Bidding in energy-only wholesale electricity markets

Professor George Yarrow, assisted by Dr Chris Decker

November 2014

1. Introduction

This Report is concerned with how efficient price discovery is achieved in energy-only wholesale energy markets and with the role that price signals play in realising efficient market outcomes. The focus is on wholesale electricity markets although some of the points are also relevant to wholesale gas markets. It addresses both theoretical and practical issues and both short-term trade-offs affecting efficiency in dispatch and longer-term matters such as efficiency in investment.

The background context is that of potential concerns about aspects of re-bidding[1] behaviour in the NEM, an energy-only wholesale electricity market. The Report is directed chiefly at general questions about bidding and price formation in energy-only markets based on pooling arrangements in which a single price is determined for all generators, and it is therefore not focused on the technical detail of the NEM arrangements. However, a number of the features of the NEM are relevant to the general questions and the discussion therefore takes them into account where they appear to cast light on the wider issues. These features include:

- Generators are required to submit initial price/quantity offers for each thirty-minute trading interval in up to ten price bands by 12.30pm the day before the settlement day, which starts at 04.00am.
- Re-bids may be submitted up until the start of the relevant five-minute dispatch interval by re-allocating offered volumes within the nominated price bands.
- AEMO periodically publishes aggregated information to inform participants' decision-making including 30 minute and 5 minute pre-dispatch schedules detailing expected demand and prices, together with (to each generator on an individual basis) details of the expected level of generation.
- AEMO also publishes following-day information about final bids and re-bids used for dispatch purposes.

The specific matters to be addressed are:

- How short-term and long-term efficiencies are achieved in the bidding process.
- The role of efficient price discovery in this process.

[1] We adopt the specialised terminology that generators make 'bids'. In other commodity market contexts what is here called a 'bid' would be called an 'offer', with the term 'bid' being applied to the actions of buyers rather than sellers.

- How generator re-bidding can contribute to the iterative process of efficient price discovery.
- How re-bidding strategies can result in inefficient market outcomes.
- The incentive for generators to participate in re-bidding strategies that result in inefficient outcomes.
- The distinction between re-bidding that may lead to inefficient outcomes and market power.
- Possible options to limit the extent to which generators can engage in re-bidding behaviour that results in inefficient outcomes, including identification of any trade-offs from restricting the ability of generators to re-bid and taking account of the desirability of preserving the efficacy of price signals.

2. Energy-only markets: basic features

Energy-only markets like the NEM are based upon spot pricing of electricity in which prices and volumes are determined by equilibrating bids with demand requirements. In pay-as-bid markets, prices are determined transaction by transaction on a continuous, bilateral basis. In electricity pools bids are aggregated and a single price is determined pricing-period by pricing-period.

In the latter (pooling) case rules are required to translate bids into prices, and these rules can vary. For example, the market price can be determined by the highest-price bid that is accepted[2], which is the standard way of doing things, or alternatively by the lowest-price bid that is not accepted. These rules, usually embedded in a computer programme, can affect price determination, but, as indicated, this aspect of energy-only markets is not the central focus of this Report and the detailed characteristics of the rules, important though they are, will be noted only in so far as they are relevant to more general points.

In practice and in current circumstances, demand side participation is typically fairly limited in wholesale markets based on the determination of a single price for all suppliers. Nevertheless, small changes in demand-side responsiveness can have potentially significant effects on prices, particularly during peak periods. Most regulatory authorities therefore tend to be keen to facilitate or promote demand-side participation in wholesale markets.

As in any competitive commodity market, changes in circumstances can be expected to lead to shifts in the relevant, supply and demand bid functions[3], giving rise to fluctuations in prices and volumes. Demand (at a given price) varies substantially by time of day and time of year and tends to be relatively insensitive to price, particularly over short periods. Supply conditions tend to be more stable, although plant outages (both planned and unplanned) and plant operating problems are recurrent events that give rise to shifts in bid functions from one period to the next. Supply may also become highly inelastic as demand approaches available capacity and ‘peaking plant’ with high operational costs is called on to generate

[2] Or by a demand-side bid when it is ‘marginal’.
[3] The bid functions are the combinations of prices and volumes submitted to the system/market operator.

electricity. Putting these elasticity/responsiveness points together, when market conditions are tight prices become highly sensitive to shifts in supply and/or demand, i.e. only small movements in volumes on either side of the market can cause large changes in prices.

The significance of contracts

The existence of an energy-only spot market does not mean that market participants cannot contract on a longer term basis to supply or procure electricity at more stable prices and/or to secure required levels of volume. This can be done, for example, by means of contracts for differences by which a given volume can, in effect, be supplied or procured in over a particular period at a fixed price. Alternatively specified volumes might be agreed at contract prices that are indexed to general price indices or to energy prices, including to electricity spot prices

Options contracts are also feasible: in return for a fixed payment, a buyer can contract to buy specified volumes at a specified strike price in the event that spot prices rise above the strike price. For example a contract can be struck to buy/sell 50MWh at $300 per MWh, say, in the event that the spot prices are higher than $300 per MWh. The fixed payment for the contract then provides income to the generator irrespective of whether or not the option is actually called. This is not dissimilar to a capacity payment in market designs that provide explicitly for such payments, except that that the procurement arrangements occur on a voluntary, market basis rather than being determined by an administrative or regulatory mechanism.

As well as providing a basis for hedging strategies, financial contracts at fixed prices – or more generally at prices that are not subject to the influence of the relevant parties – can have powerful effects on bidding strategies. If there are no such contracts and all transactions are on a spot basis, a generator will benefit from any spot price increase across *all* of its volume sold. If, however, a percentage of that volume is covered by a CfD, the benefit of the spot price increase to the generator will be reduced (the revenue from the contracted volume will be unchanged). The effect is to reduce the payoffs from bidding strategies aimed at influencing the market price upwards -- see Annex 1 for a more formal account.

Indeed, as shown in the Annex, if a generator is contracted to sell more than it expects to produce in a given period, for example because of non-availability of plant, there can be incentives to bid prices that are below marginal, incurred costs. In effect, the generator has to 'buy' MWhs from the spot market in order to cover its contract position and hence, to the extent that it has any price influence, it will want to use that influence to lower spot prices.

For similar reasons (i.e. a change in incentives), vertical integration between generators and downstream retailers can also be expected to influence bidding strategies.

Remuneration of fixed costs (the 'missing money' question)

One of the issues that has been raised in the analysis of energy-only markets is the question of how capacity costs and other fixed costs get remunerated, particularly (but not exclusively) for marginal generating plant that may only be operated for short periods, if at all, during any given year. The issue is an important one because it affects how spot market bidding strategies are to be analysed and interpreted.

Unfortunately, there has been some confusion around this matter in the past, in part due to recourse to over-simplified economic models that suggest that efficient market clearing requires price to be equal to marginal, *incurred* cost[4] over the relevant pricing period, most notably the costs of fuel required to generate incremental MWhs. The issues therefore merit some initial clarification.

In hypothetical conditions where the spot market cleared at prices that were always equal to the within-period, marginal, incurred costs there would indeed likely be a missing money problem. For example, plant that was system-marginal (i.e. the highest cost plant to be dispatched) in <u>any</u> particular pricing period – not just the peak period – would not be able to recover any 'availability costs', such as start-up and shut-down costs, that were incurred before and after the relevant pricing period. Within-period revenues would therefore not even cover its short-run costs, let alone make any contribution to capacity costs and other overheads.

Such under-recovery would be an issue even if, in other pricing periods in which the plant operated (i.e. periods in when it was not system-marginal), it was able to benefit from revenues in excess of within-period, marginal, incurred costs, i.e. it received very short-term economic rents.[5] Power systems are characterised by certain types of dynamic costs – costs attributable to changes in output, of which start-up and shut-down costs are an example – and these require remuneration. Further explanation is provided in Annex 2.

In general, then, it is to be expected that, for any plant that is marginal in a particular pricing period, efficient prices will be in excess of within-period, marginal, incurred costs.

A fortiori, in the assumed, hypothetical conditions (of prices being equal to short-run, marginal, incurred costs) there would be a 'missing money' problem for plant that, whenever it operates, is always or nearly always marginal for the system as a whole, i.e. peak plant. Such plant can typically only recover its capacity and other overhead costs over a small number of pricing periods, or, put another way, it will only infrequently, if at all, be in a position to benefit from economic rents.

These points are sometimes made to support arguments to the effect that energy-only market designs are *necessarily* inconsistent with efficient, competitive[6] market clearing, but such arguments are generally misguided. Short-run efficiency can be achieved in energy-market designs provided that it is recognised that pricing should reflect *economic* costs, not incurred costs. Economic costs encompass *scarcity rents* as well as such things as expenditures on fuel used to generate electricity.

To illustrate, consider a simplified situation that abstracts from dynamic costs and in which one generator operates plant characterised by a marginal incurred cost of $300 per MWh and another generator operates plant that is adjacent in the merit order and is characterised by

4 That is, actual *expenditure* made or incurred as a result of increased output.

5 An economic rent is a payment for an economic input or output that is in excess of the amount necessary to induce voluntary supply of it.

