accurate measure of avoidable cost.

dispatch, but will essentially set the electricity price as it would be in a day-ahead forward contract before various uncertainties are resolved.

This logic still holds if generators sign CfDs linked to the SMP.[25] As each generator believes that he is unable to affect SMP, CfD income will be taken as independent of his bid, so the dominant bid strategy will be unaffected. The ability to sign CfDs may allow parties to efficiently share risk.

Next, consider capacity payments assuming that LOLP and VOLL are estimated accurately. In most markets, including those with relatively inelastic demand and a steeply sloped supply schedule, there is no clear justification for additional capacity payments to suppliers. Efficient peak-load pricing implies high energy prices when supply is tight, and these prices will allow plant to recover fixed costs and will attract efficient new entry. In contrast a capacity payment as calculated in the Pool will tend to spread these high prices over more periods, and, more worryingly, may lead to excess and inefficient provision of capacity with associated fixed costs. Of course, capacity payments may be justified on efficiency grounds if the method of setting energy prices used in the Pool does not capture the value consumers place on security of supply and, because of the centralised nature of energy trading (or free-rider problems) consumers are unable to contract with generators to supply an optimal level of capacity. However, in practice, with an inelastic estimate of demand used in setting the Pool price, the Pool's methods seem more likely to overestimate the value customers place on supply rather than underestimating it.[26]

A further implication of capacity payments is that generators' incentives to bid their avoidable costs may be slightly distorted. For example, in a non-peak demand period a generator with a high avoidable cost of generating may shave his bid, as, in the likely event that he is not scheduled to generate, his capacity payment will be based on LOLP*(VOLL – BID). Of course, it is an empirical question about whether the cost of such a distortion, financially or in terms of risking inefficient dispatch, is likely to be large.[27]

---

[25] Green (1998a) also discusses the benign role of CfDs for SMP and capacity payments on the incentives of atomistic generators.

[26] Of course, it may be politically attractive for the government to maintain excess security of supply and to spread price peaks. Current generators have obvious reasons to oppose reform of capacity payments.

[27] One piece of evidence that bid-shaving may happen comes from the demand-side bidding programme. Participants who rarely set SMP (and so might behave atomistically) but do receive capacity payments, have a record of failing to reduce their production even when PPP is above their bid prices. If their bids are

A general problem suggested by this discussion is that inefficiencies could result from using poor estimates of the elasticity of demand. In fact, this affects not only how SMP is set, but also the calculation of VOLL.

As described above, the demand estimate used in setting SMP is completely inelastic, with the only implied elasticity coming from the effect demand-side bids have by deplacing generation in the supply curve. However, even if formal participants in the scheme report accurately, price-responsiveness will be underestimated. Other business customers have contracts with supply companies that are based on Pool prices and so may reduce their demand in response to announced high day-ahead prices. For its own planning the SO uses a demand estimate which has greater elasticity and it would make sense for such estimates to be used in setting prices.[28] In addition, insofar as the current structure of the Pool, relative to less centralised arrangements, discourages customers from having contracts sensitive to wholesale price fluctuations, the Pool may lead to less efficient outcomes than alternative arrangements. However, it is not obvious *a priori* that this is the case.[29]

The calculation of VOLL should also reflect customers demand elasticities. However, for VOLL the relevant issue is the size of losses a customer may suffer from being completely disconnected without warning, rather than his ability to profitably manage his demand when he knows prices day-ahead. It seems likely that some customers value supply less than the current VOLL, as illustrated by customer participation in reserve auctions. On the other hand some customers may value it rather more than VOLL.[30] This observation leads to two points. First, it would make sense, where possible, to identify large customers who would be willing and able to be disconnected at prices below VOLL so security of supply could be allocated more efficiently. Second, as VOLL was first set rather arbitrarily, it may inaccurately reflect the average value of lost load in either direction with the consequence that availability decisions may be being made

---

at the level of their avoidable costs then they should want to reduce their demands. Such observed behaviour could be rationalised by bid-shaving.

[28] Doing so would require a change in the Pool rules which is not straightforward to achieve given current governance arrangements.

[29] For example, if customers are unwilling to sign contracts because the Pool price is perceived as volatile due to generator market power, an alternative system will only be more efficient if it limits market power.

[30] Newbery (1997a) provides an interesting discussion of many issues concerning the calculation of capacity payments, and I draw on his discussion.

inefficiently. It would make sense for the SO or the regulator to undertake a more accurate examination of consumers valuation of capacity if capacity payments were to be maintained.

Miscalculation of LOLP could also lead to inefficient availability decisions. LOLP should represent the probability that there is insufficient capacity available to meet demand, and it is based on the variance of expected demand, the difference between available capacity and forecast demand and the probability of declared available capacity failing. However, the methodology currently used takes no account of the reason why capacity is withdrawn, and in particular it therefore overestimates the probability that plant would not be available at the time of maximum system stress.[31] A further reason why LOLP may be overestimated is that because there is some actual elasticity of demand, high announced PPP prices on days when excess capacity is expected to be short deters actual demand. Clear evidence that LOLP is overestimated comes from the fact that using estimated values of LOLP the probability that the system would fail at some point in any given year would be essentially certain, whereas, in fact, the system has not failed once due to lack of generating capacity since Vesting.[32] One alternative approach would be to simply give generators the value of lost load as the electricity price on occasions when there is actually deficient generation (which would incentivise participants to forecast the probability of this event accurately). This would effectively lead to the fixed costs of peaking plant being recovered on the very few occasions that the lights actually go off.[33]

Finally, consider whether entry and exit decisions will be made efficiently in the absence of market power. As already remarked, "excess" capacity payments could lead to excess aggregate entry (and excess availability). However, even without this problem, entry and exit decisions may be biased by the details of the current arrangements. At present, baseload plant, which may be inflexible, receives the same payment for its electricity as plant that sets the SMP, which may

[31] Newbery (1997a), pg., 14. It is also worth noting that distribution network failures are far more common causes of power cuts than generation or transmission failures.

[32] Newbery (1997a), pg. 16-17

[33] Of course, it may be politically controversial for generators to receive high payments on exactly those occasions when the system fails. More generally and as indicated above, there seems to be a real issue as to whether it is desirable to smooth the price spikes that would naturally occur in competitive electricity markets on those occasions when supply is tight. Of course, in the presence of market power price spikes may represent something other than naturally tight supply margins and are likely to be even more controversial.

be more flexible.[34] However, as flexible plant can gain through signing ancillary service contracts and by providing response and reserve at bid prices close to real time any distortion is likely to be small. More of a problem may be created by plant which fails not bearing the system costs which result from more expensive plant having to be scheduled to replace it.[35] This could result in investment decisions being biased towards unreliable plant. I am unaware of an empirical examination of whether this has been a problem in practice.

### 3.2. The current energy trading arrangements and the exercise of market power

A major motivation for NETA is that the current arrangements are viewed as having been susceptible to market power. This is certainly the case in theory, and there is some evidence, particularly from the early-mid 1990s, that it has been a problem in practice. Unfortunately whether it remains a problem now – with NETA imminent – has not been substantiated.

The sources of the alleged ability to exercise market power are market concentration, especially in the price setting part of the supply curve, the general nature of electricity markets and the detail of the Pool's rules. Figure 3 shows the pattern of the HHI index for total generation capacity and SMP setting from Vesting to 1999/00. At Vesting, both generation and price setting were highly concentrated, reflecting the fact that generation (aside from the capacity constrained interconnectors) was basically a triopoly and that mid-merit plant which set the SMP was an effective duopoly of National Power and PowerGen.[36] New entry, plant closure by the incumbent generators, and forced divestiture has led to a steady decline in both HHIs.[37] In 1998/99 generation capacity concentration was around 1,500, although the HHI for SMP setting was around 3,000. This later figure is high relative to commonly cited standards such as the DoJ

[34] For example, nuclear plant and many independent CCGT operations are unable to provide much flexibility either for technical reasons or because they lack the personnel to provide swift response.
[35] OFGEM (1999b), pg. 200, presents the example of gas-fired plants declaring themselves unavailable close to real time in order to sell their gas into the gas market.
[36] Nuclear Electric, operating nuclear plants with low marginal costs and limited flexibility, provided baseload power. In the first five years of the Pool, National Power and PowerGen set SMP nearly 90% of the time (Newbery (1997a), pg. 3)
[37] In 1996, Eastern Group took 6,000 MW of National Power and PowerGen mid-merit plant on 99 year leases. However, they also agreed to pay the leasors £6/MWh generated with the obvious effect that Eastern's opportunity cost (avoidable cost) of generation increased. This would have the effect of increasing Eastern's bids even if it bid "competitively", i.e., at avoidable cost levels. It is therefore questionnable whether this divestiture could have been particularly effective.

Horizontal Merger Guidelines.[38] However, the latest round of divestitures have caused both HHIs to fall further, and the SMP setting HHI quite dramatically: in the first five months of the year 2000, the price setting HHI was merely 1,338 (similar to the generation capacity HHI).[39]

Two static theoretical approaches have dominated discussion of how market power can be exercised in the Pool: the Supply Function Equilibrium (SFE) approach of Green and Newbery (1992), building on Klemperer and Meyer (1989), and the first price, multi-unit auction approach of von der Fehr and Harbord (1993).[40] Papers using the SFE approach have players bidding continuously differentiable, upward sloping supply functions (implying increasing mark-ups on marginal cost, which may itself be increasing) in situations where demand has some elasticity but is uncertain.[41] In general, the SFE approach leads to a range of possible equilibria, although the set reduces when capacity constraints are introduced. The auction approach typically uses inelastic but uncertain demand[42], and each generator has constant marginal cost. However, the generator can bid different prices (although not smooth upward sloping schedules) for electricity from each of the gensets they operate.[43] Depending on the size of expected demand relative to capacity constraints, pure strategy or mixed strategy equilibria can result.

---

[38] Department of Justice (1997) defines markets as "moderately concentrated" if the HHI is between 1,000 and 1,800 and "highly concentrated" if the HHI is above 1,800. Both the overall generation and price-setting concentration figures are relevant to considering the exercise of market power. Even if only a small number of firms set the market price, their incentives to attempt to exercise it will depend on the amount of *inframarginal* generation they operate. Hence for a given level of price-setting HHI we would, *ceteris paribus*, expect a decline in the total generation HHI (associated with less ownership by the price setting firms) to lead to reduced incentives to raise prices.

[39] OFGEM (2000b), pg. 41

[40] As I discuss below, the same approaches are now being applied to considering the Balancing Mechanism part of NETA.

[41] As Green and Newbery (1992) note, the uncertainty of demand is analytically equivalent to demand varying during the day which the bids are made for.

[42] The inelastic demand assumption leads to the device of a maximum possible price being imposed. A simple cap on price has never been imposed in E&W, although there was an annual price cap from 1994 to 1996 and in 1999 an attempt was made to prevent spikes over £60/MWh occurring with the effect that the number of spikes between £50 and £60/MWh increased (OFGEM 2000b, pg. 42).

[43] However, some of the von der Fehr and Harbord analysis in fact uses an even simpler case where there are two generators, each of whom has only one genset (or one price), although the two gensets may have different capacities. In fact this takes us naturally to an Edgeworth duopoly world, with the exception that the marginal unit sets the market price. Although this appears to be a very important simplification both in theory and in practice (where National Power and PowerGen bid upward sloping, if not perfectly smooth supply functions), the authors claim that their results generalise to the more realistic case (von der Fehr and Harbord (1993), pg. 536).

While there are differences between the two frameworks (particularly in their use of continuous or discrete strategies), both approaches agree that paying all plant the SMP gives generators controlling multiple gensets incentives to raise the system marginal price by restricting output, due to the benefit received on infra-marginal plant. With fewer players and tighter capacity constraints, raising prices becomes both more profitable and more possible. As pointed out by Wolfram (1998) large players will also tend to have more incentives to raise prices than small players, who, like those operating inflexible baseload plant, will free-ride on the market power of those setting the marginal price.

In addition to static explanations of market power, the possibility that generators engage in tacit collusion is suggested by the repeated nature of electricity market and the level of information generators have about each other's technologies and strategies.[44] The empirical question of whether generators use strategies that are tacitly collusive has not been investigated in the academic empirical literature.

Subsequent papers have focussed in more detail on the exact Pool rules and how they interact with market power. For example, Wolak and Patrick (1997) argue that the Pool rules mean that parties will not necessarily bid above cost for each unit, but may still be able to exercise market power to raise the SMP. Remember that bids for each genset can only be changed daily, but that availability can be changed hourly. Wolak and Patrick suggest that the profit-maximising strategy may be bid each genset at close to avoidable cost, but then vary availability of gensets in order to raise the system marginal price in particular half-hour periods. While the marginal unit will not make excess profits, the marginal price will be higher than it would be without the restricted availability strategy and this will raise the profits on inframarginal units. The authors suggest that this strategy can be used to raise capacity payments and uplift payments, as well as SMP. It is worth noting however that because LOLP is calculated on the basis of maximum availability in the previous seven days, variations in availability over the day will not increase capacity payments. Similarly uplift payments are based on bids, so if bids are around the level of avoidable cost, increased uplift payments would not necessarily mean more profits for generators.

[44] Bidding information is published ex-post. In electricity markets there is reasonably good information on available technologies. This was re-inforced in the early 1990s by generation having previously been under common ownership.

Of course, this does not mean that generators do not take longer-term availability decisions to try to raise LOLP, or that they do not have other strategies for profitably increasing uplift.[45]

OFFER/OFGEM's reviews of Pool pricing have also claimed to find evidence of generators manipulating capacity payments and uplift as well as SMP. Monthly averages of SMP, PPP and PSP are shown in Figure 4. A feature of these reports is that the details of the Pool rules are important. For example, the first OFFER price report in December 1991 concerned the major generators declaring plant unavailable, to increase the capacity payment, and then declaring it available on the day so it received the higher payment. In following years, OFGEM noted a pattern where SMP and capacity payments moved in opposite directions. It claimed that this is the opposite of what would be expected in competitive markets.[46] OFGEM has also associated price spikes as reflecting generators' exploitation of the details of the algorithm used to draw up the unconstrained schedule. Generators place high bids for the last few MWs of a genset's capacity, with lower bids for the rest of capacity. The cheap bids will result in the genset being included in the unconstrained schedule (below full capacity) for most of the day. In order to avoid the costs of switching on more gensets at the daily peak demand, SUPERGOAL will continue to include these gensets at the daily peak but at their full output. As a result, SMP will be very high for one or two half-hours.[47]

Of course, proving that generator's actions are caused by attempts to raise prices rather than being based on a reasonable assessment of its own costs can be hard. For example, consider the recent case of OFGEM's concern about capacity withdrawal by several generators in April and May 2000 that caused capacity payments to increase twenty-fold.[48] To show that, absent any attempt to raise prices, it would have been profitable for the generators to declare the capacity available

[45] Newbery (1995) presents a simple model showing that a generator might decide to delay the return (for example, following maintenance) of capacity to the market in order to raise LOLP. Wolak and Patrick (1997), Table 4 provides some evidence from the mid-1990s that National Power and PowerGen had lower availability than operators of comparable plant in the UK or the US. This could indicate strategic availability decisions.

[46] OFGEM (1999b), Appendix 3. OFGEM's reasoning is that if SMP is low then this implies that there should be a large amount of more expensive generation that is not operating. The reserve margin should therefore by high. However, this argument is not entirely satisfactory, as low SMP may make it unattractive to make plant available at all, especially when at some point in the year maintenance has to be undertaken. This explanation of negatively related SMP and capacity payments is consistent with generators acting as price takers.

[47] OFGEM (1999b), Appendix 3, pg. 27 and OFGEM (2000b), pg. 14. OFGEM estimates that during the winter of 1998/99 these strategies increased electricity purchase costs by £90m..

requires a very detailed analysis of a generator's costs and the plausibility of its assessment of the market. Only in cases of gross manipulation will it be possible to achieve certainty: in those where profitability is marginal it will always be possible to argue either way.[49]

Certain implications follow if generators are exercising market power. The first is, obviously, that prices will be inefficiently high, although the degree of deadweight loss may be relatively small due to the low elasticity of final demand for electricity. The second is that production will be inefficient. This is particularly the case where firms are asymmetric as larger firms, with more incentive to raise prices, will bid higher prices for similar capacity. Wolfram (1998) finds some evidence to suggest that National Power, the largest of the generators created after deregulation, tended to bid higher prices for similar plant than PowerGen its smaller mid-merit rival in the period 1992-4. Third, high prices may encourage inefficient excess entry, in particular by plant that can be bid in at low prices and free-ride on the price increases resulting from the strategic actions of the larger generators. Green (1998a) concludes that the two "dashes for gas" in the 1990s, at a time of excess capacity in the industry, raised the total costs of generation (i.e., there was excess entry). Fourth, inefficiencies that would come from the details of market rules even without market power may become more important in the presence of market power. As discussed above, capacity payments may lead to inefficiently high prices even without market power. When they can be manipulated, inefficiencies will tend to be increased.

Unsurprisingly the high level of Pool prices has been a consistent objection of OFGEM to the current arrangements, and one of its main reasons for pressing ahead with NETA.[50] In particular, OFGEM has argued that Pool prices have not followed the downward path of costs, resulting in increased profits for generators, rather than lower prices for consumers, despite falling concentration.[51] The basis for OFGEM's claims is illustrated by Figures 5-10, taken directly from OFGEM (2000a).[52] These show how average prices have stayed fairly steady over time, while fuel prices and capital costs of CCGT entry have fallen and plant efficiency has increased. The

---

[48] OFGEM (2000a), pg. 9.

[49] OFGEM decided not to pursue the case in question when Edison Mission Energy agreed to return capacity to the market in October 2000.

[50] OFGEM's assessment is that the introduction of NETA could lead to prices falling by 13% (Financial Times, NEWS DIGEST, 4 August 1999).

[51] OFGEM (1999b), pg. 171

[52] Estimates of CCGT entry costs and plant efficiency will to some extent necessarily be based on estimates, so it is possible that slightly different results would be produced by alternative assumptions.

margin of capacity over demand has remained reasonably high. The profits, measured by the rate of return on capital, of the two major generators have remained above 1992 levels in spite of their sharply declining market share.[53] Green (1998) supports that the suggestion that any decline in electricity prices since Vesting has reflected falling fuel costs, rather than declining margins for generators. Of course, while this evidence is interesting, it does not directly address the issue of how large is the mark-up above the marginal cost of production which is central to the claim of market power indicated by the HHIs above.

Wolfram (1999b) does examine the size of the marginal mark-up, by using both direct and indirect methods to estimate the marginal cost of the unit setting the SMP. Wolfram estimates that market power is only one-fifth of the level that would be expected from a theoretical SFE model. Of course, Wolfram's direct estimates will tend to underestimate market power if generators in fact use the kind of strategies suggested by Wolak and Patrick (1997) where units are bid in at the avoidable cost but availability is manipulated.[54] The finding that in the early 1990s less market power was exercised than might have been expected has been explained in three different ways. It is important to think about these, as, if they constrain prices under the Pool, they may continue to constrain prices under NETA. Unfortunately, there is no analysis similar to that of Wolfram looking at what has happened to electricity prices in the late 1990s, and examining whether they have approached theoretical levels given the less concentrated market structure.

The first explanation argues that CfD contracts have constrained market power. The greater the proportion of a generator's energy covered by CfDs, the weaker are the incentives to raise price, as for more inframarginal units the price received by the generator is fixed. When a generator is over-contracted, he becomes a net *purchaser* of power in the Pool and he should rationally want to reduce the Pool price. Of course, this effect on incentives in the Pool weakens generators incentives to sign CfD contracts, but as Allaz and Vila (1993) show in a Cournot framework the

[53] In 1990 NP and PG had 47% and 30% of generation capacity (including the interconnectors) respectively (OFFER (1998a), pg. 43). In 2000/01 the figures will be 12.5% and 14.9% respectively (NGC (2000), Table 3.3, pg. 3.25-3.27).

[54] In this case one would want to estimate the marginal cost of the unit which would be available if availability is not manipulated. This is obviously much harder to estimate and requires making a set of assumptions. The Wolak and Patrick strategy also suggests that market power may only be exercised at particular moments of the day, in which case one would ideally want to separate out the level of market power found at high daily demand and the level at low demand, as well as looking for the average level.

presence of multiple forward markets prior to contracting can result in market power being completely eroded.[55] Empirical evidence on the effects of contracts are mixed. Helm and Powell (1992) and Armstrong, Cowan and Vickers (1994) associate Pool price increases in 1991 and 1993 with the expiry of some of the initial contracts.[56] OFGEM has also claimed that in July 1999 a lower level of contract cover by National Power and PowerGen was associated with a profitable increase in Pool prices not reflecting increased costs.[57] However, Wolfram's (1999b) analysis of mark-ups in the period 1992-94 finds only a 2% change in mark-up in response to a 20% fall in contract coverage, raising doubts about the size of any contract effect.

