Cash-out review 2007

"An Independent perspective"

Nigel Cornwall

"The cash-out rules are a highly complex and often controversial area of the wholesale market arrangements", Ofgem, May 2004

1. Introduction

1.1. Purpose

The Authority's decision letter on BSC modification P205 in October 2006, in addition to approving a rule change, made clear that Ofgem is contemplating whether wide-ranging change to electricity cash-out arrangements is required. "The analysis [of P205] suggests that there may be scope for further improvements in the cash-out arrangements," it said, adding there are "fundamental issues remaining with the existing cash-out arrangements that need further industry consideration". Ofgem indicated that it intends to commence a review of cash-out arrangements in spring 2007 at the latest. To help prepare for this review, it has sought views from Cornwall Energy on the current operation of the cash-out mechanism and possible options for improvement.

An earlier review of cash-out initiated in 2004 ended inconclusively, with no specific change proposals emerging from it. The focus of the 2007 review is likely to be different. On the previous occasion the review focused on issues likely to be germane to the subsequent winter, specifically security of supply issues. The terms of reference for that review therefore excluded a number of important issues from its scope, most notably single vs. dual cash-out, pay-as-bid vs. cleared price, merits of ex post trading, as well as definitions of gate closure, balancing periods and the system/energy balancing principle. The imminent review affords the opportunity to take on board some of these wider issues, and it also allows experience since 2004 to be taken into account.

This paper sets out our preliminary views on some current issues and change options. It:

- explains some of the background to, and the design issues associated with, electricity cash-out and the development of the current rules to place some of the comments in this paper in context;
- considers deficiencies with the current tagging process;
- identifies concerns arising from negative prices;
- outlines some arguments with regard to single vs. dual cash-out;
- highlights issues surrounding the operation of the so-called "beer fund"; and
- looks at other practical issues surrounding cash-out, including price signalling and transparency.

In each case it tries to highlight issues, possible alternatives and options. It also notes where relevant some international parallels.

1.2. Development of cash-out

Cash-out is the process for pricing uncontracted trades on the electricity system. In the British system the matching of supply with demand occurs over half hourly balancing periods. Under Neta, offers and bids for incremental and decremental energy are submitted into the balancing mechanism by an hour prior to each balancing period (termed "gate closure"), and these are accepted as appropriate by National Grid as system operator (SO) on a pay-as-bid basis.

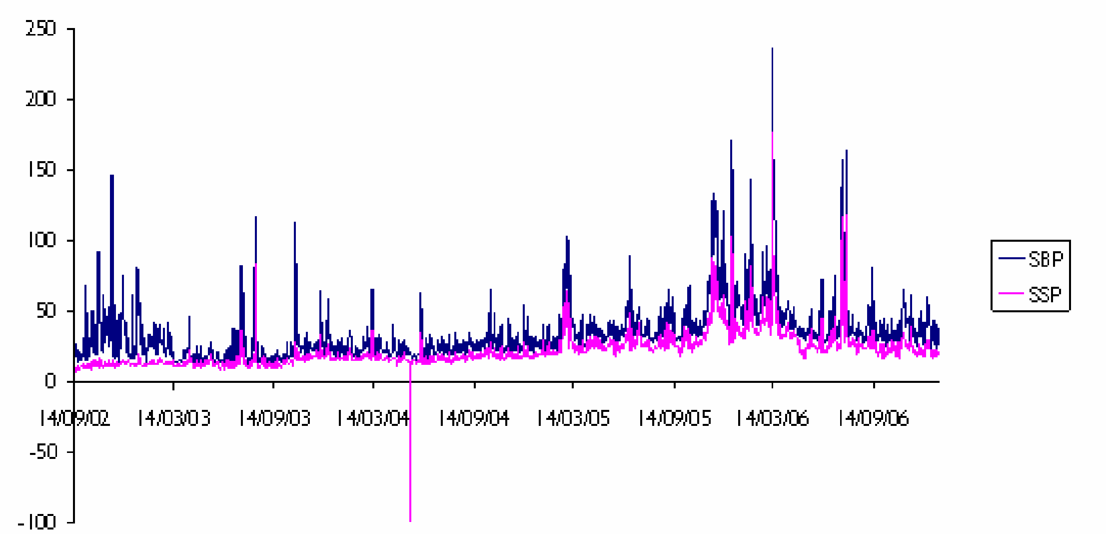
The cash-out design is based on the concept of two – or dual – prices. The main price is the price for imbalance trades *in the same direction* of the system as a whole. It will be the system buy price (SBP) when the system is short and the system sell price (SSP) when the system is long. Accepted offers are arranged according to detailed rules to calculate the main cash-out price. The second price is the reverse price, that is for trades *in the opposite direction,* to the system as a whole. Both prices are usually notified to the market fifteen minutes after the relevant balancing period.

The SSP–SBP methodology is premised on a simple concept – the desire to encourage parties to balance their positions *as an end in itself,* although there is no regulatory requirement as such that parties should do so or that they should be fully contracted. This approach means that if a trading party is out-of balance against its contract nominations[1] it will tend to see a price that is at a discount to the short-term energy market price if it is spilling (exporting) or that is at a premium to it if it is taking a top-up supply (importing) from the system. The cash-out prices that the trading party sees, and levied by the SO, are linked to actions taken by the SO that are *treated as* restoring the system to energy balance.

These balancing actions are typically sourced from higher cost plant or plant with specific dynamic characteristics. SBP, the top-up price, was on average over twice as high as the spot price in 2001 and 2002, though SBP has tended to decline over time as rule changes have removed certain types of transactions from the calculation. Also parties have become more attuned to market operation under the new trading arrangements. In particular, SBP has significantly reduced since modification P78 was implemented in June 2003, a change which set the reverse cash-out price equal to a market-based spot price rather than as previously to the weighted average of specific balancing trades. Even so, SBP was on average 14.7% over the spot price in 2005 and 17.8% over it in 2006. In the case of SSP, it was on average 17.2% below the spot price in 2005 and 18.6% below it in 2006. Average price trends are shown at figure 1.

[1] As adjusted for its bid/offer acceptances and measured against loss-adjusted metered volume.

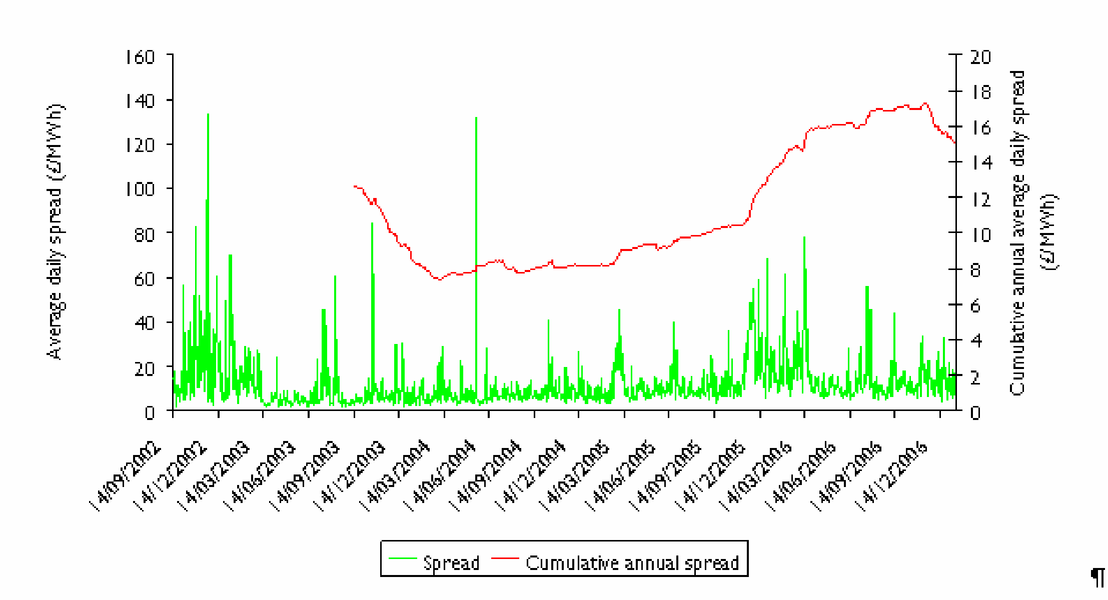
Figure 1: Average daily SBP and SSP – 2002-07



The spread between the two imbalance prices is shown at figure 2. The cumulative annual average daily spread is shown on the right axis. Some commentators have observed that the difference between the two cash-out prices is in effect "a tax" on buying and selling uncontracted electricity, though the extent of the spread is on average lesser since P78 was implemented. The fact remains dual cash-out systematically produces a situation where the main price will on average be at a significant premium or discount to traded energy prices, and the average differentials at around 15% or more over the past two years is significant.