6 If there is market power, 'monopoly rents' will be available to remunerate capacity costs and other overheads, but here the focus is on competitive markets. Market power will be discussed later in the Report.

marginal incurred costs of $400 per MWh. Suppose that demand is uncertain at the time that bids have to be submitted, but that it is expected that one or other of the two relevant bids will be system-marginal in a relevant pricing period.

In these circumstances, the lower cost generator can bid up to almost $400 per MWh and be safe in the knowledge that its plant will be dispatched (at higher prices there is risk that its bid will be undercut by the competitor). It is therefore to be expected that, in a competitive market, the lower-cost generator will bid at a price closer to $400 than to $300. The generator will, therefore, be able to earn economic rents – it would have been willing to supply at a price down to $300 per MWh but achieves a significantly higher price – which will be available to remunerate fixed costs, *even if it turns out that the higher cost plant is not dispatched.*[7]

This indicates that 'missing money' problems are not an inherent feature of energy-only markets. What would be problematic is if misguided regulatory policy *required* that bids reflected within-period, marginal, incurred costs or set an unduly low upper bound to prices.

This simple example also points toward some of the complexities that arise once the full factual context is addressed. The higher-cost generator might itself (just like the lower-cost generator) be expected to bid above its own marginal, incurred costs, implying that the lower-cost competitor could price above $400 per MWh and still see its plant dispatched; but by how much? The question is not easily answered *ex ante*: as will be emphasised later, this is something that is settled by the competitive process itself. Indeed, if it were readily answerable *ex ante*, competition would be largely redundant.

The more relevant questions for the purposes of this Report are therefore:

- How can energy market designs provide for the market to determine appropriate levels of economic rents?
- How can participants anticipate the appropriate levels of economic rents in their bidding strategies?

The first of these questions is focused on general issues of market design, the second is focused on some of the more specific issues considered in the *Stanwell* case.

3. The role of efficient price discovery

Economic concepts of efficiency

The notion of efficiency has multiple meanings, but for purposes of public policy it can be taken to be a measure of the aggregate net benefits from trading in a relevant market. The question "Can efficiency be improved?" is, for practical purposes, equivalent to the question "Can the aggregate net benefits from trading be increased?"

[7] It can also be noted that in the case of the highest (marginal, incurred) cost plant on the system, which faces no higher-cost competitor, the preferred bidding strategy will be the highest per MWh bid allowed by the market rules, which might be based on some estimate of the value of lost load.

A traditional policy-making categorisation of sources of efficiency gains is:

- Productive efficiency: can existing outputs be supplied at lower cost?
- Allocative efficiency: can available resources be re-allocated to increase the gains from trade?
- Dynamic efficiency: can available resources be re-allocated to increase the rate at which the gains from trade (the aggregate net benefits) increase over time?

In addition to these categories, efficiency in risk sharing also tends be an important consideration in energy markets, so we can add:

- Risk sharing efficiency: can risks be re-allocated in ways that increase aggregate gains from trade?

Productive and allocative efficiency assessments depend also upon the time period involved. A traditional distinction in economics is between the short-run and the long-run, based on the criterion of whether capital is fixed or variable within the relevant period. This is a crude, binary approximation to a more complex reality: in practice, the scope for varying capital increases with the length of the relevant real (historical) time interval involved.

Productive and allocative efficiency issues are usually assessed on the basis of a given state of economically valuable knowledge/information (i.e. knowledge/information that has bearing on economic transactions). Thus, in the traditional short-run/long-run distinction, the state of knowledge is held constant: there is no 'learning' as we moved from the short-run to the long-run.

In contrast, dynamic efficiency is concerned with the expansion (or the variation[8]) of this state of knowledge/information, but it too brings into play considerations of time intervals. Thus, in a particular context the interest may lie in the discovery of economically valuable information over a very short period (e.g. knowing more about how much something is worth today/tomorrow) or over much longer periods (e.g. developing new technologies and new ways of doing things that will increase gains from trade).

Efficiency effects from these various sources are entangled and there are highly significant trade-offs amongst them. For example, pursuit of short-term allocative efficiency gains can, in some contexts, be harmful for dynamic efficiency, for example because it reduces incentives for discovery and innovation. Similarly, reducing the costs of risk via risk-sharing can come into conflict with the achievement of efficiencies in other dimensions by weakening incentives: there are trade-offs between 'insurance' and 'incentives' that appear as a relevant consideration for public policy decisions on a recurring basis across a number of areas of economic life, including in electricity markets.

[8] Knowledge can be lost as well as gained.

Efficiency in dispatch

Dispatch decisions in energy markets engage all aspects of economic efficiency. A full treatment is beyond the scope of this Report, but a number of salient points can be noted for electricity system designs that rely upon the establishment of a uniform, market-clearing price (rather than on prices determined on a transaction-by-transaction basis in the manner of pay-as-bid arrangements):

- Competitive bidding tends to lead toward a least cost generating mix (i.e. productive efficiency) for any given level of required output. A price is established and, at least on a short-term basis, electricity that can be generated at lower unit cost than the market price will be supplied whereas higher cost supplies will not be made. This is a tendency only, albeit a strong one, since it is always possible for a generator to mis-estimate the economic rents that are expected to be available in a pricing period and, by bidding a price that is too high, find itself under-bid by a competitor whose own costs are actually lower. Such events are part of the competitive *process* by which prices come to be determined and which tends to lead to a least-cost *equilibrium.*
- If there is demand-side participation in the bidding process, competition also tends to promote allocative efficiency: output will be supplied when buyers are willing to pay for it, but not otherwise.
- In many circumstances (and in common with the vast majority of other economic markets), individual market participants will likely have some degree of price influence. The implications of this will be discussed in detail later, but for the moment it can be noted that it implies, as already stated, that bidding behaviour in the spot market can be influenced by contract positions.

Perhaps the most salient point to note, however, is that whilst efficient price discovery – i.e. discovery of the price at which very short-term gains from trade in the wholesale market are maximised – can be considered to be a sufficient condition for achieving a strong tendency toward efficiency in dispatch, it is probably not a necessary condition. Put more accurately, it is unlikely that efficient price discovery offers substantial advantages over, say, more centrally planned approaches in dispatch efficiency.

To clarify, what is here meant by the 'very short-term' here is a period over which plant available to the system is a given. In these conditions acquiring the information required by a system operator to achieve efficient dispatch of the available plant is not an infeasible task and the historical evidence from periods before the development of wholesale markets does not suggest that there were major, systematic efficiency failures under the earlier arrangements. On this basis, whilst efficient price discovery has a significant role to play in achieving efficient dispatch in today's arrangements, it is not a role that adds *substantially* to dispatch efficiency. It is a largely replacement role, rather than a role that moves performance to a new level.

The more distinctive role of efficient price discovery

What it is that efficient price discovery adds is more to do with the discovery of efficient levels of economic rents than with achieving efficient dispatch in the very short term. To see this, consider again the example discussed above of a step jump in the system, incurred, marginal cost of production from $300 to $400 per MWh. These two numbers do not in themselves determine an efficient price which, as discussed, may lie within a range from near the lower figure to something above the higher figure. That price remains to be discovered and its value will have implications for other decisions such as plant availability and longer term investment decisions.

One of the difficulties here is that electricity systems tend to exhibit limited demand side participation. If, for example, there was a continuum of price bids on the demand side, the 'demand curve' (i.e. the aggregation of buyers' bid functions) would 'pick out' a price in the relevant range and this in turn would determine the (very short-term) level of economic rent. For example, the lower cost producer might find that its bid was fully dispatched at, say, $350 per MWh, but that any higher price would call forth significant demand curtailment. This is illustrated in Annex 3.

In the absence of demand-side bids within the relevant price interval this will not happen in the very short-term and the question becomes: how is price to be determined in these types of circumstances?

Under current NEM arrangements it is left to generators to determine the matter via their own bid functions. Thus, the low cost generator is allowed to choose the prices at which it offers output to the system/market operator and, among things, its preferred prices will depend upon its forecasts of what demand will be in the relevant pricing period. If demand is expected to be relatively low, it may seek to price relatively low in relation to its own costs, in the hope of winning business from a yet lower cost competitor in the event that the latter over-prices its output. If demand is expected to be a little higher in the pricing period, it might prefer to price at a somewhat higher level and, in effect, seek to obtain a higher contribution to its dynamic costs and its fixed costs.

This suggests that an unconstrained generator might seek to offer something more closely approximating a smooth bid function than its own, 'stepped' cost function might suggest. In practice, however, generators in the NEM are constrained to a degree by the restriction that they offer no more than ten price bands. There is therefore a question as to whether this restriction might constrain the price discovery process in ways that that limit the discovery of efficient prices. That is, the restriction may mean that, in their initial bids, generators' may bid 'too high' in some volume ranges and 'too low' in others because, speaking loosely, they have to approximate a preferred bid function with a more stepped bid function.

