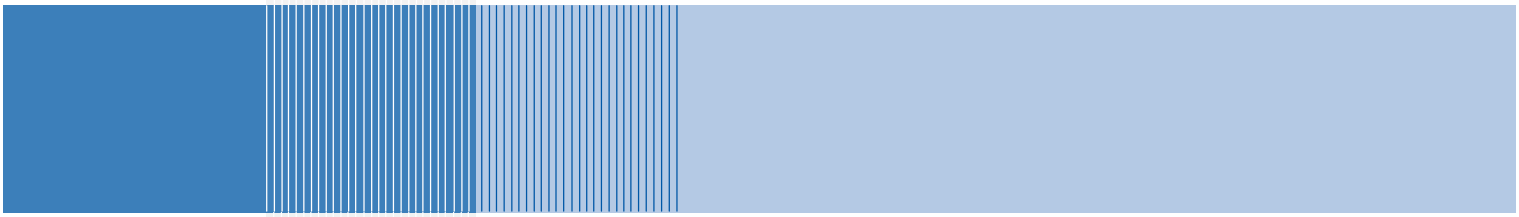
22 June 2011

Potential Generator Market Power in the NEM

A Report for the AEMC



NERA
Economic Consulting

Project Team

Hayden Green, Senior Consultant

Greg Houston, Director

Adrian Kemp, Associate Director

NERA Economic Consulting
Darling Park Tower 3
201 Sussex Street
Sydney NSW 2000
Tel: +61 2 8864 6500
Fax: +61 2 8864 6549
www.nera.com

Contents

Executive Summary

This report has been prepared for the Australian Energy Market Commission (AEMC). Its subject is a number of matters arising out of the Rule change that is proposed by the Major Energy Users (MEU). The Rule change request from the MEU seeks to constrain the contended exercise of market power by generators in the NEM. The MEU's specific concern is that, on days of very high demand, large generators are able to cause the wholesale spot price to increase by more than it should by offering prices that far exceed their costs.

The MEU rule change proposal refers repeatedly to economic concepts such as 'market power' and 'effective competition'. However, these terms are often used without a great deal of precision, and often not in the way that they are generally understood and applied in competition economics. To that end, the principal purpose of this report is:

- to describe the economic concepts of 'competition' and 'market power', the associated implications for regulatory measures that may be designed to enhance competition or mitigate market power, and how those concepts apply to wholesale electricity generation markets such as the NEM; and
- to consider the appropriate market definition for the purposes of considering the proposed Rule change, and to discuss the role and relevance of market definition in this particular instance.

In order for the economic concepts of 'competition' and 'market power' can be properly comprehended, it is helpful first to understand the (often misconstrued) concept of marginal cost, which is of central relevance to the efficiency of pricing.

Marginal Cost

Marginal cost is the added cost of producing a specified increment in output or, equally, the cost that is avoided by reducing production by a specified amount. Marginal cost can be estimated in either short run or a long run terms. The fundamental difference between short run marginal cost (SRMC) and long run marginal cost (LRMC) is the time frame under consideration and the implications of this for the extent to which a firm can adjust its production process.

- SRMC is the cost of an incremental change in demand, holding at least one factor of production – generally, capacity – constant; whereas
- LRMC relaxes this constraint and reflects the cost of an incremental change in demand assuming all factors of production can be varied.

An important distinguishing feature of SRMC is that, in the event existing capacity is insufficient to meet all demand, SRMC is represented by whatever level is necessary to curtail demand to match available supply. It therefore takes account of the costs of shortages faced by customers. The estimation of LRMC accounts for the fact that, in the long run, firms have the option of expanding their capacity in order to meet increased demand. Measuring LRMC therefore involves estimating the costs associated with undertaking a

capacity expansion sooner than would otherwise be the case in response to a change in demand.[1]

Both SRMC and LRMC can fluctuate over time and there is no *a priori* reason to expect them to be equivalent at any particular moment. However, there is a strong 'in principle' link between SRMC and LRMC over the long term. In particular, when demand is growing over time, or subject to short term fluctuations, SRMC can be expected to increase to the point at which the expected cost of curtailing demand exceeds the cost of expanding capacity to *meet* that demand, ie, when LRMC < SRMC.

Of course, market imperfections mean that the timing of capacity expansions will not always be perfect, eg, SRMC may rise above LRMC for a period if the optimal expansion is particularly lumpy, or occurs on slower than the ideal timing.[2] Nonetheless, provided that the concepts are measured over a sufficiently long timeframe, the link between SRMC, LRMC and new investment decisions should mean that, on average, there is no material difference between the value of SRMC and LRMC.

Competition and Market Power

In the context of trade and commerce, **competition** is a process whereby firms strive against each other to secure customers for their product or services. In a perfectly competitive market[3] there is a strong relationship between prices and costs. Firms can sell whatever they like at a price equal to its SRMC (which also equals the LRMC) and earn a return that is equal to that available if they were to divert their capital to its next best use. Any change in market conditions that results in prices above or below this level will be met with an immediate response that eliminates those positive or negative margins.

Of course, the distinguishing characteristics of perfect competition are seldom (if ever) seen in real markets. In the more realistic setting of a workably competitive market, prices will not always be perfectly aligned with SRMC and LRMC. However, any change in market conditions that results in prices that are significantly and persistently *above LRMC* or *below LRAC* should prompt a supply-side response *over the long term*. Specifically, in time, the rivalry that exists between suppliers should restore prices to levels that, on average, reflect the LRMC of adding capacity (or, equally, the LRAC of reducing capacity).

---

[1] The LRMC of adding capacity (and the LRAC associated with reducing capacity) will therefore be determined by the operating and capital costs associated with the optimal investment profile needed to meet the relevant increment (or decrement, as the case may be) in demand. This may comprise investment by both existing market participants and by new entrants, and, potentially, investment in different production technologies. When the term LRMC is used throughout the remainder of this report, it should be interpreted in this way, ie, as the LRMC *for the market.*

[2] Government intervention may also affect the relationship between SRMC and LRMC. For example, government taxes and subsidies can affect the economics of various investment propositions and, potentially, the LRMC of expanding capacity. Such interventions may therefore also influence the time it takes for the SRMC of curtailing demand to reach the new LRMC benchmark. The renewable energy target and a carbon tax (if implemented) are two relevant examples that are discussed in more detail in footnote 7.

[3] The theoretical ideal of a perfectly competitive market is one in which there are many buyers and sellers, homogeneous products, no barriers to entry, expansion or exit, costless transactions and perfect information. *See:* Nicholson, W. (1998), *Microeconomic Theory: Basic Principles and Extensions*, Seventh Edition, The Dryden Press, USA pp. 401 – 402 (Nicholson (1998); Pindyck & Rubinfeld (1995), *Microeconomics*, Third Edition, Prentice Hall Inc., New Jersey, USA, p. 271 (Pindyck & Rubinfeld (1995); and Case & Fair (1996), *Principles of Microeconomics*, Fourth Edition, Prentice Hall Inc, New Jersey, USA, p. 53 (Case & Fair (1996)).

Substantial market power is the antithesis of workable competition. A firm (or group of firms) can be understood to possess a substantial degree of market power when it is able to sustain prices that would *not be observed* in a workably competitive market. Specifically, a firm can only be considered to have substantial market power when:[4]

- it has the ability to *sustain* prices *above LRMC*, including an appropriate return on capital and accounting for risk; and
- it is insulated from competition by *significant barriers to entry and expansion* (as opposed to, say, minor differences in product attributes).[5]

The application of price control should therefore focus on addressing the existence and exercise of substantial market power – as opposed to, say, *temporary* pricing power, which is a common feature of workably competitive markets – and the attendant adverse consequences for economic welfare. This requires a focus on:

- genuine and enduring barriers to entry and expansion, as the fundamental source of the substantial market power; and
- the undertaking of long-term price cost tests, as evidence of the exercise of substantial market power.

The overarching criteria for imposing price control is that any such intervention must enhance economic welfare, relative to the counterfactual in which that intervention is not undertaken. To intervene in any other circumstances would be contrary to the long term interest of consumers since it would involve the needless invocation of a 'second best' solution.

Application to Electricity

Energy-only electricity generation markets such as the NEM have some characteristics that distinguish them from many other markets. The product itself is homogeneous, non-storable and has few (if any) substitutes. Suppliers are also characterised by significant variation between the costs of the different generation technologies available, eg, base load, mid-merit and peaking plants. Prices are also highly dynamic, with a new 'spot price' determined every thirty minutes. Despite those differences, a workably competitive wholesale electricity spot market functions no differently from most other workably competitive markets.

Specifically, with certain limited exceptions, if prices are significantly and persistently *above LRMC* or *below LRAC* (recognising that the measurement of LRMC and LRAC will depend

---

[4] Put another way, a firm has substantial market power when it has the ability to sustain prices that *should* attract additional investment (ie, because they exceed LRMC) – from either existing market participants or new entrants – but that do not, because such investment is prevented or delayed by various factors (ie, barriers to entry).

[5] When considering the significance of barriers to entry (and, by extension, the degree of competitive constraint provided by new entry and expansion), it is necessary to assess the time it would take for a new firm to enter the relevant market and offer customers a competitive alternative, or for existing firms to expand, following the exercise of substantial market power. The evaluation of whether these responses would be sufficiently timely will vary with the dynamics of the market. As a general rule, entry and expansion will generally provide an effective competitive constraint if they would occur in an appropriate time to deter or defeat any non-transitory exercise of substantial market power. However, the appropriate timeframe will depend on the particular market under consideration.

upon the relevant *types* of capacity that are added in each instance[6]) this should, given time, prompt a supply-side response that restores prices to these levels.[7] It follows that a generator (or group of generators) can be considered to possess a substantial degree of market power when it is not constrained by the forces of workable competition, ie, when:

- it has the ability to increase average spot prices to such an extent and with sufficient frequency that they exceed the LRMC of adding capacity, including a return on capital and accounting for risk;[8] and
- it is insulated from the forces competition by significant barriers to entry and expansion (as opposed to, say, minor differences in product attributes) that enable it to sustain average prices at that level.[9]

Any assessment of whether a generator has a substantial degree of market power consequently requires:

- a focus on genuine and enduring barriers to entry and expansion, as the fundamental source the substantial market power, noting that this might also include 'strategic' barriers to entry and expansion; and
- the undertaking of long-term price cost tests, as evidence of the exercise of market power, eg, comparisons of average spot prices to the LRMC of adding capacity, rather than comparisons of spot prices to SRMC at particular points in time.[10]

Although periods of high prices that appear to have arisen from strategic bidding conduct (such as opportunistically withholding supply in order to increase spot prices) are certainly

---

6 As noted earlier, the LRMC of adding capacity (and the LRAC associated with reducing capacity) is determined by the operating and capital costs associated with the optimal investment profile needed to meet the relevant increment (or decrement, as the case may be) in demand. This may comprise investment by both existing market participants and by new entrants and, potentially, investment in different production technologies. For example, depending upon the circumstances, the most efficient expansion profile may involve investment by both existing generators and new entrants, and a mix of technologies, eg, base-load, mid-merit and peaking plant and, possibly, additional transmission or interconnector capacity.

7 The renewable energy target and a carbon tax (if implemented) also have the potential to influence the LRMC of new generation capacity and the period of time during which SRMC and LRMC are misaligned. The principal effect of the initiatives will be to make renewable (ie, low carbon) forms of generation less expensive and carbon-intensive generation (eg, coal-fired plant) more expensive. The schemes may therefore change the *cost* and, possibly, the *configuration* of the next capacity expansion in a location. Specifically, it may be the case that, absent the initiatives, the optimal investment profile to meet the next capacity expansion would have comprised largely new coal-fired base-load plant. The effect of the schemes will be either to increase the cost of any such plant, or to change the economics of the investment to such an extent that a greater proportion of renewable energy is brought on-stream instead. In each case, the LRMC of the capacity expansion increases, relative to the state of the world in which the government intervention does not take place. In addition, because the LRMC has increased, it may take longer for the SRMC of curtailing demand to reach that new, higher, threshold.

8 Note again that this may involve engaging in strategies such as predatory pricing, albeit for the same purpose.

9 As noted above, as general rule, entry and expansion will provide an effective competitive constraint if barriers to entry and expansion can be overcome in an appropriate time to deter or defeat any non-transitory exercise of substantial market power.

10 There are also a number of other indicators of substantial market power that are not discussed in this initial report. In particular, the 'Lerner Index' and the 'Pivotal Supplier Index' are two additional measures that are commonly employed to assist in the detection of substantial market power in wholesale electricity markets. However, a detailed discussion of these methodologies is outside the scope of this initial report.

relevant to the assessment indicated above, this only applies to the extent that they have had a sustained effect on average spot prices that is likely to persist over the long term. Having now established 'what it is that we are looking for', it is helpful to define an appropriate market that can be adopted for the purposes of making that assessment.

Market Definition

The purpose of defining a market is to frame the relevant arena of competition to enable the question of interest to be addressed. In this particular case, we are interested in whether particular generators possess substantial market power that may be worth addressing by means of a market intervention, eg, by imposing price control. The market definition must therefore be capable of identifying circumstances that may necessitate *ex ante* regulation to address structural concerns or enduring market failures for which *ex post* intervention is impracticable or inappropriate.

This suggests that the relevant timeframe for defining the market will need to span at least one year and possibly two. A shorter timeframe risks overlooking relevant economic forces and, in particular, mistaking temporary pricing power for substantial market power. A further advantage of adopting such a timeframe is that it allows the entire 'demand cycle' to be accounted for in the comparison of prices and LRMC – a comparison that cannot be meaningfully implemented over, say, a three month period.

The relevant product market for the purposes of assessing the MEU Rule change proposal is likely to comprise electricity energy supplied to the wholesale electricity market. There is no need to extend that definition to include electricity derivatives, since these instruments are simply another way of expressing the price for the same underlying product. There is also no need to expand the functional dimension of the market to include electricity retailing, since the complementarities between the generation and retailing functions are not so strong as to preclude the separate performance of these functions.

The drawing of a definitive conclusion on the appropriate geographic dimension of the market is an empirical exercise, the answer to which may vary from one starting point location to another. However, in our opinion, that such a modelling exercise is likely to reveal that, in many cases, the relevant geographic market was limited to a NEM region, or combinations of NEM regions. Assuming the appropriate empirical analysis did conclude that the market is delineated by a series of NEM regions or combinations of NEM regions, this does not mean that that generators located in other regions can then be ignored in the subsequent assessment of substantial market power. Rather, the constraining effect of those competitors operating by means of relevant interconnectors would also be a critical consideration.