One important consequence of the dual price approach – arguably its key design feature – is that it better enables the SO to balance the system in short time-scales, especially given the expectation that imbalance costs will rise as the system becomes shorter. Such an approach had obvious merit given the newness of the self-commitment/decentralised despatch model implemented under Neta in March 2001. At that time there were also policy concerns regarding the ability of generators to game centralised markets, which gave rise to a desire by Ofgem to establish a strong incentive for trading parties to enter into contracts to mitigate their interest in short-term prices following the abolition of the Pool. Both factors are now significantly less important.

Figure 2: Spread between average daily SBP and SSP – 2002-07



¶

An outline of the development of the imbalance cash-out methodology has been set out on various occasions, and need not be repeated here.[2] But five aspects of the current arrangements should be highlighted as they provide the cornerstones of the current rulebook and are the key drivers of the observed prices:

- the first is the introduction of the concept of the **continuous acceptance duration limit** (CADL) by P18A, CADL tagging and the removal of certain deemed system trades from the energy price stack which it effects. Since mid 2001 acceptances of less than 15 minutes duration have been assumed to be system-related and removed from the pricing of energy imbalances;
- the second is the introduction of the removal of **de minimis trades** (defined as acceptances by the SO of less than 1MWh) by P10, again justified on the grounds that small acceptances might be necessitated by requirements to maintain system frequency (that is, they also relate to the system balance);
- the third is the replacement of the weighted average price calculation on the shorter price stack and its replacement by the **reverse price** indexed to relevant prices in the short-term energy market under P78;
- the fourth is the introduction into the pricing rules of the concept of the **net imbalance volume** (NIV) and the associated mechanism of NIV tagging for the purposes of setting the main price, which was also implemented under P78. This change also introduced a key change by using the shorter stack representative of "undo" actions and netting it off the larger stack. While the shorter stack is not assumed to comprise system trades as such, they are deemed to be not relevant to the definition of the true energy imbalance, which is reflected by the net, not the gross, imbalance volume; and
- the fifth is the more recent shift from the average weighted calculation of the main

[2] See *Electricity and cash-out review*, Ofgem (May 2004), especially pages 15-21.

price, first by the implementation of P194 initially by taking the top 100MWh (PAR100) of the stack, which in turn was overwritten before implementation by use of the top 500MWh (PAR500) calculation under P205. This change gave rise to a **partial marginal imbalance price**, which went live in November 2006.

The first four of these changes (especially in combination) have had the effect of narrowing the spread between the two imbalance prices over time. This trend has been partly counter-balanced since early 2004 with the increase in global fuel prices, which has increased the cost of all balancing actions, a tendency that has only recently started to reverse. The narrowing effect is likely to be offset *to a degree* by the introduction of the PAR mechanism as this is specifically designed to have the effect of producing imbalance prices closer to the marginal cost of acceptances made during periods when the system is under stress and/or when the NIV is shorter.

A table of the most relevant cash-out pricing rule change proposals is at Appendix A.

2. Perspectives on cash-out design objectives and properties

The basic design concepts underpinning imbalance cash-out are summarised in the following quotes from key Ofgem documents:

> "A dual cash-out mechanism provides incentives for parties to **contract ahead** to meet their customers' demands"; "[It] provides incentives for parties to contract ahead to meet their customers' demands, as those parties that are long are likely to receive a lower price for electricity than if they had been fully contracted and parties that are short are likely to pay a higher price than if they had been fully contracted"[3].
>
> More specifically: "The cash-out arrangements are designed to **target the costs of energy balancing** to the **parties who create the costs**...they do this by imposing imbalance charges on parties who are not in balance that **reflect the costs incurred by the SO in rectifying the imbalance**".[4]
>
> And again: "Generators and suppliers are not under obligations to balance and can choose to pay the cash-out price. But the cash-out price should **correctly signal to them** the [SO's] cost of [energy] balancing. In response to this signal, generators and suppliers should try to balance their own positions if they are able to do so **at lower cost than [the SO]**".[5]

We have added the emphases to highlight the key desired properties sought by Ofgem. There are at least six specific elements of this design philosophy that should be commented on as they are the main drivers of the detailed price formation rules. They are:

[3] *Electricity and gas cash-out review*, Ofgem (May 2004), page 17.

[4] P194 decision letter, page 2.

[5] As above.

- the use of the dual pricing methodology;
- what is meant by energy balancing;
- what we mean by targeting the costs of energy balancing;
- how we define trading parties who create the costs;
- what is meant by reflecting the costs incurred by the SO in rectifying the imbalance; and
- the assumption that balancing actions can be achieved at lower cost by trading parties than by the SO.

We discuss each of these below, in order to highlight what we believe are gaps between theory and practice. In some cases we also question the appropriateness of the objective in today's environment.

2.1. Dual prices

Although the basis for calculating imbalance prices has been modified over time, the dual price approach has been a constant. It has been the "most controversial and criticised feature of the new trading arrangements", according to one commentator. He also observed that dual price cash-out "was not a necessary concomitant of Neta, but the result of a separate decision as to how cash-out prices should be set".[6]

The use of two prices is not of itself controversial, a fact demonstrated by its application in a number of electricity markets internationally. A variety of different approaches are evident. Other two-price systems include France, where a single price per period is discovered, but specified premia and discounts are applied to produce two different prices for each balancing period and a spread to incentivise balancing. Nordpool, with the exception of Norway, uses two cash-out prices in a way that is conceptually similar to the British approach, but which then distinguishes between situations where the imbalance has been to the benefit or disbenefit of the system as a whole.

A number of other jurisdictions use a single price. In Texas a single market clearing price for energy is calculated immediately ahead of the balancing period, while in Norway a single imbalance price is calculated after the event ("ex post"). While details differ, NYISO, PJM and ISO-NE use a two-settlement system model,[7] which comprises a financially-binding day-ahead market and a real-time balancing market which produces a single price. Real-time prices are based on actual hourly quantity deviations from day-ahead scheduled quantities calculated every five minutes just after the beginning of each five minute period and integrated over the hour. The prices are locational.[8]

[6] *"Small suppliers in the UK domestic electricity market: Experience, concerns and policy recommendations"*, Professor Stephen Littlechild (29 June 2005), pages 50-51.

[7] Average prices between the day-ahead and real-time markets diverged by -2.2% (NYISO), -1.1% (PJM) and -0.8% (ISO-NE) respectively March – December 2003.

[8] PJM averages transmission losses, NY and New England does not. All three markets include transmission constraint costs.

Two other markets are worth noting, and both of these operate a single organised market and settlement process and do not specifically have a mechanism for dealing with balancing energy. In New Zealand there is a single price (albeit one that is applied on a nodal basis taking into account transmission constraints and losses across the system, and which also covers instantaneous reserve costs), though trading through the NZEM is not mandatory. In contrast, the Australian NEM is mandatory, and this trading system also results in a single clearing price based on aggregate supply and demand in the balancing period.

Clearly there are many legitimate design choices with regard to pricing uncontracted energy. The use of one or two (or differentiated zonal or nodal) prices is a key element, which is clearly linked to wider market design. However, we believe it is the detailed rules of price formation that give the British cash-out arrangement its unique features, especially the tagging rules, not the dual pricing approach. A further differentiating factor is the scope accorded to the SO in entering the energy markets to procure reserves outside of its balancing services contracts. Another is the complex process for allocating between energy and system based on fixed rules.