More specifically, given the wide range of possible bid prices, there may be particular issues to be addressed in periods of high demand relative to availability in that the magnitudes of the 'steps' in bid functions tend to be relatively high in these periods, such that deviations of constrained bid prices from unconstrained bid prices are also likely to be relatively large.[9]

In thinking further about what might happen in an unconstrained, competitive market it can be noted that generators can be expected to have good information about their own costs, including their dynamic aspects, but pricing strategies based around recovery of those and of other costs that are not directly related to the level of output in a particular period will depend upon them taking views on the state of demand for their output in a sequence of pricing and dispatch periods. The relevant demand is the 'residual demand', which not only depends on system demand in the relevant pricing period but also on the bid functions of other generators, which in turn will depend on rivals' costs and *expectations/forecasts* (rivals will be trying to anticipate their own residual demand functions). The information available to any generator will therefore be much more limited once it steps into areas outside its own domain.

Competitive bidding processes, then, are characterised by a continual updating of information as (always incomplete) new information becomes available and expectations are adjusted. To repeat, such updating of expectations will depend on new information about rivals' cost positions and rivals' expectations, including rivals' expectations of others' expectations, and so on in a potentially infinite regress.

Whilst generators will tend to have good information about their own dynamics, the relevant, within-pricing-period costs can and do change quickly. Even for generators, therefore, bids reflect contingent (albeit very short-term) forecasts of their own costs at the relevant dispatch times. The individual forecasts then need to be pieced together, via the spot market information provided, by the market/system operator.

Only when *all* supply-side and demand-side information is combined will the level of any scarcity rents be determined. Again, generators and demand-side bidders can anticipate/forecast the outcome and build it into their decision making, but it is important to realise that these should be recognised for what they are, *forecasts*, and that, even over short periods, forecasts in energy spot markets are only rarely 100% accurate.

In a nutshell, efficient *economic* costs are revealed in the process of price discovery; they are not something that can accurately be determined *ex ante* for the simple reason that the information required will not be fully available ahead of the price determination process itself. The effectiveness of this discovery process is an aspect of the (short-run) dynamic efficiency of the relevant market (recall that dynamic efficiency is to do with *changes* in information conditions).

In technical economic terms, what this amounts to is the result that *economic costs* and prices are jointly and simultaneously discovered/determined via the competitive process. The point can cause some difficulties because much regulation – in particular the determination of terms and conditions for use of network services – proceeds by setting prices on the basis of

[9] Although, it can also be noted that it is in such periods that demand-side bidding may be particularly helpful in improving information conditions.

costs. Moreover, because such an activity is governmental, there tends to be a heavy emphasis on seeking to rely, wherever possible, on cost estimation processes that are in some sense objective (in order to avoid decisions that might be said to be arbitrary).

Such approaches are unobjectionable given the regulatory task, but they are only so because, in the relevant circumstances, the alternative of competitive discovery/ determination of efficient prices is infeasible. That is, it becomes the best feasible option for the purpose of determining prices and costs for want of a superior alternative. However, when such an alternative is available the notion that economic costs should be estimated on an *ex ante* and 'objective' basis should properly be put to one side.

Efficiency in plant availability

The above points have been made in relation to plant that is made available to the system to supply output in any given pricing period, but the availability decision itself leads to costs that are incurred whether or not the plant is actually dispatched. The prices that are determined therefore have significant consequences for the availability decision.

In systems that are characterised by payments for capacity as well as for MWhs supplied, there is an alternative source of revenues for generators. In energy-only markets, however, remuneration is by way of per MWh payments alone and the energy-price determination process is critical in ensuring that appropriate remuneration is established. Thus, for any given set of availability decisions, if market rules are such that prices are set below economically efficient levels, one effect that can be expected is a tendency toward inefficiently low levels of availability over time. When this happens it can cause the market price to be determined at inefficiently high levels over longer time periods.

Efficiency in investment

Investment decisions are influenced by expectations of prices over rather longer periods than decisions about levels of generation for existing plant. Indeed, the relevant time periods for considering investment efficiency may extend out for many years or decades, and it is on these that we will focus.

Given this point, it is obviously the case that expectations of spot prices in the very near future will have only very damped effects on investment decisions. Nevertheless, the *price determination process* itself, partly embodied in the rules of the system/market operator, will tend to be more durable and, on that basis, the 'rules of the game' are a matter of some importance.

If, for example, there is a systematic bias toward the determination of prices that are lower than short-term efficient levels, the general expectation is that there will be a tendency toward inefficiently low levels of investment in plant: expectations that economic rents will be suppressed will feed into expectations of sub-normally low rates of return on investment.

Similarly, price determination processes that lead, in competitive conditions, to higher than efficient levels of economic rent will tend to lead to over-investment in generating plant.[10]

These points are unaffected by the ability of generators and downstream purchasers of electric power to enter into longer-term contracts, since the terms on which contracts are struck will be heavily influenced by expectations of future spot prices. Adjusting for the values associated with the different distribution of risks between the two transactional arrangements (contract and spot), it is not to be expected that a buyer would be willing to enter into a long term power purchase agreement at prices that would imply a significantly larger net present value of payments than might be expected from relying on spot purchases. Similarly, with the risk-adjustment proviso, a generator would likely be unwilling to strike a contract that yielded a significantly lower NPV of revenues than it anticipated would be available from spot sales.

4. Re-bidding and price discovery

In light of the above remarks, the role of re-bidding in efficient price discovery is easy to see: circumstances can change and significant new, relevant information about costs and demand can become available, even over short periods. Even in highly competitive conditions – where competitive pressures are sufficient to cause generators to bid close to their forecast *economic* costs – generators may wish to re-adjust those forecasts (and again it is to be stressed that such costs depend upon factors such as power station dynamics and scarcity rents). Similarly, on the demand side energy retailers or large commercial users might glean new information about their own demand conditions and/or, in the case of end users participating in the spot market, about their costs of curtailment.

The most obvious, although far from the only, illustrative example is that of plant failure (or of lesser technical problems that affect output) in the period following the initial bids but before dispatch decisions are made. Such events are a familiar feature of electricity systems and, from an operational standpoint, all systems, whatever their design, have developed methods of coping with such events. The differences among systems are therefore usually more to do with the financial implications of such events: how do they affect prices, generating costs and financial payments?

In pooling systems such as the NEM, in the absence of re-bidding re-dispatch following the event will proceed on the basis of the bid functions of generators set out in their initial bids. These bid functions will not, therefore, reflect information concerning the event that has occurred. The system/market operator will, of course, have later, better information and will re-dispatch sets when the operational difficulties become known. In broad terms, the physical responses to the event will be appropriate, even in the absence of re-bidding.

[10] The position in circumstances of significant market power is less clear cut: the excessive rents may be due to monopolistic restrictions associated with low levels of investment – they may be ‘monopoly’ rather than ‘scarcity’ rents – but the same high rents may stimulate higher investment by others. This could be cost-inefficient if the resulting investment costs are higher than they could have been if the generator with market power had done more in the way of investment itself.

What in effect happens is that there is a jump to a new, 'calculated equilibrium' position. For reasons given above, however, that 'calculated equilibrium' will be tend to be inefficient because it is based on out-of-date information (out-of-date bid functions). In particular, it is based on out-of-date assessments of the level of economic rents.

As discussed above, economic rents are things that are discovered/determined by the competitive process. In making initial bids, market participants will, explicitly or implicitly, base their calculations of *forecasts/expectations* of such rents, and these forecasts will depend on explicit or implicit assessments of how market conditions might change over the period between bids and dispatch. Subsequent events will typically invalidate the forecasts to greater or lesser extent. Re-bidding allows for the revisions of forecasts/expectations in the light of the new information and provides the system/market operator additional information on which to base dispatch decisions and, more importantly, price determination.

In other types of commodity markets the re-adjustments take place on a continuous, iterative basis. That is, bids and offers in those markets are continually adjusted in the light of changing information. When background conditions are such as to cause no major changes in expectations/forecasts, prices may be relatively stable. However, when background conditions do change in a way that falsifies forecasts or disturbs expectations, offers, bids and prices will typically change.