1. Introduction

This report has been prepared for the Australian Energy Market Commission (AEMC). Its subject is a number of matters arising out of the Rule change that is proposed by the Major Energy Users (MEU). The Rule change request from the MEU seeks to constrain the perceived exercise of market power by generators in the NEM. The MEU's specific concern is that, on days of very high demand, large generators are able to cause the wholesale spot price to increase more than it should by offering prices that far exceed their costs.

The MEU states that this is a particular concern in South Australia, where spot prices in excess of $8,000/MWh have occurred for brief periods on more than twenty occasions since January 2008.[11] The basic idea of the MEU proposal is to prevent these periods of very high prices – at least insofar as they arise from the exercise of so called 'market power'. In its Rule change proposal, the MEU defines market power in the NEM as:[12]

> 'an ability of a generator to manipulate the spot price at a regional demand less than the maximum regional demand, by either physical or economic withholding of its capacity'.

A noticeable feature of the MEU rule change proposal is that it refers repeatedly to economic concepts such as 'market power' and 'effective competition'. However, these terms are often used without much precision, and often not in the way that they are generally understood and applied in competition economics. The AEMC recognises in its Consultation Paper that, in order to assess the merits of the MEU proposal, it is necessary to reach a common understanding of the problem that it is trying to address.[13] To that end, the principal purpose of this report is:

- to describe the economic concepts of 'competition' and 'market power', the associated implications for regulatory measures that may be designed to enhance competition or mitigate market power, and how those concepts apply to wholesale electricity generation markets such as the NEM; and
- to consider the appropriate market definition for the purposes of considering the proposed Rule change, and discuss the role and relevance of market definition in this particular instance.

Because our report is based largely on 'first principles', it does not contain a detailed account of the different ways in which substantial market power might be measured[14] or of whether

---

[11] However, it has suggested that such conduct may also be a potential problem in other NEM regions.

[12] MEU, Rule change request, 23 November 2010, p.32.

[13] AEMC, *Consultation Paper, National Electricity Amendment (Potential Generator Market Power in the NEM) Rule 2011*, 14 April 2011, p.19 (hereafter: 'AEMC Consultation Paper').

[14] For example, the 'Lerner Index' and the 'Pivotal Supplier Index' are two additional measures that can be used to assist in the detection of substantial market power in wholesale electricity markets. However, a detailed discussion of these methodologies is outside the scope of this initial report.

there is any indication that such power has, in fact, been exercised.[15] Its remainder is structured as follows:

- **section two** explains the often misconstrued concept of marginal cost, which is of central relevance to the efficiency of pricing and the identification of substantial market power;
- **section three** describes the economic concepts of 'competition' and 'market power' from 'first principles' and describes the circumstances in which price regulation should be contemplated;
- **section four** explores the application of those economic concepts to electricity wholesale generation markets such as the NEM;
- **section five** considers the appropriate market definition for the purposes of considering the proposed Rule change;
- **section six** concludes; and
- **appendix A** provides a more detailed description of the concept of marginal cost and some of the challenges that arise in its estimation over both the short and long term.

[15] For example, although our report identifies the importance of barriers to entry and expansion – including strategic barriers – to the identification of substantial market power, it does not provide a comprehensive assessment of precisely what those barriers are likely to be in this instance. Such detailed consideration is beyond the scope of this initial report, but would, of course, be a vital component of any subsequent assessment of whether generators possess, and have exercised, substantial market power.

2. Marginal Cost

Before the economic concepts of 'competition' and 'market power' can be properly comprehended, it is helpful first to understand the (often misconstrued) concept of marginal cost, which is of central relevance to the efficiency of pricing. In very simple terms, marginal cost is the additional cost that a firm incurs (avoids) by increasing (reducing) output by a specified increment.[16]

Marginal cost can be estimated in either short run or a long run terms. The fundamental difference between short run marginal cost (SRMC) and long run marginal cost (LRMC) is the time frame under consideration and the implications of this for the extent to which a firm can adjust its production process. This section explores SRMC and LRMC, as well as the relationship between them. Appendix A provides a yet more detailed description of these costs concepts and their estimation.

2.1. Short Run Marginal Cost

In the short run at least one 'factor of production' is fixed, ie, a firm cannot instantaneously add new production lines to its factory. It is therefore not possible for a firm to increase the quantity of a product that it is supplying by expanding its existing capacity. The only way that firms can increase supply is to use their *existing* capacity, ie, to produce more with what they already have. SRMC can therefore be thought of as the cost of meeting an incremental change in demand, *holding capacity constant.*[17]

This is often construed simply as the operating and maintenance costs associated with providing the product. At times, that can be correct, but *not always.* When an incremental change in demand can be met through increased supply from existing capacity, the SRMC *will* be equal to the operating and maintenance costs associated with producing those additional units. However, at other times, SRMC can be significantly above the marginal operating and maintenance expenditures incurred serving incremental demand.

Specifically, an important but often overlooked element of SRMC is that, in the event that supply *cannot* expand to match the incremental change in demand, SRMC rises to whatever level is necessary to curtail demand to match supply. Specifically, in situations where there is an increased risk of shortages, the costs associated with this demand side component can cause SRMC to rise *well above* variable costs. Importantly, it is during these periods of scarcity that firms are able to make a contribution to their *fixed costs*, which do not vary with output over the short-term and are therefore not a component of SRMC.

---

[16] Strictly, marginal cost is the first derivative of a firm's production cost function, with respect to output. However, its practical application involves the measurement of the change in a firm's cost of production when its output changes by a specified increment and is often also referred to as incremental cost or avoidable cost (where the specified change involves a reduction in output). For the purposes of this report, we have taken the concepts underpinning marginal, incremental and avoidable cost to be synonymous, since their technical distinctions have no consequences for the matters at hand. For further discussion *see:* Kahn, A, (1988), *The Economics of Regulation, Principles and Institutions, Volume 1* (MIT Press), p.66 (Hereafter: 'Kahn (1988)').

[17] It can also be specified as the cost that would be avoided by having to meet a slightly reduced level of demand.

Kahn (1988) offers the example of a bridge that is contemplating charging a toll. The incremental operating, maintenance and capital costs caused by each additional vehicle on the bridge are practically zero but, as Kahn observes:[18]

> '[W]hat if charging a zero toll would, at certain hours of the day, produce such an increase in traffic that cars lined up for miles at the bridge entrance and a crossing took an hour instead of a few minutes? In that event, the SRMC of bridge crossings, at those times, is not zero. It can be envisaged in terms of congestion: the cost of every bridge crossing at the peak hour is the cost of the delays it imposes on all other crossers. Or it can be defined in terms of opportunity cost: if A uses the bridge at that time, he is taking up space that someone else could use; therefore, the cost of serving him is the value of the space or capacity to others who would use it if he did not.'

In other words, in times of scarcity, the cost of serving one customer must, by definition, include the value foregone by other customers who cannot be served as a consequence. For example, if Sydney's water supply began to run low, continuing to supply some customers may mean placing restrictions on the usage of others. The costs imposed by those restrictions may be very high, and may include costs such as plant losses in residential gardens and parks, reductions in agricultural output, diminished quality of golf courses and higher production costs for breweries. All of those costs form a part of the SRMC of serving one customer in circumstances where that implies restricting supply to another.

Although SRMC can be estimated as at any particular point in time, its magnitude varies from one point in time to another. Its application in the context of decisions affecting the future (such as, following Kahn's example, whether to build a second bridge to relieve congestion) therefore relies as much on probability and expectation as on fact. As Appendix AA.1 explains in more detail, a forward-looking SRMC is the sum of the various additional costs arising under different scenarios (holding capacity constant), multiplied by the probabilities of these scenarios occurring. Formally, the expected SRMC is given by:

- the SRMC when supply exceeds demand (ie, operating and maintenance costs), multiplied by the probability that supply exceeds demand; *plus*
- the SRMC when supplies are less than demand (ie, *including* the costs of shortages) multiplied by the probability that supply is less than demand.

To summarise, SRMC can be defined as the cost of an incremental change in demand, holding capacity constant. Importantly, its estimation takes account of the potential costs of shortages faced by customers. In the event supply cannot expand to match demand, SRMC rises to whatever price level is necessary to curtail demand to match available supply.

2.2. Long Run Marginal Cost

In the long run, all factors of production are variable and so incremental changes in demand no longer need to be met from current capacity alone. Rather, firms have the option of expanding capacity in order to meet an incremental increase in demand and, equally, of reducing their capacity in order to meet a slightly reduced level of demand. LRMC can

[18] Kahn (1988), p.87.

therefore be thought of as the cost of supplying a specified, permanent increment in demand, allowing for future augmentations in supply.[19]

In most industries it is not practicable to add capacity in very small increments.[20] Rather, there are often 'economies of scale' associated with augmentations. For example, once a business has purchased land it may make sense to construct a two storey office building, even if not all of that space will be used right away. This is because adding the second storey now will be much cheaper than building it later. Taking the analogy one step further, it is likely to be yet more expensive (in unit cost terms) to add capacity 'room by room'.

In other words, capacity is often added in 'lumps' rather than very small increments. The likely effect of a permanent increment in demand is therefore to *bring forward* the time at which a planned future 'lump' of capacity needs to be added – by firms that are already in the market and/or by new entrants. The LRMC is therefore the costs – both operating and capital costs – associated with undertaking that expansion *sooner than would otherwise be the case* in response to the incremental change in demand, and the associated congestion costs.[21]

This implies that where capacity must be added in 'lumpy units' (rather than in very small increments), this gives rise to *time-dependent fluctuations* in LRMC. Specifically, the LRMC of supply in such a market will be relatively low when capacity utilisation is low and the next capacity expansion is some distance in the future, but will rise as capacity utilisation increases and the timing of the next expansion is nearer. Specifically, as Appendix A.2 explains in more detail:

- in the time period immediately following a capacity expansion, the LRMC of the next increment to capacity is low because the value of any potential deferral of that future capacity requirement is relatively low due to the effect of discounting; and
- as spare capacity declines over time and the need to invest in new capacity approaches the LRMC of the next increment to capacity increases, because the value created through any potential deferral is closer in time and so less (negatively) affected by discounting.

In other words, LRMC *changes over time* as new capacity is added. This is because the cost today of, say, bringing forward by one year a $1m investment that would otherwise have taken place in 12 months' time is much greater than the cost today of that same one year

---

[19] Note that the LRMC of adding capacity (and the LRAC associated with reducing capacity) will be determined by the operating and capital costs associated with the optimal investment profile needed to meet the relevant increment (or decrement, as the case may be) in demand. This may comprise investment by both existing market participants and by new entrants, and, potentially, investment in different production technologies. When the term LRMC is used throughout the remainder of this report, it should be interpreted in this way, ie, as the LRMC *for the market.*

[20] The exception is industries in which assets are highly mobile and capacity can be added in very small increments. In these circumstances, any level of demand can be met by quickly adding (or subtracting) capacity, ie, there is never any need to curtail demand. Of course, such industries are rarely seen in practice. We explore this in more detail below.

[21] To be clear, LRMC does *not* equal the total operating and capital costs associated with that expansion. This is because an incremental increase in demand does not generally result in investment that would otherwise never be required; rather it brings forward the timing of an expansion.

rescheduling applied to a $1m investment expected to be made in 10 years' time, because of the time value of money.[22]

In summary, LRMC reflects the cost of serving an incremental change in demand in a market, assuming all factors of production can be varied. Importantly, because LRMC is a long run concept, it accounts for the fact that firms have the option of *expanding their capacity* in order to meet an incremental increase in demand. Measuring LRMC involves estimating the costs involved with undertaking a capacity expansion *sooner than would otherwise be the case* in response to that change in demand.

2.3. Relationship between SRMC and LRMC

The previous sections explained that SRMC is the cost of an incremental change in demand, holding capacity constant, whereas LRMC reflects the cost of meeting that change in demand assuming capacity can vary. Unless assets are highly mobile and capacity can be added in very small increments – conditions that are rarely seen[23] – there is no reason to expect SRMC and LRMC to be the same *at any particular point in time*. However, there is still a strong 'in principle' link between SRMC, LRMC and capacity expansion decisions.

Specifically, when demand is growing over time, or subject to short term fluctuations, SRMC can be expected to increase to the point at which the cost of curtailing demand exceeds the cost of expanding capacity to *meet* that demand, ie, when LRMC < SRMC. In the first instance, medium term demand growth can only be met through increased risk of congestion, or the need for demand curtailment during short run peaks. However, there eventually comes a 'tipping point' at which the expected SRMC of *curtailing* demand increases beyond the expected LRMC cost of expanding capacity to *meet* that demand, at which point new investment takes place.

Exactly the same principles apply to a market in which demand is *declining* over time. In the first instance, declining demand can be met by firms continuing to supply the market with their existing capacity. However, there will again be a 'tipping point' at which the long run costs that would be *avoided* by reducing or redeploying capacity exceed the SRMC of continuing to supply the product at the current level of capacity, at which point capacity is redeployed to other markets where returns are more attractive.

Of course, in practice, it is often very difficult to time capacity expansions and reductions to coincide perfectly with the emergence of inefficient levels of demand curtailment, ie, when scarcity is either too common or too infrequent. This is particularly the case when capacity

---

[22] Put another way, the *value* today of *deferring* by one year a $1m investment expected to be made in 12 months' time is much greater than the value today of that same one year deferral applied to a $1m investment expected to be made in 10 years' time.

[23] When these conditions are present, *there is no distinction* between SRMC and LRMC since, by definition, there is no difference between the short run and the long run. Any level of demand can be met by quickly adding (or subtracting) capacity and so the need to curtail demand never arises. In these circumstances, SRMC and LRMC are always equivalent, and constant at all times. Of course, industries that exhibit such characteristics are rarely seen.

must be added and withdrawn in large increments that alter substantially the supply/demand balance. There may therefore be times when:[24]

- SRMC is *above* LRMC for a period as the market waits for new capacity to come on-stream; and
- SRMC is *below* LRMC for a period as the market waits for redundant capacity to be re-deployed elsewhere.

However, such instances of 'misalignment' are neither unexpected, given the imperfections that can affect real world markets, nor a cause for concern, provided that they are transitory. Even accounting for such periods, there is no reason to expect SRMC to differ materially from LRMC, on average, provided they are properly defined and assessed over a sufficiently long timeframe. Equally, although both SRMC and LRMC can fluctuate over time (as explained in more detail in Appendix A), there is no reason to think that either will diverge over the long term.