The second argument is that the threat of entry has acted to cap prices. David Newbery has suggested in a series of papers in the late 1990s that entry is effectively contestable, and that time-average prices, as a result, are effectively capped at a level to discourage, or at least slow, entry of new baseload plant.[58] In face of a time average cap, generators will tend to raise prices more in periods of high demand, assuming that entry of peaking plant is unlikely. To commit themselves to bidding aggressively, generators may sign CfDs that erode their market power. A very important implication of this argument is that changing the rules by which the market works (e.g., changing the way capacity payments are calculated or the way bids are made) is only likely to have a very limited effect on average prices.

However, there are at least three problems with the contestability story. First, even though it is possible for entrants to agree contracts and funding quickly, they cannot actually enter and start producing electricity for at least two years. Therefore the alternative story where the incumbent generators maintain high prices in the short-run and accept that entry will erode market share in the medium-term is just as plausible *a priori* as the story where the entry threat keeps prices low. Second, the empirical evidence that prices are around the entry costs of new plant comes from

[55] Green (1996) and Newbery (1997b) contain discussion and models of these effects.
[56] OFFER (1992) found that average Pool revenues at the very start of the Pool were probably below average avoidable costs, in part reflecting the fact that generators revenues were almost entirely given by their contracts.
[57] OFGEM (2000b), pg. 17.
[58] This reflects a significant change from Green and Newbery (1992) where it was argued that "(t)he electricity pool is certainly not a contestable market" (pg. 947). This partly reflects experience of entry in the 1990s where new entrants were able to sign long-term contracts for the gas input and electricity output (via CfDs) giving them certainty over their costs and revenues, allowing highly geared capital structures. Contractual obligations (take-or-pay, sometimes with resale restrictions) and technical limitations on plant meant that they were effectively committed to baseload strategies.

simple calculations from the early 1990s.[59] There is no good evidence that Pool prices have fallen as entry costs have fallen. Figures 5-8 would, in fact, suggest that Pool prices have not followed entry costs. There is also no indication that the policy of restricting consents for new gas-fired plant, adopted by the government, in 1997 was associated with increased prices, although this could be explained by the number of consents already given. Third, if entry deterrence was the reason for keeping prices low, then it does not appear to have been a successful policy![60] From 1990/91 to 2000/01 21GW of CCGT capacity was introduced.[61]

The third argument is that prices have been capped by actual or threatened regulatory intervention. This would lead to a cap on prices similar to that caused by the threat of entry, although with less incentive to distort the balance between time-average and demand-weighted prices assuming that the regulator looks at both.[62] Between February 1994 and February 1996 National Power and PowerGen accepted an explicit cap on both time-average and demand-weighted average prices.[63] As Newbery (1997a) records in each year they were able to come within 1% of both price caps, which is suggestive that for these years at least the regulatory constraint was binding.[64] Wolfram (1999b) also finds some evidence that prices were responsive to regulatory investigations, although (again) we do not have formal evidence that the regulatory constraint has been binding since 1996.

One important issue which has led to regulatory scrutiny has been moves towards vertical integration between generators and major supply companies, including the RECs, which control distribution networks and run large electricity supply businesses. From standard theory, vertical integration, by ownership or by contract, is only a problem when there is market power at one level of the industry and scope for imperfect competition at the other. In the E&W electricity

---

[59] Wolfram (1999b), pg. 815
[60] Of course, it might be that entry would have been even greater if prices had been higher. However, the level of entry is large enough to make it seem intuitively unlikely that entry deterrence would have been more profitable than exploiting market power before new entrants could start to operate.
[61] NGC (2000), pg. 3.11. Total UK capacity in 2000/01 is just under 70 GW.
[62] However, regulatory scrutiny would of course lead generators to favour techniques for raising prices which are harder to detect. Wolak and Patrick (1997) argue that this is another advantage of their availability strategy. However this does not necessarily seem compelling when it is considered that the regulator has access to availability declarations and it seems unlikely that systematic withdrawal of capacity for a few half-hours could be explained for technical reasons given the advantages of keeping plant running.
[63] OFFER (1998a), pg. 26

industry there is an incentive for vertical integration in order to hedge the Pool price while avoiding the transaction costs of contracts. However, as well as market power in generation it has also been suggested that market power is a problem in supply, especially to small and residential customers who are less likely to change suppliers. While initial attempts at integration (for example in 1995, National Power and PowerGen tried to merge with Southern Electric and Midlands Electricity respectively) were blocked by the government, vertical integration has subsequently occurred with several generators (including PowerGen, EdF and the Scottish generators) taking over RECs. As yet there has been no formal analysis looking at whether these developments have affected wholesale prices.

### 3.3. Current trading arrangements and transmission issues

Transmission issues have been dealt with in a rather *ad hoc* way since the start of the Pool. As a result, transmission is priced inefficiently with distorted signals for short-run production and long-run investment. However, in a reasonably small country such as E&W, the costs associated with these inefficiencies are not huge in absolute terms and have fallen since the SO/TO was provided with incentives to reduce them.

Losses are currently charged to customers on an average, actual loss basis that does not discriminate geographically, with the effect that those large customers and electricity supply companies whose demand helps to reduce losses do not receive cheaper electricity. Their demand may therefore be inefficiently low and, in the longer term, locational decisions may be made inefficiently. Efficient charging would require that customers face full marginal losses, as happens in NordPool and New Zealand. The zonal marginal loss scheme held up by judicial review would have partially moved E&W towards this form of efficient pricing, although there would have been an alleviation (scaling) factor, scaling losses back towards their average zonal levels, as also happens in California.

The fact that different generators make different contributions to losses is also not considered in determining the merit order unless bids are equal.[65] This has the effect that the merit order may

---

[64] Newbery (1997a), pg. 23. Figure 4 suggests that although there was a demand-weighted cap seasonal fluctuations in prices was greater in 1994-1996 than in the preceding or following periods.

[65] Taking losses into account would amount to partial application of marginal loss factors to marginal units, as those with high loss factors in mid-merit might not be called. Baseload units would have no loss factor applied, even though their effect on losses is the same as mid-merit plant.

not minimise the amount of generation required by the whole system, and hence may not minimise total energy costs. This also creates inefficient locational signals for investment in generation.

Are distortions associated with losses likely to be significant? In 1998/99 transmission (not distribution) losses amounted to 5.1 TWh or 1.76% of total electricity requirements (with an implied financial value of £132m).[66] A review of zonal marginal loss factors suggests that the efficiency costs associated from inefficient dispatch or location could be significant. Table 2 shows the predicted effect of 100 MW of new 2006/07 (NGC (2000)) on transmission losses at peak demand for NGC's generation zones.[67] New generation in Peninsula (South-West England) will give 13% more effective generation than new generation in the North. Of course, many of the gains of more efficient generation could only be realised by building new plant with its associated capital costs. Simply re-allocating generation amongst existing plants would tend to increase production costs (assuming the current merit order is correct) even if the cost of losses falls.

**Table 2: Effect on system losses at peak demand from adding 100 MW additional generation in a generation zones[68]**

| Zone | Change in loss (MW) | Zone | Change in loss (MW) |
|---|---|---|---|
| North | + 6 | South Wales | - 6 |
| Humberside | + 5 | Central England | - 6 |
| Rest of Yorks and Notts | + 2 | Estuary | - 4 |
| East Lancs | - 1 | Outer London | - 4 |
| West Lancs | + 3 | Inner London | - 6 |
| North Wales | + 2 | South Coast | - 5 |
| West Midlands | - 2 | Wessex | - 5 |
| Rest of Midlands and Anglia | 0 | Peninsula | - 6 |

*Source: NGC (2000), Table 6.5, pg. 6.15*

[66] OFGEM (1999a), pg. 88. Losses on lower voltage distribution networks are very much higher.
[67] Figures are similar to those predicted for 1999 in 1994 (Green (1998b), pg. 8).
[68] See Map 1 for location of these zones

Congestion is currently dealt with by paying plant that is constrained-on to produce (or constrained-off not to produce) after the unconstrained merit order and SMP have been determined. Even if plants act competitively and bid their avoidable cost the effect will be to inefficiently reward plant which is constrained-off, as this plant receives both the difference between its bid and SMP and also the availability payment. This could lead to inefficient availability decisions and inefficient locational investment decisions. In the presence of generators realising that they are likely to be constrained-on or –off the costs of congestion and associated distortions are likely to be exacerbated, as constrained-on plant has an incentive to raise its bid (exploiting its local market power) and constrained-off plant to lower its bid (both to make sure that it is included in the unconstrained schedule and to increase the difference between its bid and SMP). However, apart from some evidence of constrained-on plant submitting very high bids in the early 1990s, there is a lack of good evidence that this kind of behaviour has been observed in practice.[69]

Inefficient reward of constrained-off generators is mirrored by inefficient recovery of congestion costs from consumers. These are smeared equally across all customers through TSUoS charges, and hence those customers whose demand reduces congestion (and hence reduces the total costs of energy production) do not have their consumption subsidised. With inelastic demand the induced inefficiency may not be that great, as only very limited congestion would be eased by changing consumption decisions by more efficient pricing (instead, production has to change).

However, in fact, estimated congestion costs have fallen quite dramatically over the 1990s. Based on the increase in electricity costs resulting from using the constrained schedule rather than the unconstrained schedule, congestion costs fell from £225m in 1993/94 to £21m in 1998/99.[70] The small size of these numbers is surprising since local generation HHIs are substantially greater than national HHIs, as illustrated in Maps 1 and 2, based on NGC's generation zones and REC

---

[69] The two examples cited in the literature are National Power's Fawley oil-fired plant on the South Coast and PowerGen's Ferrybridge B power station (OFGEM (2000b), pg. 13, suggests that PowerGen made £88m in 1991/92 in this case). For more details see Armstrong, Cowan and Vickers (1994), pg. 308 and Newbery (1995), pg. 58.

[70] OFGEM (1999a), pg. 72. However OFGEM notes that the methodology for calculating constraint costs is somewhat arbitrary and that different methodologies would give higher numbers. It should also be noted that these numbers are deviations from costs based on SMP, which, in the presence of general market power, may not represent efficient production or pricing.

zones respectively.[71] In fact these HHIs are at levels such that if boundaries were congested predictably we would surely expect congestion costs to be higher than they are at present (i.e., we would expect generators to be able to exercise market power). Relatively low costs partly reflects the fact that not too many links in E&W are congested predictably, and the incentives of the SO to reduce congestion costs.

Green (1998b) provides estimates of the (medium-term) costs that result from inefficient transmission pricing in E&W for the late 1990s.[72] He uses a 13-node model reflecting inter-zonal congestion and losses to compare prices and welfare under optimal prices and various alternatives including the current system. It is assumed that re-dispatch given constraints is efficient and that generators bid their avoidable costs. The welfare losses of the current system amount to 0.6% of generation costs. This reflects the fact that physical congestion is reasonably rare in E&W, as well as the assumptions of competitive bidding.

Of course, TNUoS charges to give geographically differentiated signals and these may tend to offset some of the long-run inefficiencies resulting from failures to price losses and congestion more efficiently.[73] Current TNUoS charges are shown in Table 3, and comparison with Table 2 and the fact that congestion is mainly on flows from north to south, suggests that the pattern of charges is likely to bring more efficient decisions. However, TNUoS charges are not designed to capture costs caused by losses or congestion, so will not offset them perfectly. Instead, the Investment Cost Related Pricing methodology is designed to provide a rough estimate of NGC's long-run marginal cost of transmission capacity in different zones, rather than to capture opportunity costs from running the system.

Based on the size of transmission losses, it is apparent that the capital costs of the system could be reduced from locating generation and demand more efficiently. In practice it is not clear that more efficient prices would necessarily achieve a different distribution of plant given the costs

[71] Of course, the relevant HHIs for congestion depend on the exact location of the congestion, and how it divides up the transmission system. Therefore the numbers presented are only intended to be illustrative, although the NGC generation zones are largely coincident with critical boundaries in the transmission system.
[72] The final demand elasticity with respect to the generation price used is -0.25 which is described as a "medium-term" elasticity. Capital costs are not considered and the pattern of generation and transmission capacity is taken as given.
[73] There will be no effect on short-run production decisions as TNUoS charges are fixed for all decisions apart from peak consumption.

associated with transporting fuel, acquiring land and meeting environmental restrictions. However, given the amount of entry and exit in the last 10 years (25 GW of generation opened and 21.1 GW closed) more efficient signals would have been worth having if they could be achieved at reasonable cost. Within a centralised system like the Pool it is not obvious that the costs associated with a more efficient (though centrally calculated) pricing system would have been particularly large. As discussed below, with a decentralised system, the creation of more efficient transmission pricing might be considerably more expensive.

# 4. Description of the New Electricity Trading Arrangements

The principal characteristic of NETA is that most production will be determined by self-scheduling following decentralised trading. This reflects the underlying ideology of NETA: that decentralised, market based outcomes will be more efficient than ones that are centrally determined. However, close to real time the system will become centralised to allow the SO to balance the system. OFGEM's effort has concentrated in designing how this will be done and the method for calculating imbalance prices (applied to the difference between contracted and actual physical volumes). Even though these mechanisms will only deal with marginal amounts of energy, it is important to understand them in detail because they will influence the prices that emerge in the preceding decentralised markets. I first describe the physical, then the financial arrangements before turning to interim and future proposals for transmission. Appendix 1 provides a comparison between the organisations of the Pool, NETA, NordPool and the Californian market.

## 4.1. NETA in the absence of transmission issues[74]

### 4.1.1. How NETA will work for a given ½ hour period

The simplest way to describe NETA is to explain the timetable and procedures through which electricity is traded and demand and supply matched for a particular ½ hour period.

OFGEM expects that contracts for a ½ hour of electricity will be traded ahead of Gate Closure between generators, electricity supply companies, large customers and financial participants. Trade could be via either long-term bilateral contracts or short-term exchange-based instruments.[75] The SO will be able to sign two kinds of contracts. First, it will be able to trade energy like any other participant. Second, it will be able to sign ancillary service contracts with firms who agree to supply balancing services (or other forms of ancillary service) in real time at given prices. The fact that the SO will be able to trade energy is significant because it will allow the SO to attempt to deal ahead of real time with situations where parties decide not to contract for their full requirements. This will reduce the amount of balancing required in real time.

---

[74] This description is based on OFGEM (1999b), OFGEM (1999c) and OFGEM (2000f)

By 11am on the day before the half-hour in question all significant (over 50 MW) generators and suppliers will provide an "Initial Physical Notification" (IPN) outlining their expected generation and consumption, by BM unit.[76] This notification is indicative only, but allows the SO to start planning requirements and, for example, to invoke ancillary service contracts involving plants that requires greater than 3½ hours of warning. All parties (including the SO) can continue to trade after this point.

A key moment is Gate Closure, 3½ hours before the period in question.[77] By Gate Closure, trading parties must have submitted three types of information. First, a Final Physical Notification (FPN) must be submitted to the SO, with a minute-by-minute profile of what each BM Unit expects to produce or consume, to the nearest MWh. This is the sense in which NETA involves self-scheduling by participants, as opposed to centralised dispatch by the SO. Second, contracted volumes (but not prices) must be submitted to the organisation running the Settlement Process. It is worth noting that, in contrast to NordPool and California, there is no requirement for participants to submit balanced schedules with adequate physical contracts to meet their expected physical requirements.[78] This means that participants could choose to deliberately be out of balance (in either direction) at Gate Closure, with the potential that they might do so strategically in order to manipulate prices. This possibility is discussed in the next section. Third, BM units can choose to submit (several sets of) bids and offers to deviate from their FPNs. Such bids and offers are required if the BM Unit is to participate in the Balancing Mechanism but this is voluntary (by not participating a party effectively offers an infinite price to deviate from its FPN).

In the remaining time before real time, and during real time itself, the SO balances actual demand and supply by accepting bids and offers, and by calling on the appropriate contracts, attempting to minimise balancing costs while recognising system constraints. This procedure is the Balancing Mechanism. The latest bids and offers accepted will be available to all participants, who might

---

[75] OFGEM has decided not to procure the services of a power exchange as it deemed there was sufficient interest from a number of operators with experience in other markets. However, the volumes currently traded on APX are negligible. It is also not clear if bilateral markets will be transparent.
[76] A BM unit is equivalent to a genset on the supply-side and a Grid Supply Point on the demand side. For more details, see OFGEM (1999b) pg. 63
[77] OFGEM hopes that it will be possible to move Gate Closure further towards real time once participants, and in particular the SO, have experience of the BM (OFGEM (1999c)).

change their production or consumption decisions based on this information. An "offer" gives a price (£/MWh) at which a generator is willing to increase output or a customer to reduce demand (i.e., an offer is to increase net electricity supply to the system). A "bid" gives a price at which a generator will decrease generation (which could well be negative) or a customer will increase demand (i.e., a bid is to decrease net supply). Bids and offers cannot be changed after Gate Closure and acceptance of a bid or offer is firm on both the SO and the generator/customer, although the generator/customer also has to state a price at which it will accept the reversal of an accepted offer or bid.[79,80] To see what this means, imagine that 2 hours before the period the SO believes that there will be excess supply of electricity. He may therefore accept a bid from a generator (call him X). However, there is an unexpected decrease in supply, when a major genset (owned by Y) has an outage. The SO must either accept X's reversal offer, or, if it is cheaper, accept an offer from another generator (Z) or accept customer W's offer to decrease her consumption. As demand and supply vary over the half-hour the SO may accept a variety of bids (taking electricity off the system) and offers (putting electricity onto the system) from the same BM unit as long as they are compatible with the technical characteristics of the unit.

Of course, it is possible that there are insufficient bids or offers to balance the system. In this case, the SO can use "deemed" bids or offers to order plant to change production or to curtail supply. Essentially this is the moment when the market mechanism has failed and the SO has to act to maintain the physical security of supply. Obviously, it is hoped that such situations will be rare.

The description so far implies that bids and offers are only accepted to achieve energy balance. However they may be accepted for at least three other reasons. First, they may be accepted to relieve transmission constraints (see below). Second, they may be accepted as "arbitrage" bids or offers. For example, imagine that generator X places an offer to supply an extra 50MW for ½ hour at £20/MWh and that customer W places a bid to take 50MW for ½ hour at £25/MWh. In

---

[78] It should be noted that even though balanced schedules are supposed to be submitted in California, there exists the possibility that participants in fact plan to produce or consume differently to their schedules.
[79] OFGEM has indicated that it would like, as a matter of principle, to allow re-bidding of unaccepted bids and offers (OFGEM (1999c), pg., 41)

this case it would profitable for the SO to accept X's offer and W's bid simultaneously with no net effect on net system balance, but with an increase of 50 MWh in the total amount of electricity supplied. In effect, the SO is simply substituting for a trade which it appears should have been made before Gate Closure. Third, equal volumes of bids and offers may also have to be accepted in order to ensure that the system is able to meet demand fluctuations and outages, i.e. to create additional reserve. For example, in order to be able to increase production rapidly some plant must be running below full load.[81] Hence, some bids may be accepted soon after Gate Closure to provide flexibility. To maintain energy balance, offers would also have to be accepted.

At the end of the period, physical meter readings are supplied to the operator of the Imbalance Settlement Process. These are aggregated (as explained below) and then compared to the contract positions notified at Gate Closure to determine how far parties are "out-of-balance", so they can be charged for these volumes.

#### 4.1.2. Financial and pricing aspects of the Balancing Mechanism and Imbalance Settlement Process

Participants will pay each other according to their contracts. In advance it is difficult to know exactly what form such contracts will take (for example, which prices financial contracts will be linked to), although there is already evidence of some parties trading risk by transferring responsibility for their imbalance volumes. I now present an explanation of imbalance pricing and payments for BM services. As will soon be appreciated, this is a complicated subject. Appendix 2 provides three simplified examples which I hope illustrate the arithmetic involved.

The BM will operate on a "pay-as-bid" (PAB) basis, so an accepted bid (offer) will receive its own bid (offer) price even if more expensive bids (offers) were also accepted. This represents a further move away from the present Pool arrangements, which are based around a uniform clearing price. Contracted balancing services will be paid according to the negotiated contract

---

[80] Although bids and offers formally cannot be withdrawn or changed, there is some scope for participants to restrict the number of bids and offers that can be accepted by changing their maximum import and export limits. For example, a generator producing 800 MWh and with offers for further 100, 200, 300 and 400 MWh can effectively remove the last two offers by specifying a maximum import quantity of 1000 MWh. This can be done after Gate Closure.

[81] Of course, some flexible plant may deliberately choose not to operate at full output as it hopes to be paid for extra power by having an offer accepted. Where there is competition between flexible plant it is more likely to have its offer accepted if it does not have to have a bid accepted as well.

terms. There is no separate payment for simple participation in the BM (akin to a capacity payment) or for providing flexibility which it is assumed will be included in any bid/offer price or contract terms.