There are a number of points that are worth noting in this respect:

- It is *changes* in expectations that trigger the adjustments. Whilst these may be triggered by some change in the relevant economic environment that is readily observable, and can therefore be said to be 'objective' in some sense, the shifts could also occur in the absence of such a change. Thus, since a market participant may have based bids and offers on an expectation that something 'out there' will happen, an observation that it doesn't happen can equally well disturb expectations (e.g. the failure of a central bank to change interest rates in circumstances where a change was anticipated). Bids, offers and prices can therefore change in the absence of an immediately obvious, 'objective' correlate.
- Whilst a major event can give rise to a sharp and near instantaneous adjustment in bids, offers and prices, there is typically an iterative process of adjustment that takes place. This is the discovery process by which the expectations/forecasts of all market participants, which typically differ, exert their influence on market prices. Thus, if an event happens (or doesn't happen), any particular market participant may have a good idea of its general implications, but may be less clear about precisely how others will interpret the event (or non-event). Since the expectations and behaviour of others are material factors in determining how the participant will want to behave, there is some learning to do. The iterative process might be quite short in duration, particularly when all are responding to a familiar type of event, but it can be more protracted when the issues are more complex and uncertainty is greater.
- In cases where there is an 'objective' change in circumstances due to a plant failure, it is not just the plant-owning generator immediately affected by the turn of events that may reasonably want to change its bidding behaviour. The loss of plant will lead to a

shift in the residual demand curves of other generators, giving rise to changes in their bids too. Moreover, the bidding adjustments made by the generator directly affected are affected by its contract position. If the plant-owning generator finds that it has less contracted output to offer, it may bid more aggressively. Indeed, as shown in Annex 1 it may even choose to bid at lower than marginal incurred costs. Other generators may then wish to adjust their own bids in response.

Re-bidding in electricity markets replicates the discovery and adjustment features of other types of market. In very broad terms, its potential 'value-added' will depend on the value of the extra information that it is capable of discovery and it is on this question that policy approaches towards it may turn (we will address the other side of the coin, possible costs or inefficiencies that may arise from re-bidding in the next section).

Cost-benefit assessment of re-bidding in the NEM is beyond the scope of this Report, but a few of the particular characteristics of the market that are relevant to such an assessment:

- For market participants making bids, perhaps the most difficult, most uncertain (*ex ante*) consideration to be taken into account when deciding how to bid is economic rent, and this is where there is most to discover/learn. The economic value added by re-bidding is therefore likely to be higher where 'rent discovery' issues is of greater significance.
- For thermal generating capacity, lumpiness and very short term inflexibilities in production (operating dynamics) are sources of rent discovery problems.
- Rent discovery is made more difficult by limited demand-side participation in wholesale electricity markets. This means that the system/market operator lacks information that is directly relevant for efficient pricing.
- The effectiveness of rent discovery is also a function of the rules used to determine market prices by the system/market operator: as is the case for many administrative mechanisms, boundaries and discontinuities between what is acceptable and what is not acceptable may be introduced for reasons of administrative expediency.[11]

Finally in this section it can be noted that re-bidding arrangements can also have implications for the distribution of risk among market participants. Since the risk distribution is also affected by contractual positions / hedging strategies, the effects are complex. One possibility is that re-bidding could make spot prices more volatile because they are influenced by new information that becomes available between initial bids and dispatch, i.e. to new influences that add to the variability of prices. Even if that is the case – which is not inevitable: it depends on the precise nature of the adjustments that occur – the effects may be offset or mitigated by adjustments in contracting/hedging positions. That is, there are other ways in which the preferred distribution of risk can be attained.

5. The potential for inefficiencies arising from re-bidding

[11] In the case of the NEM, for example, the rules exclude the possibility of a participant submitting a smoothly upward-sloping bid function.

Particular sets of economic arrangements only rarely come with benefits but with no cost and re-bidding options are no exception. This section of the Report serves to highlight some of the sources of potential costs.

Limited responsiveness

If the discovery of efficient prices relies upon an iterative process in which market participants respond to changes in new information, including information about how other market participants are bidding, it follows that very late re-bids – bids very close to dispatch decisions – are likely to be less valuable to the discovery process because they leave less time for the iterative process of price discovery to play out. This applies particularly to very late bids that follow a period of inactivity in re-bidding (a very late bid that is responsive to earlier bids is more likely to be part of a normal iterative sequence). The issue is nowadays familiar from behaviour in internet auctions with a defined and fixed end time for the posting of bids.

The immediate question that arises concerns the materiality of the effects and here it is relevant that the dispatch period in the NEM is only five minutes long and that re-bids can be made up until the beginning of the period. This effectively means that the response period cannot be shortened to less than five minutes, since responses can always be made in time for the next dispatch period. The iterative process can therefore be slowed, but only to a limited extent.

In periods of high demand relative to available capacity it is possible that the effect of a very late re-bid on prices may be significant, even if it is sustained for only a few minutes, but assessment of effects is complicated by the fact that, in the relevant circumstances, the observed change in the market price may not be a good measure of any inefficiency that is caused by the conduct. It may be the case, for example, that the efficient price is significantly higher than the price that would have prevailed in the absence of the re-bid, and it is the (undiscovered) efficient price that is the appropriate benchmark for assessing the actual (post-bid) price, not the price that would have been set by the system/market operator 'but for' the re-bid. Thus, it may be the case that the re-bid takes prices 'too high' for a short time period by leading to an over-shoot relative to the efficient price, but it is the overshoot that matters, not the difference between the price following the re-bid and the 'but for' price.

Assessment is further complicated by the fact that overshoots and undershoots are a normal feature of iterative discovery processes. Markets are never fully efficient and deviations from efficient, equilibrium prices are an inevitable part of the discovery process. Experimentation always has some costs.

Unlike in financial markets, adjustment of bids in wholesale electricity markets has potential effects on production decisions as well as on prices and these may give rise to more material concerns about efficiency. A change in the bid of one generator may induce another not only to change its own bid, but also to change its own production plans for the relevant period. Indeed, if a re-bid leads to a change in dispatch schedules – i.e. not just to the re-pricing of the prevailing schedule – adjustments in the output of other generators will be required. Such production adjustments typically involve time lags and costs, and costs tend to be higher the shorter the time period over which adjustments have to be made. Late re-bids therefore tend to impose higher costs. The point will be developed a little more in the next section, but here it can be noted that there can be a potential 'externality' (an effect on third parties unaccompanied by compensation/payment from the party responsible for the effect) in situations in which re-bids can be made without cost to the party making them. In general, 'external effects' are potential sources of inefficiency.

One again, however, the situation is complicated by other factors. For example, there is a potential concern that in relation to fast response plant. A jump in price caused by a late re-bid may lead to a ramping up of output from fast response plant. The appropriate reward for this would be the market prices established in the relevant dispatch periods, i.e. those in which the fast response plant was generating, but the NEM rules would determine the price on the basis of average prices in pricing periods for some part of which the plant was not dispatched. Hence, in circumstances in which the re-bid was very quickly reversed, the market price fell back again and fast-response plant was ramped back down again, all within the same trading interval, the fast-response plant would tend to be under-remunerated.[12]

On the arguments set out earlier this could, over the longer term, imply that less fast-response capacity became available to the system, which would in turn tend to increase the payoffs from very late re-bidding and to increase the frequency with which it occurred. Whilst the effects of reduction of capacity of fast response plant on the system might be mitigated by adaptations of other plant that are designed to increase responsiveness, this does not affect the underlying point. Such adaptations would themselves be under-remunerated and would require extra costs to be incurred. That is, the incentive structure would continue to be 'tilted' against 'responsiveness' irrespective of the precise effects to which this would lead.

Possible effects linked to frequency of price-signals

Re-bidding increases the frequency of bidding in general and one of the potential pitfalls of this is that it can create opportunities for signalling among competitors that lead to higher-price outcomes. There has been a considerable amount of theoretical, empirical and public policy work on the issue in other markets.

The relevance of this work in the current context is arguable. On the one hand its significance could be low because the frequency of bidding in the relevant spot market would already be high even without re-bidding possibilities. There is a daily, day-ahead bidding process before any issues of re-bidding arise, so re-bidding frequency effects are incremental to an already high base. In addition, most major energy markets are continuously monitored by regulators and co-ordination among competitors is discouraged by the sanctions available in the event of findings of collusive behaviour.

[12] Consider for example a re-bid that spikes prices in the last dispatch period of a trading interval, a fast response set is dispatched in the first two dispatch periods of the subsequent trading interval, and a reversal of the original re-bid takes prices back down again in the remaining four dispatch periods of the trading interval. Rather than being remunerated at the high prices in the two dispatch periods in which the fast response set was operating, the price achieved would be the (lower) average of the high and low price dispatch periods.

On the other hand, it is possible that, given the frequency of day-ahead bids, incentives for aggressive pricing are already subdued and any further weakening of them could have significant effects.

The question may, in the end, reduce to one of materiality: signalling in re-bids aimed at co-ordination may exist, but its implications may be of a lower order of magnitude than in the kinds of context where such tendencies have been found to be a significant policy or enforcement problem. In bidding markets the problematic contexts tend usually to be associated with relatively infrequent, non-repeated auctions, such as for radio spectrum or for major procurement exercises. In these cases, competing bidders start from a position of very limited information about the bidding strategies of their rivals and have only a limited period of time in which to learn. The payoffs from signalling can be expected to be much higher in such circumstances and, on the basis of the empirical evidence, appear in a number of contexts to have been sufficient to overcome the deterrent effects of competition law.

Sequencing of bids and deception issues

There are also possible issues raised by the 'two-stage' nature of the bidding process, an issue encountered in a past RPI study of reverse e-auctions in UK NHS procurement. In that case it was concluded that arrangements in which initial posted bids were followed by an 'open-outcry' auction process were liable to affect the opening bids (compared with what those bids would have been in a sealed-bid tender). Specifically, compared with a sealed bid tender, the opportunity afforded by re-bidding could lead to higher posted bids, since any initial bids could later be adjusted. Put simply, the initial bid could be used as a cost-free punt at securing the relevant contract at a higher price.