Of course, market structure cannot be seen in isolation, and interaction with industry structure also needs to be taken into account. The dual cash-out price arrangement, because smaller non-integrated ones are more exposed to imbalance than larger more vertically-integrated ones, systematically favours larger players than smaller players. They therefore pay a larger share of the net imbalance revenues.[9] The Neta cash-out arrangements are generally regarded as compounding powerful pressures to consolidate and re-integrate, and aspects of the trading arrangements could well need to be reviewed as a consequence of the degree of change in industry structure that has occurred since 2001.

2.2. Defining energy balancing costs

The concept of the energy balance is central to the formation of imbalance prices under Neta, and they have been designed to try to reflect the cost of matching supply and demand for energy only. This feature means that cash-out prices should exclude the costs of non-energy actions (for instance, the costs of system services such as reserve procurement) and the effect of locational and operational constraints. If this separation is achieved, the conditions should arise that permit efficient arbitrage between cash-out prices and forward energy markets.

Post implementation of P78 the main price only (and not the reverse price) is derived using eligible balancing actions taken by the SO to alleviate the net imbalance volume. What constitutes “eligibility” is a moot point. In reality there can be no fixed definition of the costs associated with energy balance (as opposed to the system balance), and there is no definitive way of classifying them in all instances. Some of the SO’s contract

[9] The average imbalance of the six largest integrated players on four sample days (two stress and two benign) was roughly a third of that by the six largest independent suppliers. The days selected were 9 March, 13 March, 15 April and 18 July 2006.

commitments (through its balancing services, energy purchases or balancing mechanism acceptances) cannot be described as falling precisely into either category, although certain general types of action have certain attributes that can make them more closely related to either energy or system balance. Thus energy balancing actions are non-locational and would include reserve holding and acceptances associated with maintaining the operating margin. System balancing actions include, but are not limited to, frequency control and the alleviation of locational transmission constraints.

The allocation process for determining what actions are attributable to which balancing activity is carried out after the event, and it is done through application of a series of approximations and adjustments which taken together constitute the tagging rules. The end point of this process is a position whereby the volume of actions left in the energy stack represented by the NIV in any balancing period are deemed to represent energy trades and are the basis for calculating the main imbalance price.

Tagging rules presently apply to:

- de minimis trades under 1MWh;
- short duration trades defined as acceptances of less than 15 minutes; and
- NIV tagging (where the shorter stack is offset against the longer stack).

All these elements are removed from the price stacks assembled for each balancing period based on balancing mechanism acceptances on the grounds that they capture actions that do not relate typically to the energy balance, along with arbitrage trades.

Other adjustments to the stacks also relate to items defined as energy rather than system from actions taken under balancing service contracts transacted ahead of gate closure. In these instances adjustments are made to the stacks after the system actions have been removed. The BSAD and ABSVD methodology statements produced by the SO under its transmission licence set out how information on such balancing actions must be submitted for the purposes of determining cash-out prices, which is then processed under the BSC. The main adjustments[10] are:

- reintroduction and reordering of CADL volumes;
- insertion of the system BSAD volume; and
- incorporation of the energy BSAD volume with the associated prices.

The drafting of these rules implies considerable precision in the definition of these adjustments as either energy or system. As we have noted, in reality some balancing trades are not specifically allocatable, and some are actioned for a mix of reasons or are contingent on circumstances. To illustrate this mismatch between theory and practice, any trade over 1MWh lasting over 15 minutes and remaining in the energy stack, but which is not the cheapest available offer/bid at the point of acceptance, is likely to be

[10] In all there are eight BSAD components, six relate to electricity balancing actions and two relate to system balancing.

attributable to some form of system constraint, but the cash-out rules do not reflect this.

Concerns about deficiencies with the existing tagging rules have been a feature of all recent cash-out modification decisions from Ofgem. One of the primary arguments for rejection of P136/7 was that the proposals "could create distortions in the market as cash-out prices would have been set by a small and possibly unrepresentative volume of accepted offers".[11] Then, in its P194 decision, Ofgem noted "that the effects of tagging imperfections could be increased as a result of the implementation of P194 pricing". Further on in that decision letter it notes that, although implementation of P205 significantly reduces the impact of imperfect tagging on cash-out prices, "it appears that actions taken for system constraint reasons can still contribute to the calculation of the price". As a consequence, a key issue at the heart of the P205 decision was a recognition that replacing a PAR100 mechanism with PAR500 would reduce the potential distortions to cash-out prices relative to the then PAR100 baseline. It is "on balance appropriate to accept the proposal as these distortions are likely to be detrimental to both the economic and efficient operation of the system and to effective competition," said Ofgem. Its conclusion was that the benefits of reducing price distortions caused by imperfections in the tagging mechanism under P205 outweighed the potential detriment resulting from any reduction in the signal to balance under P194.

Analysis by EDF Energy[12] during assessment of P205 suggested that the potential distortions under a PAR100 approach, as evidenced by the increase in negative SSPs, were real, but that PAR500 would mitigate this effect. Other analysis produced by Elexon showed that the PAR100 methodology would have changed cash-out prices in 83% of periods in 2005-06, whereas the PAR500 would have changed cash-out prices in only 23% of periods. But in periods when the system was shortest, the PAR500 price was within about £3/MWh of the PAR100 price. Ofgem considered this modest gap to be a minor disadvantage compared to a position where the main price could be routinely skewed by the inclusion of expensive system actions.

According to the P205 decision letter, National Grid subsequently confirmed that in 56 of the 63 periods in the PAR100 data set it had taken actions to resolve system constraints at the Cheviot boundary [and that these appeared in the energy stack], which suggested that the problem was even more pervasive than suggested by EDF Energy's analysis. "In our view", Ofgem concluded in the light of this information, "this provides strong evidence that cash-out prices would be polluted by actions taken for system reasons under the PAR100 methodology." And "the apparent reduction in the impact of imperfect tagging brought about by the proposal [P205, because it deepens the stack] would be beneficial to competition as compared to the baseline."

Taken in the round these statements and the supporting analysis provide compelling evidence that the current rules do not result in the accurate identification in many

[11] P136/7 decision letter, page 3.

[12] A presentation of the analysis is at www.elexon.co.uk/documents/modifications/205/P205_UrgentConsultation_Attachment3C_FurtherAnalysis_14thAugust2006.pdf. See especially slides 7-9.

periods of the true energy balance.

There are several options for enhancing the tagging process. Any alternative mechanism would need to address by whom it was actioned as well as how. In terms of how:

- a resource-based approach could be applied perhaps based on types of acceptance or instruction, perhaps by type of balancing service;
- specific locational trades could be excluded;
- any trades accepted out of price order could be excluded;
- some form of cap and floor tagging could be applied (which might also deal with "sleeper bids" (see below)); or
- some combination of the above.

In terms of who and when, options include:

- permit the SO to carry out discretionary tagging in real-time, subject to some form of ex post audit;
- permit the SO to carry out tagging in accordance with a defined procedure after the event specified under a compliance statement developed under the transmission licence;
- develop some form of automated tagging process applied as part of the clearing process by Elexon or one of its agents under a procedure specified in the BSC; or
- some combination of the above.

A different approach, also feasible, might be to construct an ex post unconstrained schedule for each balancing period by examining all eligible trades and resolving this volume in the cheapest possible way allowing for dynamic parameters. A similar approach is already applied in Spain, and one is also proposed for the All-Ireland market. More radical options may well exist, but they would require revisiting other elements of the market design (for instance changing definitions of gate closure or balancing period).

No method will be perfect, but a key enabler to allowing a rational assessment process could well be for the SO to begin classifying acceptances as they are taken and then communicate this information to trading parties.

2.3. Targeting costs

The terminology applied with regard to "targeting" and "price signalling" is misleading in some important respects. In fact the actual cost of imbalances is recovered not through the energy imbalance prices at all, but through the mechanism of balancing services use of system charges (BSUoS), which smears the actual costs of balancing the system across all grid users. But trading parties in energy imbalance do "see" the costs associated with energy balancing actions through the energy imbalance prices albeit only once they have been calculated after the event and even though the payments and receipts are then cancelled out.