In the Settlement Process, participants (apart from the SO) will be paid and charged for electricity on the basis of the difference between their *ex-ante* contract volumes (i.e., volumes agreed by Gate Closure), modified by accepted BM bids and offers, and metered physical volumes.[82,83] Each BM unit controlled by a participant is allocated to that participant's Production or Consumption account, according to whether, in the previous year, the BM unit was a net producer or consumer of electricity (most obviously, a genset will typically appear in a firm's Production account).[84] The contract volumes associated with each unit in a particular account are added together (of course, for a particular half-hour some of these volumes could have different signs), and these are compared with the sum of the metered volumes for the same set of BM units. This will be done on a half-hourly basis.[85] Thus each party will have an imbalance volume on a Production account and one on a Consumption account. These cannot be netted off against each other.[86] Note that this method of allocating volumes means that even if parties are physically in-balance in the sense

[82] An accepted bid or offer counts as a balanced contracted trade. For example, a generator might have contracts and a FPN to supply 2,000 MW at gate closure, have an accepted bid for 500 MW and actually produce 1,800 MW. In this case he will be deemed to have spilled 300 MW (1,800 MW – (2,000 MW – 500 MW)) onto the electricity system for which he will receive the SSP. The fact that trades with the SO after Gate Closure will not lead to imbalances, means that resolution of congestion will not cause imbalances.

[83] The SO is not charged imbalance prices if it is out of balance in its energy trading, in order to prevent giving it perverse incentives to manipulate imbalance prices by choosing certain bids and offers.

[84] A participant can allocate one of its BM units to the account of another participant for Settlement Process purposes. For example, an electricity supply company might allocate one of its consumption meters to the consumption account of a generator. This allows participants to agree contracts with some risk sharing properties. In addition "aggregators" may provide trading services, including BM and Settlement Process participation, for small generators and final customers.

[85] The use of half-hourly periods as the basic time unit for contracts and production will continue from the Pool. In some other markets (California, PJM, Finland) there has been move towards calculating at least imbalance prices more frequently.

[86] An important reason for this restriction is to prevent vertically integrated firms having an advantage in being able to net off demand and supply imbalances *ex-post* in a way which would be not be available to firms only active in generation or only active in supply. However, prior to Gate Closure a firm can organise to trade between its two accounts. Also, large generators can net off over and under production from different gensets.

that they produce and consume the same amount, they may be liable to imbalance prices if physical volumes differ from contracted volumes.[87]

For a net deficit of electricity on either account (i.e., the party has generated less than it was contracted for, or has consumed more than it was contracted for) the System Buy Price (SBP) is charged for each KWh. This price is the volume-weighted average of accepted offers relating to the ½ hour period in question. Similarly the System Sell Price (SSP) which is paid to those who provide surplus electricity is the volume-weighted average of accepted bids. These prices will include relevant costs from any contracts invoked by the SO. The use of weighted averages is partly necessitated by the use of a reasonably long settlement period (½ hour). With a shorter settlement period the possibility that marginal and average prices might be significantly different would be less important.

However, not all accepted bids and offers count for the calculation of these imbalance prices! In particular, bids and offers accepted for arbitrage reasons and those accepted to deal with transmission constraints and to maintain voltage are excluded.[88] However, a certain quantity of trades on both sides designed to allow part-loading of plant to provide reserve will be included. When the system is short of electricity it will be bids accepted for this purpose that set the SSP. Price setting is illustrated in Figure 11 for a case in which the system has excess electricity, so a greater volume of bids than offers are accepted.

Although out of balance participants are charged imbalance prices, this mechanism is not used to directly recover the net costs incurred by the SO in balancing the system, and the costs of its various energy and ancillary contracts. These costs include those from transmission related and frequency reserve related trades which are not included in the calculation of imbalance prices.

---

[87] For example, imagine that generator X contracts with customer W for 1,000 MW in the half-hour period. However, X only produces 800 MW and W only consumes 800 MW and we ignore the possibility of transmission congestion issues. Then X would be liable to pay the SBP for 200 MW and W would receive the SSP for 200 MW. If, as expected by OFGEM, the SBP is higher than the spot price and the sell price lower then the combined parties would suffer a net loss even though they cause no physical imbalance on the system. Note that they would still appear from OFGEM's description to be liable even if they notified their FPNs at the 200 MW level. The obvious rejoinder to this point is that if they are aware of the difference between contracted and expected volumes at Gate Closure then W should sell X a contract for 200 MW electricity, which would remove their deficits.

[88] This leads to the important question about how these trades designed to maintain transmission and system balance are to be identified, as any accepted bid or offer is likely to affect these balances. The preferred route at the moment, at least until a new regime for transmission access is introduced, is to designate the most expensive bids and offers as being those related to transmission and system balancing.

These are recovered via smeared Balancing Service charges from all participants based on their metered volumes, subject to NGC's incentive scheme. Any net surplus or costs from the Settlement Process is also returned or charged to all participants (whether out of balance or not) on this basis.

Provision has been made to ensure that firms do not have perverse incentives to not respond to accepted offers or bids. This is important as non-response would make it harder to balance the system. For example, imagine a generator (X) with a FPN and contracts to produce 2,000 MWh. Then the SO accepts X's offer to produce another 500 MWh at £10 per MWh, but the System Buy Price is (expected by X to be) £6 MWh. If X fails to respond he still receives £5,000 (as an accepted bid is paid for by the SO whether the participant responds or not), but he only pays £3,000 to buy the electricity at the imbalance price. If X's avoidable cost of production is more than £6 MWh then it would be profitable for him not to respond, but the SO would then have to accept an offer from another generator. To avoid this incentive, X is made liable for the extra £2,000 difference between his offer and the System Buy Price (only, of course, in the case where the buy price is less than the offer). In this case, if he does not respond to the acceptance he effectively receives no payment for it at all. As long as an offer which was originally profitable (i.e., no less than avoidable cost of extra production) remains so in real time, it will be rational for X to respond when the offer is accepted.

There is also provision for an Information Imbalance charge to be levied on firms whose FPNs differ (in the absence of accepted bids/offers) from their metered volumes. However, on the basis that NGC does not expect deliberate mis-reporting of FPNs to be a significant problem, the charge has been initially set to zero.

## 4.2. Transmission-related issues and NETA

NETA itself is not designed to deal with transmission, and initially little will change. However, OFGEM is proposing radically different arrangements for the future, and it is important to understand how they will interact with NETA.

### 4.2.1. Transmission issues at the start of NETA

The transitional arrangements will in many ways represent only small departures from the current system, which reflects the small role transmission considerations have played in designing NETA.

Participants will be responsible for purchasing transmission losses, which will be calculated on an ex-post (actual) average loss basis.[89] 45% of losses will be charged to generators and 55% to demand-side participants, and will be included in their physical volumes for comparison against contract quantities.[90] There will be no geographic differences in charges for losses. Two innovations are worth noting. First, both sides of the market will be charged for losses. Second, as imbalance prices will be charged for out of balance quantities, participants will be penalised (assuming the energy spot price is between the SBP and SSP) for being inaccurate in their assessment of losses which will depend on everyone else's output as well as their own. At present customers are simply charged to same price for losses as they pay for all energy whether they predict them accurately or not. However, imbalances due to inaccurate predictions of losses on the transmission network seem likely to be small.

For congestion management there will be no trading of access rights, with the BM and ancillary contracts being used instead. TNUoS and TSUoS (included in BSUoS charges) charges will be maintained. The former will be calculated on the same basis as at present. TSUoS charges will recover the cost of bids and offers accepted in the Balancing Mechanism which are identified as being for the purpose of relieving transmission and ensuring system stability. As described above, these are not being charged to those in energy imbalance, instead being smeared across all participants. As at present, TSUoS charges will not be differentiated on a geographic basis.

### 4.2.2. Proposals for future transition arrangements[91]

As future arrangements are still open for active discussion, I describe the system which OFGEM favored in its December 1999 consultation paper, which, for transmission access, is based around decentralised trading of physical rights.[92]

---

[89] Making a participant responsible for losses means that the metered volumes reported to the Settlement Process will be adjusted to reflect losses and that this new number will be used for the comparison with contract volumes. For example, if a customer takes 97 MWh from the system and losses are 3% then to be in balance he must have a contract for 100 MWh.

[90] The asymmetry reflects the different position of meters relative to the transformers on the Grid Supply Points of generators and customers.

[91] Originally it was hoped that these arrangements would be introduced from April 2001, but it seems unlikely that this date will be achieved.

[92] OFGEM (2000a), pg. 13, indicates that OFGEM continues to favour a system with market-based trading of access rights.

For transmission losses OFGEM would like to move to a system not dissimilar to that currently delayed by judicial review (which raises an obvious question about its feasibility).[93] Participants on both sides of the market would be responsible for purchasing losses (probably based on the 45:55 split mentioned above) which would be charged on an annual and zonal basis, using (possibly scaled) marginal transmission loss factors. These would be based on ex-ante predictions of losses, calculated using the previous year's records, with the effect that a participant who knew her physical demand or generation would know exactly the level of contract cover she would need to be in balance for the settlement process. Full charging for losses on a marginal basis should lead to a net surplus of energy, assuming that participants aim to be in balance.[94] This would result in the SO having to accept a certain number of bids in the Balancing Mechanism (or acting as a seller of electricity is markets before real time), even if everyone was in balance (for settlement purposes) relative to their contracts. *Ceteris paribus*, this will also lead to a financial surplus (assuming that bids are non-negative), which the SO could use to set against other transmission charges. A scaling down of marginal TLFs might be used to reflect the effect of constraints on losses (at constraints, marginal losses are zero) or to mute zonal differentiation so that the scheme could be phased in. An appropriate scaling could also be used to eliminate the expected net surplus, while maintaining zonal differentiation.

To deal with congestion, OFGEM favours a system of physical access rights.[95] Generators and demand-side participants would purchase physical entry and exit rights respectively with an entry right defined on a zone-to-reference point basis (vice-versa for exit rights).[96] Purchasing a right would give a generator (customer) the legal right to put a certain quantity of electricity into (out of) the system (with compensation if this right was curtailed) and exempt the generator (customer) from paying other congestion-related charges.[97] Rights would be sold by the SO through an auction, with the volume of rights calculated on a "top-down" basis with the effect

[93] This description is based on OFGEM (1999a), pgs. 87-98

[94] In the simplest case, imagine that losses are geographically uniform with average losses of 2% and marginal losses of 4%. In this case, to be "in balance" participants would aim to trade 2% more energy than is actually required.

[95] This description is based on OFGEM (1999a), pgs. 70-86.

[96] Rights will be physical in the sense that those producing or consuming without rights will be financially liable for some form of settlement charge.

[97] Of course, other transmission charges would be applied in order to recover the costs of the transmission network.

that the volume of capacity is likely to be over-estimated.[98] Auctions would take place either every six months or annually, possibly with multiple rounds to aid information revelation. For generation or consumption which helps to relieve constraints (so we would expect efficient transmission prices to be negative), participants will effectively be bidding lowest subsidies for "obligations to supply".[99] Secondary trading will then be possible, and near to the day in question the SO will either sell more rights if capacity is likely to be greater than was originally forecast or, more likely, will be active in repurchasing rights for capacity which is not expected to be available. Markets for secondary trading are likely to cease at Gate Closure for energy markets. In real-time the SO will balance the system, as in the interim period, using the BM. This will be the method for dealing with unexpected *inter*-zonal congestion and any *intra*-zonal constraints. In the BM, the SO would simultaneously trade energy and transmission (when the SO accepts an offer for additional generation it will have to simultaneously sell the generating BM unit an entry right).

OFGEM has not, as yet, provided information on the settlement process which will work to make sure that participants do not produce or consume without rights, and do fulfil their obligations to produce when they have received a subsidy. Presumably there will have to be some form of disclosure of rights holdings at Gate Closure which could then be compared to metered volumes. Dual prices might be used to incentivise participants to be in balance, as is intended for NETA's energy arrangements. OFGEM has also raised the possibility of having "use-it-or-lose-it" which could force those who are not planning to use rights they hold to divest them prior to Gate Closure. One way of doing this would be to allow the SO to sell further quantities of interruptible transmission rights if it believed that holders of the firm transmission rights were not going to use them.[100] Alternatively, a dual price arrangement may be deemed to provide sufficient incentives. A related question is whether participants with several BM units within the same zone would be

[98] Forecasts will of course have to depend on inputs, outputs and capacity for the entire system, and not just the zone itself, because of the nature of loop flow. As discussed below, this is likely to make forecasting much harder and may open the possibility of gaming by participants with generation or consumption in several zones.

[99] It is not clear whether under a reformed transmission rights system certain generators who are frequently constrained-on would have bilateral contracts similar to "Reliability Must Run" contracts in the US. This would be consistent with the use of bilateral contracts under the transitional arrangements where there is local market power so that auctions would not be competitive.

[100] The sale of a class of interruptible rights might also be one way round the problems of making accurate forecasts in the presence of loop flow.

able to net off generation and consumption (to determine a net level of energy exports) and whether they will be able to total their generation and consumption (so they will not have to specify ownership of rights for a particular BM unit). Totalling would be consistent with the energy arrangements but netting would not.

## 5. Review of the existing literature on NETA

The existing literature has been sceptical about whether NETA will provide any improvements on Pool outcomes, even though the Pool is recognised to be imperfect.[101] Analysis has focussed on two issues: the role of PAB in the BM and performing simulation exercises to estimate how well NETA will work. Simulations have raised the issue of whether the detailed rules of NETA create possibilities for lucrative gaming of the system. While I argue that the gaming strategy identified in the literature will not be a problem in practice, the possibility that there are profitable gaming strategies must remain a major concern. OFGEM and participants should be ready to redesign the rules as soon as gaming strategies emerge.

The introduction of PAB has been justified by a number of theoretical models, but criticised by other authors as being likely to distort the merit order and, in particular, disadvantaging independent baseload operators. Under PAB all players have to try to predict the "going rate" so that they can get the best prices possible. This may be complicated, especially for small players.[102] Newbery (1997), Green (1999a) and Wolfram (1999a) all argue that this could result in inefficient dispatch if baseload generators overestimate the market price. As a result, it could deter entry by small participants and give more power to portfolio generators.[103]

The issue of PAB, as opposed to SMP, pricing has been the focus of several recent theoretical working papers which have been explicitly motivated by NETA: in particular, Fabra (2000), Federico and Rahman (2000) and Garcia Diaz (2000). These papers do not take up the question of whether small players may be disadvantaged due to information asymmetries raised by Newbery, Green and Wolfram. Instead they try to find whether prices will be higher if an electricity auction is run by PAB, rather than SMP. Fabra shows using a von der Fehr and Harbord type of model with inelastic demand and equally efficient firms. She shows that SMP makes collusion easier to support in the dynamic version of the game. Garcia Diaz uses the static version of the same game, but assumes that firms know demand for certain. A perfectly symmetric duopoly leads to PAB and SMP being revenue-equivalent. Asymmetry in cost or

[101] For generally sceptical comments see Green (1998a), Wolfram (1999a), NERA (1999) and Harbord and McCoy (2000).

[102] As discussed below this will be true in the BM where bids and offers have to be made before Gate Closure. However, in markets operating prior to Gate Closure it will obviously depend on the degree of liquidity and price reporting which emerges.

capacities makes PAB better for productive efficiency and for customers (lower transfers to firms). Federico and Rahman use the SFE approach with uncertain and elastic demand, and derive results for the polar cases of perfect competition and pure monopoly (which they suggest could represent the collusive outcome). Under PAB, even perfectly competitive firms will bid above cost and this induces allocative inefficiency, although demand pays lower average prices for power. Under monopoly, SMP allows the monopolist to perfectly discriminate across different demand outcomes (i.e. set the optimal price for each demand realization). This is not possible under PAB, where the price received for a particular unit is the same under all demand realizations. With PAB there is an incentive to raise bids on low cost units (which cannot free-ride when realized demand is high as they do under SMP), but less incentive to raise bids on high cost units which are unable to have an inframarginal effect.

The general conclusion from these papers is that PAB may help to undermine market power when demand is uncertain, even if it does not produce fully efficient outcomes. However, the value of these models in assessing NETA is limited. First, they do not capture the issue of asymmetry of information amongst participants which more general surveys have thought important. It is intuitive that this asymmetry could result in productive inefficiency, but if small participants are risk averse this might also lead them to bid conservatively, leading to lower prices on average. In turn, this might encourage larger participants to signal their actions to correct the asymmetry. This would be an interesting and relevant issue to explore further. Second, it is unclear whether the models are supposed to represent the entire physical electricity market (in which SMP is currently used for the Pool) or simply the BM, for which the models seem more suited. In the BM it would appear appropriate to inelastic or almost inelastic demand, as prices will only be revealed in or close to real time. Third, if the model is supposed to represent the BM, then the models are severely limited by their failure to consider the implication of how imbalance prices are set under PAB, which will affect the position of the demand curve. Instead it is assumed that demand will be similar under SMP and PAB.

A related point is that within the BM the position of the demand curve facing participants making offers will depend on their own imbalances. For example, in the case where there is a large demand for electricity this may indicate that several participants are short on electricity, and hence will have to buy electricity at the SBP (the average of accepted offers). Generators in this

---

[103] Newbery (1997a), pg. 26, Wolfram (1999a), pgs. 11 and 12, Green (1998a), pg. 19

position would clearly prefer a low SBP, even though they would still have an incentive to secure a high price for their own accepted offer. Putting a high offer into the BM would have the effect of limiting the risk facing a generator.[104] The models presented abstract from this feature that may play an important role in determining incentives, especially if generators and customers are not price takers.

However, the most important issue missed is that overall electricity prices will depend on the relationship between the BM and the preceding markets for bilateral contracts and electricity forwards and futures, in which OFGEM and participants expect the bulk of power to be traded. For example, the presence of sequential markets may undermine market power as in the Allaz and Vila (1993) model. Alternatively, local market power in the BM may lead to all generators achieving higher prices in earlier markets. Analysis should start with the BM (in this sense the papers start at the right point), but then work backwards to understand how the structure and possible imperfections in the BM will feed back through earlier markets, and how the desire to manipulate earlier markets will effect strategies in the BM. To understand this, consider some simple examples.

Imagine that there is scope for portfolio generators (controlling a large amount of capacity) to exercise market power in the BM, in the sense that SSPs above cost can be secured and high offers accepted. This may reflect the particular rules in the BM or the fact that because it operates close to real time only a subset of plant can provide balancing services, so competition is limited.[105] In this case, however liquid earlier markets may potentially be, portfolio generators will have incentives to withhold capacity from them, instead supplying power in the BM, unless prices in the pre-Gate Closure markets increase to the expected levels of BM prices. In equilibrium, the market power in the BM will flow backwards and raise prices in earlier markets. From the point of view of the portfolio generator considering his strategy the fact that by withholding capacity until the BM or otherwise committing himself to raise prices in the BM increases prices in earlier markets effectively restores some of the infra-marginal effect of his bids, even though the BM itself operates on a PAB basis.

---

[104] In the event that he is out of balance and there is a high SBP then he will have a lucrative offer accepted. Of course, for a generator whose contract is in absolute MWh, being short of power this may indicate plant failure in which case he may be unable to fulfil his offer.
[105] In countries where generation is primarily hydro-electric, the vast majority of plant is able to provide flexibility close to real time. This is not necessarily true in a thermal dominated system.

On the other hand, if the generators who can set prices in the BM only control a small amount of capacity and there is a surplus of generating capacity trying to sign contracts in earlier markets, then competition between baseload generators may lead to earlier prices significantly below the SSP. Flexible generators may exercise market power in the BM but this will be "local" market power and will not have much effect on average energy prices.

Bower and Bunn (2000) provide a simulation model which aims to compare prices under four alternative systems: SMP with daily bidding, SMP with half-hourly bidding, PAB with daily bidding and PAB with half-hourly bidding. The first option is taken as representing the current system, the final as representing NETA. Players are allowed to learn when developing their strategies, but never know other player's strategies. Forward markets exist – and players contract in them to achieve target utilisations – but prices in them are not modeled explicitly. Bower and Bunn find that SMP with daily bidding produces lower prices than PAB with hourly pricing, as daily bidding reduces the ability of generators to discriminate across demand periods and with PAB baseload generators may overbid, increasing the amount of demand for mid-merit generators. This approach therefore does capture the asymmetry of information missed by the more theoretical models discussed above. However, interactions with earlier markets are not captured in any detail, and their models of "NETA" and the Pool are highly simplified. For example, in the Pool although bids are submitted daily, availability allows the supply curve to be varied half-hourly. For NETA, the method of calculating imbalance prices as weighted average and how this may affect willingness to trade in the BM is ignored.