The corresponding issue in electricity is that the potential for re-bidding serves to weaken the 'firmness' of the initial bids. As well as weakening the information content of initial bids, there is a possibility that a generator might reason that it can make initial bids a little higher since, in the event that they are undercut and plant that it anticipated would be dispatched is not scheduled to be dispatched, they can be lowered if necessary.

Again, however, the problem appears likely to be less significant in a context in which 'opening bids' are made with high frequency and where any pattern could easily be detected by a regulator. As far as we are aware, re-bidding in the NEM, when it occurs, does not conform to a pattern in which initial bids are undercut and prices fall. However, if that is correct, it is not by and of itself decisive evidence of the absence of the inefficiency. In theory, at least, initial prices could be elevated, but not subsequently undermined by re-bidding, in which case what would be observed would be very little re-bidding activity.

Weakening the 'firmness' of initial bids also provides greater opportunities for the kinds of deceptive conduct that are familiar in financial markets and which can be labelled 'market manipulation'. The issues here are somewhat different from the traditional concerns of competition law and policy since, although they can be said to involve notions of economic efficiency, the way in which inefficiencies come about is not the same. More specifically, inefficiencies arise from deceptive conduct aimed at influencing the expectations of other market participants.

This is a deep and difficult area – see the complexities of financial markets regulation – and here it will simply be noted that lack of firmness in initial bids coupled with re-bidding opportunities increases the scope for deceptive conduct. An initial bid could provide the system/market operator (and hence other market participants, via their interpretations of information later released by the system/market operator) with false information about a generator's intentions, which can later be exploited by, say, a very late re-bid.

The inefficiency/harm caused arises from the degradation in the reliability of the information that is made available to market participants. Competitive markets can be described as 'discovery processes' and deceptive conduct serves to impair the effectiveness of information discovery. The problem is ubiquitous in market systems and it would be utopian to believe that it can be eliminated entirely. The relevant policy tasks are therefore chiefly to do with initial assessment of the materiality of the problem in a given context and, if the problems are found to be sufficiently material, with the development of methods of mitigation.

'Regulatory arbitrage' / gaming of the rules

Another relevant issue in the current context might be re-bidding that seeks to take advantage of opportunities to increase revenues afforded by the relevant rules. For example, in short time periods when demand is hovering just below a large upward step in the bid function the NEM rules allow a generator to re-allocate a quantum of output from a lower to a higher price band. This moves the 'step' in the bid function to the left (in diagrammatic terms) such that the system/market operator continues to dispatch the same level of output from the generator, but at a substantially higher price.

The tactic is constrained by a number of factors, some of which are set out in the *Stanwell* judgment. For example, if the system/market operator is sitting on a lower price bid that can be called on to replace the re-allocated quantum, the re-allocation may be unprofitable on account of lost volume. If it is the case that there is no alternative bid in play – i.e. if demand increased a little it would be the generator's higher price bid that would be accepted next – there remains a question as to whether or not the volume re-allocation is inefficient, which in turn raises the question of what is the relevant efficient price.

At first sight the answer to the second question seems clear enough: it is the lower of the prices in the two bands, since demand is insufficient to warrant dispatch of output higher than was originally bid in the lower band. However, the banding itself is constrained by the market rules. It may well be the case that the higher price would be inefficient, but it could also be the case that, in the relevant conditions, the lower of the prices is also inefficient on account of being 'too low'.

Whilst it may reasonably be concluded that regulatory limits on the number of price bands can give incentives to 'manipulate' prices in the neighbourhood of the discontinuities that they create, it is much more difficult to conclude in any particular case that, *given the discontinuities*, the behaviour that is observed is actually inefficient. As indicated, the market rules force approximations and the efficiency issue reduces to a comparison between two, different 'approximations'. In the end, whether harm is done may depend upon whether there is a systematic tendency toward the less good approximation.

Gaming of the rules is an issue that has attracted considerable attention in the analysis of electricity markets, but the particular type of conduct at issue – moving output from one price band to another in a generator's bids – is just one example of a much more general phenomenon that occurs when, usually in the name of administrative expediency, boundaries are created that do not correspond with clear-cut distinctions in economic realities. An administrative regime might, for example, place different burdens on larger or smaller firms, and base the distinction between the two categories on some bright line measure of the size of the firm. Incentives then tend to be heavily distorted in the neighbourhood of the boundary. For example, a firm may choose not to expand on account of a discontinuous upward jump in regulatory costs that would be encountered if did expand.

Full analysis of the efficiency implications of administrative arrangements that create such 'threshold discontinuities' will therefore properly take account of the trade-offs between administrative expediency and the effects of the distortions of incentives that may be caused. Among other things, this will likely have to take account of what might be called the 'population density' in the neighbourhood of the threshold, which affects the frequency with which distortions of incentives might serve to affect behaviour, as well as the magnitude of the harm caused in the problem cases and the administrative benefits of the thresholds themselves.

6. Incentives for inefficient re-bidding

The general incentive for inefficient re-bidding is the increased profitability that may result. Assessment of incentives therefore involves examination of the various ways in which this might come about.

Late re-bids to take advantage of non-responsiveness

A re-bid may trigger both pricing responses (re-bids by rivals) and physical responses (e.g. a change in the dispatch decisions of the market/system operator). Price reactions by rivals can be very quick, down to a period of a few minutes under the NEM rules, but physical responses may take longer, particularly if it involves calling on plant with slower response times. The payoffs from inefficient re-bidding can therefore be expected to be higher in the latter type of circumstances.

Non-responsiveness is much less of an issue during the early part of the period following the submission of initial, day-ahead bids. Re-bids early on in the period likely allow adequate time for adjustments by other market participants and, in relation to achieving efficiency in dispatch, by the system/market operator. Significant incentives for inefficient re-bidding that is aimed at exploiting non-responsiveness are therefore likely to occur as the time to dispatch shortens to an interval in which dynamic constraints on adjustment, such as ramp rates, start to have a material effect. In general, the incentives tend to be at the maximum at the last possible moment before dispatch, i.e. the last moment that a re-bid can be submitted.

Perhaps the biggest risk of harm/inefficiency occurs from late re-bidding becoming a normal feature of a market. In such circumstances, anticipations that re-bidding is more likely to take prices up rather than down may create a risk that initial bids will come to reflect the lower responsiveness of market participants at the end of the re-bidding period, rather than the greater responsiveness at the beginning. To the extent that a likelihood of such an eventual outcome is taken into account by a late re-bidder, the effect will be to further strengthen the payoffs from, and therefore the incentives for, the bidding strategy.

For completeness, it can be noted that the directional effect of declining responsiveness is not necessarily always toward higher prices, and that there can be incentives for late re-bidding that is efficient, not inefficient. In economics, the theory of contestable markets provides a vivid illustration of how non-responsiveness can provide incentives to undercut prices that are above 'competitive' levels by making hit-and-run pricing more profitable. This depends, however, on the ability of the price-cutter to adjust its own production very quickly and is probably only feasible in relation to highly flexible plant.

Regulatory arbitrage

The AEMC's consultation document raises the possibility of generators being able to use the disparity between the 5 minute dispatch interval and thirty minute settlement interval to increase the level of price. As discussed above, there appears to be a potentially adverse effect on fast response generators in circumstances in which the market price first rises then falls within the pricing period. If fast-response generators are under-rewarded on account of the averaging that takes place over a pricing period, it can be expected that other types of plant will be over-rewarded (by the same calculation), so there does appear to be an incentive for inefficient re-bidding associated with the disparity between dispatch and pricing periods.

The materiality of the incentive effect is not immediately obvious, however. If fast-response volumes are small in relation to the total volume supplied to the system, the revenue transfer to other types of plant might be expected to correspond with relatively small movements in the market price. It is beyond the scope of this Report to seek to quantify the any such effect.

Trading on information

Information relevant to the determination of market prices has economic value and traders naturally seek to use their information in the most profitable way. Re-bidding strategies will therefore be influenced by the relevant incentives, whether they are efficient or inefficient, since re-bidding both makes use of information available to the generator concerned and also provides new information to other market participants. As indicated above, there may be incentives to delay bids in order to squeeze out extra revenues. As also indicated, whether or not this is a 'problem' will in part depend upon whether it serves to offset biases in arrangements that tend toward the suppression of returns from economic rents or to raise prices above efficient levels.

Assessing informational issues is a difficult area for public policy, as revealed in the complexities of insider trading rules. Policy on insider trading is fraught with controversies, with some holding that its efficiency effects are benign, particularly if such trading is carried out at relatively low levels (on the ground that it provides incentives for the acquisition of information) and others taking a zero-tolerance position. There is more agreement on the proposition that enforcement is exceedingly difficult, and that seems a relevant consideration for any consideration of how 'undue delay' in re-bidding might be addressed in electricity markets. Thus, it would be difficult in practice to distinguish between conduct that was simply the result of a strategy that said "wait until better information is available before re-bidding" and one that was based on deception or intentional delay aimed at exploiting declining responsiveness.