It is important to qualify this concept of cost targeting in other ways, because of other factors, such as:

- the process for the construction of the price stacks (and therefore the method of targeting) and the derived prices has changed significantly on several occasions through introduction of new tagging rules, application of netting and most recently introduction of PAR values. This situation suggests based on experience to date that there is no "correct" formulation and also that perceptions of what is appropriate have developed in the light of experience of the trading arrangements;
- the energy/ system split and tagging is now irrelevant for the purposes of calculating the reverse price (the price for imbalances against the direction of the market as a whole in a particular balancing period), and the reverse price calculation has been deliberately delinked from the SO's costs through indexation to prices realised in the short-term energy markets; and
- because the imbalance cashflows in a balancing period are made whole (i.e. are cleared to zero) through application of the Residual Cashflow Reallocation Cashflow (RCRC) mechanism, this has the effect among other things of resulting in adjustments to trading parties' energy accounts that change the net impact – and therefore costs – of the energy imbalances they incur.

Nevertheless recent Ofgem decision letters indicate the concept of targeting costs remains key. But because imbalance prices do not reflect a clear separation of energy and system, we conclude that some costs are inappropriately targeted and distort the prices paid by those in imbalance.

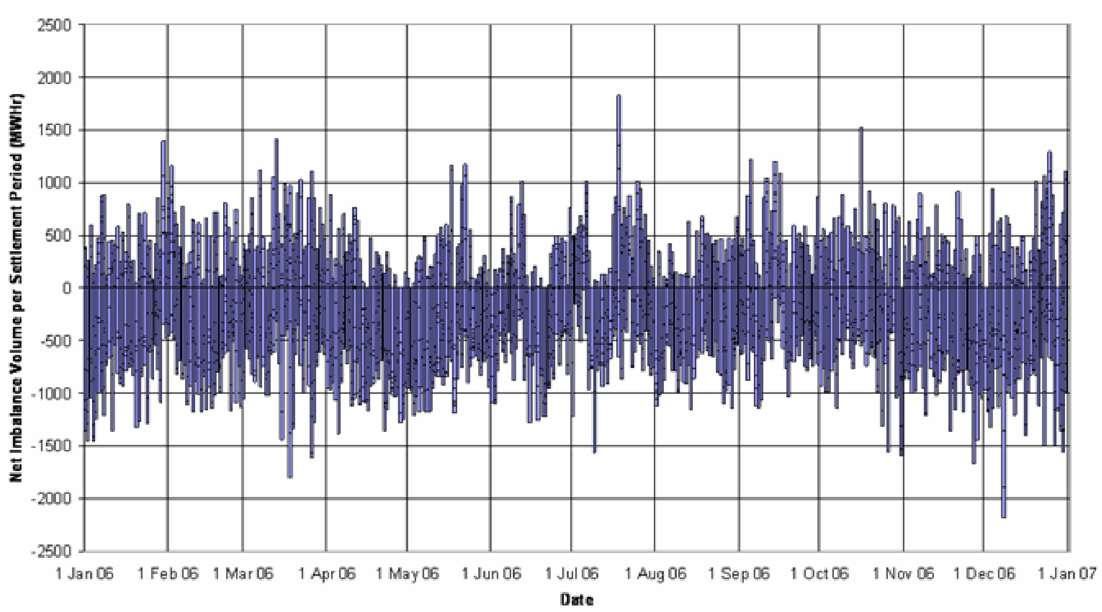
2.4. Parties who create the costs

The concept of causer pays is central to the imbalance cash-out design concept under Neta. But what does or should this mean?

All electricity systems are prone to trading parties being in physical imbalance against their contracts. Even with strong incentives and efficient operation there will always be a requirement for the residual system balancer, and an integral part of its role will be to ensure the energy balance in real-time both nationally and locally. Trading parties do not (indeed cannot) meet their commercial intentions and contract notifications in many instances. The total energy imbalance volume as a percentage of contract volume for all production accounts (i.e. generation assets) in 2006 was 0.13% (i.e. long by 0.13%). The corresponding figure for all consumption accounts (i.e. suppliers) was -1.16% (i.e. long by 1.16%). Figure 3 below displays the trend in total account energy imbalance volume on a settlement date resolution.

Analysis of the underlying data suggests the production accounts tend to move around the zero line whereas the consumption accounts are more consistently long. Generators have more control over their assets and while they try to remain long the amount is usually small. Most conventional generators can usually avoid a shortfall over the balancing period except in the event of unforeseen operating problems or plant breakdown. In practice they can also over-generate at times when SBP is the main price knowing they will see prices that are likely to reflect a premium; at times when SSP is the main price they have the insurance of receiving spill prices.

Figure 3: Imbalance volumes - 2006



In contrast suppliers have less control over their customers' demand, and they are also incentivised to be long but must be long by a larger amount to be safe. As a result, the task for the supplier tends to be more difficult and riskier. It is commonplace for a forecast error on demand of about 3% on supply reflecting the sensitivity of demand to temperature, and there is limited controllability of demand except in exceptional price conditions for certain large metered customers. This situation does not mean that all suppliers are impacted to the same extent; scale suppliers are likely to have a diversity benefit from a spread of load. The extent of the net exposure to imbalance is also reinforced by operation of the "beer fund", which also tends to favour scale players.

Setting aside these differences, any trading party who is consistently short or long over a sequence of balancing periods may know in advance roughly what its exposure in terms of volumes to imbalance will be. However it might flip between exposure to the main or reverse prices depending on the position of the system, and it would have little real ability to understand the respective levels of prices other than by reference to complex historical correlations. Thus the issue for trading parties is not how to avoid imbalance but to what degree they see it and whether this imbalance in a particular balancing period is with or against the system. Some players, especially smaller players, do not attempt to forecast imbalance prices because of the complexities and uncertainties involved, and there is no provisional or real-time price available in the British trading system to help them manage the risk of imbalance exposure.

While a participant will not know in advance what the cost of imbalance will be, the operation of the rules means that on average it is somewhat better to be long rather than short. This situation will almost always be the case as the cost of imbalance when the system is long is represented by bid prices of generators being backed off, which typically relate to the avoidable fuel costs of generation. The average SSP was

£26.70/MWh during 2006. In contrast a short system will see the full incremental cost of a marginal generator brought onto the system. The average SBP during 2006 was £39.75/MWh, giving rise to an average spread – a price differential – of over £13/MWh. Over this period the system has been long 73.9% of the time; it has been short 26.1% of the time. This characteristic has been manifest throughout Neta.

A long system is good for the SO as it creates free reserve, which is one reason presumably why the SO seems to prefer the incentive properties of the current pricing arrangement. But this situation also creates a position where the volume of necessary balancing actions on the system is exaggerated most of the time because of the length of the system in most balancing periods. Many balancing actions relate to backing off surplus generation. Again, the costs that arise are then targeted onto parties in imbalance. Combining these points, cash-out routinely creates a situation wherein all physical trading parties see an imbalance cost that is unavoidable. In addition suppliers, especially non-integrated suppliers, see a further artificial cost premium produced systematically by the dual price system. This creates, on our view, a strong argument in favour of some form of tolerance band,

The increased risk of higher prices under part-marginal pricing rules could well aggravate pressures to over-nominate contract quantities and result in a larger volume of inefficient balancing trades being transacted. It also creates a larger *de facto* rental to generators through a greater volume of decremental trades though the balancing mechanism than might otherwise have been needed.

2.5. Trading parties reflecting the costs incurred by the SO

Few trading parties would disagree with the basic premise of the desirability of cost-reflectivity in securing efficient balancing arrangements - the issue is how best to achieve delivery of this objective while also taking into account other objectives set and practical limitations. But there are important definitional issues about *what we actually mean by the term "cost-reflective"*. Do we mean *cost replicating* or simply *representative of the costs*? Should it be the *average cost* or a *marginal cost* (or as at present some mix of the two)?