London Economics (1999) also provide results of simulation studies, this time using actual people rather than computers as players. These were performed on behalf of OFGEM. Rules in the BM were varied to test their effects, but all were simplified versions of those that will actually be used (for example, the averaging method of calculating SBP and SSP will be used). Several interesting results were found. First, dispatch tended to be close to efficiency, with better performance under PAB (2% above optimal costs) than SMP (5-6%). Similarly, average prices were closer to efficient levels (defined in terms of cost recovery of marginal generating units) for PAB with dual imbalance pricing (5.6% above) than SMP (11.2%). Note these results differ from those suggested by Bower and Bunn (2000). Second, BM prices appeared to be more volatile than PX (spot market pre-Gate Closure) prices, with dual imbalance pricing appearing to raise PX prices. Third, the gross volumes of imbalances were relatively large, even when a large spread was imposed between dual imbalance prices. Net imbalances were smaller, although balancing 5% of the market with a 40% imbalance would still require significant activity by the SO (for example, 5% is more than NordPool's SO has to balance in its regulation market). Of course, it is necessarily to be slightly sceptical about how much we can learn from this kind of simulation with simplified rules and scenarios, especially when experience in the Pool and in other markets such as California suggests that market manipulation can depend on the details of the system.

Serious issues have been raised about the ability to game the system. Harbord and McCoy (2000) describe a strategy that allowed the gaming of the London Economics system, which they claim would allow a gaming of the BM generally. They propose the following generator strategy: buy a large volume of power in the PX (e.g., far greater than the system demand for the half-hour). Then submit an FPN to generate a reasonable volume and put in a massive bid in the BM to reduce your production. Other producers will produce their full output to meet their contract requirements and the SO will accept the huge bid immediately, along with other bids to achieve system balance. The SSP will be high, and you profit on all your contracted power (which of course you never planned to take) greater than the actual system demand. This profit more than compensates for the prices you paid buying the contracts and the high bid paid on your (small) production. However, this strategy ignores the fact that a massive bid would count as an arbitrage trade for the SO and so would not affect the SSP unless there was not a corresponding volume of offers at lower prices.[106] This would mean that you could lose a huge amount of money with this strategy, in the case that the actual SSP was lower than the PX price.[107,108] However, although the gaming strategies suggested appear not to be a problem in practice, if there are other unilateral gaming strategies which would be profitable, this would be a serious issue. In general, experience in markets such as California indicates that it is very hard to design a system where no

[106] OFGEM (2000f), pg. 17

[107] Exactly the same objection applies to the other gaming strategy played by the same player, which did not involve selling or buying huge quantities. In this case, the generator sells some power forward and then under produces, at the same time as putting in an offer for a small volume with a huge negative price (he will pay to produce) which is accepted. This created a negative imbalance price that meant that the generator was paid a large amount for being short. Again, a negative offer would, as long as there were corresponding volume of positive bids, count as an arbitrage trade.

[108] In addition, the ability of the SO to trade energy prior to Gate Closure could mean that, if the SO recognises what is happening, that it would not have to accept the bid for energy balancing purposes in the BM.

rules can be manipulated. Hence during the initial months and years of NETA it is reasonable for both OFGEM and participants to expect that the rules may have to be developed and refined.

A more serious potential problem, in the sense that it seems more robust to whatever kinds of detailed rules were used, identified by London Economics, is the possibility of several players withholding capacity and driving up price.[109] This is a basic strategy used to drive up prices in any market (apart from in cases where there are capacity constraints anyway), including the current Pool and it should not come as a surprise that such strategies could raise prices. The questions for whether this could be a problem in practice are, as usual, the size of individual generators' capacities relative to total market demand and, more generally, the ease with which collusion can be maintained.

[109] London Economics (1999), pg. 53

## 6. Analysis of the NETA arrangements

As with the Pool, inefficiencies will arise from the details of the mechanisms used and from the possibility of market power. The overall success of the reforms will also depend on how much power is traded through the BM. If the SO has to do a lot of trading close to real time then we will have returned in large measure to a centralised system., but without the potentially efficient uniform price auction used by the Pool. The other main point is that the proposed arrangements for transmission seem likely to be expensive and may exacerbate rather than reduce problems of local market power.

### 6.1. The NETA energy market in the absence of market power

In the absence of market power, I consider three questions. First, how much trade is likely to pass through the BM? Second, are prices (and quantities) likely to be efficient? Third, what effect will the NETA arrangements have on market structure and investment decisions?

OFGEM intends that the BM should only deal with unexpected variations in output and demand (as well as constraints) close to real time. If, instead, BM volumes are large this could create problems. It will be harder for the SO to manage the system, and it will be required to engage in more trading in advance, with the effect that the SO's decisions may have a large effect on market prices. In addition, greater volumes may induce price volatility and make BM prices, determined by a subset of flexible plant, more important in determining overall average energy prices.

With price taking behaviour the level of imbalances will not be determined by strategic withdrawal. However this does not mean that participants could not deliberately decide to pass through Gate Closure out of balance. Deliberate imbalances are relevant under NETA in a way they are not in NordPool and California where participants are supposed to submit balanced schedules.[110]

To understand the incentives to be in balance under NETA consider three slightly different situations.[111] First, consider a case with no transactions costs in either the BM or the markets

[110] As mentioned above, this does not mean that participants in NordPool or California always report accurately.

[111] It is also worth noting that London Economics (1999) simulations indicated that dual imbalance prices reduced but did not eliminate deliberate imbalances.

operating prior to Gate Closure and participants who make their production/consumption decisions (apart from responding to accepted bids/offers) by Gate Closure, although they may be subject to outages or demand shocks. Participant would want to be in balance if the price available in the futures market is greater than the expected SSP and below the expected SBP.[112] If the price went outside these bounds, participants would start trying to buy or to sell and the price would return to inside them. Hence, in this case the assumption that players will want to be in balance should hold. Second, if transaction costs are introduced in the futures market then participants may be prepared to remain deliberately out of balance past Gate Closure depending on the difference between the expected imbalance and futures prices and the relative size of the transactions costs. Expectations of price volatility may also encourage participants to contract prior to Gate Closure when they are not able to take flexible production or consumption decisions.[113] Third, consider the case where participants are heterogeneous, with some able to decide on their production/consumption, independent of whether their acceptance of bids and offers are accepted. Imagine a flexible customer of this sort. As well as submitting bids and offers to the BM, the customer might decide to increase his consumption as more information becomes available on the probable level of the SBP (remember information on the latest offers accepted will be available). Even if it is expected to be higher than the futures price at Gate Closure, there will be a certain set of outcomes when it will turn out to be a better price. This logic may mean that flexible participants may deliberately put themselves out of balance after Gate Closure.[114]

A more general problem is that the operation of the BM and Settlement Process do not encourage participants to engage in their own balancing where this is possible, but instead requires that all balancing should be done centrally through the SO in the BM. For example, if a vertically

[112] The SBP will always be greater than the SSP because arbitrage trades are not counted in the calculation of these prices.

[113] With price taking behaviour, volatility is likely to be determined by the volume of imbalances itself and by the extent to which the SO contracts for reserve in advance at given prices. Some participants have suggested that the SO may contract too much, with the effect that development of earlier markets is retarded.

[114] Generator flexibility in the light of changing prices could also increase imbalance costs through dramatic decisions taken to try to be in contract balance, even though these actions may increase the degree of system imbalance. For example, suppose that a generator agrees contracts for 500 MWh at Gate Closure but intends to produce 750 MWh based on expectations of the SSP, and gives this FPN. However, in real time it emerges that the SSP will be very low. In this case, the generator may dramatically reduce his production in real time, requiring the SO, *ceteris paribus*, to accept a corresponding level of offers.

integrated participant knows that he will fail to meet his contract generation volume, he cannot remedy this by cutting his own consumption. Such action would instead face him with imbalance prices on both his production and consumption accounts, which would not cancel, even though he helps to maintain system balance. A similar issue applies to combinations of independent generators and suppliers who might want to sign contracts to manage their own imbalances. Not allowing this kind of action will increase imbalance volumes, and would seem to block a potentially efficient behaviour. Overall, it would appear that volumes traded in the BM are likely to be (inefficiently) greater than they would be if participants were required to submit balanced schedules or were allowed to self-balance.

Even if no participants take deliberate actions to be out of balance, trades are likely to be required in the BM to deal with shocks, transmission issues and the need to provide adjust outputs to provide the system with flexibility. A 1999 NGC estimate of these volumes is included in Table 4, made under the assumption that participants aim to be in balance.[115] For comparison, average E&W demand in 1999/2000 was around 33,000 MW, so the average gross requirement is expected to be less than 10% of total demand and the net volume would be far smaller.

Are prices likely to be efficient? Relevant prices include those received by participants making bids and offers in the BM, imbalance settlement prices and prices in earlier markets. Of course, as described in the last section, all these prices are related and are, to large extent, determined together. In particular, the expectation of prices in the BM and Settlement Process will influence feasible prices in earlier markets. It is important to note that even if generators are small and competitive in the sense that in other market settings they could not raise prices, they will still exploit simple gaming opportunities that may exist through the details of the market rules.

It is possible to identify a number of ways in which prices may not be fully efficient even with price taking behaviour. First, if the restrictions on efficient self-balancing mean that the volume of trades is inefficiently high then, non-horizontal bid and offer curves, *ceteris paribus*, the SBP will be too high and the SSP will be too low. In addition, co-ordinated actions of the type described above that do not affect system balance would pay imbalance prices, which, in itself, will not be efficient. Second, PAB will lead even price-taking firms to bid above their costs, meaning that prices determined by these bids cannot be cost reflective. Of course, if a lot of firms have reasonably homogenous costs the mark-ups are likely to be small. Third, those out of balance pay SBP and SSP rather than the marginal cost to the system of imbalances. In the absence of other considerations, it is the marginal (net) cost that is the efficient imbalance charge.[116] Of course, the attractiveness of a single price is reduced if it increases the required volume of balancing trades. Finally, these distortions in BM prices are likely to feed through to earlier prices even with price taking behaviour. However with competitive behaviour it is unlikely that the distortions resulting from the design of the BM will be great. Of course, on occasion electricity prices may still be high, reflecting the shape of the underlying supply curve, but this is an entirely different issue.

It is important to consider how NETA will affect incentives for entry and integration, and the long-term plant mix. The removal of capacity payments means that entry and exit decisions will have to be taken purely on the basis of revenues from selling energy. For peaking plant this means that costs must be recovered from energy prices in the few hours of peak demand, and this may require marginal prices in the BM or futures markets to go very high. However, in E&W the initial effect may be to reduce the capacity margin with the consequence that system security could be reduced, or require the SO to subsidise such plant through contracts.[117]

The structure of the BM and imbalance prices may give greater advantages to flexible and reliable plant than exist under the Pool. With NETA, flexible plant will tend to receive higher prices by being able to participate in an additional market (the Balancing Mechanism) and by being able to respond to optimal prices. Unreliable plant will face greater risk of being exposed to unfavourable imbalance prices. These forms of generation may also be more attractive to back risk sharing contracts with electricity supply companies, as they are less likely to be unfavourably out of balance on their own account. Hence, in the long-run we could see a shift towards more flexible and reliable generation.[118]

---

[115] The estimates were made when Gate Closure was expected to be 4 hours, rather than 3½ hours before real time.

[116] In particular, if a generator supplies electricity to the system in excess of his contract volume then there is no reason in theory why he should not receive the marginal value to the system of this power.

[117] The current level of system security in the UK could be thought to be excessive, reflecting in part the problems with capacity payments described above.

[118] Of course, at the same time there may be forces (such as government requirements on the amount of power which must be purchased from renewable sources, some of which are very unreliable such as wind power) which point in the other direction.

It is not obvious that the move to NETA will necessarily lead to greater participation by demand-side participants, many of whom already have incentives to be price-responsive if they are able to do so. However, as price elasticity is largely ignored for the purpose of setting Pool prices, a move to bilateral contracting may help to make the existing elasticity more effective. In order to deal with market risk, vertical integration (or, alternatively, contractual innovations) may be encouraged. Even though generation and consumption totals cannot be netted off for the purpose of calculating imbalance charges, we would expect the SSP to be high when the SBP is high, creating a natural hedge through vertical integration. In addition, horizontal integration may also be encouraged by the ability of flexible plant to provide cover when inflexible plant fails, and in addition to share the transaction costs of trading, which are largely fixed, over a larger amount of capacity. The transaction costs of trading in several markets may be significant for small generators.[119] At present, small generators typically sign long-term contracts with RECs and then operate on baseload, apart from taking plant out for maintenance when it is not bid into the Pool. Many do not have any active trading staff. Under NETA, they will require more active management of contracts to deal with the possibility of expected or unexpected outages. Alternatively, they may sub-contract out their risk management to large generators, electricity supply companies or specialised aggregators in return for a fee.

In summary, in the absence of market power and without gaming opportunities, BM prices are unlikely to be at their exact efficient levels, but the distortions are unlikely to be great. Two concerns remain: first, the volumes the SO has to balance in the BM may be large, especially if transaction costs are large, and second, the details of market rules may contain opportunities for gaming. Both issues may require NETA rules to be changed soon after implementation, and it would be useful for participants to understand that this may be the case.

## 6.2. NETA energy market and market power

If generators are able to exercise market power under the Pool, it is likely that they will also do so under NETA, as now appears to be recognised by OFGEM (OFGEM (2000a)). The main consequence will be the same: higher prices. However, higher BM volumes could also be a problem, making it harder for the SO to balance the system, and prices more sensitive to SO actions.

[119] The same arguments would apply to relatively small customers who are connected directly to the Grid.

As outlined already, NETA is unlikely in practice to encourage significant extra demand-side elasticity, relative to what already exists but is inadequately considered, which would restrain prices in either the Balancing Mechanism or earlier markets. Without changes in market structure, any reduced scope for the exercise of market power will have to come from the changing of market rules. The elimination of capacity payments, especially given the way that LOLP is calculated, will remove one route by which prices may have been increased in the past, although it may have the consequence of reducing capacity and increasing price volatility in the long-term (which may be efficient). However, there are a number of ways in which it appears that NETA will be no less attractive for market power to be exercised than the Pool.

First, as only flexible generating plant and consumption will be able to operate in the BM close to real time, there will be an increase in market power relative to a situation where all plant was able to compete, as in setting PPP through the Pool. There also appears to be an opportunity for portfolio generators to engage in some strategic withdrawals of capacity from the BM even after Gate Closure, using the maximum export and import limits. This will tend to exacerbate market power problems. Participants are not supposed to be able to change their bid/offer schedules after Gate Closure. However, a portfolio generator can steepen his offer schedule by reducing the maximum export limit on some of his plant with relatively low offers. This may be profitable if it appears that a lot of offers will be accepted.

Second, the hope that PAB might help to restrain prices by removing profits on infra-marginal units may be misplaced. If generators take into account the effect of increasing prices in the BM on prices in earlier markets, then some inframarginal effect will be restored, encouraging higher bids. Generators active in the BM may also be able to attempt to induce volatility in imbalance prices to increase prices for forward contracts and to increase the premium they can earn on providing risk-management services as aggregators.[120]

The most obvious consequence of market power would be that energy prices are too high. In fact, there is some encouraging but limited evidence that prices are expected to fall. EfA prices for baseload energy for 2001/02 are below the level for 2000/01 although the volumes currently being traded are small.[121] If market power also encourages greater volatility of prices and higher

[120] If large generators act as aggregators this could essentially bring a large amount of now-independent generation under their control. This would have to be a concern for the regulator.
[121] OFGEM (2000a), pg. 46

transactions costs for participants, especially if liquidity is also impaired, then this will tend to reduce entry by small participants, encourage perhaps excessive investment in flexible plant and discourage further direct demand-side participation.[122] Of course, as may be the case at present, the threat of regulatory intervention may act as a binding constraint on prices. In this case introducing NETA will not achieve very much.[123]

What will be the effects of market power on the costs of balancing the system? If generators withhold capacity until the BM, which may be particularly attractive when the supply margin will be tight this may tend to increase the amount of trading the SO has to do in the BM. This will not necessarily be changed by allowing the SO to trade in advance, as it will depend, absent regulatory intervention, on the willingness of the generators to trade with the SO. The SO may only be able to trade at high prices, reflecting the generators market power.

### 6.3. NETA and interim transmission arrangements

The interim transmission arrangements will have very similar effects to the current arrangements, both in themselves and in how they will affect the energy market.

The interim arrangements for charging transmission losses (splitting actual average losses between generators and electricity supply companies 45:55, with no geographical differentiation) will, like the present arrangements, not provide efficient locational signals for either short-run production decisions or long-run investment decisions.[124] The SO will only pick bids and offers in the BM to minimise losses if there is a tie in terms of other characteristics, so the choice of bids and offers to accept may not minimise system costs (achieve productive efficiency).[125] There will be no incentive (apart from the very marginal effect from the total of smeared charges) for

[122] The traditional entry route for small generators of signing long-term contracts with RECs is also becoming less attractive to the RECs as supply-side competition has developed.

[123] OFGEM (2000a), pg. 46 states "we do not believe that it would be desirable or feasible to pursue structural remedies, such as further plant divestment, to the extent that would be required to address the problem [of potential market power]".

[124] As long as participants agree contracts having recognised that the split will be 45:55 (rather than 100% on suppliers or 40:60) this should have no effect on outcomes as explained above. The use of actual losses will tend to mean that all participants are slightly out of balance, unless they predict losses with perfect accuracy. However the errors involved are likely to be small.

[125] If bids and offers were chosen to minimise true energy costs this would amount to implicit charging for marginal losses on BM bids and offers.

participants to trade efficiently to reduce losses in earlier markets.[126] The use of actual losses is likely to result in all participants being slightly out of balance, although net system imbalances may be reduced, relative to using ex-ante SO forecast, if participants can predict losses accurately.

Transitional congestion arrangements will lead to similar perverse incentives as at present, with the consequence that balancing volumes may be increased above the level there would be under more efficient pricing. As described above, under the Pool, plant that is likely to be constrained-on has an incentive to bid high and plant that is likely to be constrained-off has an incentive to bid low. Similar strategies will be optimal under NETA, with the effect that constrained-off plant will make profits and the volume of balancing trades will be increased.[127] Plant likely to be constrained-on should not contract before Gate Closure (when it is competing with all generators), and instead will place high offers in the BM. Constrained-off plant should sell before Gate Closure as long as prices are above avoidable costs and then place low and possibly negative bids into the BM, which the SO will have to accept.[128] The incentives for such exploitation of local market power created by transmission constraints could actually be made worse by the elimination of capacity payments which currently give an incentive for constrained-on plant to shave their bids. In addition, the fact that the BM will operate closer to real-time, whereas the Pool operates day-ahead, may also strengthen the local market power of certain generators. The SO may want to sign contracts with plant that is constrained-on or -off on a predictable basis, although bilateral contracts are unlikely to prevent such generators capturing rents. Of course, as noted above, current congestion costs are not that great and there is no obvious reason to expect them to be significantly higher at the start of NETA.

The arrangements will also continue to charge for congestion in an inefficient way, by smearing the costs amongst all customers via TSUoS charges even when some customers provide helpful

---

[126] Of course, in the presence of concentrated local generation markets there may be good reasons for not wanting to encourage this kind of trading as it may exacerbate local market power.

[127] NGC's estimate of the average volume of trade required to deal with constraints is fairly small (see Table 4). Increased strategic behaviour would make this an underestimate.

[128] Pre-Gate Closure contracting under NETA is also reflected in the incentives of the two kinds of plant to sign CfDs linked to SMP before the Pool opens. For example, a generator which knows that it will be constrained-on does not reduce its risk by signing a CfD, as the price it receives in the Pool is determined by its bid and not by SMP (also, in the event that SMP was very high and above its bid, it would lose by having to compensate the other party to the CfD).

consumption. As at present this will be partly offset by locational TNUoS charges with a roughly similar pattern to more efficient TSUoS charges.

### 6.4. NETA and transmission - suggested future arrangements

For transmission losses, OFGEM wants to move towards a system of (possibly scaled) zonal marginal loss charging. From the point of view of efficiency, moving to marginal loss pricing is sensible, while zonal charges make sense as long as losses within zones are reasonably homogenous, an assumption which may be violated when there is intra-zonal congestion (at the point of congestion marginal losses are zero). However, charging for marginal losses, without an appropriate scaling factor, will lead, *ceteris paribus,* to an energy surplus of around 2%. As a proportion of trade in the BM the percentage may be much larger, and could increase the costs of balancing, as well as increasing the potential for market power from flexible plant offering bids. It will also need to be considered how the system for handling transmission losses will interact with trading of inter-zonal congestion rights. In particular, if losses are predicted ex-ante, but rights prices reflect actual congestion then there is the possibility that participants will face charging for losses and congestion, when only one actually occurred. This has led some participants to suggest that the SO should include an adjustment for losses in the way that transmission rights are traded (for example, by adjusting bids for transmission rights to take account of losses). However, this is likely to be complicated and leave considerable discretion to the SO which a lot of participants are also anxious to avoid.

The most interesting part of the future arrangements concerns the trading of physical transmission rights to deal with congestion, in separate markets from energy. One of the apparent motivations for reforming how transmission is dealt with is that under the current system pockets of local market power resulting from transmission constraints distort the potentially competitive energy market. In consequence, liquidity is reduced and outside participation discouraged. The result is higher general energy prices. Removing this distortion justifies expensive reforms, even when the current costs that can be directly attributed to congestion are relatively small. A further motivation for reform is that in the longer-term congestion may become more of a problem. This is most likely to be the case if the flow of power into southern England from the French interconnector ceased or changed direction. This may become more likely with the integration of continental European energy markets, especially if E&W prices fall.