Contract positions and vertical integration

As discussed earlier, bidding strategies, including re-bidding, are influenced by a generator's contract position and/or by the extent of vertical integration. Thus long positions in the wholesale market establish incentives to bid higher, short positions (e.g. when the generator is contracted to supply more than it can produce) establish incentives to bid lower, in both cases relative to a counterfactual position in which there are no fixed price contracts.

In wholesale electricity markets that are surrounded by well-developed contractual markets, contract positions can change quickly, via trading in those financial markets. Bidding incentives can therefore change quickly too, in ways that are not captured by conventional indicators of price influence such as measures of market concentration and of pivotality. The incentive effects, and hence the potential volatility of such effects, exist whether or not re-bidding is allowed, and they affect re-bids as well as initial bids.

7. The relationship between inefficient re-bidding and market power

The relationship between 'inefficient re-bidding' and 'the exercise of market power' is, to a substantial extent, one of semantics and influenced by the meaning ascribed to each of these terms. Accordingly, and as discussed in this section, appreciating differences in the conceptions and understandings of these terms is important when it comes to developing appropriate public policies and practices for the NEM.

Different conceptions of inefficient re-bidding and market power

In discussions of the proposed rule change the term 'strategic re-bidding' has sometimes been used to refer to situations where a generator has a profit incentive to submit a re-bid so as to change the wholesale price even in the absence of new 'objective' information. Referring to such conduct as 'strategic bidding' is potentially confusing however in that the label then encompasses types of conduct that would not ordinarily be called 'strategic'. It is better therefore to think in terms of efficiency, where the term 'inefficient bid' can be used for any action that does more harm than good, i.e. others are harmed and that harm exceeds the benefits of the conduct to the bidder.

Given this, the first question to address is: does inefficient re-bidding automatically equate to an exercise of market power, i.e.: are 'inefficient re-bidding' (in this sense) and 'the exercise of market power' synonymous, or can the two forms of conduct be distinguished. Here, it is relevant to note that different conceptions of market power are used in economics.

One definition of market power is that it amounts to the ability of a business, or group of businesses acting collectively, to *influence* the market price.[13] Another is that amounts to an ability of a firm to set a price in excess of marginal cost. An immediate point applies to both definitions: neither is of much practical use without further development because they imply that market power almost ubiquitous and hence that the state of there being no market power is a theoretical construct that is almost never observed. This is unhelpful in making distinctions between economic sets of circumstances that do, or realistically may, actually exist.

What matters much more is the *degree* to which prices can be influenced by one party or group of co-ordinating parties, or to which prices can be set above some relevant measure of economic costs, since there is evidence that a high degree of influence tends to be linked, albeit not in an exact way, to inefficiency, i.e. to harm done, which is the matter of ultimate concern. The relationship is not exact because price influence can have beneficial effects as well as harmful effects – indeed modest levels of price influence are generally beneficial, hence their ubiquity – and the balance of advantages and disadvantages is sensitive to the particular details of the relevant context.[14]

For these reasons the term market power in competition law and public policy generally appears with a qualifying adjective such as 'significant' or 'substantial' so as to focus on the issue of interest, the degree or level of such power. Moreover, usage of the term itself only tends to occur when the degree of market power is thought to lie above some (often fuzzy) threshold level. Thus, a market in which individual participants each have only limited price influence would typically be described as "competitive", not as a market characterised by low levels of market power, even though it can be maintained that the latter description is accurate.

In relation to the follow up question, when is market power to be considered significant or substantial?, the implicit benchmark is once again the potential for its exercise to cause inefficiency/harm. Since public administration has costs of its own, engagement with market power issues tends only to occur where the potential for harm is sufficiently high to warrant the incurring of such costs.

In markets where prices are relatively stable over time the typical approach is to examine whether prices are elevated for a sustained period, usually measured over a period of a year or more. In electricity markets, where prices can spike to very high levels for short periods and where significant harm might become an issue over a much shorter period, a different approach is warranted, though the principles remain the same. The question is simply whether the net present value (NPV) of potential harm is sufficiently high to warrant the attention of public regulation, taking account of the costs of such regulation.

[13] Stoft (*Power System Economics*, p 318) notes: "*Market power is the ability to affect the market price even a little and even for a few minutes. This definition may sound harsh, but it is not. It is simply a definition without punitive implications.*"

[14] Price influence is central to the discovery processes that drive economic adaptation and progress.

In this context, it can be noted that the AEMC, in its recent assessment of a rule relating to generator market power, adopted a definition of 'substantial market power' which referred to a (relatively) long-term measurement of power, involving an examination of the ability of a generator or group of generators to increase *annual* average wholesale prices above *long-run* marginal costs and to sustain prices at that level because of significant barriers to entry.[15] This is consistent with a normal competition law approach and there is no conceptual difficulty in extending the definition to encompass assessments of rather shorter term price movements, or periodic but recurring spikes in prices, which lead to a deviation between the NPVs of revenues and costs of equivalent value to that implied by the AEMC test as currently specified. In both cases the NPVs of the returns from above-cost pricing, which is the underlying measure of the potential for harm, would be the same.

Given these points, references to market power are best confined to situations in which the potential for inefficiencies in bidding processes are at a level consistent with general policy thresholds. Lesser degrees of price influence can then be referred to simply as 'price influence' and their implications for efficiency can be assessed on that basis. All re-bids in the NEM involve price influence, or more strictly anticipated price influence, in this sense – they would not be made if they were expected to have no effect on the market price. The substantive policy questions are, therefore, not to do with definitions, but rather with the causes and effects of inefficient re-bids and with the materiality of the effects.

Transient price influence or pricing power

In its final rule determination on generator market power in the NEM, the AEMC referred to individual instances of re-bidding using the notion of 'transient pricing power'. This was defined as the ability of a generator to bid electricity into the market at prices above short-run marginal costs in situations of tight supply and demand. The manifestation of transient pricing power takes the form of occasional price spikes.

On one interpretation of terms this type of behaviour might be considered an example of 'inefficient re-bidding'. However, in its final decision on the market power rule change the AEMC drew the conclusion that occasional price spikes were an inherent feature of a workably competitive market since they allowed firms to recover their fixed costs. In short, the exercise of transient pricing power was seen as generally *efficient*, because it allows firms to recover their *economic* costs.[16]

In terms of the earlier discussion in this report, this accords with the notion that the ability of a generator with marginal cost of $300 per MWh operating in a context where the next lowest cost set on the system has a marginal cost of $400 per MWh can legitimately price up to $400 per MWh without this raising any policy issues (it is how we would expect a business with lower costs than a competitor to behave in a normally competitive market). The lower cost generator has pricing power/price influence, but it is not problematic.

[15] AEMC, *Potential Generator Market Power in the NEM*, p 20.

[16] The AEMC did, however, qualify this point by noting that individual instances could in some circumstances be harmful and that appropriate remedies lie in the CCA and the good faith provisions of the rules (see further the discussion below).

Deceptive conduct / market manipulation

Market manipulation based on deceptive conduct also involves the exercise of price influence and the quantum of harm caused may be greater than or lower than the kinds of threshold implicit in competition law and policy. Even if the harm is substantial, however, in this case it would not be usual to talk of this (at least outside the classroom) as a market power problem. The reason is probably that to do so would draw attention away from the main issue, which is the deception/manipulation itself.

Thus, even where a specific re-bid has exactly the same price effects irrespective of the *cause* of the ability to exercise price influence, the regulatory response might appropriately be very different according to the *nature* of the cause. The reason for this is that some of the potential effects of the behaviour are not just limited to the immediate effect on the market price: they can be much wider. Thus, the re-bid may be caused by a fortuitous and transient market opportunity and its effect may be short lived. In contrast, deceptive conduct may impair the efficacy of the market discovery process by casting doubt on the reliability of market information more generally. The potential consequences may therefore be much more damaging in the longer term.

Precisely because the harmful effects of deceptive conduct can be long term in nature and relatively diffuse (raising issues of trust and confidence in the reliability of market information) they are not readily susceptible to the kind of "effects analysis" that is conducted in many economic evaluations. The focus of policy therefore tends to be much more on the conduct itself and on the motivations and intentions lying behind it. This is rather different from the way in which most market power issues are addressed in economics.

In summary

A 'fundamentalist' linkage of inefficient price bidding to the notion of market power on the basis that the latter is to be defined as the ability to affect the market price is unhelpful because it distracts attention from some key distinctions that are important for economic evaluation. These include distinctions:

- based on the degree of market power that is exercised or is available to be exercised;
- between efficient and inefficient price influence (the equation of price influence with market power introduces a cognitive bias toward the proposition that price influence is bad, since market power in other contexts is usually linked to concerns about potential economic harm); and
- between different causes of price influence (e.g. transient system conditions, a strategy based on deception) which may be associated with very different longer-term economic effects.