In various recent decision letters Ofgem has made it clear that it wishes to reflect and allocate the marginal costs incurred by the SO even if this increases the overall cost of imbalance to trading parties and the size of RCRC cashflows. Some in the market have argued that changing the pricing rules so that the price signal is set to reflect the SO's marginal costs, and thereby over-stating recovery of its actual costs, is inappropriate, especially in a system where pay-as-bid applies on the input side. This is an argument with which we have some sympathy.

We would also note that:

- the term "cost-reflective" has meant differing things at different times[13];

[13] In earlier decision letters Ofgem seemed to be more concerned in recovering the actual costs of the SO.

- implicit in the term is the precept that the costs are reflective of energy, not system-related, costs;
- distortions inherent in actions that have the effect of polluting the energy stack are amplified under marginal approaches; and
- more marginal approaches also increase distortions inherent in the RCRC true up.

Setting aside economic theory (which can be used for and against marginal pricing depending on the market context of its use), there are persuasive arguments that a strong marginal signal should only be contemplated where the imbalance pricing rules enable discovery of an unconstrained energy price.

2.6. At lower cost than the SO

The design concepts that underpin the cash-out rules suppose that trading parties can achieve energy balance at lower cost than the SO. A key assumption invoked here is that the SO as a monopoly provider is disadvantaged in what it can provide relative to competitive players. By *correctly signalling* to market participants the SO's *cost of balancing energy*, which is higher, they face real incentives to enter into efficient trades themselves. This aspect of the design of cash-out is given emphasis in both the P194 and P205 decision letters.

We have difficulty with important elements of these assumptions. The market rules presume a SO, which is legally obligated to balance the system efficiently and in the interests of the consumers, but which also has a strong profit incentive to do this at lowest cost. And it can do so because it sees the whole physical system, including the complexities and interactions with system balancing, a position which is not rationalised in terms of the accounting between energy and system balance until after the event. Under the market design, the SO already has under its control a large portfolio of balancing service contracts, which has the effect of taking significant quantities of plant off-market and confers on the SO options for balancing over which only it enjoys a right.

Under the British model, the SO also has the right to enter energy markets on a non-speculative basis if it believes such action will enable it to realise lower costs for procuring reserves than otherwise. It uses this facility routinely, and in 2005-06 it contracted *net* volumes of 1.05TWh in this way at a *net* cost of £34.2mn. It is not the purpose of this paper to explore the adverse implications of the SO entering energy markets. But we note that the SO has tended increasingly to contract on a disclosed basis with counter-parties for locational trades, giving some trading parties preferential information. It is also relevant that many generators would prefer to trade with it partly for credit reasons. The SO could be characterised as competing for volumes with trading parties who are out of balance.

These energy markets are also prone to illiquidity, and the SO is able to trade in the light of the better knowledge that it holds. It does not follow therefore that many participants – especially those that do not hold diversified portfolios, and who are not integrated – are able to balance their own portfolios at lower cost than the SO and if so to what degree. We are not aware of any analysis that compares the costs to trading parties for imbalance energy against the costs incurred by the SO across the various options open to it. Trading parties already have strong incentives to identify and transact efficient trades. Structural distortions such as one-hour gate closure embedded in the market structure already constrain them and their choices. Ofgem in its analysis on repricing bids and offers in the P194 regulatory impact assessment suggests that market information for even scale players is very imperfect.

Taken as a whole, these factors suggest that in many circumstances the lowest cost option may well be with the SO. This position is, of course, the starting point for the argument that electricity systems should be managed by a single entity, even if the "pooling" in its loosest sense is restricted to residual balancing activity as under the Neta model. The costs are recovered through BSUoS, and the SO can profit from efficient purchases through its incentive scheme. We conclude that there is a strong argument for acknowledging this natural purchasing role by the SO and excluding a portion of its energy purchase costs from the cost of imbalance energy. This is of course another way of suggesting that there should be a tolerance band for cash-out.

3. Other considerations

3.1. Negative prices

The risk of greater incidence of negative SSPs under marginal imbalance pricing relative to averaged approaches to setting the main price has already been noted. The risk of negative spill prices arises because of trades remaining in the longer stack that reflect negative bid prices. This situation might arise where the SO makes a number of constrained-on acceptances (that is, out of price order) and where the acceptances are not tagged out because they do not fit the criteria established under the rules or where the shorter stack does not remove such trades.

An extreme example of negative prices was provided by the Damhead Creek "incident" on 18 May 2004, although the immediate circumstances triggering the event – an emergency instruction – have been addressed through a BSC rule change. This change had the effect of transferring any such instruction from the energy stack. But the effect of including a high negative bid in the stack (-£9,999) on that occasion had an immense impact on one trading party who was spilling onto the system in otherwise benign conditions.[14]

There are a number of situations arising from system balancing where highly negative bids could be called. During the discussion on Damhead Creek it was noted that about one-third of generators used "sleeper bids", although analysis by the SO in support of BSC standing issue 18[15] suggested the problem was much more endemic. The risk has always existed (and been recognised), but the frequency and the impact under the P205

[14] The effect of including the acceptance in the energy price was to change SSP from +£14.95 in one balancing period to -£5,870.84. As a consequence a positive payment of £321 transposed to a charge of -£125,213.

[15] www.elexon.co.uk/documents/BSC_Panel_and_Panel_Committees/BSC_Panel_Meetings_2005_-_099_-_Papers/99_001e.pdf

baseline is now greater than under the P78 baseline as a result of its part marginal effect.

Any cash-out arrangement that can generate negative prices in circumstances where the system is net long at a particular point in time lacks credibility. If the potential for this deficiency is not addressed, and sleeper bids – which can rise to -£99,999 – set system prices, it could significantly damage the credibility of the trading arrangements as a whole.

3.2. "Sleeper" bids

This issue of sleeper bids arises where a trading party wishes to signal its reluctance to be pushed off the system by submitting a high priced negative bid. In various decision letters Ofgem has stated concerns about this practice, not least because of it inflationary effect on imbalance costs.[16] The P194 letter restates the concerns, which "would on average have a larger impact under P194", an observation which applies to a lesser degree under P205.

The Elexon report from the issue 18 group set out three possible solutions to address the defect:

- establishing a threshold price – if the acceptance price breaches a defined threshold price then the SO would determine whether the acceptance made was for system or energy balancing reasons. If the SO determined that the action was for system reasons, the volume would enter into the pricing calculation "at a replacement price reflective of the actions which would have been taken if the system action had not been taken"[17];
- use of bilateral contracts – the SO could develop contracts with trading parties to cover issues such as compensation and risk carried by a trading party for an instructed plant shutdown. This mechanism might apply for settlement periods following an acceptance during which the party might be exposed to energy imbalance prices and any other contractual obligations they could not fulfil; the contract might have the effect of smearing the costs associated with the acceptance over a number of balancing periods rather than that in which the initial decision were taken; and
- development of more dynamic transmission access arrangements – if an appropriate regime for compensating for loss of transmission access[18] were developed by the SO outside of the BSC, this might result in parties submitting lower priced bid/offers more representative of the costs with decrementing output. In such instances bid acceptances could be removed from the BSC and the energy price stack.

[16] See especially decision letters on P173 and P175.

[17] The basis for setting the replacement price was not addressed.

[18] Cusc currently covers provision for compensation for generators if there is either a planned disconnection or an unexpected disconnection form the transmission system. Payments for a planned disconnection are paid to the disconnected party through rebates of TNUoS payments only, while for unplanned events the generator is compensated using the market index price for the first 24 hours of an event or fault, with compensation being paid based on TNUoS charges for periods in excess of 24 hours.

These alternative approaches are not exhaustive. Given the number of occasions on which this and related issues have been flagged by Ofgem and debated by the industry, it is surprising that more focussed debate has not been initiated under Cusc or by National Grid's Transmission Charges Methodologies Forum (where the issue rightly belongs).