However, there is a general problem with this reasoning, as well as specific problems with the scheme proposed. Suppose for the moment that congestion is a major problem. Energy and transmission are entirely complementary and physical participants will have to trade (corresponding quantities) in both markets. The creation of a separate market for transmission which is distorted by market power will necessarily lead to distorted purchases in the energy market as physical participants are concerned about the price for delivered electricity (energy plus transmission).

Of course, in the absence of local market power, the proposed scheme could provide an efficient way of dealing with inter-zonal congestion, and in particular an improvement to the current and interim arrangements which reward constrained-off generation. Suppose that it is possible to estimate the amount of transmission capacity available. The price of entry rights in an export-constrained zone should be bid up to the price at which the marginal plant (given the constraint) in the zone is just willing to produce given its costs and the energy price. Similarly in an import-constrained zone, the subsidy associated with the entry right should be bid up to the level at which the marginal generator is just willing to produce. Constrained-off plant will receive no compensation for not being able to run (assuming it is not initially given rights below their market price). Similar arguments apply for exit rights. Total electricity prices should be the similar to those that would emerge from creating competitive local energy markets with limited amount of trade between markets.

However, without market power, a complicated scheme like the one proposed (with its attendant transaction costs) is not necessarily required to get efficient outcomes. For example, bids and offers into the BM or a day-ahead market could be used to work out local energy prices and clear markets and all energy passing across the constraint could be charged a tax based on the market clearing prices in different zones. With day-ahead trading such a system would be similar to the one used in NordPool. Of course, one might be hesitant about using the BM, as currently designed, operating so close to real time, for this purpose as it might lead to an inefficient selection of plant changing its production and setting prices. It would also increase the traded volume.[129]

[129] Only flexible plant can participate in the BM, whereas it might be efficient to get less flexible plant to reduce (or increase) its production which would require trade to happen the day-ahead.

In the presence of local market power the suggested arrangements, like any other system, are unlikely to be unable to mitigate all the resulting distortions. A detailed example of how market power could be exercised is presented below, but generally generators with local market power in the export-constrained zone may be able to reduce their bids for entry rights to reduce the price of those rights for their plant which does produce.[130] On the other hand, even with this distortion constrained-off plant may do less well than it does at the moment or would do under the interim arrangements (where all of the constrained-off plant is rewarded and exporting plant receives the national energy price). Similarly, customers with market power in the same zone could try to drive up their consumption subsidy. Of course, these problems are not peculiar to the proposed arrangements: Johnsen, Verma and Wolfram (1999) find the effects of local market power under transmission constraints in NordPool. Trying to use the BM to determine local energy prices would be subject to many of the problems discussed above, as well as allowing constrained-off plant to pay very low bids as is possible with market power under the interim arrangements.

The proposed arrangements will not mitigate distortions in the energy market created by local market power. Therefore the value of the proposed reforms depends on a comparison of distortions under the proposed and current systems, and the cost of creating and operating the new system. Given that the current costs of transmission constraints are small, the results of this cost-benefit analysis cannot be assumed. The proposed system is likely to be complicated and costly to operate, and may be subject to gaming by participants who can exploit the SO's problems in predicting the volumes of transmission capacity, especially given the nature of loop flow. In this case, it might have made more sense to deal with transmission problems through developing the NETA energy markets in a different way (for example, by requiring the day-ahead submission of bids which could be used to clear local energy markets). A number of alternatives have been precluded by trying to develop NETA first and then trying to deal with transmission as an add-on.

First comes a general problem with any system of physical access rights: in the presence of loop flow, it is very hard to calculate the volume of rights that can be sold without knowing the exact distribution of inputs and outputs in advance. When links are congested the volume of power that can be put in at one node will depend on actions at all other nodes. Under NETA, these will only

[130] If generators in a zone could collude perfectly (or if they are a monopolist) then they would bid for a volume just below that sold by the SO, making the price of entry rights in an export-constrained region equal to zero.

be available to the SO at Gate Closure, whereas rights are supposed to be sold up to year before this point. The situation is made more complicated by the fact that the limits on capacity are not only thermal, but also relate to voltage limits and depend on the set of contingencies used when planning the system. If the volume of capacity is not correctly estimated then prices will not tend to be efficient. This is one reason why financial transmission rights systems, where the SO can exploit all the capacity actually available close to real time, may be preferable. Of at least equal concern is the possibility that participants may be able to game the system, by changing their production decisions to make sure that the SO is always wrong in its capacity estimate, profiting when the SO has to buy back rights the day ahead. When generators control plant in a number of different regions this kind of gaming could be straightforward and seems likely to be potentially profitable.

Second, even if inter-zonal congestion can be dealt with efficiently, intra-zonal congestion will remain a problem. It will continue to be dealt with through the BM or via contract, allowing constrained-off plant to profit. If REC zones form the basis of congestion rights (the only option unless substantial costs are incurred to change metering arrangements) about 50% of congestion will be intra-zonal. In addition, the demand and supply decisions of different generators/customers within a REC zone can have very different effects on inter-zonal constraints. This will make the calculation of inter-zonal capacity even harder.

A further problem with using inflexible zones, based on administrative convenience rather than congestion, is that it may, by itself, help to create or increase local market power. This is important given that current congestion costs are not great. For example, imagine a zone that would be relevant for dealing with transmission congestion, but which is split between three REC zones. Concentration when the REC zones are combined might be reasonably low. However, within each REC zone concentration might be quite high, and these concentration levels will be important in the trading of transmission entry or exit rights for each of the zones. The possibility that individual REC zones might have high HHIs is illustrated in Map 2. For the purpose of comparison, remember that the national HHI based on registered capacity is less than 1,500.

Third, transaction costs associated with access trading could be large in comparison with current congestion costs. The number of markets which participants – including the SO – will have to act in will increase significantly. As with energy trading, the costs of transmission trading could be largely fixed and hence fall disproportionately on small players. In addition, variable rights prices and volatile imbalance prices will increase participant risks. In contrast, vertically integrated participants may naturally hedge against this risk as, within a zone, entry and exit rights should have values with opposite signs. Participants and the SO may find it hard to predict prices because of the nature of loop flow meaning that prices may depend on all activities in the system. These issues have led participants to favour a simpler system than has been currently suggested.

Fourth, as already suggested, a system of transmission rights (physical or financial) raises the question of how these may affect local market power (in the energy market). As discussed by Joskow and Tirole (2000), issues concerned with physical withholding of rights would be mitigated by some form of "use-it-or-lose-it" rule, an option which appears to be under consideration by OFGEM.[131] Similarly, participants buying production obligations would have to pay some form of penalty for failing to supply. However, these rules on their own do not entail that participants could not take decisions in the energy market to influence the value of their rights, or could not use rights to influence energy prices. For example, imagine the typical three node arrangement shown in Figure 12. Generation at point B is not competitive (the simplest case is monopoly) and by producing more they can ease the congestion on the line between A and B. Hence we would expect generators at B to receive a production subsidy (a negative price for their entry rights). Consider their incentives when deciding production. The more they produce, the lower the energy price: this occurs both directly through their own production and by allowing an increase in the production of producers at A. If producers at A have a positively sloped aggregate supply schedule and bid competitively then when the generators at B produce more they also reduce the value of their per MWh subsidy and reduce the entry price for generators at A. In general there is no reason for the energy price or the transmission price to be at the efficient level, and in this example there is no need for withholding or failure to use rights (B simply buys a limited number of obligations to supply and produces this amount). This form of strategic decision not to produce will make the SO's task of correctly estimating the volume of capacity even harder. Furthermore, as suggested above, a generator with plant in different zones

[131] It is not clear whether it is envisaged that dual imbalance prices would provide the necessary deterrent or whether there would be a formal mechanism by which rights which a holder did not plan to use would be re-auctioned. One possibility mentioned to me is that the SO would sell additional interruptible transmission rights for capacity which it did not expect the rights holder to take-up.

could make decisions that would always make the SO wrong in its estimations and try to profit from these mistakes.

When B increases her production there is also an effect on the price of the exit right for consumers at point A, which in the simple example B does not take into account. However, in the case that the generators at B are integrated with an electricity supply company, itself operating in an imperfectly competitive market, at A there may be an additional reason for generators at B to hold back on production. On the other hand if the generator at B is integrated with the electricity supply company at C or at B itself, again in imperfectly competitive markets, then the incentive to restrict output would be reduced. In the light of vertical integration in the UK, the issue of how transmission rights and integration combine to influence the incentives for market power should perhaps be considered in more detail than it has been at present.[132]

Finally, it is very unclear how imbalance prices for transmission rights will or could be calculated from the BM. Even if the costs of transmission related trades can be separated, an important issues is how to charge participants who do not use some of their transmission rights. Often this will impose no costs on the system (there turns out to be excess capacity) so should arguably not be penalised. On the other hand, there will be other occasions when it led to increased congestion. It may be particularly hard to deal with cases when at the time of purchasing the right, the participant was not being paid for an "obligation to supply". These are the kinds of details that will have an important effect on how the system works in practice.

[132] The take-over of RECs by the Scottish generators and EdF provide possible examples of the simple issue illustrated here. In contrast in California reforms have been associated with utility divestitures.

## 7. Conclusions

My analysis suggests that NETA is unlikely to represent a dramatic improvement over the Pool. This is not because the Pool does not have its problems. Prices are set using forecasts of demand that ignore price elasticity. The justification for capacity payments is questionable in theory, while in practice they are poor measures of what they are supposed to represent. Market power has almost certainly been a problem, which may have been made worse by the existence of capacity payments giving additional incentives to withdraw capacity when the supply margin is tight.

NETA aims to improve on the Pool by removing capacity payments and allowing demand elasticity to play a role in determining prices. OFGEM also hopes that passing power to decentralised markets will allow the most efficient arrangements to evolve. However, close to real time the system will still be centralised with complicated mechanisms and price-setting rules. These rules may allow gaming opportunities, in which case the rules should be changed as soon as these opportunities become apparent. More generally, the structure of the BM would appear to give scope to portfolio generators to exercise market power, especially as they control much of the flexible plant that will be active in the BM. The effects of this market power will roll back into the earlier markets even if they are decentralised. Moreover, if the SO has to be very active in balancing the system, prices will continue to be determined by a central operator rather than pure market forces. The same will be true if there is frequent regulatory intervention to deal with perceived pricing problems. Of course, whether the problems of NETA turn out to be greater or less than those of the Pool is an empirical question, and it will be interesting to watch what happens.

While the arrangements for NETA have already been agreed in detail, those for the transmission access are only just starting to be developed. The current system gives inefficient signals and allows some scope for certain generators to exercise local market power, although the costs from this should not be exaggerated. In the initial phase of NETA the situation will remain broadly the same. The proposals for the long-term are more worrying. Even in theory, a physical access rights system works well only under assumptions that capacity can be accurately estimated and participants lack market power. In practice, these assumptions are unlikely to be met and in addition it seems likely that transaction costs will be significant. In consequence, the proposed system could be susceptible to a number of kinds of manipulation which will distort both the transmission rights market (possibly at the expense of the SO) and the energy market. The imposition of a "use-it-or-lose-it" rule on its own will not remove this danger.

These problems raise an obvious question of whether there is an alternative set of arrangements which would improve on the current situation without entailing the complexity and dangers of what has been proposed. OFGEM should begin this search at once. One alternative would have been to design NETA in such a way that transmission could have been dealt with simultaneously to arranging energy balancing ahead of real time. This has been precluded by the way OFGEM has dealt with NETA first and transmission second. Unless an alternative can be devised, OFGEM and participants should do a detailed analysis of whether they should proceed with the current proposals.

**Glossary and Abbreviations**

| | |
|---|---|
| Bid (NETA) | Generators and customers can place bids into the BM to take power from the transmission system. For generators a bid is to *reduce* generation, for consumers a bid is to *increase* consumption. |
| BM | Balancing Mechanism, starting 3½ before real time, through which the SO will attempt to balance the system. |
| BM Unit | Metered physical generation or consumption units which are used for bidding into the BM and for accounting in the Settlement Process. For a generator they correspond to gensets, for consumers grid supply points. |
| BSUoS | Balancing Service Use of System charges which will be used to recover the costs incurred by the SO through the Balancing Mechanism. In the initial phase of NETA this will include costs resulting from dealing with transmission constraints. |
| CCGT | Combined Cycle Gas Turbine generating plant, the most common type of generation to enter the UK market since 1990. |
| CfDs | Contracts for Differences, financial instruments used to hedge Pool prices, negotiated between the parties directly, usually lasting at least a year and for large amounts of energy. |
| E&W | England and Wales |
| EFA | Electricity Forward Agreement, financial instruments used to hedge Pool prices, traded through brokers, usually for small amounts of energy. |
| FPN | Final Physical Notification, submitted at Gate Closure to the SO under NETA, detailing the BM unit's production or consumption intentions. |
| Genset | Generating unit for which bids are submitted into the Pool. There may be several gensets in a single power station. |
| HHI | Herfindahl-Hirschman Index of market concentration, equal to the sum of squared market shares. |
| IPN | Initial Physical Notification, submitted day-ahead to the SO under NETA, giving an initial indication of generation and consumption by BM unit. |
| LOLP | Loss of Load Probability, used to calculate capacity payments under the Pool. Intended to represent the probability that the system will fail due to lack of generating capacity. |
| NETA | New Electricity Trading Arrangements |
| NGC | National Grid Company, which acts as SO and TO |
| Offer (NETA) | Generators and customers can place offers into the BM to add power to the transmission system. For generators an offer is to *increase* generation, for consumers a bid is to *reduce* consumption. Note that OFFER refers to the former name of the regulator, OFGEM. |
| OFGEM | Office of Gas and Electricity Market, which acts as regulator. |
| PAB | Pay As Bid, used to describe an auction system where participants whose bids are accepted pay or receive their bid prices rather than the marginal price. |

| | |
|---|---|
| PPP | Pool Purchase Price, the sum of the system marginal price and the capacity payment, received by all generators producing who were also included in the unconstrained schedule |
| PSP | Pool Selling Price, the sum of PPP and Uplift, paid by wholesale customers on their metered electricity supply. |
| PX | Generic term used to refer to an electricity spot market operating before real time |
| RECs | Regional Electricity Companies, privatised in 1991 controlling large regional electricity supply businesses and distribution networks. |
| SBP | System Buy Price, calculated as the volume-weighted average of accepted offers in the BM, excluding those deemed to be arbitrage offers or related to transmission constraints |
| SFE | Supply Function Equilibrium, an approach to modeling the Pool and certain other electricity markets, where participants can bid smooth price-volume functions |
| SMP | System Marginal Price, calculated in the Pool as the marginal price where supply meets expected demand. More generally refers to pricing systems or auctions where buyers pay and sellers receive the marginal price. |
| SO | System Operator, with responsibility for making sure that the electricity and transmission systems are balanced in real time. |
| SSP | System Sell Price, calculated as the volume-weighted average of accepted bids in the BM, excluding those deemed to be arbitrage offers or related to transmission constraints. |
| SUPERGOAL | Computer algorithm used under the Pool to draw up the merit order of plant which can produce given quantities of electricity at minimum cost |
| TNUoS | Transmission Network Use of System charges, zonally differentiated charges paid by generators and customers designed to recover the capital costs of the transmission network. |
| TO | Transmission Owner, the owner of the transmission network (NGC in E&W) |
| TSUoS | Transmission Service Use of System charges, non-locationally differentiated charges designed to recover the cost of resolving transmission constraints, subject to SO incentives |
| UPLIFT | Component of PSP recovering the costs of energy balancing and USAV payments, subject to SO incentives |
| USAV | Unscheduled Availability payments, capacity payment received by generators who are available but not included in the unconstrained schedule |
| VOLL | Value of Lost Load, used to calculate capacity payments in the Pool, intending to represent the loss to final customers of losing their electricity supply without notice. |

## APPENDIX 1

### Table 5: Wholesale energy arrangements

| | GENERATION CHARACTERISTICS | PRE-DAY AHEAD | "DAY AHEAD" OR PRE-GATE CLOSURE | DESPATCH AND REAL TIME BALANCING | IMBALANCE PRICING |
|---|---|---|---|---|---|
| **E&W Pool** | Pre-dominantly thermal, with range of marginal cost characteristics | Financial contracts only, in the form of CfDs or EfAs. Not organised by SO. | Generators submit non-firm multi-part bid schedules. SMP and Capacity Payments based on forecast demand. | Central dispatch by SO<br><br>SO manages system balance and dispatches plant. | Not relevant as based on central dispatch. |
| **E&W NETA** | ″ | Physical and financial contracts. Markets not organised by SO. | Generators and customers must submit contract position to Settlement Operator and FPNs to SO. Voluntary submission of firm bids and offers to BM. Gate Closure 3½ hours before real time. | Self-dispatch<br><br>SO manages system by accepting bids and offers in BM and calling on ancillary contracts. | Differences from contracted volumes charged at dual imbalance prices calculated from the BM. |
| **NordPool (Norway)** | Pre-dominantly hydro (99% Norway), including ownership by electricity supply companies | Physical and financial contracts cover majority of electricity. Financial markets organised by SO.. | Participants may submit simple price/quantity schedules for net demand and supply to day-ahead market. Sets market clearing price. 3% of energy traded in this form. | Self-dispatch<br><br>SO manages system using voluntary bids and offers in the Regulation Market. | Differences from submitted contracted volumes charged at system marginal price from regulation market. |
| **California** | Majority thermal, but 25% hydro and 11% renewables | Physical and financial contracts may be signed. | 80% of energy traded in PX[133]: generators and customers submit simple demand and supply schedules. Trade at PX MP. PX and other SCs submit balanced schedules and voluntary deviation bids to SO. | Self-dispatch<br><br>SCs can revise schedules up to an hour before real time. SO balances system using voluntary deviation bids in the Balancing Market and ancillary contracts. | Differences from submitted schedules settled at system marginal price from regulation market. |

[133] Until 2002, the three largest vertically integrated generators must sell all their power through the PX

## APPENDIX 1

### Table 6: Transmission Arrangements

| | METHOD FOR RESOLVING CONGESTION | PRICING OF CONGESTION | PRICING OF LOSSES | ASSOCIATED OTHER CHARGES |
|---|---|---|---|---|
| **E&W Pool** | SO schedules deviations from the unconstrained schedule in the Pool or using contracts. | Constrained -on, -off plant paid based on bids. Recovered from consumers via smeared TSUoS charges and SO incentive scheme. | Actual, average losses recovered from customers without geographical differentiation. | Connection charges and geographically differentiated TNUoS charges |
| **E&W NETA (Suggested)** | Generators and supply companies purchase physical transmission entry and exit rights based on zones. SO alleviates expected congestion by buying rights in secondary trading. Unexpected and intra-zonal congestion solved in Balancing Mechanism. | Physical transmission entry and exit rights priced in auctions and secondary trading.<br><br>BM trades related to unexpected congestion might be recovered from TSUoS charges. | Predicted zonal marginal losses (perhaps scaled down) to be recovered from generators and customers. | Unclear, subject to review. |
| **NordPool (Norway)** | SO divides Norway into zones based on predicted major congestion. Participants must submit schedules for generation and demand in each zone. SO matches demand and supply to solve expected congestion. Unexpected and intra-zonal solved in Regulation Market. | "Capacity charges" on all power in each zone creates differences from the system price at times of congestion. SO has revenue surplus based on flows across constraint, given to TO and used to reduce other system prices. | Predicted nodal marginal losses charged symmetrically on generation and consumption, but capped at +/- 10%. SO responsible for purchase. | Admission fee based on capacity and peak load with no geographic discrimination. Power charge based on the net volume at the node. |
| **California** | State divided into four zones for congestion purposes. Expected inter-zonal congestion resolved through adjustment bids submitted to PX and other SCs. Unexpected and intra-zonal congestion managed by accepting adjustment bids in the Balancing Market, subject to the need to keep each SC in balance. | Expected inter-zone congestion: marginal SC bids and offers used to value transmission, with charge levied on all flows. Charge returned to owners of Financial Transmission Rights (FTRs). The costs of accepting bids and offers in the Balancing Market are recovered from all SCs in the zone. | Participants responsible for purchasing predicted nodal marginal (scaled to average) losses. | Distribution companies charged for sunk and operating costs of the transmission network. |

## Appendix 2: Examples of the Balancing Mechanism and Imbalance Settlement

In this Appendix, I present three simplified examples to illustrate the discussion of how NETA will work in the first part of Section 4.

### Basic Structure

There are three generators (A, B and C), with A controlling two gensets which count as separate BM units. There are three customers (X, Y and Z), with X controlling two Grid Supply Points which also count as separate BM units. To keep units and calculations simple, the examples assume that the basic unit of time is the hour rather than the half-hour.