2.4. Summary

Marginal cost is the added cost of producing a specified increment in output or, equally, the cost that is avoided by reducing production by a specified amount. The fundamental difference between SRMC and LRMC is the time frame under consideration and the implications of this for a firm's ability to adjust its production process. Specifically:

- SRMC is the cost of an incremental change in demand, holding capacity constant; whereas
- LRMC relaxes this constraint and reflects the cost of an incremental change in demand assuming all factors of production can be varied.

An important distinguishing feature of SRMC is that, in the event that current capacity is insufficient to meet all demand, SRMC rises to whatever level is necessary to curtail demand to match available supply. It therefore takes account of the costs of shortages faced by customers. This element of SRMC is often not fully appreciated.

The estimation of LRMC accounts for the fact that, in the long run, firms have the option of expanding their capacity in order to meet increased demand. Measuring LRMC therefore involves estimating the costs associated with undertaking a capacity expansion sooner than would otherwise be the case in response to a change in demand.

Both SRMC and LRMC can fluctuate over time and there is no *a priori* reason to expect them to be equivalent at any particular moment. However, there is a strong 'in principle' link between SRMC and LRMC over the long term. Specifically, when demand is growing over time, or subject to short term fluctuations, SRMC can be expected to increase to the point at

[24] Government intervention may also affect the relationship between SRMC and LRMC. For example, government taxes and subsidies can affect the economics of various investment propositions and, potentially, the LRMC of expanding capacity. Such interventions may therefore also influence the time it takes for the SRMC of curtailing demand to reach the new LRMC benchmark. The renewable energy target and a carbon tax (if implemented) are two relevant examples that are discussed in more detail in footnote 51.

which the cost of curtailing demand exceeds the cost of expanding capacity to *meet* that demand, ie, when LRMC < SRMC.

Of course, market imperfections mean that the timing of capacity expansions will not always be perfect, eg, SRMC may rise above LRMC for a period if the optimal expansion is particularly lumpy, or occurs on slower than the ideal timing. Nonetheless, provided that the concepts are measured over a sufficiently long timeframe, the link between SRMC, LRMC and new investment decisions should mean that, on average, there is no material difference between the value of SRMC and LRMC.

3. Competition, Market Power and Regulation

This section discusses from 'first principles' the economic concept of 'competition' and the closely related concept of market power, and describes the circumstances in which the application of administered price controls might be contemplated.

3.1. Competition

In the context of trade and commerce, competition is a process whereby firms strive against each other to secure customers for their product or services. Competition limits the extent to which a firm can ignore market signals:

- by producing goods and services that consumers do not want to buy; and/or
- by attempting to sell those commodities at a price that exceeds significantly the cost of producing them.

In general terms, if the forces of competition are sufficiently strong, a business that sought to increase its price, or to reduce the quality of its product to any significant extent (other than to reflect increases in the costs of supply) is likely to find that:

- buyers switch to alternative products that are cheaper or of a superior quality; and/or
- alternative suppliers alter their production plans in order to compete.

Below we consider the relationship between prices and costs in competitive markets. We begin by considering 'perfectly competitive' markets, before turning our attention to the more realistic setting of 'workable' competition.

3.1.1. Perfect competition

Perfectly competitive markets exhibit the most vigorous competition that can be conceived. Clark (1940)[25] explains that the relationship between prices and costs in this 'ideal setting' can offer important insight into the outcomes that can be expected in more realistic market settings. Perfectly competitive markets have the following characteristics:[26]

- many buyers and sellers – sellers can always find a buyer and vice versa;
- suppliers can enter the market, exit the market and/or expand production without incurring additional costs, ie, there are no 'barriers to entry, exit or expansion' arising from, say, 'sunk costs'[27];

---

[25] Clark, J.M. (1940), 'Toward a Concept of Workable Competition', *American Economic Review*, Vol. XXX, p. 241.

[26] Nicholson, W. (1998), *Microeconomic Theory: Basic Principles and Extensions*, Seventh Edition, The Dryden Press, USA pp. 401 – 402 (Nicholson (1998); Pindyck & Rubinfeld (1995), *Microeconomics*, Third Edition, Prentice Hall Inc., New Jersey, USA, p. 271 (Pindyck & Rubinfeld (1995); and Case & Fair (1996), *Principles of Microeconomics*, Fourth Edition, Prentice Hall Inc, New Jersey, USA, p. 53 (Case & Fair (1996)).

[27] Sunk costs are those that cannot be recovered once incurred, eg, a firm may need to purchase highly specialised assets that have no alternative uses. If a firm faces or has incurred significant sunk costs it may be less inclined to enter or to exit the market. Such costs may constitute a potential barrier to entry (since firms may be less willing to incur those costs) and to exit (since firms may not be prepared to forego those costs by exiting a market).

- identical products – the characteristics of products do not vary across suppliers, and so customers are indifferent about the supplier from which they buy the product;
- perfect information – prices and product attributes are assumed to be known to all consumers and producers at all times, and so it is not possible for a seller to change its price without everyone else in the market knowing about it immediately; and
- transactions are costless – buyers and sellers incur no costs in making an exchange.

In a perfectly competitive market, any level of demand can be met by suppliers instantaneously expanding or reducing their capacity. There is therefore no distinction between the SRMC and LRMC (as they are defined in sections 2.1 and 2.2 above) of supplying the market, since supply-side adjustments are instantaneous. This dynamic produces a strong relationship between prices and costs, ie, every individual seller in the market is a *price taker*.

Specifically, a firm can sell whatever it likes at a price equal to its SRMC (which will also equal the LRMC). If a firm tries to increase its price above its SRMC its customers will know straight away that they can get a lower price for the same product from another supplier, ie, customers are assumed to be perfectly informed, and products are identical and available from many sellers. Similarly, there is no point in the firm reducing its price since it will not cover its costs and earn below-normal returns.

Moreover, any change in market conditions – such as an increase in demand or a change in the cost of an essential input – that results in prices above or below this level will be met with an immediate response that eliminates the positive or negative margins. This is because:

- if all firms in the market are seen to be setting prices above SRMC (say, due to an increase in demand), new firms will immediately enter the market 'chasing' the resulting profits, until the point at which prices are realigned SRMC;[28] and
- if all firms in the market are seen to be setting prices that are below SRMC (say, due to a reduction in demand), firms will reduce their production and divert their resources to other markets where the returns available are more attractive.[29]

To summarise, in a perfectly competitive market there is a strong relationship between prices and costs. Firms can sell whatever they like at a price equal to their SRMC (and LRMC) and earn a return that is equal to that available if they were to divert their capital to its next best use, eg, to reinvest in another market. Any change in market conditions that results in prices above or below this level will be met with an immediate response that eliminates those positive or negative margins.

---

[28] Recall that entry, expansion and exit from the market are all costless, so new firms are free to enter the market and existing firms can ramp up their production in infinitely small increments.

[29] Under perfect competition, there is no need for firms to take time to ramp down production or to arrange an orderly exit, eg, to discharge contractual obligations and sell equipment. Rather, they are assumed to be able to reduce their sales immediately and in infinitely small quantities.

3.1.2. Workable competition

Although perfect competition serves as a useful reference point, its distinguishing characteristics are seldom (if ever) seen in real markets. Sellers generally are not pure price takers, parties are almost never perfectly informed and there are almost always some barriers to entry or expansion. Economists therefore typically speak of a market being at least 'workably' or 'effectively' competitive. For example, Professor Maureen Brunt has described workable competition as:[30]

> '…a situation in which there is sufficient rivalry to compel firms to produce with internal efficiency, to price in accordance with costs, to meet consumers' demand for variety, and to strive for product and process improvement.'

When competition is workable (but not perfect) firms will often adjust their prices over the near term to reflect changes in market conditions, and in the underlying *trend* in SRMC. In particular, it is common for firms increase their prices when supply is scarce (and when SRMC is increasing) and to reduce their prices when it is plentiful (and when SRMC is decreasing). Indeed, it is only by increasing their prices when scarcity emerges that firms are able to recover their fixed costs. For example:

- when the demand for hotel rooms in Sydney is high (as is often the case during major events, such as New Years Eve), it is common for the price of rooms to increase substantially to reflect the elevated SRMC of rationing demand for that scarce capacity; and
- when the demand for hotel rooms in Sydney is low, it is common for the price of rooms to decrease significantly, reflecting the relatively modest SRMC of supply during those periods (which would be limited largely to the cost of cleaning the room, electricity and other variable costs).

However, it is usually infeasible or undesirable for firms in workably competitive markets to perpetually adjust their prices to ensure that they reflect SRMC *at all times*. First, it will usually be too difficult or expensive for firms to compute the changing SRMC of supplying an additional unit – including the potential congestion costs. For example, it would not be practicable for McDonalds to set a unique price for every Big Mac that it sells, based on the SRMC of the ingredients at the time of sale. Indeed, the transaction costs involved in doing so (eg, reprogramming software, amending signage, etc) would outweigh the benefits of doing so.

Second, even if such a computation was practicable, there may be other, negative consequences from such a pricing approach. For example, daily fluctuations in McDonalds' prices may not be well received by its customers, and would make nationwide 'price-based' advertising campaigns very difficult. There are therefore good reasons why prices in workably competitive markets might be 'sticky' over the near term and depart materially

---

[30] Brunt, M (1970), 'Legislation in search of an objective', in J.P.Nieuwenuysen (ed.), *Australian Trade Practices: Readings*, Melbourne, Cheshire, p.238.

from SRMC at any particular moment in time. In practice, prices may be affected by the following factors:[31]

- firms often set their prices around focal points (eg, a price of $1.99 is more common than $1.79) and so short-term changes in costs or demand may not lead to a price change, since the producer will often decide to keep the product at the same price point; and
- firms may have long-term, fixed-price contracts with their customers, which allows for a stable, certain price over an extended period – they must therefore wait until those contracts expire or are renegotiated before prices can change.

Moreover, because there are always frictions that impose costs on entry and exit decisions in workably competitive markets, and because new capacity cannot be added in infinitely small units, prices that depart from SRMC will not prompt an immediate supply side response. Rather, it will take time for firms to ramp up and down their production, or to enter and exit markets. Such actions are also unlikely to be contemplated simply because prices in a market appear to be temporarily misaligned with SRMC. For example:

- a prospective new entrant to the Sydney hotel market who sees high prices leading up to the Christmas and New Year period will not respond by quickly constructing a hotel to take advantage of those high prices because:[32]
  - it is not simply not possible to construct a hotel in that timeframe, eg, to find a site, obtain planning approvals, arrange financing, undertake construction, etc; and
  - that investment decision will not be based solely on one period of high prices – rather, it is the expected returns over a much longer time horizon that are relevant (see further discussion below); and
- an existing hotel chain that experienced a temporary period of low prices due to reduced demand is unlikely to respond in the near term by reducing its number of rooms or by exiting the market, because:
  - it is unlikely to be possible to reduce capacity in that timeframe in any event, eg, it will take time to find a buyer for the assets it no longer wants and it may also have received bookings for several months in advance; and
  - exit or capacity reduction decision will not be based on the occurrence of one period of low prices – again, it is the expected returns over a much longer time horizon that are relevant consideration.

For these reasons, it is unremarkable to observe prices in workably competitive markets that are misaligned with the SRMC of supplying the product in question. Moreover, for exactly the same reasons, prices can be above or below the *LRMC* of supply. It follows that there are myriad potential price outcomes in such markets that are consistent with workable competition *at a particular point in time*. However, that is not to say that there is *no*

---

[31] *See:* Johnson & Leonard (2008), 'Frictions and Sticking Points: Applying the Textbook Model to the Analysis of Cost Pass-Through in Indirect Purchaser Class Actions' in *Antitrust Insights*, Winter 2008.

[32] Equally, existing hotels are not going to respond by adding more rooms.

relationship between the prices that are observed and the underlying costs of production *over the long term*.

Although prices may depart from SRMC and LRMC in the short term, there are limits to the extent that prices can depart from *LRMC* once a supply-side adjustment is possible. Specifically, once firms are able to respond to changes in demand- and supply-side factors by expanding or reducing their capacity, one would not expect to see prices that are significantly and persistently *above the LRMC of adding capacity*, or *below the long run avoidable costs (LRAC) associated with reducing capacity*. This is because:

- if average prices exceed the LRMC of adding capacity (eg, because prices frequently increase to reflect the increased risk of congestion, or the need for demand curtailment) then, over the long term, firms will expand and/or new entry will occur as they 'chase' the resulting profits; and
- if average prices are less than the LRAC associated with reducing capacity (eg, because prices frequently decrease to reflect an abundance of capacity) then, over the long term, firms will redeploy their capital to other markets where returns are superior.

Of course, as section 2.3 explained, it can be difficult in practice to time capacity expansions and reductions to perfection – particularly when those increments must be lumpy. There may therefore be times when prices are *above LRMC* for a period, as the market waits for new capacity to come on-stream (from existing market participants and/or new entrants). Equally, there may be times when prices are *below LRAC* for a period as the market waits for redundant capacity to be redeployed.[33] Nonetheless, provided that competition in the market is at least workable, these periods of 'misalignment' should only be temporary.

In summary, in a workably competitive market there is still a strong relationship between prices and costs. Unlike the theoretical ideal of perfect competition, prices will not always be perfectly aligned with SRMC (and LRMC). However, any change in market conditions that results in prices that are significantly and persistently *above LRMC* or *below LRAC* should, in time, prompt a supply-side response. Specifically, the rivalry that exists between suppliers should restore prices to levels that, on average, reflect the LRMC of adding capacity (or, equally, the LRAC associated with reducing capacity).

3.2. Market Power

Substantial (as opposed to temporary[34]) market power is the antithesis of workable competition. Where substantial market power exists, a business is no longer adequately constrained by its competitors, and production decisions can be made with less (or no) regard to the needs of consumers, or the potential reaction of rivals. In this circumstance, the crucial

---

[33] Note that there may be other special circumstances in which the price for a good is persistently below LRAC. For example, a firm might set the price for a good below its LRAC in order to sell an increased quantity of another complementary product for which margins are very high. Razors and razor blades or printer and ink are potential candidates. Provided that the revenue from sales of *both* products is greater than the combined LRAC, such pricing is consistent with workable competition.