3.3. Dual versus single price

Industry arguments supporting a single cash-out price have not been vocal, but this is probably a reflection of the expectation that this issue is off Ofgem's agenda. The P74 original and alternative modification proposals contemplated a single price after a fashion, but these proposals were rejected at the same time as P78 was endorsed, and a variant whereby licence exempt generators would see a single price was rejected as part of consideration of P95. On both occasions in its decisions Ofgem referred to the reduced incentives to balance that a single price would give rise to.[19] Since then the matter fell outside the 2004 cash-out review on the basis that change of this nature was too wide-ranging to be implemented ahead of winter 2004-05.

As we have noted, the case for maintaining a spread *for the sake of it* by using two prices is weaker now than previously. One argument we have already referenced against a dual cash-out price is it distorts the market because it increases the balancing costs of smaller suppliers relative to larger ones. Smaller players, especially suppliers, resort more frequently to short-term trades and ultimately to use of the cash-out mechanism. To quote Professor Littlechild again: "The dual cash-out policy thereby handicaps smaller suppliers in their energy purchasing. By the same token, it also handicaps smaller generators by reducing the amount they are paid for surplus energy at times when the system is short".[20] "This is another measure of the artificial premium on balancing that is paid disproportionately by small suppliers," he concluded.

The effect flows not just from dual pricing as such but from the detail of the construction of the price formation rules adopted in Britain and the significant spreads they give rise to. The same effect might be achieved through a single price with an automatic adjustment where a short party pays the single price plus an uplift and a long party is paid the single less a discount, as in France.[21] A single cash-out

[19] Of the two decisions, the P74 letter is the more informative: "While it is difficult to value the actual cost imposed by the Party being out of balance, to assume that the cost is zero by adopting a single cashout price would be even more arbitrary. Consequently, it is appropriate that participants who are spilling electricity should receive a lower price for their electricity than if they had been fully contracted since they may be imposing costs on the system. Conversely, participants on whose behalf the SO has to procure the flexible delivery of electricity at short notice should pay the full cost of power delivered over short timescales. The use of a dual cashout price regime incentivises participants to balance their own positions by Gate Closure and hence the actions that the SO has to take are minimized." All we would note is that cost would definitely not be zero.

[20] Littlechild (June 2005), page 50.

[21] in France the cash-out price equivalent to SBP is set equal to the weighted average price paid to all other generators that were called upon by the system operator in real-time to make up the shortfall. An additional

price (or two prices with a fixed incentive to contract derived from a single price) might thus remove a distortion against smaller players and which also artificially favours vertical integration.

A single price could have other advantages too:

- it would encourage parties to look more to the short-term energy markets instead of relying on cash-out;
- it could increase liquidity by providing a marker against which trading parties could price their contracts; and
- it would remove any incentive for parties to influence the determination of spot market prices in order to influence reverse cash-out prices.

Professor Littlechild has also floated the possibility of a hybrid approach to dual pricing. It would, for instance, be possible to improve the basis of the reverse price without immediately changing from dual to single cash-out prices. For example, the price in the reverse direction could be set equal to halfway between the spot price and the cash-out price in the main direction. This shift might reduce the extent of the present distortion and in doing so arguably make the reverse price more cost-reflective. Over time, and provided there was no adverse impact on the SO's ability to balance the system, the reverse price could be set closer to the main price and eventually equal to it.

Professor Littlechild has also put forward a proposal for quantity premium bands. This is a complex issue, which could if implemented balance actual cost recovery with graduated charges, including a marginal signal on some balancing costs.[22]

A lot would depend on the detailed rules on construction of a single imbalance price. Many of the criticisms noted above would apply to a single price if it were based on the current main price methodology. But as we have already seen a number of alternative approaches to constructing a single energy price are available, seem to have merit and are worthy of consideration.

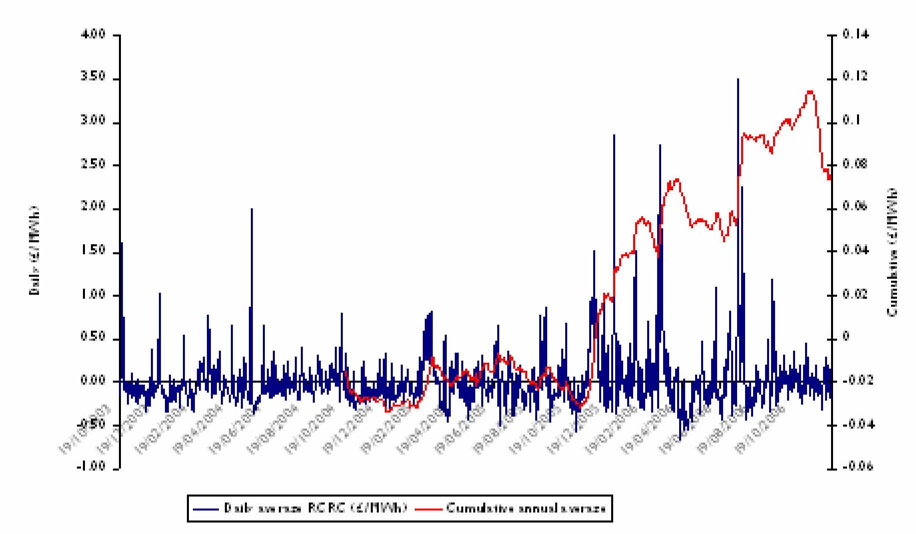
3.4. RCRC

The cash-out rules, because of the use of two prices, lead to the accumulation of a surplus or deficit in each balancing period. If the receipts and payments match, it is coincidental. The balance is distributed to trading parties proportionate to each energy account's share of total energy summed over the balancing day, and this is usually termed the "beer fund". Daily average values October 2003 to December 2006 are shown at figure 4.

5% 'uplift' is applied by the SO to the imbalance price. Excess payments are recycled to all "balance responsible parties" on a per MWh basis within the relevant trading period.

[22] www.electricitypolicy.org.uk/pubs/misc/littlechildimbalance.pdf . This is not explored further here as the issue is explored further in Professor Littlechild's *Perspectives on cash-out*.

Figure 4: Daily average RCRC - October 2003 to December 2006



The working of the "beer fund" and its interaction with the cash-out pricing rules raises a number of issues:

- the cashflow arises purely as a function of the operation of the cash-out rules;
- it has become much more significant in money terms and consistently more positive (i.e. a surplus) over the past year;
- it is funded disproportionately by smaller market participants to the extent that their imbalance volume usually constitute a larger proportion of their total electricity purchases;
- it is returned to all participants pro rata to size on the basis of their metered volume in each balancing period irrespective of their own balance position;
- taking the last two points together, it encourages larger and more vertically-integrated players than would otherwise be the case; and
- the RCRC charge is only known after the event and cannot be forecast, so there is no ability to weigh up the probability of the exposure to cash-out in tandem with the likely RCRC adjustment.

Ofgem has acknowledged that there could be a distortion of incentives created by cash-out from the operation of the RCRC mechanism or that it reduces them.[23] In particular the balance (usually a repayment) is rebated independently of whether the party is in balance.

There are various other ways in which the RCRC fund could be redistributed, including:

- rebating proportionate to whether a participant is reducing or increasing its imbalance volumes;
- netting the surplus off BSUoS across the market;

[23] Ofgem (May 2004), page 31. Also see the P194 decision letter.

- making a single netting payment against trading charges; and
- should the rules be changed to permit it, giving any surplus monies to charity.

Of course the issue of the redistribution would not apply were a single price that recovered actual energy imbalance costs employed.

It is important to recognise that the RCRC mechanism is independent from the incentivising effect of the two prices. We do not consider that it is appropriate to distribute the residual cashflow in a way that interacts with the imbalance prices, to correct perceived deficiencies with other aspects of the rules.

3.5. Standing reserve

At present option fees for standing reserve are profiled according to *expected* patterns of utilisation. These fees feed into cash-out via the buy price adjustment under the BSAD methodology, which in turn is added to the main price following the volume weighted calculation of energy-related balancing mechanism trades and forward trades. Where negative reserve is procured, the relevant BSAD parameter is the sell price adjustment, and which is reflected in cash-out price when SSP is the main price.