**Example 1: Arbitrage bids, no imbalances**

In this case all generators and customers submit FPNs equal to their contract quantities. At the level of the generator actual quantities match contracted quantities so that there are no imbalances. Generator A's genset level production differs from its FPNs but the differences cancel out when the output of its BM units are summed together.

Although there are no energy imbalances for the SO to deal with, the SO accepts offers from generators B and C, and bids from customers X and Z because the offers are at lower prices than the bids, so the exchange is profitable. These are arbitrage trades. Customers pay their bids and generators are paid their offer prices.

The following table summarises contract positions, FPNs, bids and offers, and financial outcomes for each participant. To keep things simple within hour fluctuations are ignored and quantities always vary in units of 10MWh. The actual quantities listed are prior to the acceptance of bids and offers. I assume participants always respond to acceptances. For example, if, for a generator, the actual MWh is 50 and the first offer of 10 MWh is accepted then at the end of the half-hour the meter would read 60MWh. For a customer, the meter would read 40MWh, as an accepted offer will reduce consumption. The rows of offers and bids show the prices submitted for the first 10MWh deviation from the FPN and the second 10MWh deviation from the FPN. A dash means that no offer or bid is submitted (the Balancing Mechanism is voluntary). A check (√) means that the offer or bid is accepted by the SO. Negative payments are net payments to the SO.

| | **GENERATORS** | | | | **CUSTOMERS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **$A_1$** | **$A_2$** | **B** | **C** | **$X_1$** | **$X_2$** | **Y** | **Z** |
| Contract MWh | 100 | | 50 | 70 | 120 | | 20 | 80 |
| FPN MWh | 60 | 40 | 50 | 70 | 60 | 60 | 20 | 80 |
| Actual MWh (pre-bids and offers) | 50 | 50 | 50 | 70 | 60 | 60 | 20 | 80 |
| Offer £/first 10MWh | 20 | - | 10 √ | 14 √ | 20 | 22 | 30 | - |
| Offer £/second 10MWh | 24 | - | 19 | 26 | 20 | 24 | 30 | - |
| Bid £/first 10MWh | 4 | - | 8 | 7 | 17 √ | 6 | 18 √ | 9 |
| Bid £/second 20MWh | 3 | - | 6 | - | - | - | 8 | - |
| Bid/Offer Payments (£) | - | | +10 | +14 | -17 | | -18 | - |
| Imbalance Payments (£) | - | | - | - | - | | - | - |

**Example 2: Net surplus of electricity, arbitrage bids and calculation of the System Sell Price**

In this example the actual positions indicate that there is a net surplus of electricity on the system that has to be corrected by the SO through the acceptance of bids. However, the SO first accepts arbitrage bids and offers as in the previous example, and only then accepts bids to balance the system.

The system sell price is the volume weighted average of the non-arbitrage bids used to balance the system. In this case these bids are 9,8,8 and 7 and come from generators (who reduce their production when the bid is accepted) and customers (who increase their consumption). As all prices are for equal 10 MWh blocks the SSP is 8. The SSP is applied to the imbalances (differences between "actual" and contract volumes) for generator A and customers X and Y.

| | **GENERATORS** | | | | **CUSTOMERS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **$A_1$** | **$A_2$** | **B** | **C** | **$X_1$** | **$X_2$** | **Y** | **Z** |
| Contract MWh | 100 | | 50 | 70 | 120 | | 20 | 80 |
| FPN MWh | 60 | 40 | 50 | 70 | 60 | 60 | 20 | 80 |
| Actual MWh (pre-bids and offers) | 70 | 40 | 50 | 70 | 40 | 60 | 10 | 80 |
| Offer £/first 10MWh | 20 | - | 10 √ | 14 √ | 20 | 22 | 30 | - |
| Offer £/second 10MWh | 24 | - | 19 | 26 | 20 | 24 | 30 | - |
| Bid £/first 10MWh | 4 | - | 8 √ | 7 √ | 17 √ | 6 | 18 √ | 9 √ |
| Bid £/second 20MWh | 3 | - | 6 | - | - | - | 8 √ | - |
| Bid/Offer Payments (£) | - | | 2 (=10-8) | 7 (=14-7) | -17 | | -26 (=-18-8) | -9 |
| Imbalance Payments (£) | +8 | | - | - | +16 (=2 * 8) | | +8 | - |

**Example 3: FPNs differ from contracted volumes, arbitrage bids and a variety of bids and offers accepted for system balance**

In this example, generators B and C and customers X and Z submit FPNs that are different from their contracted volumes. This might be justified based on their prior beliefs about imbalance prices and/or transaction costs in contracting ahead of real time. Actual production has some firms taking more from the system than their contracted volumes and others taking less. Hence it is necessary to calculate both the SBP and the SSP. Based on changing information about the system balance, the SO accepts a number of non-arbitrage bids and offers (as always picking the highest bids and lowest offers). The net number of bids accepted is 4, as the difference between the actual volume generated and the actual volume consumer (pre-bids and offers) is 40 MWh.

Imbalance volumes are calculated by comparing actual volumes and contracted volumes, not FPNs. Therefore B, C, X and Y all have surpluses of 10 MWh, A has a surplus of 20 MWh and Z has a deficit of 20 MWh. The SSP (applied to A-Y) is calculated as the average of non-arbitrage bids: (9+8+8+7+6+6)/6 = 7.3. SBP (applied to Z) is calculated as (20+20)/2 = 20. Note that even if A merged with Z it would still face the imbalance prices calculated below. This is because the production and consumption accounts do not net off against each other.

| | **GENERATORS** | | | | **CUSTOMERS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **$A_1$** | **$A_2$** | **B** | **C** | **$X_1$** | **$X_2$** | **Y** | **Z** |
| Contract MWh | 90 | | 40 | 60 | 110 | | 20 | 60 |
| FPN MWh | 60 | 40 | 80 | 50 | 60 | 60 | 20 | 80 |
| Actual MWh (pre-bids and offers) | 70 | 40 | 50 | 70 | 40 | 60 | 10 | 80 |
| Offer £/first 10MWh | 20 √ | - | 10 √ | 14 √ | 20 √ | 22 | 30 | - |
| Offer £/second 10MWh | 24 | - | 19 | 26 | 20 | 24 | 30 | - |
| Bid £/first 10MWh | 4 | - | 8 √ | 7 √ | 17 √ | 6 √ | 18 √ | 9 √ |
| Bid £/second 20MWh | 3 | - | 6 √ | - | - | - | 8 √ | - |
| Bid/Offer Payments (£) | +20 | | -4 (=10-8-6) | 7 (=14-7) | -3 (=20-17-6) | | -26 (=-18-8) | -9 |
| Imbalance Payments (£) | 14.6 (=2*7.3) | | 7.3 | 7.3 | 7.3 | | 7.3 | -40 (=2*-20) |

## Bibliography


Allaz B. and Vila J-L (1993) "Cournot Competition, Futures Markets and Efficiency", *Journal of Economic Theory* 59(1), 1-16

Armstrong M., Cowan S. and Vickers J. (1994) *Regulatory Reform: Economic Analysis and the British Experience*, Cambridge, MA: MIT Press

Bower J. and Bunn D.W. (2000) "Model Based Comparisons of Pool and Bilateral Markets for Electricity", *Energy Journal* 21(3), 1-29

Fabra N. (2000) "Uniform Pricing Facilitates Collusion: The Case of Electricity Markets", mimeo, EUI

Federico G. and Rahman D. (2000) "Bidding in an Electricity Pay as Bid Auction", mimeo, Nuffield College, Oxford

Garcia Diaz A. (2000) "Uniform Pricing vs. Pay-Your-Bid Multi-Unit Auctions under Complete Information", mimeo, Universidad Carlos III

Green R.J. and Newbery D.M. (1992) "Competition in the British Electricity Spot Market", *Journal of Political Economy* 100(5), 929-953

Green R.J. (1996a) "The Electricity Contract Market", DAE WP 9616, University of Cambridge

Green R.J. (1996b) "Increasing Competition in the British Electricity Spot Market", *Journal of Industrial Economics* XLIV(2), 205-216

Green R.J. (1997a) "Electric transmission pricing – An International Comparison", *Utilities Policy* 6(3), 177-184

Green R.J. (1997b) "Transmission Pricing in England and Wales", *Utilities Policy* 6(3), 185-193

Green R.J. (1998a) "England and Wales: a competitive electricity market?", POWER WP no. 60

Green R.J.. (1998b) "Electric Transmission Pricing: how much does it cost to get it wrong?", POWER WP no. 58

Harbord D. and McCoy C. (2000) "Misdesigning the UK Electricity Market", *European Competition Law Review*

Helm D. and Powell A. (1992) "Pool prices, Contracts and Regulation in the British Electricity Supply Industry", *Fiscal Studies* 13(1), 89-105

Hjalmarsson L., Midttun A. and Svindland E. (1999) "Nord Pool: Issues and Dilemmas", mimeo

Hsu M. (1997) "An introduction to the pricing of electric power transmission", *Utilities Policy* 6(3), 257-270

Johnsen T.A, Verma S.K. and Wolfram C.D. (1999) "Zonal Pricing and Demand-side Bidding in the Norwegian Electricity Market", mimeo

Johnsen T.A. (1999a) "Hydropower generation and storage, transmission constraints and market power in a deregulated electricity market", mimeo

Johnsen, T.A. (1999b) "Pricing of transmission services, fixed cost recovery, incentives for generation investment and economic efficiency", mimeo

Joskow P.L. (2000) "Deregulation and Regulatory Reform in the US Electric Power Sector", mimeo, MIT

Joskow P.L. and Tirole J. (2000) "Transmission Rights and Market Power on Electric Power Networks", mimeo, MIT

Klemperer P.D. and Meyer M.A. (1989) "Supply Function Equilibria in Oligopoly under Uncertainty", *Econometrica* 57(6), 1243-1277

London Economics (1999) "Role Playing Simulations of the New Electricity Trading Arrangements: A Report to the RETA Programme by London Economics"

NERA (1999) "New Electricity Trading Arrangements for England and Wales: A Comment by NERA on OFGEM's proposals", *Energy Regulation Brief August 1999*

Newbery D.M. (1995) "Power Markets and Market Power", *Energy Journal*, 16(3), 39-66

Newbery D.M. (1997a) 'Pool Reform and Competition in Electricity", DAE WP 9734, University of Cambridge

Newbery D.M. (1997b) "Competition, Contracts and Entry in the Electricity Spot Market", DAE WP 9707, University of Cambridge

Newbery D.M. and Pollitt M.G. (1997) "The Restructuring and Privatisation of Britain's CEGB – Was it Worth It?", *Journal of Industrial Economics* XLV(3), 269-300

NGC (2000) *National Grid Seven Year Statement*, Coventry: NGC

OFFER (1998a) "Review of Electricity Trading Arrangements. Background Paper 1: Electricity Trading Arrangements in England and Wales", February 1998

OFFER (1998b) "Review of Electricity Trading Arrangements. Background Paper 2: Electricity Trading Arrangements in other countries", February 1998

OFGEM (1999a) "NGC System Operator Incentives, Transmission Access and Losses under NETA. A Consultation Document", December 1999

OFGEM (1999b) "The New Electricity Trading Arrangements: A Consultation Document", 2 volumes, July 1999

OFGEM (1999c) "The New Electricity Trading Arrangements: OFGEM/DTI Conclusions Document", October 1999

OFGEM (2000a) "Introduction of the market abuse condition into the licences of certain generators: OFGEM's second submission to the Competition Commission", June 2000

OFGEM (2000b) "Introduction of the market abuse condition into the licences of certain generators: OFGEM's initial submission to the Competition Commission", May 2000

OFGEM (2000c) "NGC system operation under NETA: transitional arrangements. A consultation document", April 2000

OFGEM (2000d) "The New Electricity Trading Arrangements and Related Transmission Issues: An OFGEM Policy Statement", July 2000

OFGEM (2000e) "OFGEM's investigation of Edison First Power under the market abuse licence condition. Initial Findings", July 2000

OFGEM (2000f) "An Overview of the New Electricity Trading Arrangements V1.0: a high level explanation of the New Electricity Trading Arrangements (NETA)", May 2000

OXERA (1999) *New Electricity Trading Arrangements: Prospects for Market Development*, Oxford: OXERA Press

United States Department of Justice (1997) *Horizontal Merger Guidelines*, Washington

Von der Fehr, N-H. M. and Harbord D. (1993) “Spot Market Competition in the UK Electricity Industry”, *Economic Journal* 103, 531-546

Von der Fehr, N-H. M. and Harbord D. (1997) “Competition and Regulation in Electricity Markets: Economic Theory and International Experience”, mimeo, University of Oslo

Wilson R. (1998) “Efficiency considerations in designing electricity markets”, Report to Competition Bureau of Industry in Canada

Wilson R. (1999) “Market Architecture”, mimeo, Stanford University

Wolak F. and Patrick R.H. (1996) ‘Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market”, mimeo, Stanford University

Wolfram C.D. (1998) “Strategic Bidding in a Multi-Unit Auction: an Empirical Analysis of Bids to Supply Electricity in England and Wales”, *RAND Journal of Economics* 29(4), 703-725

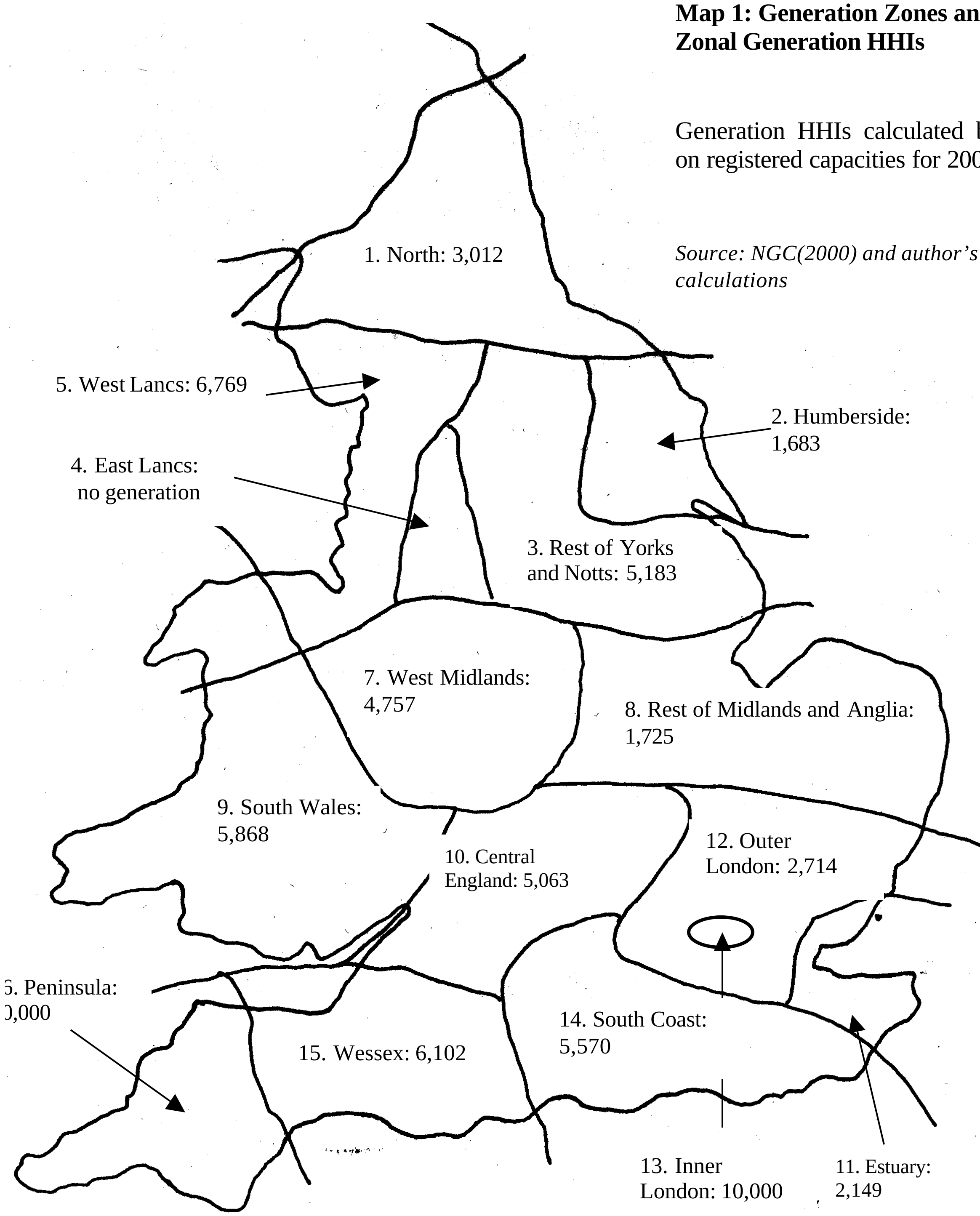
Wolfram C.D. (1999) “Electricity Markets: Should the Rest of the World Adopt the UK Reforms?”, POWER WP no. 69

Wolfram C.D. (1999) “Measuring Duopoly Power in the British Electricity Spot Market”, *American Economic Review* 89(4), 805-826

**Map 1: Generation Zones and Zonal Generation HHIs**

Generation HHIs calculated based on registered capacities for 2000/01

*Source: NGC(2000) and author's calculations*



**Map 2: REC Zones and Zonal Generation HHIs**

Generation HHIs calculated based on registered capacities for 2000/01

*Source: NGC(2000) and author's calculations*

**Table 3a : Transmission Network Use of System Charges (kW) 2000/01**

| GENERATION | | | DEMAND | | |
|---|---|---|---|---|---|
| **Zone no.** | **Zone name** | **Charge (p/kW)** | **Zone no.** | **REC name** | **Charge (p/kW)** |
| 1 | North | 8.33 | 1 | Northern | 0.95 |
| 2 | Humberside | 5.19 | 2 | Norweb | 5.77 |
| 3 | Rest of Yorks and Notts | 4.09 | 3 | Yorkshire | 4.98 |
| 4 | East Lancs | 2.37 | 4 | Manweb | 5.53 |
| 5 | West Lancs | 4.15 | 5 | East Midlands | 7.74 |
| 6 | North Wales | 6.33 | 6 | Midlands | 9.34 |
| 7 | West Midlands | 1.63 | 7 | Eastern | 9.95 |
| 8 | Rest of Midlands and Anglia | 1.59 | 8 | Swalec | 15.21 |
| 9 | South Wales | -4.64 | 9 | Seeboard | 10.91 |
| 10 | Central England | -0.93 | 10 | London | 14.49 |
| 11 | Estuary | 0.82 | 11 | Southern | 13.45 |
| 12 | Outer London | -0.06 | 12 | South Western | 16.67 |
| 13 | Inner London | -11.00 | | | |
| 14 | South Coast | -3.70 | | | |
| 15 | Wessex | -5.68 | | | |
| 16 | Peninsula | -10.35 | | | |

*Source: NGC (2000), Table 8.1, pg. 8.17*

**Table 3b : Transmission Network Use of System Energy Consumption Charges (p/kWh) 2000/01**

| REC Zone | TNUoS Charge (p/kWh) | REC Zone | TNUoS Charge (p/kWh) |
|---|---|---|---|
| Northern | 0.135716 | Eastern | 1.251901 |
| Norweb | 0.790859 | Swalec | 1.880786 |
| Yorkshire | 0.713735 | Seeboard | 1.517293 |
| Manweb | 0.862185 | London | 1.893716 |
| East Midlands | 1.061887 | Southern | 1.798764 |
| Midlands | 1.324182 | South Western | 2.226393 |

*Source: NGC (2000), Table 8.1, pg. 8.18*

**Table 4: NGC Estimates of volumes required to secure the system from 4½ hours before real time (MW)**

| | **Average requirement** | **99% confidence interval** |
|---|---|---|
| Demand forecasting errors | +/- 400 | +/- 1,800 |
| Plant failures and re-declarations | + 450 | + 1,100/- 2,400 |
| Failure to follow dispatch instructions | + 150 | + 200/ -400 |
| Transmission constraints | +/- 350 | +/- 2,000 |
| Frequency response | +/- 1,000 | +/- 1,700 |
| **Total** | + 2,350/- 1,750 | + 6,800/-8,300 |