[34] Firms may have temporary market power, perhaps because there are small differences in the products of competing firms or because the firm has transitory advantages that will be competed away as other firms innovate. This type of market power should not raise concerns.

resource-allocation function of competition is undermined, to the detriment of economic efficiency and social welfare.

Most definitions of substantial market power emphasise the discordant nature of market power and competition. For example, Werden defines market power as the ability of a seller to 'profitably ... maintain prices above competitive levels by restricting output below competitive levels.'[35] This interpretation is also consistent with the definition articulated in the seminal judgment *Queensland Wire Industries Pty Ltd v Broken Hill Pty Co Ltd*, in which market power was described as:[36]

> 'the ability of a firm to raise prices above supply costs without rivals taking away customers *in due time*, supply cost being the minimum cost an efficient firm would incur in producing the product …' (emphasis added)

A firm (or group of firms) can therefore be understood to possess a substantial degree of market power when it is able to set prices that would *not be observed* in a workably competitive market. Specifically, drawing upon the material set out in the previous sections, a firm can be taken to have substantial market power when:[37]

- it has the ability to *sustain* prices *above LRMC*, including an appropriate return on capital and accounting for risk;[38] and
- it is insulated from competition by *significant barriers to entry and expansion* (as opposed to, say, minor differences in product attributes).[39]

It is particularly important to distinguish substantial market power from *temporary pricing power*, which can enable firms to elevate prices above long run costs for short periods. To be substantial, market power must enable a firm to set prices above LRMC for a *sustained* period of time. Indeed, the previous sections described a number of circumstances in which prices might increase above LRMC for a short period in a workably competitive market, including:[40]

---

35 Werden (1996), 'Identifying Market Power in Electric Generation', Public Utilities Fortnightly, 15 February 1996.

36 *Queensland Wire Industries Pty Ltd v Broken Hill Pty Co Ltd* (1989) 167 CLR 177.

37 Put another way, a firm has substantial market power when it has the ability to sustain prices that *should* attract additional investment (ie, because they exceed LRMC) – from either existing market participants or new entrants – but that do not, because such investment is prevented or delayed by various factors (ie, barriers to entry).

38 Substantial market power may involve the power to do things other than raise prices, eg, to engage in 'predatory pricing' by lowering prices below SRMC or various forms of exclusionary conduct. However, these actions are all, ultimately, directed at maintaining or expanding the firm's market power so as to enable it to charge prices that exceed LRMC.

39 When considering the significance of barriers to entry (and, by extension, the degree of competitive constraint provided by new entry and expansion), it is necessary to assess the time it would take for a new firm to enter the relevant market and offer customers a competitive alternative, or for existing firms to expand, following the exercise of substantial market power. The evaluation of whether these responses would be sufficiently timely will vary with the dynamics of the market. As a general rule, entry and expansion will generally provide an effective competitive constraint if they would occur in an appropriate time to deter or defeat any non-transitory exercise of substantial market power. However, the appropriate timeframe will depend on the particular market under consideration.

40 Note that these circumstances are not mutually exclusive.

- when demand *temporarily* increases and prices rise above LRMC to reflect the reflect the increased risk of congestion, or the need to curtail demand for that scarce capacity, until such time as demand returns to 'normal' levels; and
- when *steady growth* in demand causes prices to rise systematically above LRMC to reflect the elevated SRMC of managing that scarce capacity, until such time as firms are able to expand their capacity.

Firms may also be in a position to engage in short term *opportunistic conduct* in workably competitive markets. Specifically, they may be in a position to increase prices above LRMC by withdrawing capacity and manufacturing scarcity. However, provided competition is workable, this again only amounts to temporary pricing power. Specifically, in time the conditions allow that temporary power to be exercised will dissipate (ie, demand will fall) or, if such prices endure, they will prompt entry and expansion.

It is for these reasons that Justice French concluded in *Australian Gas Light Company (AGL) v Australian Competition and Consumer Commission (ACCC)* that temporary pricing power 'does not amount to an ongoing ability to price without constraint from competition.'[41] In particular, the existence of such power does not mean that prices will exceed LRMC, on average, over the long term. Indeed, most economists would agree that, provided that barriers to entry to a market are low, then the exercise of temporary pricing power is not something that compromises the long-term interests of consumers.

It is for precisely this reason that competition regulators tend to assess the existence of market power over a sufficiently long time period to observe substitution possibilities at work. Unless transitory pricing power persists for a sufficient period of time (or occurs with sufficient frequency) for average prices to be sustained above LRMC, it does not amount to substantial market power. The assessment of substantial market power consequently requires:

- a focus on genuine and enduring barriers to entry and expansion, as the fundamental source of that power, ie, barriers that will persist over the long-run and hinder the entry and/or expansion of competitors; and
- the undertaking of long-term price cost tests, as evidence of the exercise of market power, ie, comparisons of average prices over the long run to the LRMC of expanding capacity over the same period.

If a firm has substantial market power, in the sense described above, it will be in a position to sustain prices above the LRMC of supply (including the applicable cost of capital) in the long run without prompting entry and/or expansion from rivals. Put another way, it will have the ongoing ability to set prices without constraint from workable competition.

---

[41] *Australian Gas Light Company v Australian Competition and Consumer Commission (No 3)* [2003] FCA 1525, paragraph 493.

3.3. Application of Administered Prices

In economic terms, the central objective of controlling prices through the application of regulation is to promote outcomes *consistent* with workable competition, when competition is *not* workable.[42] However, the application of price control represents an intervention into the competitive process that is a 'second best' solution to workable competition. Because price controls cannot flawlessly mimic the competitive process they can only improve what would otherwise be, and then not necessarily so.

The overarching criteria for imposing price control is that any such intervention must enhance economic welfare, relative to the counterfactual in which that intervention does not take place. This has two important implications. First, a prerequisite for such an intervention is a finding that substantial and enduring market power exists and has been exercised. In particular, price control should be limited to instances where:

- prices have been sustained above the LRMC of supply, rather than where prices have exceeded such levels only for short periods; and
- there are enduring barriers to entry and expansion that prevent any market power from being diminished over time.

If substantial market power does not exist, there is no need to *promote* workable competition since, by definition, it *already exists*. Imposing a 'second-best' solution is likely only to introduce needless additional costs. An intervention also serves no purpose if substantial market power exists but has not been acted upon. It will again simply impose unnecessary regulatory costs for no perceivable benefit, and so result in a deterioration in overall economic welfare.

Second, the application of price control must improve the prospects for the market structure – and consequently enhance economic welfare – relative to the counterfactual of not intervening. This requires the benefits of intervening to be weighed against the additional costs that inevitably would arise, ie:

- there will inevitably be administrative costs associated with designing and implementing the regulated prices; and
- there may also be other costs arising from the effect that intervention may have on the incentives of firms to invest, and the attendant consequences for dynamic efficiency.

To be warranted, the benefits of introducing regulation must not be outweighed by these additional costs. Indeed, if the costs incurred in attempting to replicate workably competitive prices would outweigh the benefits to be achieved, the intervention would not be in the long term interests of consumers. Specifically, it would not enhance economic welfare, relative to the counterfactual in which that intervention does not take place.

---

[42] Professor Alfred Kahn explains that: 'the single most widely accepted rule for the governance of the regulated industries is to regulate them in such a way as to produce the same results as would be produced by effective competition, if that were feasible. Microeconomic theory provides regulators with a set of principles that, if followed, will produce optimum results, by widely accepted criteria of optimality.' See: Kahn, A., 1988, *The Economics of Regulation, Principles and Institutions, Volume 1 – Economic Principles*, Massachusetts Institute of Technology, p.17.

3.4. Summary

In the context of trade and commerce, competition is a process whereby firms strive against each other to secure customers for their product or services. In a perfectly competitive market there is a strong relationship between prices and costs. Firms can sell whatever they like at a price equal to its SRMC (which also equals its LRMC) and earn a return that is equal to that available if they were to divert their capital to its next best use. Any change in market conditions that results in prices above or below this level will be met with an immediate response that eliminates those positive or negative margins.

However, the distinguishing characteristics of perfect competition are seldom (if ever) seen in real markets. In the more realistic setting of a workably competitive market, prices will not always be perfectly aligned with SRMC and LRMC. However, any change in market conditions that results in prices that are significantly and persistently *above LRMC* or *below LRAC* should prompt a supply-side response *over the long term*. Specifically, in time, the rivalry that exists between suppliers should restore prices to levels that, on average, reflect the LRMC of adding capacity (or, equally, the LRAC of reducing capacity).

Substantial market power is the antithesis of workable competition. A firm (or group of firms) can therefore be understood to possess a substantial degree of market power when it is able to sustain prices that would *not be observed* in a workably competitive market. Specifically, a firm can only be considered to have substantial market power when:

- it has the ability to *sustain* prices *above LRMC,* including an appropriate return on capital and accounting for risk; and
- it is insulated from competition by *significant barriers to entry and expansion* (as opposed to, say, minor differences in product attributes).

The application of price control should therefore focus on addressing the existence and exercise of substantial market power – as opposed to, say, *temporary* pricing power, which is a common feature of competitive markets – and the attendant adverse consequences for economic welfare. This requires a focus on:

- genuine and enduring barriers to entry and expansion, as the fundamental source of the substantial market power; and
- the undertaking of long-term price cost tests, as evidence of the exercise of substantial market power.

The overarching criteria for imposing price control is that any such intervention must enhance economic welfare, relative to the counterfactual in which that intervention is not undertaken. To intervene in any other circumstances would be contrary to the long term interest of consumers since it would involve the needless invocation of a 'second best' solution.

4. Application to Electricity Generation Markets

This section discusses the application of the economic principles described hitherto to 'energy only' wholesale electricity generation markets, such as the NEM. It begins by describing some of the distinguishing characteristics of such markets.

4.1. Characteristics of Electricity Generation

The electricity sector is characterised by a homogeneous, non-storable commodity-type product that has few (if any) substitutes. These attributes deprive consumers of some of the usual means for adjusting to variations in price and supply, eg, storing the product,[43] switching to close substitutes and so on. Suppliers are also characterised by significant variation between the costs of the different generation technologies available, ie:

- base load plants (such as coal, solar and wind), have relatively low operating costs, but this intrinsic, short run cost advantage is offset by relatively high capital (fixed) costs (ie, the cost per unit of potential output) and reduced ability to vary output in the short term (ie, 'stopping' and 'starting' such plants is not straightforward);
- mid-merit plants, typically in the form of combined cycle gas turbines (CCGT) have higher running costs, but mid-range capital (fixed) costs; and
- peaking plants, typically in the form of open cycle gas turbines (OCGT) have relatively low capital costs, a high degree of short-term controllability (ie, 'stopping' and 'starting' such plants is easy) but relatively high running costs.

The way that prices are set is also a distinguishing characteristic. Section 3.1.2 explained that, in most workably competitive markets, prices do not continually change – primarily because of the associated transaction costs and customers' aversion to price volatility. The NEM is an exception. Prices in the NEM are highly dynamic and are set in a way that reflects the fact that:

- demand for electricity is highly variable and must be met at (almost[44]) all times, ie, it is highly undesirable for the 'lights to go out' at any time;
- output must change very rapidly, and by large amounts within the course of a day in order to meet that variable demand; and
- a suite of technologies is required to meet that variability efficiently, ie, typically a combination of base-load, mid-merit and peaking plant (see above descriptions).

Scheduled generators in the NEM are required to submit 'offer prices' for their capacity for every 5-minutes of the day. From all offers submitted, the Australian Energy Market Operator (AEMO) determines, through a centralised process, the generators that will be called upon to produce electricity based on the principle of meeting demand in the most cost effective way, ie, generators are dispatched in 'merit order'. Prices are set as follows:

---

[43] We note that hydro electricity is sometimes considered to be a storable form of electricity.

[44] Strictly, the reliability target is unserved demand of 0.002 per cent.

- a 'dispatch price' is determined every five minutes, based on the offer lodged by the most expensive generator that must be dispatched in order to meet prevailing demand in that period – the 'marginal generator'; and
- six dispatch prices are averaged every 30-minutes to determine the 'spot price' for each trading interval for each of the five regions of the NEM, ie, the spot price is determined 48 times per day.

Because the NEM is an 'energy only' market, the only way that a generator can be paid for investing in plant is by having that capacity dispatched to produce electricity. It cannot be paid for having plant that is not being used, even if the existence of that capacity offers security of supply benefits. This sets the NEM apart from other wholesale market arrangements that *do* include payments to generators for simply offering capacity, such as the Western Australian market.

4.2. Competition in Generation

The unusual features of the electricity generation market give rise to highly variable SRMCs. The wholesale market design is directed towards promoting competition between generators that produces prices that reflect those variable SRMCs. Specifically, the expectation is that, most of the time, generation plant should be 'dispatched' according to its economic merit order, as given by the ascending SRMC of running each type of plant (as determined by the respective operating and maintenance costs – the cost of curtailing demand during times of congestion is discussed subsequently).

Although generators are permitted to offer their capacity at any price (subject to a $12,500/MWh market cap[45] and the Cumulative Price Threshold (CPT)[46]), the existence of competing offers by alternative plant owners normally constrains the prices that generators can bid. For example, a base load plant that bids substantially above its operating and maintenance costs (or withholds capacity – see discussion in section 4.3) risks not being dispatched and being forced to incur the expense of shutting down and restarting its plant. For this reason, generators can *generally* be expected to offer to supply the market at a price that reflects their short run operating and maintenance cost and are *generally* scheduled to run in line with their economic 'merit order'.