8. Possible options for changes in the bidding rules

Given the nature of this Report the following remarks are limited to discussion of the broad characteristics of a few potential, alternative modifications to existing arrangements that might be considered in the event that it is concluded that there are material problems to be addressed: detailed evaluation of the options is not attempted. The options listed are not necessarily mutually exclusive and it would be possible to combine features of some with features of others.

Re-bids should be objectively justifiable

This is the approach of the rule change proposal submitted by the South Australian Minister, the main elements of which are to:

- place a greater information burden on generators to demonstrate that re-bids have been made in good faith;
- provide that a variation to a bid or re-bid must not be made unless in response to a significant and quantifiable change in price, demand or other data published by AEMO and must be made as soon as practicable after the change comes to the bidder's attention;
- provide that the non-fulfilment of a trader's subjective expectation regarding a re-bid is not a change that justifies another re-bid;
- insert a requirement for participants to provide the AER (on request) with accurate and complete data and information to substantiate compliance; and
- allow the AER to assess the intention of a participant by having regard to all of the bids and re-bids that the participant has substantial control over.

It can be noted that, whilst the discussion earlier in this Report suggests that the timing of re-bids might be expected to have significant implications for their effects and consequences – relatively early re-bids do not appear to give rise to the issues that are potentially problematic in relation to very late re-bids made very close to dispatch – a general, 'objective justification' approach is arguably not targeted on a clearly identified problem. It seems to cast the policy net too wide. It could, however, be modified so as to capture only re-bids made within a period close to dispatch, save perhaps for re-bids made in such a period which are responsive to re-bids prior to the commencement of the period.

To be operational, such an approach likely requires the development and publication of rules on what constitutes an objective justification for re-bids being made in good faith and, more specifically, how a trader's subjective expectations and objective expectations will be distinguished (an issue that was a major focus of the *Stanwell* judgement). This will therefore require rules on how boundary cases will be adjudicated, including how the different types of expectations (subjective and objective) will be defined.

It may also be difficult to assess technical claims under this proposal. Even if re-bidding is limited to actions that can be justified by changes in AEMO information, there may still be a question of whether a re-bid is sufficiently related to the change in information as to 'justify' the particular action.

Finally, placing a greater information burden on generators to substantiate that all re-bids have been made in good faith will likely entail very high administrative and compliance costs. It follows that for such a shift to be justified on efficiency grounds it is necessary to be able to show clear, significant and substantiated prospects of efficiency gains associated with the rule change (implying that the magnitude of the problem to which the rule is addressed would have to be significant or potentially significant).

Fixed/hard 'gate closure' ahead of dispatch

This option would see the introduction of rules that disallow re-bids made within a certain time period prior to dispatch. So, for example, rather than being allowed to submit bids right up until the 5 minute interval prior to dispatch, it might be required that any re-bids be submitted up to, say, 60 minutes (or longer) in advance of the relevant dispatch period.

The approach is targeted on late re-bids, which is where the potential for inefficiencies is greatest. Compared with the 'objective justification' approach, it is permissive of re-bids ahead of gate closure, but is restrictive in relation to the possibility of re-bids post gate closure (these are disallowed, irrespective of whether or not they might be objectively justified). The relevant trade-off to be considered is therefore between the possible losses arising from prohibition of some re-bids that would have been efficiency enhancing had they been evaluated and the much simpler and lower-cost administration/enforcement requirements.

Gate closure arrangements are to be found in other jurisdictions and the relevant duration between gate closure and dispatch shows significant variation from case to case. In Britain, for example, since 2002 gate closure has been 1 hour prior to dispatch, but was longer before that. Direct read-across is difficult, however, because of design differences among systems. Thus in Britain there is continuous, bilateral, pay-as-bid trading that allows for adjustments to positions up until gate closure.

The most obvious concern about this option is that, in a single-price market like the NEM, introducing gate closure might simply lead to a flurry of bidding activity just ahead of closure, such that other market participants do not have time to respond. That is, in terms of price responsiveness, the problem is simply pushed back to an earlier time. On the other hand, the interval between gate closure gives the system/market operator more time to respond to re-bids in making its dispatch decisions (e.g. by beginning a ramping process earlier), which will tend to weaken some of the incentives for late re-bidding.

It can also be noted that gate closure cannot be expected to eliminate all potential issues associated with the disparity in the lengths of trading intervals and dispatch intervals. Bidders would still be able to profile their dispatch offers over the (longer) trading intervals, although the approach might mitigate potential problems to a degree that is linked to the length of the time interval between gate closure and dispatch: the longer that interval the greater the likely mitigation. If, therefore, there were interest in introducing gate closure only a short time before dispatch, it might be worth examining whether further mitigation might be achievable by fixing gate closure at a time other than the start/end of trading intervals. The two do not necessarily have to coincide.

'Soft' gate closure

This approach starts with a designated gate closure period, but allows for variations in that period contingent on specified events having occurred, an obvious example of which might be a re-bid that is made within, say, 15 minutes of the designated gate closure time. That is, such a re-bid would trigger an extension of the remaining time in which a responsive re-bid could be made. This would reduce the opportunities to exploit non-responsiveness in re-bidding by other market participants before gate closure. On the other hand, it would also potentially allow for the re-bidding process to continue right up to dispatch, thereby re-establishing the possibility of exploiting physical non-responsiveness.

Randomised gate closure

Randomised gate closure is an option that is perhaps of more theoretical than practical interest, if only because of likely psychological resistance to the notion of randomness as a policy instrument. The point of it is to counteract incentives to make last-minute bids (to which others have difficulty in reacting) by making it impossible to know when the last minute will actually occur. Thus, it might be specified that gate closure will occur between 15 and 90 minutes ahead of dispatch, with the precise time in that interval being determined by a random number generator.

Like the other gate closure options, the approach would have the effect of prohibiting some re-bids that reflect significant changes in the market and/or are capable of increasing economic efficiency.

Asymmetric restrictions on re-bids

In this option re-bids would be treated differently depending on whether or not they will raise or lower prices. A variant of the option was considered at the first drafting of the National Electricity Code and would have involved placing restrictions on the ability of generators to submit re-bids that had the effect of shifting volume into higher price bands within three trading intervals prior to dispatch. However, generators would have been able to submit re-bids that shifted volume into lower price bands up until the time of dispatch.

The approach is based on the implicit assumption that all or nearly all re-bids that have the effect of raising price are harmful in some way or another, whereas re-bids that reduce prices are beneficial. This could be the case, but there is no reason in principle to think that it will always or usually be the case. A re-bid that raises the market price may, as explained earlier, move the market price to a more efficient level.

The asymmetric treatment of re-bids may also have unintended consequences for initial bids. If only downward adjustments in price are allowed in the re-bidding process, there could be a tendency for generators to make higher initial bids, since re-bidding prices down will be an option but re-bidding prices up will not.

Mitigating the externality

Another option might be to continue to allow re-bidding, including late re-bids, but to reflect any assessed risk of harm in a compensatory payment to be made in the event of a re-bid. The payment could be to the market/system operator. This is a variant of a classic approach to reducing inefficiencies associated with actions by one party (which benefit that party but which cause a greater harm to others) by 'pricing' the harm caused.

There are a number of sub-options associated with the approach. The payments could be levied on all re-bids on an averaged basis, which would be simple but might be poorly targeted. Alternatively, payments could vary according to some correlate of assessed, harmful effects. For example, there might be no payment for re-bids that are significantly ahead of dispatch, but positive payments for late bids, perhaps rising in value the closer to dispatch the re-bid is made – a variant that is not unlike gate closure, but is less restrictive in its treatment of post gate-closure bids.

A further refinement would be a requirement for compensatory payments for re-bids that had the effect of raising the market price, whose value might be linked to the magnitude of the price increase multiplied by the output of the generator concerned (so that bigger price effects trigger bigger payments), although in this case there might be the unwanted consequence of increasing incentives to increase initial bids (see the earlier remarks on the effects of asymmetric approaches).

A 'shelf-price' approach

Many markets (e.g. the local supermarket) have the feature that the supplier posts prices (makes offers) and the final decision whether to purchase or not, and in what quantities, is left to the customer. That is, there is a sequence to decisions: supplier first, customer second.

An analogue of this situation in wholesale electricity might be:

- Generators post initial bids and are allowed to re-bid up to a designated gate closure time ahead of dispatch.
- Demand-side participants, but only the demand side, can re-bid at any time up to dispatch.

The trade-offs are the same as for other gate closure options: avoidance of the enforcement costs associated with approaches that rely on distinguishing between 'good' bids and 'bad' bids at the cost of preventing some 'good' re=bids very close to dispatch. Compared with simple gate closure, the approach adds an additional protection against inefficient re-bidding by giving demand-side participants an additional option, to re-bid post gate closure.