*Source: OFGEM (1999b), Table 6.1, pg. 61*

**Figure 1 : Bid structure in the Pool**

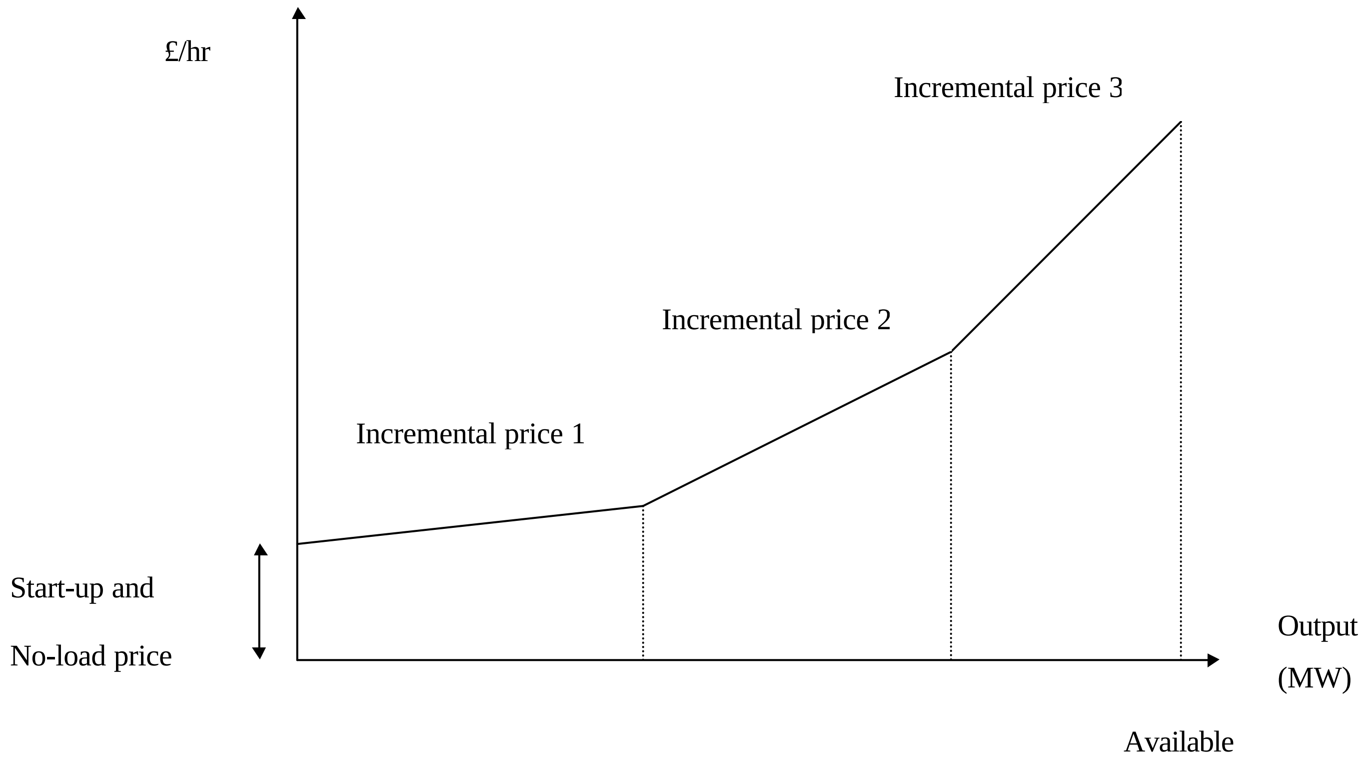


**Figure 2: Electricity Forward Agreement Trades**

*Source:OFGEM (2000a), Figure 4.2, page 35*

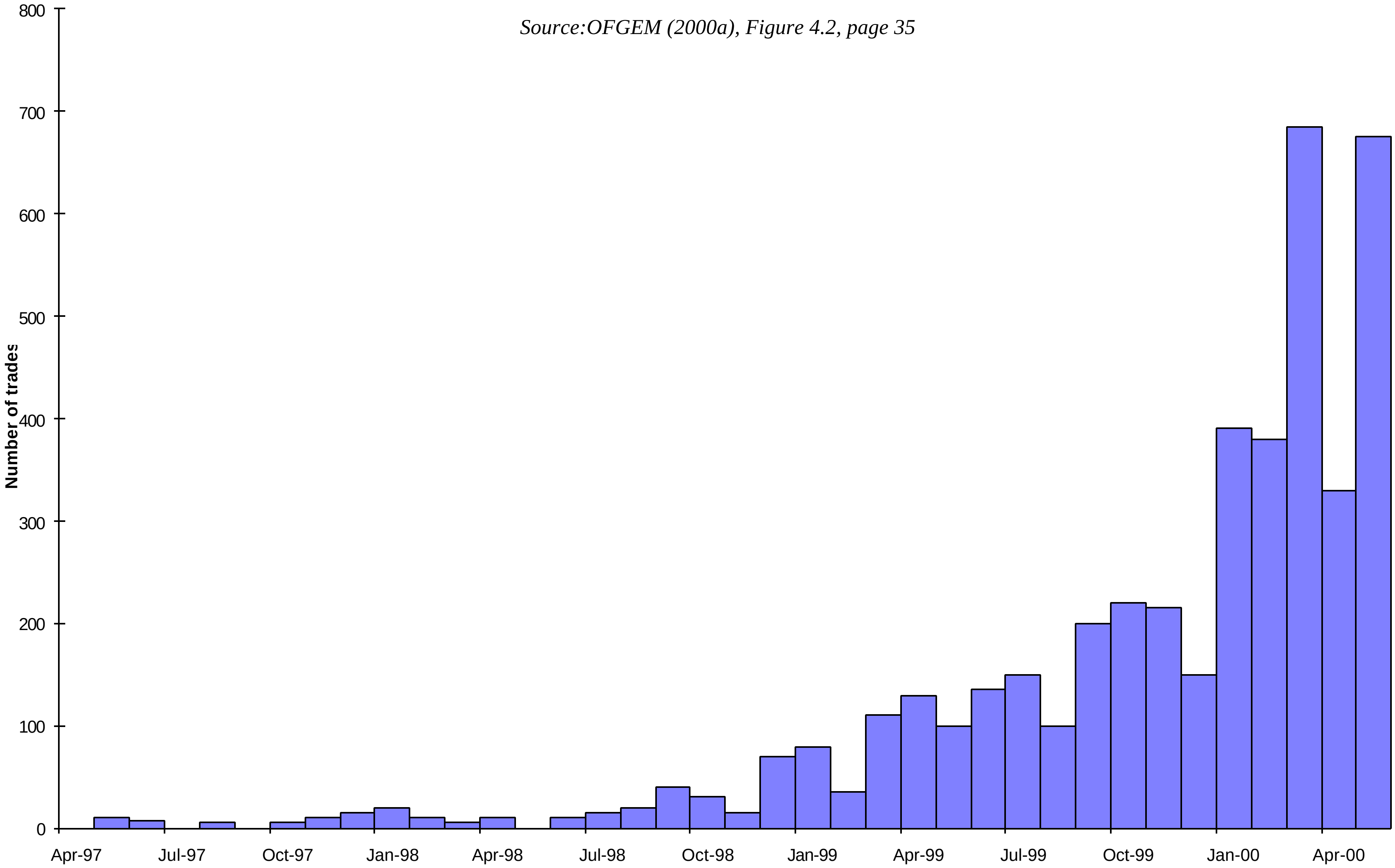


**Figure 3 : HHIs for Generation Capacity and SMP setting 1990/91 to 1999/00**

*Source: OFGEM (1999b), figure 14.3 and OFGEM (2000), Figure 5.1*

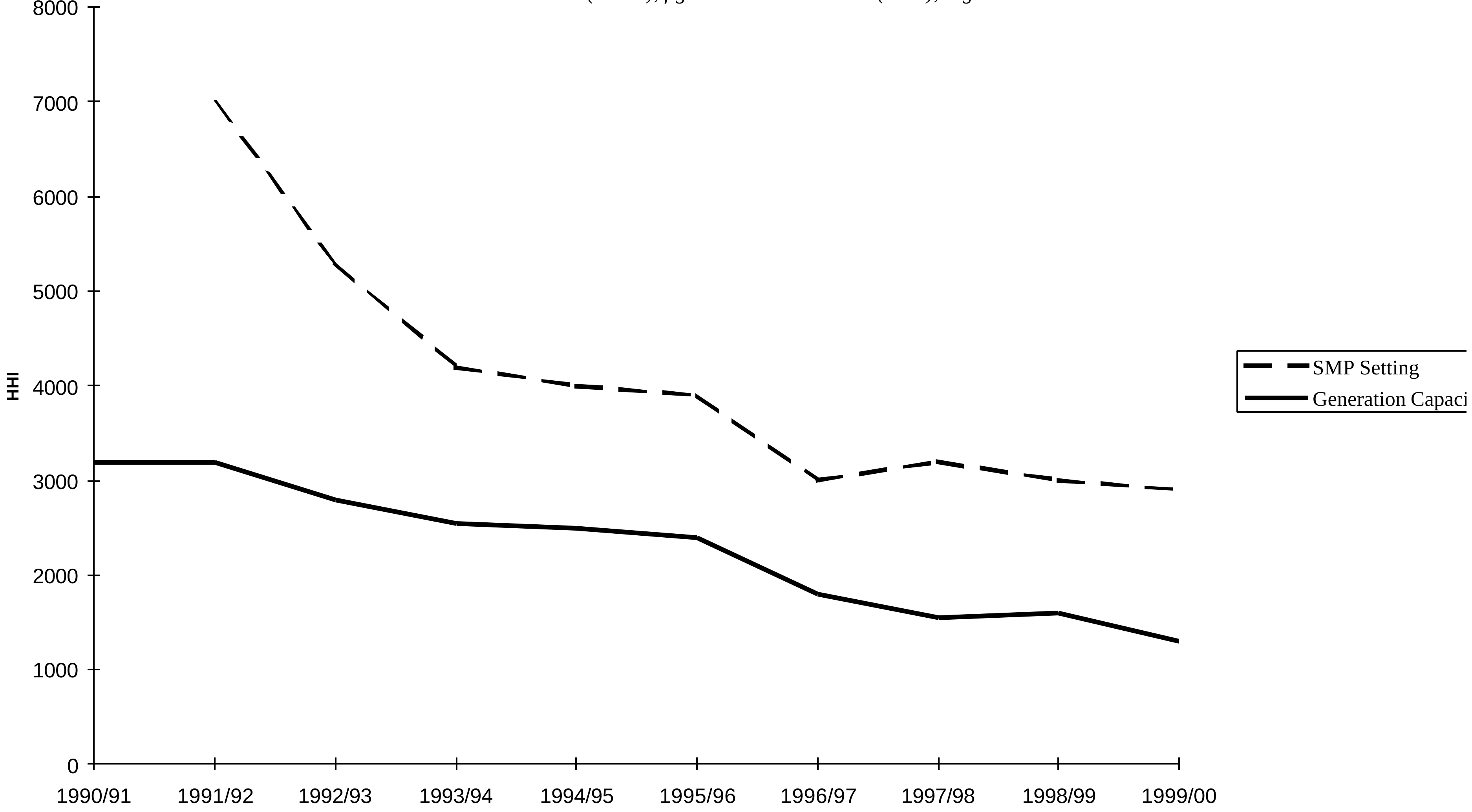


**Figure 4: Components of Monthly Average Wholesale Prices in England and Wales (1990-2000)**

*Source: ESIS Ltd. Historical Pool Price Information, available at http://www.esis.co.uk/market/average.asp*