The SO should allocate these fees according to actual usage, and it is unclear given the emphasis on cost-reflectivity why National Grid has not introduced change proposals in this area.

3.6. Treatment of forward trades

During the consideration of P136/7, a number of changes were canvassed with regard to the treatment of energy balancing actions in BSAD. In particular Barclays Capital as a feature of P137 proposed that:

- individual trades should be represented in the NIV in a disaggregated format;
- there should be incorporation of non BMU-specific delivered standing reserve into the BSAD methodology; and
- there should be modification of the treatment of option fees for standing reserve contracts as noted above to target the BM units that have standing reserves called.

The SO consulted on these changes in parallel to the proposed BSC changes and concluded that the changes should be made but only *if* either P136/7 were adopted. Despite general support for these aspects of the proposals and supportive comments by Ofgem in its decision letter, these changes have not subsequently been pursued after the two modifications were rejected. These points seem no less relevant today – if anything they are more relevant with the introduction of a part-marginal approach to calculation of the main price.

3.7. Regional considerations

Analysis by Ergeg shows that by applying different cash-out pricing methodologies used in neighbouring countries to the same hypothetical power station would produce different cash-out prices. In particular, it shows that the British and French approaches which both construct two prices nevertheless can produce quite different prices. These differences between markets can distort trading within regional markets, and the implication of the Ergeg work is that significantly different cash-out regimes would act as a barrier to regional market integration.[24]

4. Conclusions and recommendations

The P205 decision, which signalled the latest cash-out review, would seem to represent a very strong statement that cash-out needs to focus more on the practical than the theoretical, a stance with which we strongly concur, and that the design objectives may not have delivered intended results. It also provides a clear signal that the sticking plaster approach applied to the cash-out rules over the recent past to correct recognised deficiencies needs reappraisal.

In evolving the cash-out arrangements to date significant – possibly a disproportionate – emphasis has been placed on cost-reflectivity and targeting at the expense of other design considerations, despite real and undesirable competitive impacts. There is a considerable body of evidence that pure energy costs are still not being appropriately signalled. Compared with differentials seen in other electricity markets between the real-time mechanism and short-term forward markets, the differentials in price are significant and are almost certainly not representative of real differences in energy costs.

Looking at the design features we highlighted:

- dual prices;

a dual pricing system was devised with particular outcomes in mind. The issues it sought to address are much less relevant now. More specifically, nearly six years into the operation of the new market model:

  - there is an extended track record of sound system management under the new trading environment;
  - soft landing mechanisms associated with implementation of Neta have been superseded (three and a half hour gate closure; use of PGBT contracts etc.); and
  - industry structure is markedly different, with significantly increased consolidation and integration, with the market dominated to a greater degree by a handful of large two-sided players who to all intents and purposes self-balance.

- energy balancing;

there is no accepted or codified method for classifying types of balancing action, and the costs in any particular balancing period are derived from a complex series of interventions in the price stacks both in terms of subtractions, additions and netting. It is widely recognised that the main price is frequently polluted by system actions and,

[24] www.ergeg.org/portal/page/portal/ERGEG_HOME/ERGEG_RI/ERI/UK_Ireland/Papers%20and%20Written%20Consultation/20061101_Paper_IG2-5%20Imbalance%20Pricing_FINAL_.doc

despite recurring concern on this point, there are no simple ways for correcting the situation without major change to the cash-out rules;

- targeting the costs of energy balancing;

because certain types of trading party are more prone to systematic imbalance, the method applied for targeting costs loads costs onto non-integrated and smaller players, a characteristic amplified by the recycling of the RCRC fund. The signalling value of such an approach – even if it were accurate – is greatly diminished as the actual costs are not known until after the event;

- parties who create the costs;

the present rules make no allowance for the fact that gate closure with no ex post trading imposes unmanageable risks on trading parties.[25] The rule book also assumes that parties can usually achieve a state of organic balance, but this is not in the nature of the electricity system given the need to achieve instantaneous matching of supply and demand;

- reflecting the costs incurred by the SO; and

even if the energy balance could be identified with precision, some of the trades represented by it could be distorted by actions taken elsewhere by the SO, either directly as it seeks to optimise its purchase costs over a series of balancing periods or indirectly through its wider interventions in energy markets. Further, once the costs have been identified, there is scope for genuine debate about how they should be reflected to trading parties (that is, whether they should be on an average or marginal or part-marginal basis);

- achieving balancing at lower costs than the SO.

where trading parties anticipate they are likely to be in imbalance they are likely to take reasonable steps available to them to manage the situation. But the SO has better information and a view of interactions across the wider system. It is the only party that knows what interventions it might make for system balancing purposes. There are strong reasons to suppose it can balance costs at least as efficiently as many trading parties, a situation that could well improve the more the system is under stress because of the options it holds.

The accumulation of these factors means that while each of these design criteria has merit in an abstract sense the ability to deliver them in practice – especially through the very complex cash-out rules as they stand – is acutely compromised. A significant degree of latitude is required in interpreting how these parameters should be applied.

A real problem is that the cash-out rules rely on application of a series of mechanisms that have arisen out of judgements to address different problems, and in this instance the

[25] It is also a notable contrast to the gas market where there is no gate closure and parties are able to trade out their imbalances for up to 15 days after the end of the month in which the relevant gas day occurs.

whole would seem to be significantly less than the sum of the parts. Worse than that, the current mechanism taken in the round could be producing the wrong signals, undermining cost-reflectively and further distorting competition. The symptoms of some aspects of this problem have been addressed, but in a manner that has not always been consistent. Despite several promptings and some workstreams that have not been concluded, the underlying causes have not been remedied. Overall we believe that the current arrangements are not functioning in their intended manner.

The three main factors giving rise to these distortions and deficiencies warrant further comment and are:

- the approximate nature of tagging rules and the imprecise energy/system split;
- continuing interactions between the energy market, the transmission access regime and the management of transmission constraints through the balancing mechanism; and
- the complexity of the pricing rules and their lack of transparency.

We set out specific recommendations with regard to each below.

4.1. Tagging does not achieve the intended results

The current tagging rules have become particularly convoluted and do not deliver a representative energy stack in many situations. There are at least four compounding factors:

- the SO does not classify acceptances as they arise and trades are tagged after the event;
- an ex post tagging process means the market is always in catch up mode, and it has no reliable way of knowing in real-time what operating conditions will do to cash-out prices;
- tagging is applied through the use of proxy rules that are applied to trades that have certain characteristics, while trades that should be tagged out as they arise from system constraints are disregarded. Approximations per se are not necessarily bad (viz. half hour trading periods, use of profiles), but the basic judgment that de minimis, CADL and NIV tagged trades are representative of all system trades is not correct; and
- reserve option fees are added back on an expected rather than actual basis, and there is scope for legitimate debate on differing treatment of other BSAD adjustments.

The defects inherent in tagging also have implications for moving to a more marginal price, which is the clear stated aim of the SO. The actual working of the part-marginal signal represented by P205 cannot be evaluated as yet because there has been limited time for it to operate, and there have not been any real system stress days since it was implemented.[26] While it is also clear that the P194 signal was considered to be more at risk

[26] During November, December and January PAR had been active in price setting in 25% of settlement periods partly because of a relatively benign winter to date. However, the price was only subject to an adjustment in 3% of settlement periods. The average price adjustment was approximately £8/MWh with a

of being incorrect than the P205 signal based on cash-out in its current form, there are still acknowledged imperfections in the current pricing baseline. Sharpening price signals to get purchasers to improve the overall reserve on the system as an end in itself, by reflecting the marginal and not actual costs, is also likely to have detrimental competitive outcomes, especially if they are the wrong costs. Increasing the marginal signal should only be contemplated where energy costs can be correctly identified and where there is an isolatable causer above acceptable tolerance limits. Even then there are arguments as to why this may not be appropriate.

The current approach to the energy/ system split warrants a fundamental overhaul. As an urgent interim measure the SO should be requested to find some mechanism for classifying trades as they are called. At the very least this should identify acceptances taken out of price order or those required for specific locational reasons. This “fix” might initially be done on a sampled basis while enduring changes are identified, but a start needs to be made. The objective should be to signal this information to the market in real-time as soon as possible.