In the current circumstances of the NEM, the extra option is unlikely to have a material impact on outcomes because demand side participation is highly limited. The option does, however, have value for buyers and its existence might work as a factor that would encourage greater demand-side participation in the NEM over time. Given that an active demand side is something of a Holy Grail in the context of wholesale market liberalisation – it would move wholesale electricity markets toward a structure where the balance of (collective) power between buyers and sellers was less unequal – even a small contributions to the quest might make a 'shelf-price' option worth exploring in greater detail.

Market manipulation or market abuse rules

The earlier discussion identified two different pathways by which inefficiencies might eventuate:

- where the individual instance(s) of re-bidding involve prices deviating significantly from underlying *economic* costs, but no deception is involved; and
- where the motivation or purpose behind re-bidding behaviour, taken in conjunction with initial bids, is in some way deceptive or misleading and aimed at artificially changing market conditions (market manipulation).

In discussions surrounding the good faith provisions of the rules, these different aspects of unwanted re-bids may sometimes have become somewhat entangled. Whereas the first pathway engages the traditional concerns of competition law and would point to reliance on the same principles in the event that it is concluded that competition law alone is not sufficient to deal with identified, material problems, the second pathway engages the traditional concerns of financial markets regulators, which are of a different, though not always entirely dissimilar, nature. Part of the current difficulty may be that whereas the notion of good faith in trading is more traditionally associated with issues of deception, the notion is being expected in the NEM context to serve two masters. Moreover, experience in financial markets suggests that simple good faith requirements are by and of themselves not sufficient to address the kinds of problems encountered.

These are complex areas and it is well beyond the scope of this Report to give even indicative suggestions of the type made in relation to gate closure. There may, however, be merit in considering whether a more targeted approach could be appropriate, based on seeking to make clearer distinctions between the market manipulation issues and questions about the significance and effects of transient pricing influence and market power. It could be asked:

- Would the NEM rules benefit from the inclusion of a specific 'market manipulation' provision in the NEM?
- If so, could/should the 'good faith' bidding rules be broadened to encompass a wider range of unwanted conduct, or would it be better to develop market manipulation provisions from scratch?
- Alternatively, might there be a case for unified 'market abuse' provisions that encompass both types of problem, but based on a common notion of 'unacceptable harm' to other market participants and enforced on an *ex post* basis?

Review of other aspects of the NEM rules

The discussion in earlier sections of this Report suggested that some of the issues that appear to be causing concern to regulators and market participants are closely linked to specific aspects of the market rules. In particular, potential problems have been identified that are linked to (a) the restriction on the number of price bands for bidding, which introduces more significant 'steps' into bid functions and may thereby complicate efficient price discovery, and (b) the disparity between dispatch and pricing periods, which may be a source of unwanted 'external effects' among generators (see the earlier discussion of fast response plant).

Recognising that all electricity systems tend to rely on simplifications and approximations in the name of administrative efficiency, it may nevertheless be worth revisiting some of the detail of bidding rules to examine whether the balance between the achievement of administrative and market efficiencies is still appropriate, particularly in light of developments in information technologies that may enable better approximations than were feasible in the past.

Annexes

Annex 1

Contract positions and bidding

For illustrative purposes it is assumed in this Annex that demand and bid functions are continuous in price. That is, the analysis abstracts from the step jumps in these functions that are discussed in the text.

Consider a generator facing a residual demand curve:

$$q(p) = D(p) - S(p)$$

where:

p is price,

$D(p)$ is system demand at price p

$S(p)$ is the volume of output bid by other generators at price p.

Assume that the generator has contracted for a volume x at a fixed price π. The profit function is then given by:

$$\pi x + p\,[q(p) - x] - C(q(p)),$$

where $C(q(p))$ is the cost of producing output level $q(p)$.

The first term in this expression is the revenue from supplying the contract volume at the contract price, the second term is the profit from uncontracted sales on the spot market.

The condition for maximising profit with respect to price can be characterised as follows:

$$\frac{p - C_q}{p} = \frac{\theta}{\eta}$$

where:

C_q is marginal cost,

θ is the proportion of total output supplied that is uncontracted, and

η is the elasticity of the residual demand curve.

When no output is contracted this yields the standard, textbook result that the price-cost margin, the difference between price and marginal cost as a fraction of price, is equal to the reciprocal of the relevant price elasticity of demand. However, when some output is contracted/hedged it can be seen that the price cost margin is lower. For example, if two thirds of volume is sold under contract, the price-cost margin is only a third of what it would be in the absence of the contract.

Three points are worthy of note here:

- Contracted volume has the same effect as an increase in the elasticity of demand. In the case of two thirds of output being contracted, the effect is equivalent to a tripling of the demand elasticity.
- When all the output supplied is contracted price is equal to marginal cost: it is 'as if' there were perfect competition in the market.
- When the contracted volume exceeds the total volume supplied, i.e. θ is negative, the profit-maximising price will be less than marginal cost.

The intuition behind the last of these three results is that the generator has a short position in the spot market: it is effectively a net buyer of electricity in that market. As a net buyer, its interests lie in keeping spot prices low. It therefore uses any price influence it has to *reduce* the spot price, which means putting extra volume on to the spot market even if that volume itself costs more to generate than is recouped from its sale in the spot market.

Roughly then, other things equal a higher contracted volume will tend to induce more aggressive bidding strategies in the spot market.

Annex 2.

Cost functions

Many economic models of power systems tend to assume that the operating costs incurred in a given pricing period are a function only of the output supplied in that period. In effect, there is an implicit assumption (in these models) that cost functions are inter-temprally separable and of the form:

$$F + \Sigma C_t(q_t)$$

where q_t is output in period t, C_t is cost in period t and the summation is over a sequence of pricing or dispatch periods.

It is, however, more realistic to assume that cost functions are of the form:

$$F + \Sigma C_t(\dots, q_{t-2}, q_{t-1}, q_t, q_{t+1}, q_{t+2}, \dots).$$

That is, the level of output in any period t may affect costs in periods other than t. We have used the shorthand expression 'dynamic costs' to refer to this situation.

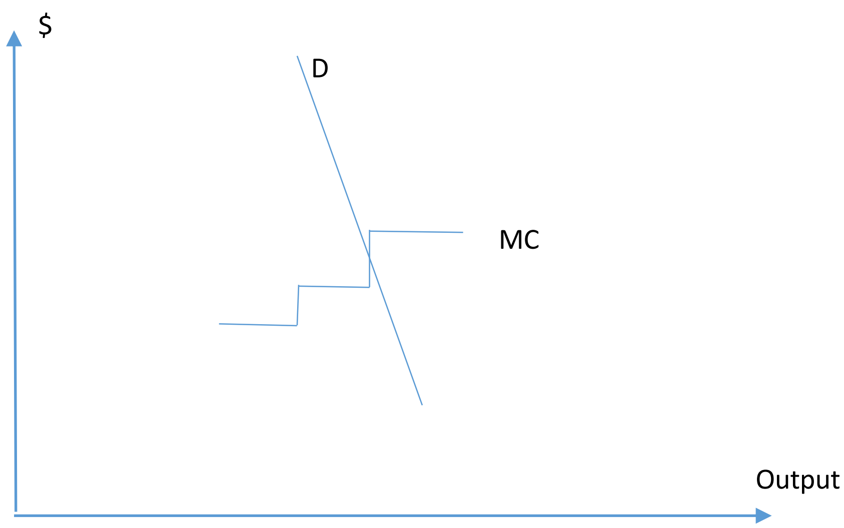
The most frequently cited examples are start-up, shut-down and ramping costs. A decision to start-up a generating set in order to provide output in a particular period will also affect costs in subsequent periods, until such time as the set is closed down. It will mean that start-up costs in those subsequent periods can be avoided. Similarly, a decision to shut down a set will affect costs in later periods. Any subsequent decision to supply output will require that start-up costs are incurred whereas a decision not to shut-down means that those start-up costs can be avoided.

Constraints on ramp rates, which are most significant for certain types of thermal plant, imply that the costs of supplying a given level of output in one period will depend on output in the previous period. Indeed, physical constraints may mean that some levels of output in the period will not be feasible even though they are, over a longer period, well within the capacity of the relevant plant. In economic terms, this means that, at the physical limit, the cost of incremental output becomes infinite and that the level of output at which this discontinuity occurs is a function of output in the previous period.

Annex 3

Illustration of economic rent

In the diagram, D shows a system demand curve and MC shows an incurred marginal cost stack (dynamic cost issues are ignored in this illustration). As shown, demand conditions are such as to pick out a price between adjacent levels of MC. Efficiency in production would require that the lower cost set, but not the higher cost set, is dispatched. The efficient price is then higher than the MC of the marginal plant dispatched, and the difference is an economic rent available to remunerate other costs.



The vertical steps in the cost function are caused by indivisibilities in production, but similar effects can be introduced by market rules. Even if the MC curve were smooth, the NEM market rule that limits generators to bidding over ten price bands would introduce discontinuities in the aggregated bid function. A generator that wanted to bid close to its marginal cost function would, in these circumstances, have to approximate a smooth function with a stepped function.