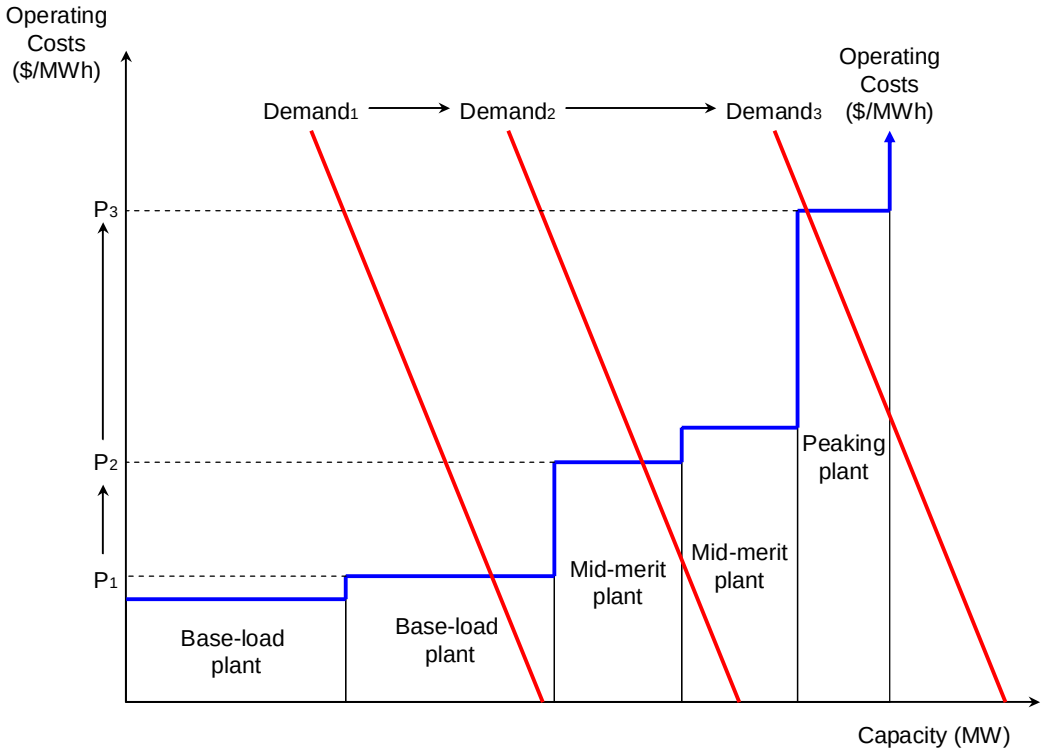
Figure 4.1 below illustrates that, although a generator may offer its capacity at a price sufficient to cover only its operating and maintenance cost, the price that it actually *receives* during a half-hour period is equal to the offer of the last generator that is dispatched in order to meet demand (the marginal generator). This means that generators with lower running costs (base load and mid-merit plant that is 'infra-marginal') will make a profit from the market prices set at the highest bid that enables them to make a contribution to their fixed investment costs. But how does the *marginal generator* cover its investment costs? The answer is no different from that in any other workably competitive market.

---

[45] Prior to 1 July 2010, the market cap was $10,000/MWh.

[46] An administered price cap (APC) of (typically) $300/MWh is imposed by AEMO whenever the sum of 336 consecutive trading interval prices (ie, 7 days) exceeds the CPT, which is currently $186,000. Once invoked, the APC remains in place until the end of the trading day during which the rolling sum of prices falls below the CPT. In other words, the CPT also has the potential to constrain the prices at which generators offer their capacity.

Figure 4.1
Economic Merit Order



Specifically, when there is a possibility that the existing generation capacity will not be able to meet demand, prices in the market must rise to reflect the increased SRMC of curtailing that excess demand. In situations where there is a risk of shortages, the costs associated with this demand side component can cause prices to rise *well above* the operating and maintenance costs of the marginal generator. It is during these periods of scarcity that those generators are able to make a contribution to their *fixed costs*. Indeed, this is the *only way* that such plants can cover their capital costs in an energy-only market.

The expected spot price is therefore based on a probabilistic assessment of possible future outcomes and the costs they entail. Specifically, it is the sum of the various additional costs arising under different scenarios, multiplied by the probabilities of these scenarios occurring. Formally, the expected spot price is derived using exactly the same formula as was set out in section 2.1 above, ie:

- the SRMC of the marginal generator when supply exceeds demand (ie, operating and maintenance costs), multiplied by the probability of that scenario occurring; *plus*
- the SRMC of the marginal generator *plus* the SRMC of curtailing excess demand when supply is less than demand multiplied by the probability of that scenario occurring.

In electricity generation markets, the cost of curtailing demand is termed the 'value of lost load' (VoLL), and reflects the amount that customers would be willing to pay to avoid a disruption to their electricity service. For large industrial users (eg, an aluminium smelter)

that amount may be very high. However, in the absence of active demand-side bidding, the regulator has to set this price. In the NEM, this price is the market price cap of $12,500/MWh. The expected spot price in the NEM can therefore be expressed as follows:[47]

Expected Spot Price = [(1 – LOLP) x SMC] x [LOLP x $12,500]

Where:

LOLP	=	Loss of load probability
SMC	=	System marginal cost, ie, the SRMC of the marginal generator
$12,500	=	Market price cap

When the probability of shortage is low, prices can be expected to resemble the operating and maintenance costs of the marginal generator (often a base-load or mid-merit plant). However, as the probability of a shortage begins to increase (which will happen once demand starts to approach the 'outer limits' of the merit curve), spot prices start to increase above SRMC and begin approaching the market price cap. In the extreme scenario in which a shortage is certain (ie, if the LOLP=1), the expected spot price *is* the market price cap and a price of $12,500/MWh should transpire for the period in question.

Periods of high prices are necessary to cover generation costs in the aggregate, to ration demand and, critically, to provide an ***inducement for new investment*** by firms chasing those high prices.[48] Indeed, when scarcity in the market causes spot prices to increase high enough, or frequently enough that the average spot price exceeds the LRMC of constructing additional capacity[49] over that timeframe then:

- firms already in the market have an incentive to expand their generation capacity so as to take advantage of those periods of high prices; and
- new firms have a stronger incentive to enter the market and offer new generation capacity, chasing those high prices.

In other words, provided that the electricity market is workably competitive, the period over which spot prices rise to reflect the increased risk of congestion, or the need to curtail demand, is finite. Specifically, once the cost of that curtailment (as represented by SRMC) has risen to a level that exceeds the costs of adding capacity (as represented by LRMC), entry and expansion can be expected to occur over the longer-term to *meet* that demand.

---

[47] Hunt & Shuttleworth (1996), *Competition and Choice in Electricity*, Wiley, p.173.

[48] In practice, these signals will be complemented by administrative planning functions. For example, the AEMO *Statement of Opportunities* also sets out information about future generation capacity requirements, based on load forecasts, in order to ensure that generation reliability standards are met.

[49] Recall that the LRMC of adding capacity (and the LRAC associated with reducing capacity) is determined by the operating and capital costs associated with the optimal investment profile need to meet the relevant increment (or decrement, as the case may be) in demand. This may comprise investment by both existing market participants and by new entrants, and, potentially, invpestment in different production technologies. For example, depending upon the circumstances, the most efficient expansion profile may involve investment by both existing generators and new entrants, and a mix of generation technologies, eg, base-load, mid-merit and peaking plant and, potentially, transmission and interconnector capacity.

In exactly the same way, there is a limit to the extent to which persistently *low* (or negative[50]) spot prices can persist without precipitating a reduction in capacity. Indeed, if surplus capacity causes market causes spot prices to fall below the LRAC associated with reducing capacity then, over the long term:

- generators in the market will have an incentive to reduce their capacity, eg, base load plants may decommission units; and/or
- generators will have an incentive to exit the market and redeploy their capital to areas in which returns are more attractive, eg, entire power stations may shut down.

In this respect, a workably competitive wholesale electricity spot market functions no differently from most other workably competitive markets. Specifically, any change in market conditions that results in prices that are significantly and persistently *above LRMC* or *below LRAC* should, in time, prompt a supply-side response that restores prices to these levels. This relationship between prices and costs is the same as that described in general terms in section 2.3.

Of course, one complication discussed in section 2.3 is that this supply-side adjustment process cannot necessarily be expected to be *perfect*. Because generation capacity cannot be added or removed in 1MW increments, it can be difficult to time 'lumpy' capacity expansions and reductions to coincide with the theoretical 'trigger points' described above. There may therefore be times when:[51]

- average spot prices (and SRMC) are *above LRMC* for periods, as the market waits for the next increment of capacity to come on-stream; and
- average spot prices (and SRMC) is *below LRAC* for periods, as the market waits for redundant capacity to be redeployed.

In other words, prices that diverge from LRMC (or LRAC) for significant periods of time may *still be explicable* in an electricity generation market. However, as sections 2.3 and 3.1.2 explained, provided that competition in the market is at least workable and the concept of LRMC is properly understood, these periods of 'misalignment' should still only be temporary. We explain the importance of adopting a longer-term perspective in section 5.3.

The more challenging complexity is that the supply and demand conditions that lead to high spot prices in a well functioning workably competitive spot market are also the conditions in

---

[50] It is not uncommon to see some plants in the NEM lodging *negative* bids (to a market 'floor' of -$1,000/MWh) to ensure that they are dispatched, eg, wind generators may lodge such bids 'when the wind is blowing'.

[51] The renewable energy target and a carbon tax (if implemented) also have the potential to influence the LRMC of new generation capacity and the period of time during which SRMC and LRMC are misaligned. The principal effect of the initiatives will be to make renewable (ie, low carbon) forms of generation less expensive and carbon-intensive generation (eg, coal-fired plant) more expensive. The schemes may therefore change the *cost* and, possibly, the *configuration* of the next capacity expansion in a location. Specifically, it may be the case that, absent the initiatives, the optimal investment profile to meet the next capacity expansion would have comprised largely new coal-fired base-load plant. The effect of the schemes will be either to increase the cost of any such plant, or to change the economics of the investment to such an extent that a greater proportion of renewable energy is brought on-stream instead. In each case, the LRMC of the capacity expansion increases, relative to the state of the world in which the government intervention does not take place. In addition, because the LRMC has increased, it may take longer for the SRMC of curtailing demand to reach that new, higher, threshold.

which market participants can have the strongest incentive to engineer price spikes through creating *artificial shortages*.[52] These incentives and the manner in which they can be acted upon are the subject of the following section.

4.3. Strategic Withholding

It is worth emphasising at the outset that the strategies that are discussed in this section assume that the existing portfolio of generation assets is *fixed*. As soon as one takes a longer term perspective and allows for the possibility of entry and expansion, the analysis may change significantly. With that important qualification, there are various portfolios of generation assets that can confer the ability profitably to affect the spot price in certain circumstances, even if a firm has only a modest market share.[53]

This can be achieved by either 'physical' or 'economic' withholding of generation capacity[54] that *would otherwise be dispatched* in order to create *artificial* scarcity in the market (rather than true 'competitive scarcity'[55]) that must then be curtailed through high prices. The former involves a generator not offering all of its capacity and the latter involves it offering some of its capacity at a price that exceeds the operating and maintenance costs of the likely marginal generator. The objective and consequences of the two strategies are the same and so the distinction is not important for the purposes of this paper.

There are a number of different withholding strategies that can be employed by generators to engineer a shortage of supply. For example, withholding can involve a low-cost producer (ie, a base-load or, possibly, a mid-merit plant) withholding part of its capacity so as to increase the price at which the remainder is dispatched. Figure 4.2 illustrates the implementation of such a strategy and the potential effect upon the spot market price.

---

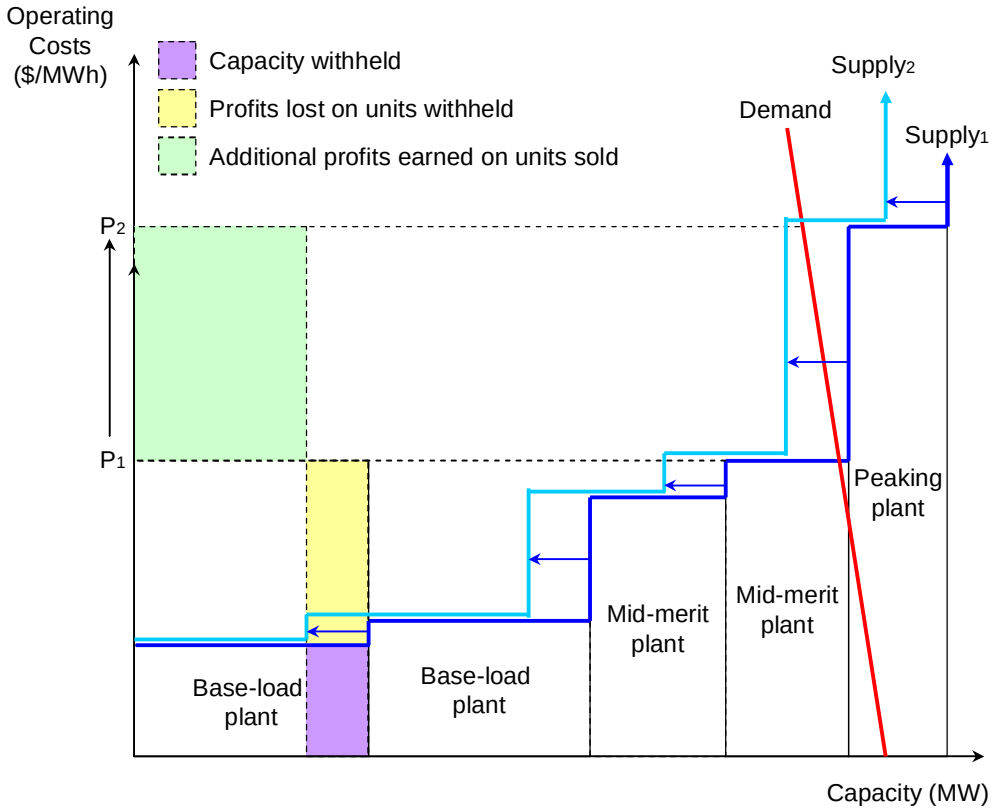
52 *See:* Joskow, P (2007), 'Competitive Electricity Markets and Investment in New Generating Capacity', *The New Energy Paradigm* (ed: Dieter Helm), Oxford University Press.

53 It is for this reason that market shares are often not particularly revealing when assessing the market power of electricity generators. *See for example:* Public Utilities Commission v FERCE 462 F.3d 1027 2006 U.S App at 1039.

54 There are also other ways of affecting prices, such as attempting to engineer transmission constraints, but such strategies are not the principal focus of the MEU Rule change proposal and so are not discussed further in this paper.

55 *See:* Joskow, P (2007), 'Competitive Electricity Markets and Investment in New Generating Capacity', *The New Energy Paradigm* (ed: Dieter Helm), Oxford University Press.