4.2. Transmission constraints should be dealt with outside the balancing mechanism

The management of transmission constraints through the balancing mechanism was seen at Neta go-live as a transitional feature of the new market. However, there has been no real progress on developing a more enduring solution since that time. While problems with tagging would remain, they could well be more manageable if a solution for dealing with limitations on transmission access were found outside of the energy market and specifically outside of the balancing mechanism. It is not clear why, given its emphasis on creating strong performance incentives for trading parties, the SO seems disinterested in or opposed to establishing the same disciplines on itself.

Sleeper bids aggravate the problems associated with transmission constraints and remain a real potential problem under the current market design and for the cash-out rules, and the issues flagged concerning them in Ofgem decision letters disappointingly have not been progressed adequately. The impact of sleeper bids could still exert a badly distorting effect on cash-out prices, especially against a background where transmission constraints have increased significantly of recent, and could aggravate the risk of negative prices.

Negative prices in this context are not the result of a market-based process but anomalies in the rules. If negative prices were to become a more recurrent feature of cash-out, this could significantly further undermine the credibility of the trading arrangements.

4.3. Cash-out requires reduced complexity and improved transparency

Explaining the cash-out rules to industry participants is a challenge. It is time-consuming and quickly reveals that the process is not intuitive. One by-product of the complexity is that “real knowledge” of how the methodology works in detail is held by a handful of

---

maximum adjustment of approximately £73/MWh. The majority of the price adjustments were in the range £1 - £5/MWh. In contrast if PAR had been set to 100MWh PAR tagging would have been active in 80% of periods, with price adjustments would have taken place in 720 periods.

people who dominate industry debates. In turn this complexity of itself, legitimately we think, is seen by some as a barrier to entry,

The mystique surrounding cash-out has been reinforced by the fact that the various rules have been developed sequentially and changes have not necessarily been mutually reinforcing. This point is illustrated by two fundamental shifts in the construction of prices represented by P78 (and the introduction of first the NIV tagging concept) and then P194/205 (and the move away from weighted average to part-marginal pricing for the main price). The first eliminates the link with costs; the second tries to amplify it.

There are other important issues regarding transparency. Significant weight was placed in the initial design objectives of Neta and assessment reports of proposed modifications to date on the assumption that trading parties can see price signals that have been created and can determine whether and how to react. The price formation process starts after gate closure, but initial prices are only visible 15 minutes after the balancing period. Cash-out prices are also inherently volatile. What happens in one half hour is no guarantee of what might happen in the next.

In this regard the British market is out of step with other electricity markets, which have established provisional prices or can generate real-time prices or permit ex post trading. Until we have real-time information, participants will have no ability to respond to these prices dynamically. This can be expected to have a continuing debilitating effect on (among other things) demand-side response.

The present limitations on price transparency and timing also highlight the importance of making available information on forward energy trades and balancing service actions as they occur and in a useable form.

4.4. Scope of review must be broad

The focus of recent change proposals to the cash-out rules has remained on mechanistic aspects of the rules, with some trading parties, particularly smaller operators, targeting what they consider needs to be done to neutralise perceived defects in the process. This is a long way from addressing the question of what is the "right" cash-out pricing mechanism. Ofgem attempted in 2004 and again in 2005 to address this question by initiating a more rounded assessment of cash-out in both electricity and gas, but unsuccessfully.

The timing constraints that applied to the previous review are not relevant, and Ofgem should adopt a suitably broad canvas for its latest cash-out review. It follows from the conclusions set out above that it should address:

- the system/energy split, and whether there are more appropriate tagging mechanisms. In considering this issue, weight should be given not only to the importance of achieving accurate cost separation and allocation but also recognising the (legitimate) practical limits of achieving this;
- the merits of dual verses single cash-out prices in their various forms, and whether a single price method might be one approach that would allow simplification and

enhancement of the current pricing rules. There is much more experience internationally of other imbalance pricing arrangements than during the Neta design phase and a wealth of available documentation, which could be considered;

- allocation of the settlement surplus/deficit, and whether the current arrangements for recycling under- or over-recovery payments remain appropriate given they can influence trading party behaviour. However the objective should be to neutralise the impact of the cashflow and not further sharpen balancing incentives;
- the role of the SO ahead of gate closure and how its actions are fed into pricing of energy imbalances. As a minimum the SO's interventions should be on a fully disclosed basis so that its judgments on the state of the system can be shared with trading parties; and
- previous changes identified as desirable, including BSAD disaggregation and changes to the treatment of standing reserve option fees, should be dusted down but not before wider reform has been properly considered. To proceed otherwise with changes in isolation could potentially increase the problems examined in this paper.

The SO should also be invited to consider:

1. the classification of acceptances in real-time, distinguishing between energy, transmission constraints and other system balancing actions;
2. how an ex post unconstrained schedule might be constructed as an alternative to current tagging arrangements, which could be used to produce a single price;
3. what level of adjustment to such a price (up and down) might be needed to create sufficiently strong incentives for trading parties to contract;
4. how (and at what level) a tolerance band might be set to deal with "natural" levels of energy imbalance on the system across all trading parties;
5. alternative methods for dealing with transmission constraints outside the balancing mechanism;
6. how to take forward the stalled workstream on sleeper bids, and how a replacement price should be constructed for acceptances of this type;
7. options for constructing a replacement price when prices go negative or establishing a zero-price floor; and
8. consider how provisional or indicative prices could be made available to market participants.

We believe that a solution based around these eight points would provide more efficient cash-out prices, retain incentives to contract thereby maintaining security of supply and eliminate competitive distortions thereby removing a significant barrier to entry.

Nigel Cornwall

February 2007

Appendix A – Landmark cash-out change proposals

When	Proposer	Number	Title	Effective from
27/03/01	Electricity Direct	P3	Correction of price spikes generated by *de-minimis* purchases	–
02/05/01	National Grid	P10	Eliminating price spikes caused by truncating effects	11/05/01
23/05/01	National Grid	P18A	Removing/mitigating the effect of system balancing actions in the imbalance price	23/09/01
05/04/02	National Grid	P78	Revised definitions of system buy price and system sell price	25/02/03
05/04/02	London Energy	P79	Revised rules for default energy imbalance pricing	–
01/08/03	National Grid	P135	Marginal system buy price during periods of demand reduction	–
01/08/03	National Grid	P136	Marginal definition of the 'main' energy imbalance price	–
01/08/03	Barclays	P137	Marginal definition of the 'main' energy imbalance price	–
08/08/3	Innogy	P138	Contingency arrangements in relation to demand control	–
26/08/05	National Grid	P194	Revised derivation of the 'main' energy imbalance price	02/11/06 but overwritten by P205
26/05/06	Utilita	P201	Energy imbalance tolerance band	–
07/06/06	Bizz Energy	P202	Energy imbalance incentive band	–
04/07/06	Good Energy	P205	Increase in PAR level from 100MWh to 500MWh	02/11/06

Appendix B - Acronyms

ABSVD	Applicable Balancing Services Volume Data
BMU	Balancing Mechanism Unit
BSAD	Balancing Services Adjustment Data
BSC	Balancing Settlement Code
BSUoS	Balancing Services Use of System Charges
CADL	Continuous Acceptance Duration Limit
Cusc	Connection and Use of System Code
Ergeg	European Regulators' Group For Electricity And Gas
ISO-NE	North Eastern Independent System Operator
MWh	Mega Watt Hour
NEM	National electricity Market of Australia
Neta	National Electricity Trading Arrangements
NIV	National Imbalance Volume
NYISO	New York Independent System Operator
NZEM	New Zealand electricity Market
Ofgem	Office of Gas and Electricity Markets
PAR	Price Average Reference Volume
PGBT	Pre-Gate Closure BMU Transaction
PJM	Pennsylvania-New Jersey-Maryland, a power pool
RCRC	Residual Cashflow Reallocation Cashflow
SBP	System Buy Price
SO	System Operator
SSP	System Sell Price
TWh	Terra Watt Hour