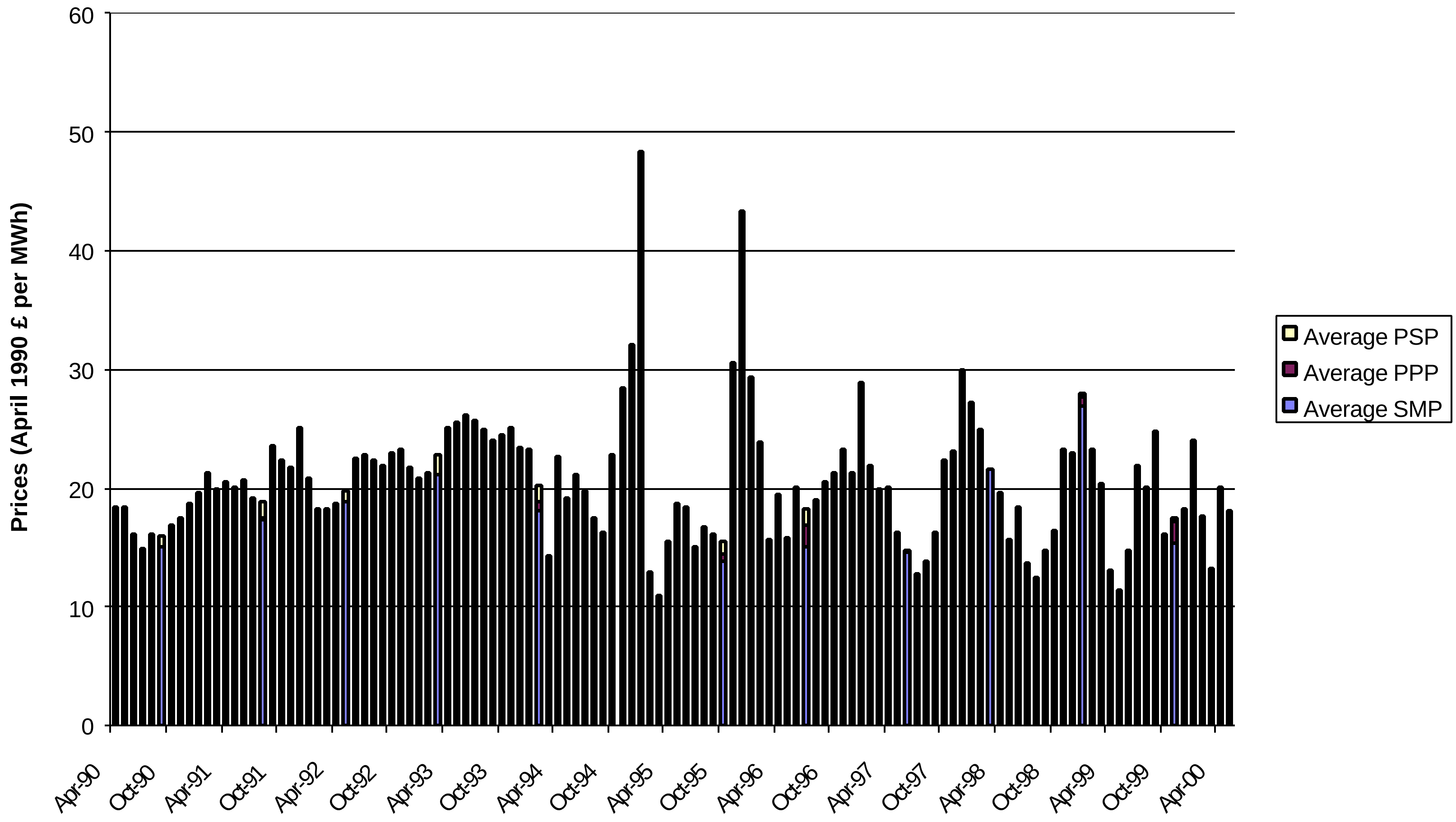


**Figure 5: Evolution of wholesale prices (12 month moving average)**

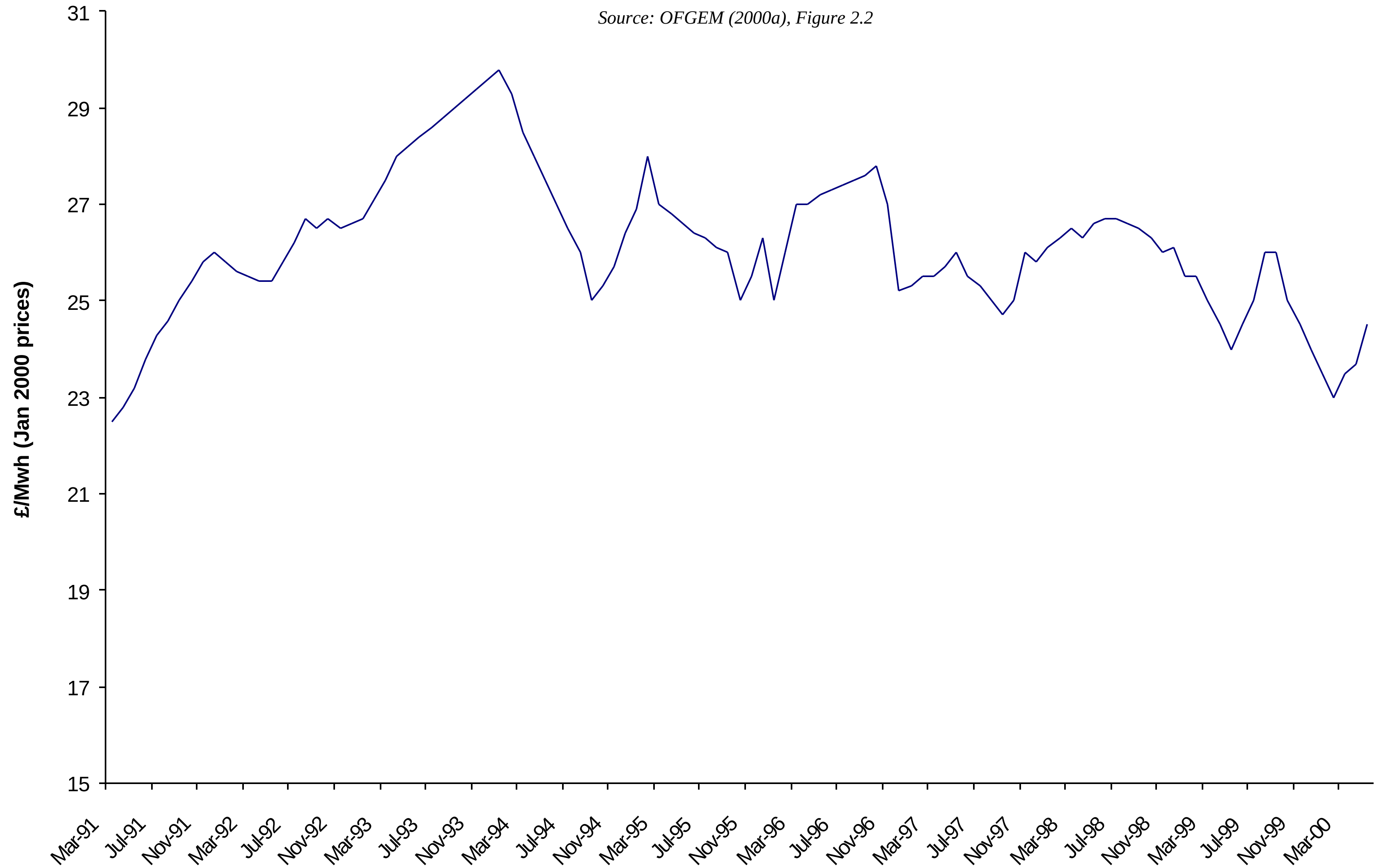


**Figure 6: Average fuel prices**

*Source: OFGEM (2000a), Figure 2.3*

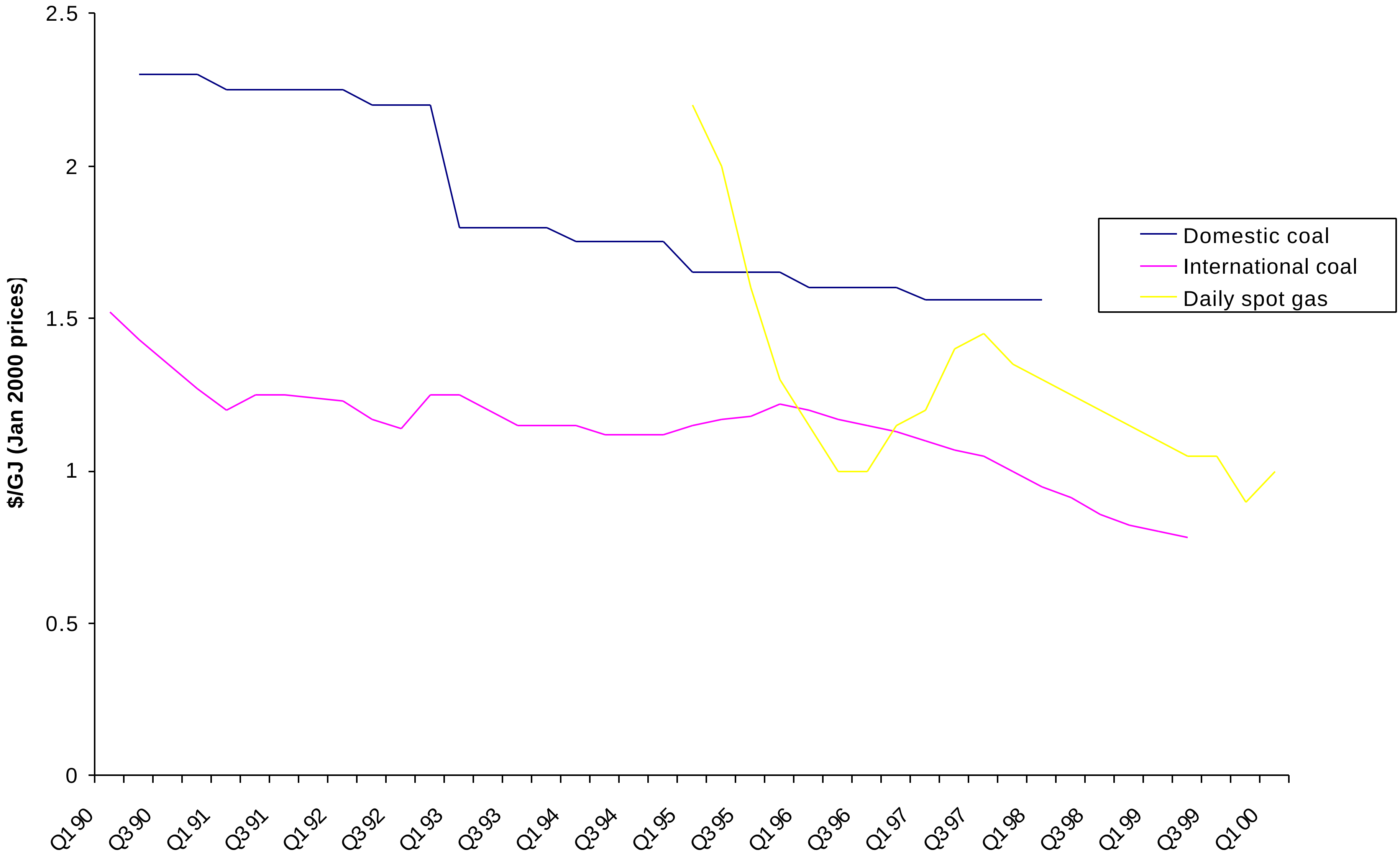


**Figure 7: Average plant efficiency**

*Source: OFGEM (2000a), Figure 2.4*

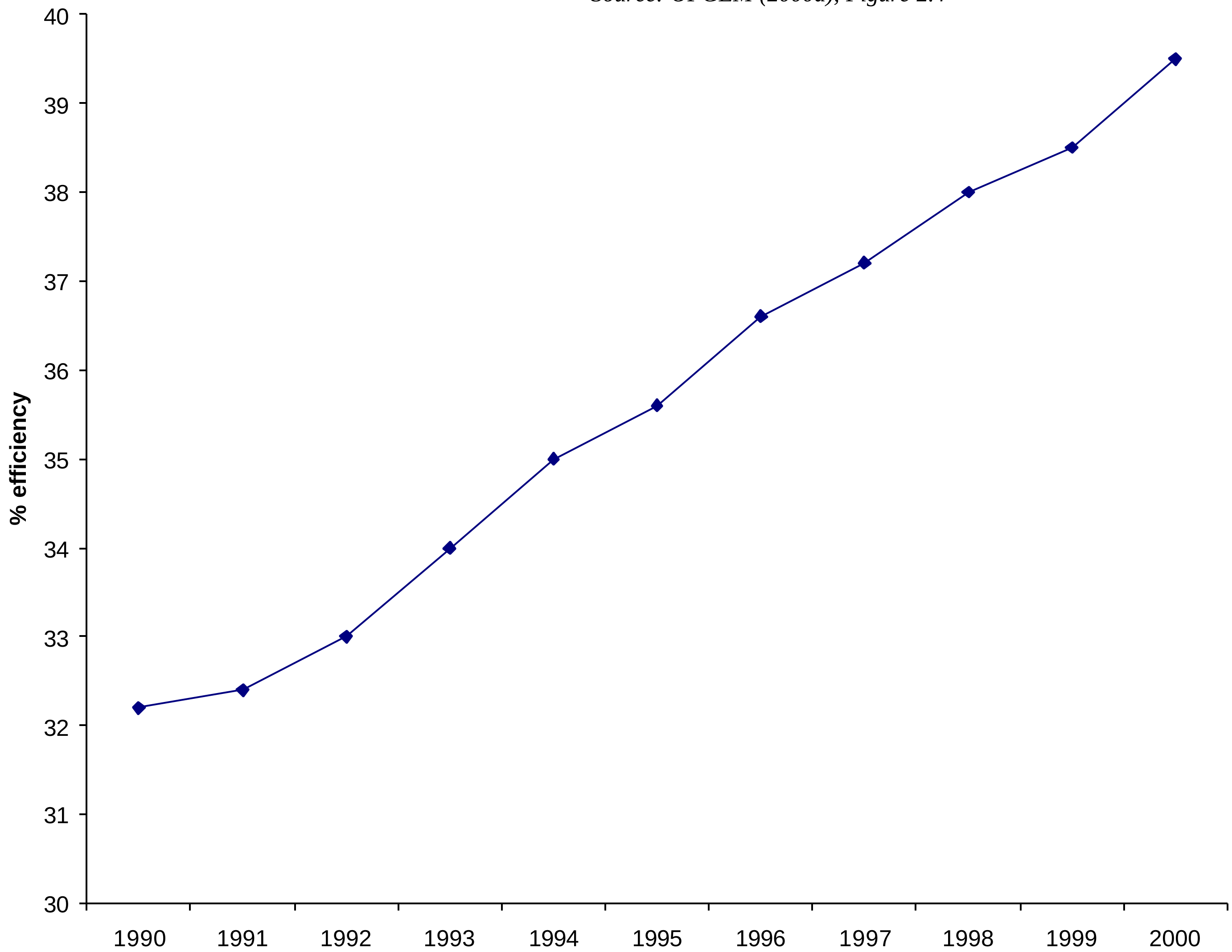


**Figure 8: Estimated CCGT Capital costs**

*Source: OFGEM (2000a), Figure 2.5*

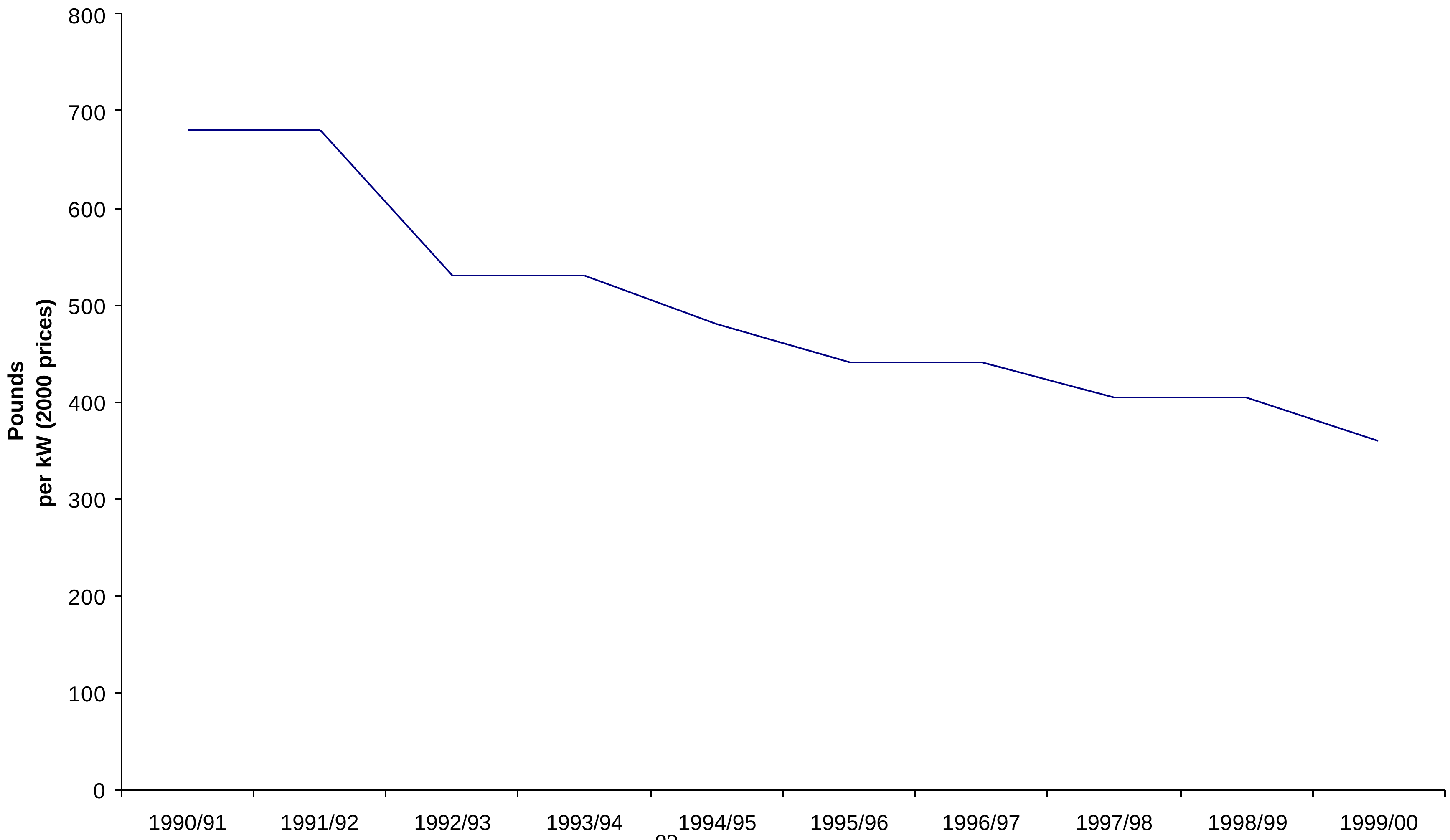


**Figure 9: Supply-demand margin**

*Source: OFGEM (2000a), Figure 2.6*

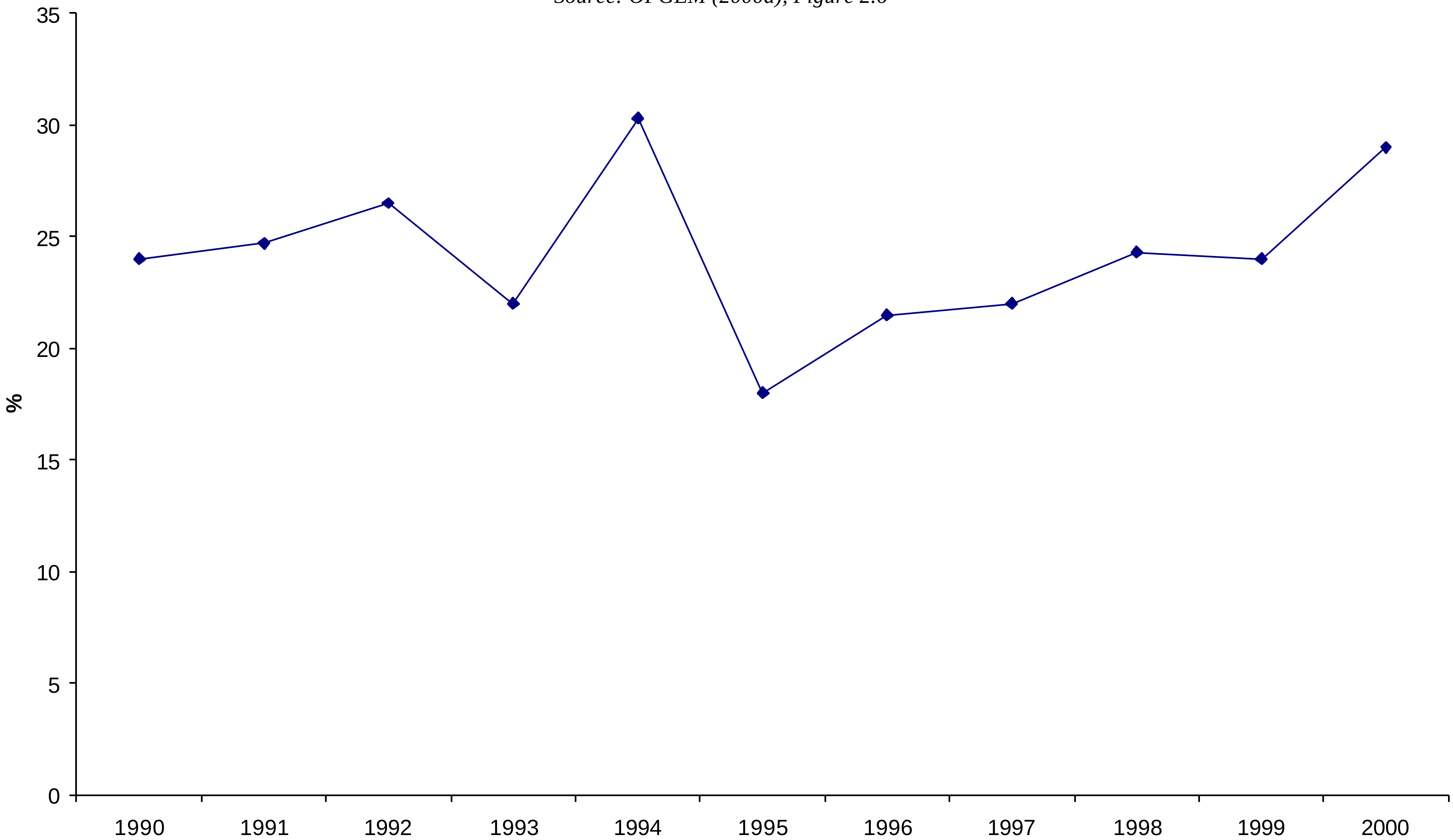


**Figure 10: Return on capital employed for National Power, PowerGen and British Energy**

*Source: OFGEM (2000a), Figure 2.13*

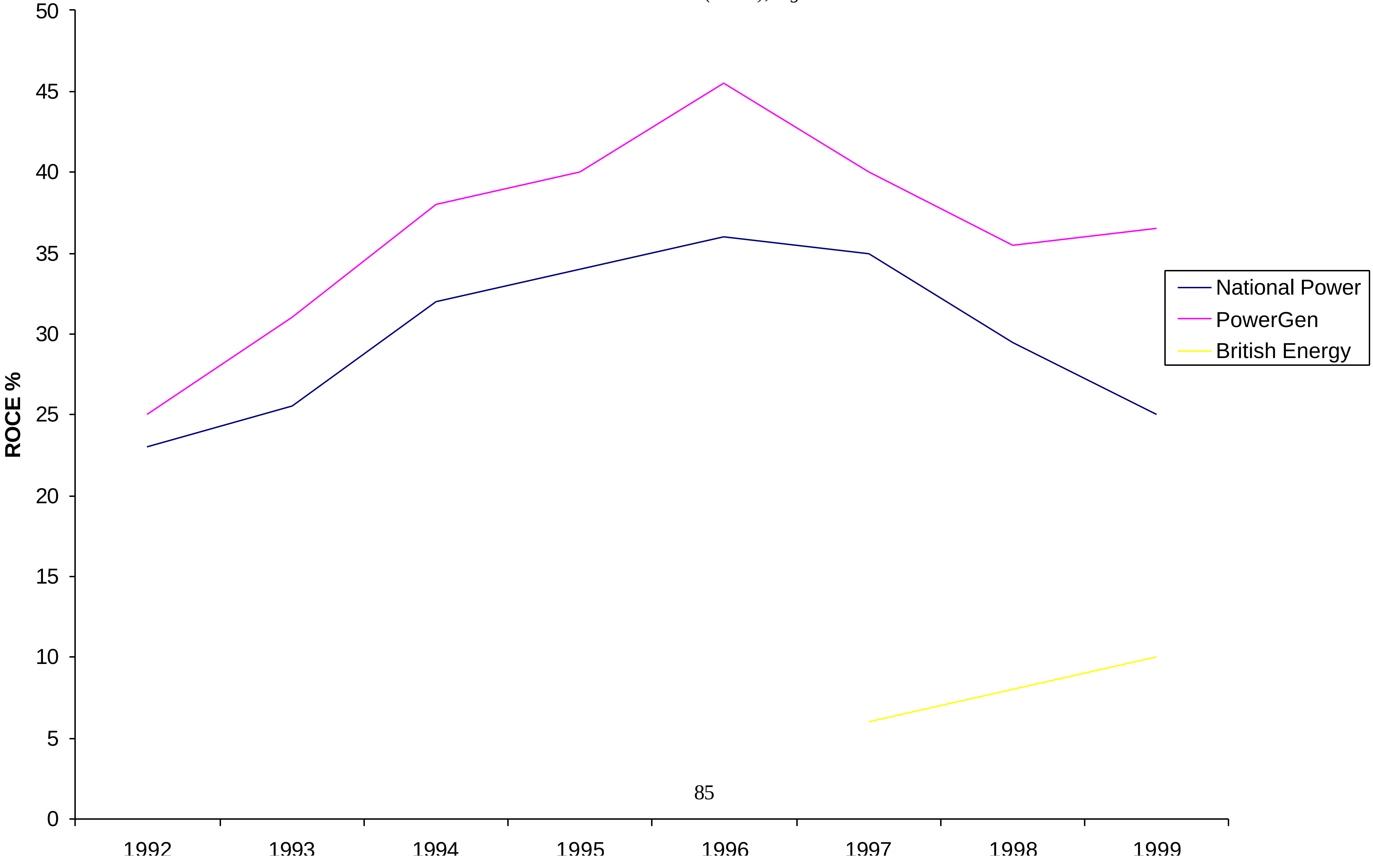
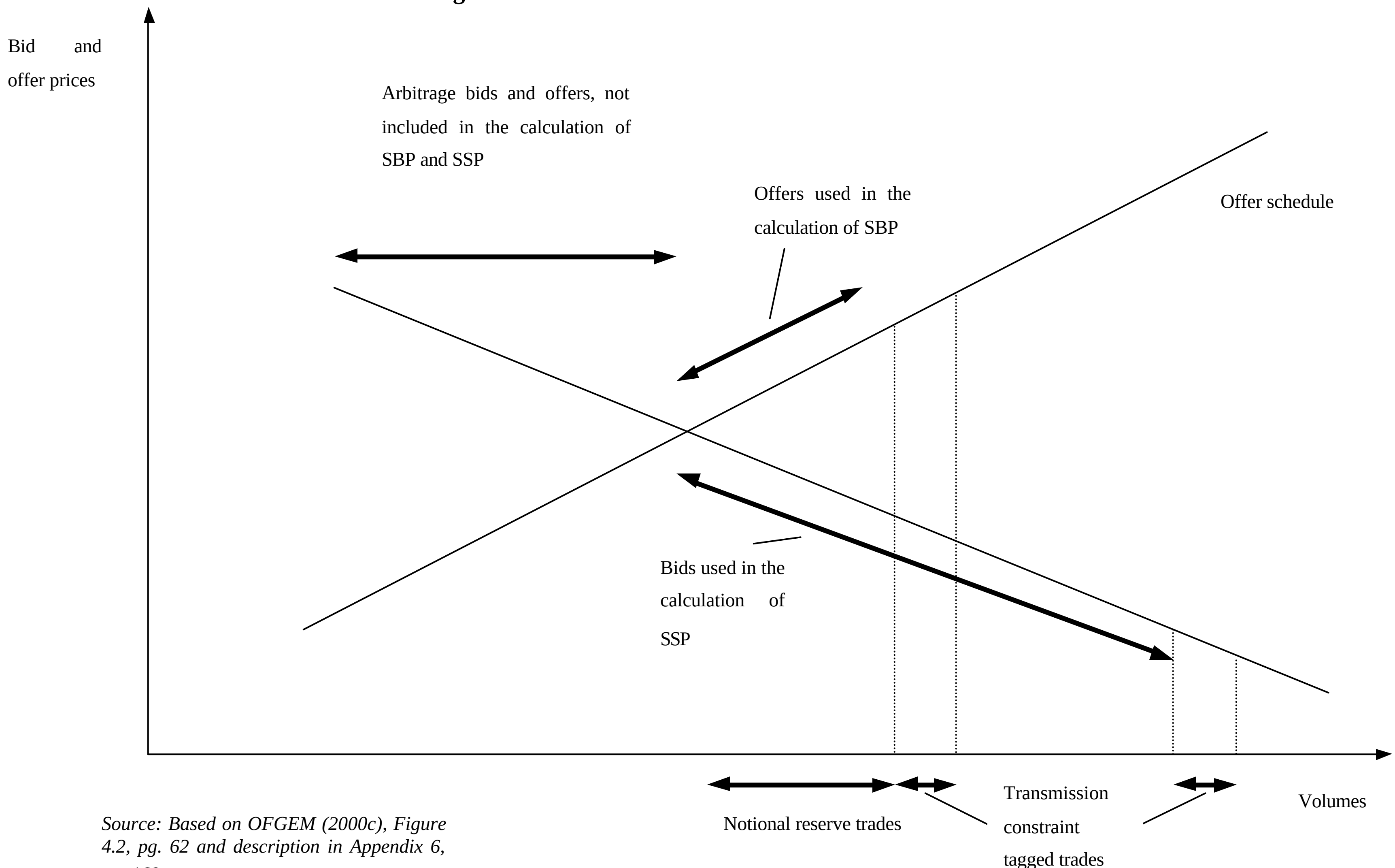


**Figure 11: Determination of SBP and SSP**




*Source: Based on OFGEM (2000c), Figure 4.2, pg. 62 and description in Appendix 6, pg. 169*

**Figure 12 : 3 node example for transmission market power**