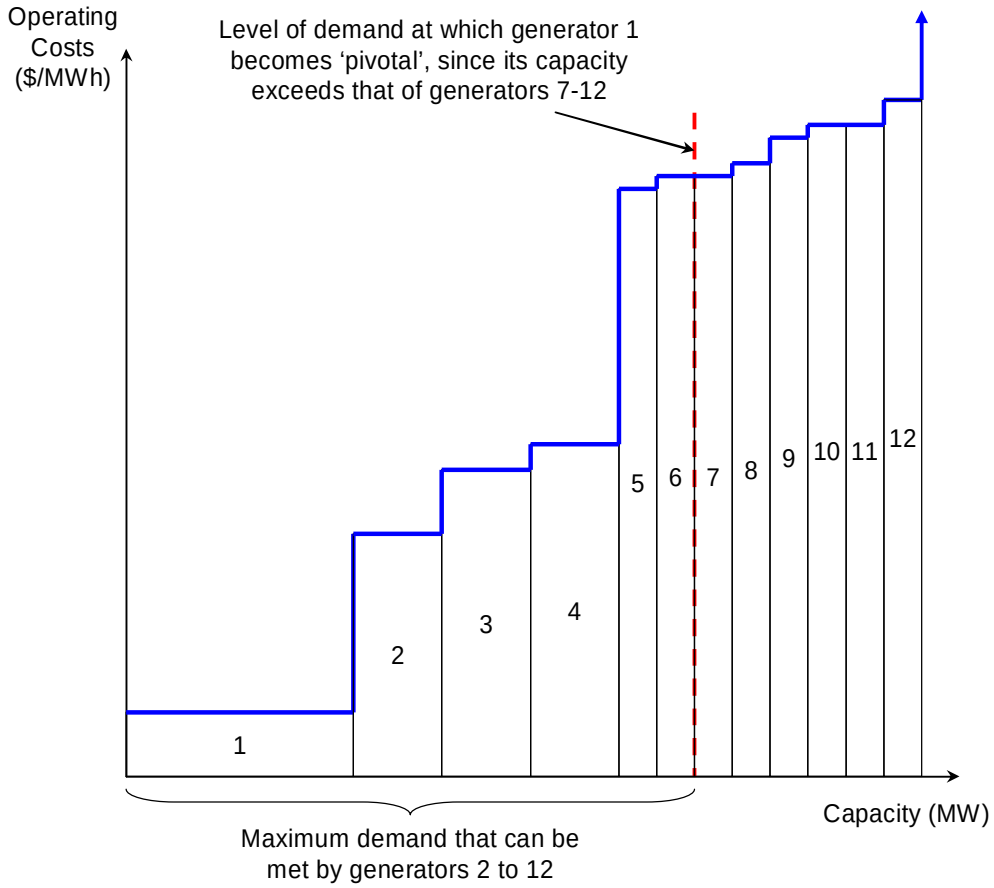
Figure 4.2
Withholding by a Low-cost Producer



This strategy can be implemented most effectively when projected demand increases to a point at which the generator knows that there is a high probability that load will not be able to be served without it being dispatched, at which time it effectively becomes the marginal (or 'pivotal') supplier.[56] When this occurs, the generator can conceivably bid some (even all) of its capacity at $12,500/MWh and be reasonably confident of receiving that price for the trading interval. Generator 1 becomes 'pivotal' in Figure 4.3 when projected demand increases to the point at which that load cannot be served by generators 2 to 12 above.

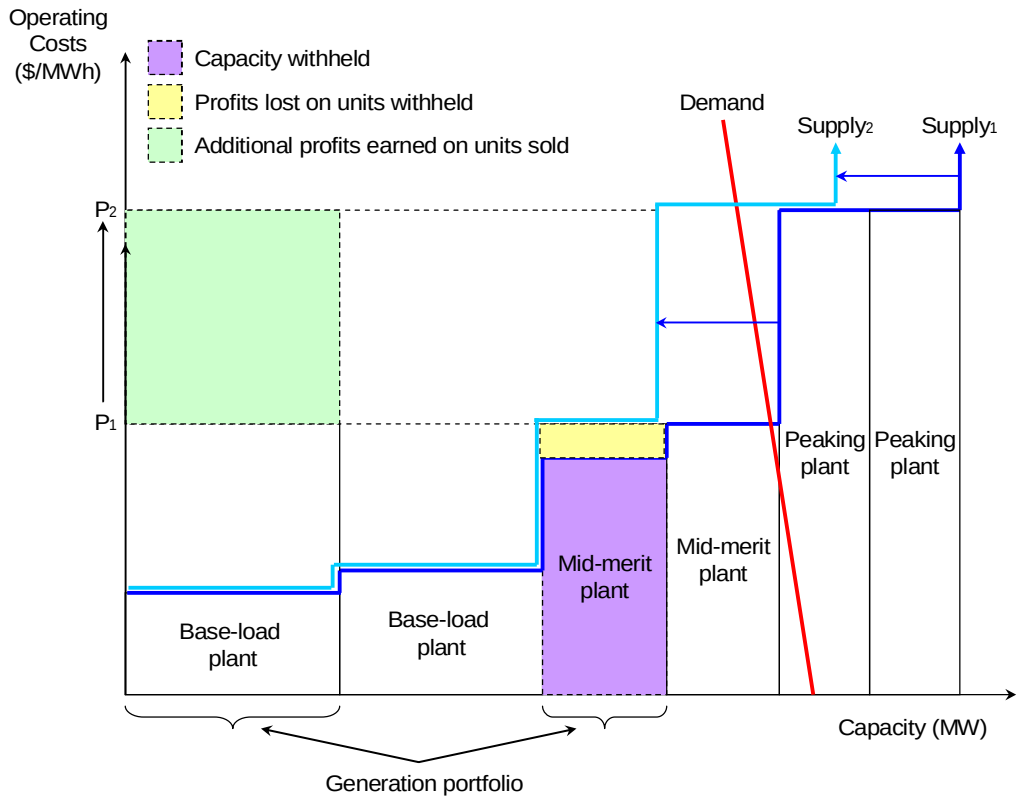
[56] This is effectively how MEU defines 'dominant generators' in its proposed Rule change

Figure 4.3
'Pivotal' Generation Unit



Strategic withholding can involve the coordinated use of multiple generation units to engineer a shortage. For example, a generator that owns both base load plant and mid-merit or peaking plant might withhold the latter in order to produce a shortage and benefit the former. Specifically, it might withhold the capacity of a mid-merit plant in order to benefit from the higher price it may consequently receive for its base load capacity. Figure 4.4 illustrates the implementation of such a strategy and the potential effect upon the spot market price.

Figure 4.4
Withholding by a Single-Owner Portfolio



The successful implementation of any withholding strategy depends on the concurrence of a number of factors, including:

- most critically, whether the slope of the 'merit curve' or 'supply curve' is 'steep' or 'flat' around the market clearing price, since this ultimately determines the magnitude of any price increase – the 'shape' of the merit curve in electricity markets can therefore be particularly conducive to such conduct at high levels of demand; as well as
- the production costs of the low-cost suppliers that potentially could restrict output to increase profits, since this affects the profits on those units withheld and those sold – these will be lowest for base-load and mid-merit plants;
- the elasticity of demand forecast demand, since any contraction in demand in response to a price increase mutes the effect of such conduct – because the demand for electricity is highly inelastic, there will often be little such response;
- the extent to which a reduction in supply by a low-cost supplier might be offset by increased supply by other low-cost so as to reduce any price effect – of course, this is not a possibility when a generator is 'pivotal';
- the total output supplied to the market by the withholding generator, ie, the greater the total output sold by the withholding generator, the greater the additional profit on units sold at the higher price; and

- the hedging position of the withholding generator, ie, if a significant proportion of its sales during the period in question are at a pre-determined contract price, this may reduce the impact of any increase in the spot price on its profitability (the relevance of hedge contracts to the assessment of market power is explored in section 4.4.2).

If such strategies are successfully executed, short-term inefficiencies result, ie, there are productive inefficiencies because lower-cost plant is not dispatched and there may be allocative inefficiencies associated with unserved demand. However, whether such conduct can reasonably be classified as the exercise of substantial market power that has long term dynamic efficiency consequences is another matter. It may indicate the exercise of such power, but it may not, as we explain below.

4.4. Substantial Market Power in Generation

Section 3.2 explained that a firm can be understood to possess a substantial degree of market power when it is able to set prices that would *not be observed* in a workably competitive market. Because workably competitive wholesale electricity spot markets function no differently from most other workably competitive markets, the test that should be applied to identify substantial market powers is also the same.

4.4.1. Substantial market power in the spot market

Substantial market power is the ongoing ability of a firm to raise prices above competitive levels without rivals taking away customers in due time.[57] A generator (or group of generators) can therefore be considered to have a substantial degree of market power over electricity spot market prices (hedge prices are discussed in section 4.4.2) when it is not constrained by the forces of workable competition, ie, when:[58]

- it has the ability to increase average spot prices to such an extent and with sufficient frequency that they exceed the LRMC of adding capacity, including a return on capital and accounting for risk;[59] and
- it is insulated from the forces competition by significant barriers to entry and expansion (as opposed to, say, minor differences in product attributes) that enable it to sustain average prices at that level.[60]

Like in any other market, it is important to distinguish substantial market power from *temporary pricing power*. Indeed, the previous sections described a number of unexceptional circumstances in which prices in a workably competitive market might increase above the

---

[57] *Queensland Wire Industries Pty Ltd v Broken Hill Pty Co Ltd* (1989) 167 CLR 177.

[58] As noted earlier, there are a number of additional indicators of substantial market power that are not discussed in this initial report. In particular, the 'Lerner Index' and the 'Pivotal Supplier Index' are two additional measures that are commonly employed to assist in the detection of substantial market power in wholesale electricity markets. However, a detailed discussion of these methodologies is outside the scope of this initial report.

[59] Note again that this may involve engaging in strategies such as predatory pricing, albeit for the same purpose.

[60] As noted above, as general rule, entry and expansion will provide an effective competitive constraint if barriers to entry and expansion can be overcome in an appropriate time to deter or defeat any non-transitory exercise of substantial market power.

LRMC of new capacity for short periods, without jeopardising long term dynamic efficiency. In a similar way, average spot prices in the NEM may exceed for short periods the LRMC of adding new generation capacity, ie:[61]

- when demand for electricity *temporarily* increases (eg, on a very hot day) and the spot price in that trading interval rises above LRMC to reflect the elevated SRMC of curtailing demand for that scarce capacity; and
- when *steady growth* in demand results in spot prices rising to curtail demand with sufficient frequency that the average spot prices exceed the LRMC of adding capacity, but will fall once firms expand their capacity to meet that demand, ie, the timing of new capacity will not always be 'perfect'.

Generators in the NEM may also be able to engage in strategic withholding of the form described in section 4.3 to increase prices above LRMC by manufacturing scarcity. However, *provided that competition is at least workable*, this again only amounts to temporary pricing power that should not be a cause for concern. This is because, in time:

- the conditions that allow that temporary pricing power to be exercised will (by definition) dissipate, eg, demand will fall and spot prices will decline; or
- if that temporary pricing power is exercised with sufficient frequency that average spot prices exceed the LRMC of adding capacity, this will prompt a response from rivals, ie, the barriers to entry and expansion that prevented an immediate supply-side response will be overcome in the longer term.

It is important to be cognisant of *all of these possibilities* when diagnosing substantial market power, since they do not call for the application of price control. As section 3.3 explained, there is no need for administered prices to be a surrogate for workable competition in these circumstances because *it already exists*. For these reasons, in order confidently to conclude that a generator (or group of generators) possessed a substantial market power that may justify the application of price control, one would need to be satisfied of a number of things.

First, those generators that control a portfolio of assets that may give rise to incentives to influence the market price by engaging in strategic bidding conduct would need to be identified. That analysis would necessarily consider the positions that those assets occupy on the 'merit curve' and the frequency with which the generator could conceivably influence the market price by engineering scarcity. It will be particularly important to consider whether:

- there are any generators that control a significant proportion of infra-marginal (eg, base-load) capacity that would be well placed to benefit from any sharp increases in spot prices if some of that capacity was withheld (see Figure 4.2 above); and
- there are any generators that control both base-load and mid-merit/peaking plant that would be in a position to withhold the latter so as to increase the profitability of the former if the spot price increases sharply.

---

[61] Note that these circumstances are not mutually exclusive.

Second, one would need to show that any generators so identified had actually acted upon that incentive. The primary complication here is that the supply and demand conditions which lead to high spot prices in a workably competitive spot market are also the conditions when the exercise of substantial market power is most likely.[62] It follows that one would need to be satisfied that any price spikes that are identified were caused by artificially generated shortages rather than from genuine 'competitive scarcity' or from legitimate occurrences such as unexpected outages.

Third, one would need to establish that the instances in which that influence had (or was likely to have) been exercised had (or was likely to have had) a material and lasting effect on average spot prices (or on hedge contract prices, as section 4.4.2 explains) that will not be undone by subsequent reductions in demand or through capacity expansion. Specifically, it would need to be shown that average prices had been increased above the LRMC of adding capacity, and that those prices are likely to persist in the future. This requires consideration of:[63]

- whether the circumstances that transpired to facilitate the high average prices (eg, extreme temperatures, interconnector constraints, etc) can be expected to occur in the future with sufficient frequency to warrant an intervention – if they will not, then there is no need to incur the costs of applying price control; and
- whether there are any barriers to entry or expansion that would prevent potential competitors from exercising a constraint on the pricing conduct of a generator over the longer term – if barriers to entry are low, then the competitive response of rivals should reduce average prices over time (provided that there are no strategic barriers).[64]

It is only if these conditions are met that one can be satisfied that a generator (or group of generators) possesses ongoing ability to influence spot prices without constraint from competition[65] that *may* warrant the imposition of price control. If any of the conditions are *not* met, then prices that are above the LRMC of supply are more likely to be indicative of temporary pricing power that does not call for intervention. The final step is then to consider whether administrative pricing would enhance economic welfare relative to the counterfactual of either not intervening, or intervening in a different way.

This section has described the incentive and ability that a generator (or group of generators) may have to exercise substantial market power so as to influence average *spot prices*. However, the spot price is not the only price that market participants pay to procure electricity, or that generators receive to supply it. Generators and retailers may also enter into

---

62 *See:* Joskow, P (2007), 'Competitive Electricity Markets and Investment in New Generating Capacity', *The New Energy Paradigm* (ed: Dieter Helm), Oxford University Press.

63 As noted earlier, although these matters are of critical importance to the identification of substantial market power, it is beyond the scope of this initial report to provide a comprehensive assessment of the applicable barriers to entry and other facilitating factors.

64 Note that these barriers may also be 'strategic' in nature, eg, if potential entrants perceive that high prices are the product of artificial scarcity, then they may be disinclined to enter the market if doing so would stop those high prices from occurring.

65 *Australian Gas Light Company v Australian Competition and Consumer Commission (No 3)* [2003] FCA 1525, paragraph 493.

various forms of hedge contracts that involve the exchange of funds by reference to the spot price during specified periods. The following section considers whether a diagnosis of substantial market power requires contract market positions to also be taken into account.

4.4.2. Relevance of hedge contracts

The existence of hedge contracts does not itself add to or detract from the problem of substantial market power. Specifically, hedge contracts do not *create or extend* substantial market power, and they do not enable counterparty customers to *avoid the consequences* of substantial market power. This is because the price of hedge contracts is determined primarily by the balance of expectations as to the level and volatility of future wholesale spot price outcomes.[66]

It follows that if a generator is able to exercise substantial market power over the spot price, then the price of hedges can be expected to adjust to reflect the higher levels of expected future spot prices. Put simply, a customer cannot avoid the consequences of intermittently high spot prices by entering into a long-term hedge contract with a generator that is *causing those high prices to occur*. That generator will simply demand a contract price that reflects its ability to influence the spot price if it so chooses, ie:

- the customer can either choose to remain unhedged, and be forced occasionally to pay the very high spot prices that result from the exercise of substantial market power; or
- the customer can enter a hedge contract, and pay a price in which those high prices it would otherwise be forced to pay are 'averaged' over the life of the contract.

In either scenario, the *expected* price that the customer must pay for electricity over the period is the same, regardless of whether a side contract is struck. However, the potential complication is that, if a significant proportion of a generator's sales are to be at a pre-determined contract price (recognising that a generator with substantial market power is unlikely to ever be *fully* hedged[67]), this *may* reduce the incentive that it subsequently has to exercise its market power by influencing the spot price – at least for a period.

This is because the profits that the generator can earn from any increases in the spot price that it subsequently engineers are lessened, as section 4.3 explained. Put simply, it may not have to exercise its substantial market power to manipulate the spot price because it has *already exercised* that power in striking the contract price. It follows that there may *conceivably* be periods during which substantial market power is being exercised, but which is not manifested in the *average spot price*. In other words, it may be *useful* also to gather hedge

---

66 If this were not the case – and the price of hedges was out of line with expectations of future market prices – then profitable arbitrage opportunities would arise to close the gap.

67 If a generator with substantial market power hedges all of its capacity, it exposes itself to substantial spot market risk in the event that it cannot deliver that capacity. For example, if one of its generating units experiences an unplanned outage, this may result in a material increase in the spot price during that period – particularly if it increases significantly the probability of the market price cap being invoked. If it is has entered into contracts for the exchange of funds by reference to its unavailable capacity at, say, $60/MWh, and the spot price during that trading interval is $1,000/MWh, then it must effectively procure that capacity at the prevailing spot price ($1,000/MWh) and sell it at the contract price ($60/MWh). It is the potential for such losses that induces generators to be cautious about entering into hedge contracts representing all of their potential capacity.

market information to identify substantial market power. However, in our opinion, garnering such information may not be strictly *necessary*.

This is because periods during which average spot prices and contract prices are misaligned cannot persist indefinitely. Indeed, if a significant period elapses without the generator demonstrating its ability to engineer high spot prices, it is highly likely that customers' perceptions of future spot prices will begin to change. In particular, they may begin to *lower* their expectations of future spot prices and, as a consequence, reduce the amount that they are prepared to pay for hedge contracts. The generator will consequently need to start affecting the spot price, so as to 'remind' customers of its substantial market power, and to re-calibrate future spot price expectations. For this reason, we are not convinced that a diagnosis of substantial market power will necessarily require contract market positions to also be taken into account.

To summarise, *in principle* there may be periods during which the exercise of market power will not be reflected in spot prices. However, that will not necessarily be the case over the long term because, unless the potential effects of a generator's substantial market power are signalled to customers through the spot market, the price that those counterparties will be prepared to pay for hedge contracts can be expected to decline. In our opinion, it may not therefore be necessary to modify the framework for identifying substantial market power by including the additional step of considering hedge contract prices.

4.5. Summary

Energy-only electricity generation markets have some characteristics that distinguish them from many other markets. However, despite those differences, a workably competitive wholesale electricity spot market functions no differently from most other workably competitive markets. Specifically, with certain limited exceptions, if prices are significantly and persistently *above LRMC* or *below LRAC* this should, given time, prompt a supply-side response that restores prices to these levels.

It follows that a generator (or group of generators) can be considered to possess a substantial degree of market power when it is not constrained by the forces of workable competition, ie, when:

- it has the ability to increase average spot prices to such an extent and with sufficient frequency that they exceed the LRMC of adding capacity, including a return on capital and accounting for risk;[68] and
- it is insulated from the forces competition by significant barriers to entry and expansion (as opposed to, say, minor differences in product attributes) that enable it to sustain average prices at that level.

Any assessment of whether a generator has a substantial degree of market power consequently requires:

[68] Note again that this may involve engaging in strategies such as predatory pricing, albeit for the same purpose.

- a focus on genuine and enduring barriers to entry and expansion, as the fundamental source the substantial market power, noting that this might also include 'strategic' barriers to entry and expansion; and
- the undertaking of long-term price cost tests, as evidence of the exercise of market power, eg, comparisons of average spot prices to the LRMC of adding capacity, rather than comparisons of spot prices to SRMC at particular points in time.

Although periods of high prices that appear to have arisen from strategic bidding conduct are certainly relevant to the assessment indicated above, this only applies to the extent that they have had a sustained effect on average spot prices that is likely to persist over the long term. It is also unlikely to be necessary to consider the price of hedge contracts, since a generator's market power must ultimately be signalled to customers through the spot market.

Having now established 'what it is that we are looking for', it is possible to define the parameters of that search with more confidence. Put another way, it is feasible to define an appropriate market that can be adopted for the purposes of making that assessment.

5. Market Definition

This section considers the appropriate market definition for the purposes of considering the proposed Rule change. It begins by discussing the purpose of defining a market before stepping through the approach that is typically adopted to delineating market boundaries for the purposes of assessing market power.

5.1. Purpose of Market Definition

Defining a market involves delineating between those parties that are likely to have a substantial effect on the business whose conduct is at issue and those parties that have a less-immediate effect. It frames the 'relevant arena' of competition and enables the *real* question of interest to be answered – in this particular case, whether particular generators possess substantial market power that may be worth addressing in some way.

A necessary part of this process involves delineating between potential substitutes – on both the demand-side and the supply-side – that are consequently ruled 'in' or 'out' of the market even though, in reality, no such 'bright line' may exist. It follows that it is generally advisable to 'err on the side of caution' by defining the market sufficiently broadly to ensure that potentially relevant constraints are not eliminated, ie, a market that is 'too narrow' can often lead to more problems than one that is 'too broad.'[69]

Nevertheless, even a narrowly specified market is not a cause for concern provided that all relevant competitive constraints (or lack thereof) are properly identified, once that market has been defined. For example, if a producer must incur significant sunk costs to expand its capacity before it can impose a competitive constraint on a firm attempting to exercise market power, it may not constitute a close supply-side substitute for the purposes of defining the market (see below). Rather, this response is more likely to constitute entry.

However, if such entry would be likely to impose a constraint on the price and output decisions of the firm whose market power is in question within the relevant timeframe (see below), it is still relevant to that subsequent – and more critical – analysis of competitive effects. Fisher (1991) expressed this succinctly when he observed that: 'At base, what matters more than defining the market perfectly is identifying the economic forces that constrain a firm's pricing.'[70]

Here, the objective is to assess whether individual generators in the NEM have the ability to exercise substantial market power so as to increase average spot prices to such an extent and with sufficient frequency that they exceed the LRMC of adding capacity. The market definition must therefore be capable of identifying circumstances that may necessitate *ex ante* regulation to address structural concerns or enduring market failures for which *ex post* intervention is impracticable or inappropriate.

---

[69] *See for example:* Brunt, M (1990), '"Market Definition" Issues in Australian and New Zealand Trade Practices Litigation', *Australian Business Law Review,* Vol 18, pp.86-128 (hereafter: 'Brunt (1990)').

[70] Fisher, F (1991), 'Diagnosing Monopoly', *Industrial Organization, Economics and the Law* (John Monz, ed).

5.2. Framework and Approach

The concept of substitutability is central to the process of defining the bounds of antitrust markets. In general terms, a market is the field of actual and potential transactions between buyers and sellers amongst whom there can be strong substitution. Specifically, the market should comprehend the range of business activities and the geographic area within which, if given a sufficient economic incentive:

- buyers can switch to a substantial extent from one source of supply to another ('demand-side' substitution); and
- sellers can switch to a substantial extent from one production plan to another ('supply-side' substitution).

It is particularly important to distinguish the second of these two effects – supply-side substitution – from *new entry* into a market. Competition regulators do generally[71] take into account the former when defining a market, but not the latter. Specifically, in order for potential supply-side constraints to be taken account in *defining a market*, any reconfiguration of production and supply should be able to occur:[72]

- within a relatively *short period of time*; and
- without incurring *significant sunk costs.*

The logic of these criteria is that, if a seller must incur significant sunk costs and/or take a long time to expand its production facilities before it can respond to a price increase, this response is more likely to represent *entry* than supply-side substitution. However, if a seller must incur significant sunk costs, but would impose a sufficiently timely constraint on price and output decisions, such entry would still be relevant to the subsequent analysis of competitive effects or market power.

Against that background, the process of defining the boundaries of a market can be interpreted as establishing the smallest area of product, functional and geographic space within which a hypothetical profit maximising monopolist could successfully impose a small but significant and non-transitory increase in price (a 'SSNIP'). The establishment of market boundaries should start by considering the product, geographic and functional areas of supply by the firm whose conduct is in question.[73] One then asks whether a hypothetical monopolist

---

[71] With some minor differences, this is the approach taken by the ACCC, the New Zealand Commerce Commission, the European Commission and the UK Office of Fair Trading. The exceptions are the US Department of Justice (DOJ) and the Federal Trade Commission (FTC). The DOJ and the FTC do not take supply-side substitution into account for the purpose of defining the relevant market in the context of mergers. The US *Horizontal Merger Guidelines* are clear that only demand--side will be considered when defining markets. Consideration of supply-side constraints – including the existence of 'rapid entrants' – is left until the subsequent assessment of competitive effects. *See:* US Department of Justice and the Federal Trade Commission, *Horizontal Merger Guidelines*, August 19, 2010, p.7 In our opinion, as a matter of principle supply-side substitution is a relevant consideration when defining antitrust markets. In this respect, we favour the approach taken by regulators in Australia, New Zealand, the UK and Europe, rather than the more constrained methodology articulated in the US Guidelines.

[72] *See for example:* Parr, N., Finbox, R. & Hughes, M. 2005, *UK Merger Control: Law and Practice*, 2nd edn, Sweet & Maxwell Ltd, p299.

[73] *See:* Brunt (1990), p.105.

could profitably impose a SSNIP on those products, usually of between 5 and 10 per cent above the price level that would apply under conditions of workable competition, and assuming that the prices of all other products remain constant.

Of course, in wholesale electricity markets, it is common to experience wholesale spot price movements that are many times higher than 5 to 10 per cent and which, in some cases, may persist for a relatively short period of time. Indeed, such fluctuations are a necessary feature of a well functioning, workably competitive spot market. For this reason, the more relevant question is whether a hypothetical monopolist could increase *average* spot prices by between 5 and 10 per cent above the workably competitive benchmark (as represented by the LRMC of adding capacity, and accounting for the cellophane fallacy), over the relevant time period (which is discussed in section 5.3).

A SSNIP is only feasible when all current and potential sources of close substitutes for the firm's products have been included in the defined market to which it is applied. If, following a SSNIP, consumers would switch their demand to other products, and/or alternative suppliers – potentially in other geographic locations – would alter their production processes and serve significant volumes of the monopolist's customers, the exercise would not prove profitable. The relevant market would consequently need to be expanded to include those alternative products and additional sources of supply, since they constitute close demand- and supply-side substitutes.

Market definition can therefore be approached systematically by starting with the narrowest possible set for each of the product and geographic market dimensions and then progressively widening those dimensions to incorporate additional products and geographical areas, until the boundaries of the market ultimately are established. The relevant market will be the narrowest set of products and geographical areas that enables a hypothetical monopolist controlling that group of products and geographic area profitably to sustain a small but significant and lasting price increase.

Although the SSNIP approach is generally accepted as appropriate for delineating relevant antitrust markets, it is important to recognise its limitations. Most notably, the inquiry involves testing whether a SSNIP can be imposed so as to increase prices by a small amount *above the competitive level*. The test can therefore be difficult to apply empirically if the prevailing market price *does not* (or is unlikely to) represent a price that would be observed in a workably competitive market. In particular, problems arise when a SSNIP test is applied empirically to prices being charged by a firm that has a substantial degree of power in a market, and so may already be exercising that power to inflate its price. It is important in these circumstances to avoid what is known as the 'cellophane fallacy'.[74]

[74] The cellophane fallacy is named after the case *United States vs El du Pont de Nemour and Co* 351 US 377 (1956). In that matter, Du Pont was the sole seller of cellophane wrapping paper, but claimed that its prices were constrained by other wrapping products, since an increase in its prices would induce a sufficient number of customers to switch to other flexible packaging material to make the price rise unprofitable. However, those alternative wrapping products were found not to be close substitutes, since the proximity of those substitutes was brought about through the exercise of Du Pont's monopoly power.

Specifically, a firm with substantial market power may already have increased its prices to the point where any further increase would be unprofitable.[75] A SSNIP applied to such prices may therefore imply that a significant number of customers would switch to alternative products or that many firms would alter their production plans, and require the market definition to be expanded. However, this may exaggerate the breadth of the market, since it may only be because the firm has used its market power to inflate its price so that those other products appear to compete with it.

Indeed, it may be that if the firm were instead to charge a 'competitive price', customers and rival sellers may not be inclined to switch in sufficient numbers following a SSNIP to make a price rise unprofitable. For this reason, any *empirical* application of the SSNIP test requires sufficient quantitative data to permit the calculation or assessment of, in particular, the *competitive price* for the product in question. This is a relevant consideration in the particular circumstances highlighted by the MEU Rule change proposal – particularly in relation to the geographic scope of the market, as section 5.6 explains.

A further limitation of the SSNIP test is that its reliable application requires sufficient data to permit the calculation or assessment, in particular, of the competitive price for the product in question.[76] It is therefore important to avoid mechanically applying the framework without also considering matters of commercial common sense. In other words, although the SSNIP framework is a very useful tool for defining antitrust markets, it should not necessarily be applied to the exclusion of all other considerations.

With those important qualifications in mind, the best approach is then to 'break down' the analysis into more manageable pieces by considering each dimension separately. The conventional dimensions of a market are: the goods and services supplied (the 'product dimension'); the area over which trading takes place (the 'geographic dimension'); and the number of levels in the production chain at which the market operates (the 'functional level'). The following sections examine each of these dimensions. However, we begin by considering the *timeframe* over which substitution possibilities should be assessed.

5.3. Relevant Timeframe

Reference is sometimes made to the 'time' or 'temporal' dimension of a market. However, it is more useful to think of a 'relevant timeframe' as a critical factor in assessing the other dimensions of the market – most notably the product and geographic dimensions – rather than as a dimension in its own right.[77] Typically, the relevant timeframe is determined by reference to the period over which substitution can take place. Adopting a longer timeframe will therefore tend to result in 'wider' market definitions, on average, because market participants have more time to react to a SSNIP, ie:

---

[75] That is because the firm(s) may have already sufficiently increased prices to the point where any further increases are unprofitable. Profit maximising firms can be expected to establish prices at the point at which demand for their product is elastic, ie, the firm's "own price elasticity" – a measure of the extent to which customers would be willing to switch away from consuming the firm's own product in response to a price rise – is sufficiently high that it is not profitable to pursue any further price increase.

[76] See the observations of Justice Sackville in *Seven Network v News Limited* [2007] FCA 1062.

[77] Leuner, T (2008), 'Time and the dimensions of substitutability', *Australian Business Law Review*, p.328.

- on the demand-side, customers typically take time to realise that prices have changed and take time to change their purchases;[78] and
- on the supply-side, firms generally cannot react immediately to a price increase but, over time, they may be able to adapt capital or expand, in order to increase production.

However, the unusual characteristics of electricity generation mean that focusing on demand- and supply-side substitution possibilities does not provide much (if any) insight into the appropriate timeframe for defining the market. This is because those characteristics limit the opportunities for demand- and supply-side substitution in both the short- ***and*** the longer-term. Indeed, it is predominantly the potential for entry and expansion that would undermine any price increases imposed by a firm seeking to exercise market power and not the prospect of widespread substitution.

On the demand-side, most small consumers are not exposed to spot prices, and so have no incentive to respond to spot price increases. Larger customers who do face spot price risk may have a number of potential demand management options, including interruptible[79] or controlled[80] loads, embedded generation[81] and dual fuel.[82] Unfortunately, the likely extent of these responses and the timeframe over which they can be implemented is very difficult to measure. Indeed, the dearth of information on the scope for efficient, cost-effective demand-side management initiatives is one of the reasons why the AEMC has identified 'building the capability and capturing the value of flexible demand' as one of its three key strategic priorities for the NEM.[83]

Moreover, it is unclear whether there is ***any*** scope for supply-side substitution. As noted earlier, to be considered supply-side substitution, any reconfiguration of production and supply must be able to occur within a relatively short period and without incurring significant

---

78 *See:* Areeda P, Hovenkamp H & Solow J (2002), *Antitrust Law: An Analysis of Antitrust Principle and their Application* (2nd ed, Aspen), pp.183-184.

79 This includes loads that do not need to operate continuously and can therefore be turned off (generally subject to limitations regarding the length of time they are asked to be off, how often they are likely to be asked to be off, and the number of consecutive days they may be likely to be asked to be switched off). As a result, these tend to be discretionary loads and are often batch loads, and tend to be found in larger commercial and industrial facilities. The switching may be manual or automated, and is generally at the customer end, but in some cases the customer may allow the retailer to have control of the switch.

80 This includes loads within the customer's facility that are controlled by the retailer. These are most often in smaller customers' facilities; examples include controlled (or off-peak) hot water, controlled pool pumps, cycling of air-conditioning, and any controlled circuit arrangements whereby connected (and generally hard-wired) end use equipment can only operate during times determined by the retailer (though these can be operated in a dynamic mode they are generally operated during published times and do not change more than seasonally).

81 Embedded generation is the use of an electricity generation system that is located on the customer side of the meter (to be distinguished from 'distributed generation' which is a generator that is connected directly to the distribution network). As such, some forms of small-scale renewable energy utilisation (eg, rooftop PV arrays – but not solar water heaters) are forms of embedded generation. Embedded generation can also include the use of gas- or diesel- fired standby generators that are located within a customer's facility.

82 Dual fuel is the ability to use an alternative input energy to power a particular end use. This could involve two different pieces of end use equipment, or more commonly a specific piece of end-use equipment that can use more than one input fuel. For example, a gas-fuelled engine can be used as an alternative to an electric motor, as has been done in critical water pumping applications. Another is the use of electric back-up for solar water heating systems, particularly where the electric element is fixed to only operate during off-peak periods.

83 AEMC (2011), *Strategic Priorities for Energy Market Development: Discussion Paper*, p.7.

sunk costs.[84] Virtually any conceivable supply-side response in the generation market would therefore be considered *entry or expansion*. Even OCGT peaking plant, which can be constructed in as little as six months, still requires capital costs in the vicinity of $0.75 to $0.95m per MW and would not constitute supply-side substitution.[85]

In short, the prospect of substitution is remote over *any* reasonable timeframe. This raises the question of how the relevant timeframe can be defined in this instance. Some commentators have suggested that the absence of substitution possibilities – and the resulting potential for large price short-term price increases – means that a shorter timeframe should be adopted to define the relevant market.[86] For example, the Australian Energy Regulator (AER) proposes a three month[87] timeframe and Biggar (2011) seems to imply that the relevant timeframe could be as narrow as a single half-hour period.[88]

Neither approach is likely to be appropriate. The fact that substitution possibilities are remote does not necessarily imply that the relevant timeframe for defining the market should be truncated. Rather, it suggests that the decision as to the relevant timeframe must be guided by *some other principle*. In our opinion, the timeframe should be determined by reference to the *overarching purpose* for defining the market in the first place. This gives rise to two critical considerations, ie:

- the question of ultimate interest is in determining whether a generator has exercised substantial market power by increasing average spot prices to such an extent and with sufficient frequency that they exceed the LRMC of adding capacity; and
- the potential response would be the application of administered prices, which will have long-lasting effects, ie, it does not simply involve the application of *ex-post* penalties as might the case with an action under s.46 of the *Competition and Consumer Act 2010*.

Once these important contextual factors are taken into account, the potential hazards associated with defining the relevant market over a very short timeframe become apparent. Most notably, it gives rise to the significant risk of false findings of substantial market power that may then precipitate an unnecessary market intervention (that also applies over a much longer period). For example, the AER's proposed three month period might span one hot summer. Consideration of that period might therefore reveal:

- that a hypothetical monopolist (of a certain product and over a certain geographic area) could increase average spot prices by engineering shortages to a level that was 5 per cent

---

84 *See for example:* Parr, N., Finbox, R. & Hughes, M (2005), *UK Merger Control: Law and Practice* (2nd edn, Sweet & Maxwell Ltd), p.299.

85 SKM (2010), *Review of the Maximum Reserve Capacity Price 2010 – Power Station Elements.*

86 *See for example:* Twomey et al (2005), 'A Review of the Monitoring of Market Power: The Possible Roles of Transmission System Operators in Monitoring for Market Power Issues in Congested Transmission Systems', *The Journal of Energy Literature,* XI, 2, p.7 and Office of Fair Trading (2005), *Application in the energy sector, understanding competition law*, p.13.

87 AER, Submission to Potential Generator Market Power in the NEM Major Energy Users Rule Change Proposal, 27 May 2011, p.4.

88 Biggar, D (2011), *The Theory and Practice of the Exercise of Market Power in the Australian NEM*, 26 April 2011, p.3.

above the LRMC of adding capacity over that timeframe – indeed, the demand conditions conducive to strategic withholding are likely to occur on hot days; and

- that a generator supplying that product in that geographic location (ie, an actual generator, not a hypothetical monopolist) could also impose such a price increase over that timeframe, and that entry and expansion would not occur within that three month period so as to prevent that price increase.

Such an assessment over such a timeframe might therefore give rise to the conclusion that substantial market power was exercised during that period, the response to which may be the application of enduring price control. But what if during the following nine months of the year there were no opportunities to engage in strategic withholding and a SSNIP could *not* be imposed during that period? The answer is that if the same analysis was undertaken over the remainder of the year, then substantial market power would *not* be found, and the conclusion would be that price control would not be needed. This seems not to provide a viable framework for decisions, since:

- it is *not necessarily* a problem if the increase in average prices in the three month period was sufficiently large that average prices increased by more than 5 per cent above the LRMC over the whole year; but
- it *is a problem* if the increase in average spot prices over the entire year is *less than* 5 per cent above the LRMC, since a SSNIP would not then have been implemented over that longer period.

The application of price control represents a potentially permanent intervention in the market. In our opinion, the decision as to whether to take that step cannot be assisted by adopting a timeframe that is so short that it might lead to different answers in different circumstances. It risks undermining the essential purpose of defining the market, ie, to identify that economic forces that constrain a firm's pricing[89] so that an appropriate decision can be made as to the merits of introducing administered prices. The basic problem with adopting a short timeframe is that temporary pricing power may be confused for substantial market power.

Of course, that is not to say that short-term phenomena are not important, particularly if they occur frequently and have a significant affect on average spot prices. Indeed, the AEMC has recognised that,[90] given the magnitude of the market price cap relative to the average spot price, a small number of periods of very high spot prices during summer could have a very large effect on the average annual spot price over the long run, and may warrant the application of price control. However, to be confident of that diagnosis the assessment must be made over a timeframe that is sufficiently long to distinguish between substantial market power and temporary pricing power.

This means that the relevant timeframe for defining the market is likely to span significantly beyond three months or a half-hour trading interval. An important practical consideration is that, in order for the entire 'demand cycle' to be accounted for in the comparison of prices and LRMC, it is necessary for the timeframe to include all four seasons. This means that the

---

89 Fisher, F (1991), 'Diagnosing Monopoly', *Industrial Organization, Economics and the Law* (John Monz, ed).

90 AEMC Consultation Paper, p.20.

assessment timeframe must be *at least one year*. Anything less risks a timeframe that captures only one hot summer, and jeopardises the application of the SSNIP test.

There are also sound reasons to consider extending that timeframe to include *multiple years*.[91] By using, say, two years, the chances of an atypically hot (or cold) summer 'skewing' the results are reduced. For example, if the year in question coincides with a 'one in fifty year' heat wave in that region, this is likely to provide an unusually high number of opportunities for generators to engage in strategic withholding. An analysis undertaken over that one year period may find that a hypothetical monopolist would be well placed to impose a SSNIP when, in normal years, there may be few (if any) opportunities to engineer shortages. An analysis undertaken over the two year period may therefore find that a hypothetical monopolist could not impose a SSNIP.

This is exactly the same principle as that described above in relation to the three-month window. In short, the longer the assessment timeframe, the more confident once can be that a SSNIP is 'sustained', which, of course, is a requirement of the test.

For these reasons, in our opinion, the consideration of all relevant economic forces is likely to necessitate a timeframe of *at least* one year. However, it may be appropriate to extend that timeframe further still to span two or perhaps even three years. Adopting a longer timeframe also recognises the important practical point that the measurement of LRMC requires a timeframe in which all factors of production are variable – a period that must be measured in years rather than months.

5.4. Product Dimension

The product or group of products supplied by the firm whose market power is at issue forms the base from which the product dimension of the market should be defined. The basic product that is supplied by generators is electrical energy. That product is supplied by different types of plant that tend to operate at different times – base load, mid-merit and peaking plant – and by generators that are 'scheduled',[92] 'unscheduled'[93] and 'semi-scheduled'.[94] All such generators must be registered 'market participants' before they can supply electricity to the NEM. Of course, regardless of when electricity is supplied, or of the type of plant that provides it, the product is exactly the same.

The SSNIP test can therefore be applied by asking whether a 'hypothetical monopolist market participant' (ie, a single firm that owned all scheduled, unscheduled and semi-scheduled base load, mid-merit and peaking plant) could increase the average electricity spot

---

[91] The minimum 'increment' to the assessment timeframe is one year, for the reasons set out above. For example, an 18 month timeframe would potentially be problematic because it would either include only one summer (and so potentially understating the prospect of a profitable SSNIP) or two summers (and so potentially inflating the prospect of a profitable SSNIP).

[92] A scheduled generator must submit bids to AEMO and must follow the dispatch target instructions that it receives.

[93] An unscheduled generator is able to produce as much electricity energy as it likes at any point in time.

[94] Semi-scheduled generators are wind firms whose output is forecast using a centralised wind forecasting algorithm.

price over a one to two year period[95] by, say, 5 per cent above LRMC without that endeavour being defeated by:

- customers either reducing their consumption or switching to alternative sources of energy, ie, engaging in demand-side substitution; and/or
- firms altering their production processes to begin generating electricity, ie, engaging in supply-side substitution.

Section 5.3 explained that the potential for demand-side substitution is likely to be modest and that there is effectively no scope for supply-side substitution.[96] A SSNIP may prompt various forms of demand side responses from some larger customers (eg, installation of embedded generation or fuel substitution). Naturally, the greatest response can be expected to occur during periods in which the spot price increases substantially, ie, at times of peak demand when the risk of shortages is greatest. Most demand-side management initiatives are aimed at reducing the relevant participant's exposure to the spot price during such periods.

We noted above that the AEMC has identified demand-side management as one of its three key strategic priorities for the NEM. It is therefore *conceivable* that the opportunities and incentives for NEM participants to engage efficiently and cost-effectively in those activities may increase in time, and that capability may even be facilitated by changes to the Rules. However, the absence of reliable data on the potential for efficient demand-side management means that it is impossible to know for certain (hence the reason that the AEMC has identified it as a key strategic priority). Moreover, it is doubtful whether those activities would be sufficiently encompassing to defeat a SSNIP *at the present time or in the foreseeable future*. This implies that the relevant product market comprises electrical energy supplied to the NEM.

Finally, there is the question of whether the product market could be expanded to include electricity derivative instruments such as swaps, futures or options.[97] In our opinion, these various financial instruments are not 'products' per se. Rather, they are instruments that provide another means of expressing the price of the *same underlying product*, ie, electrical energy. Moreover, as section 4.4.2 explained, the price of derivative instruments is linked inexorably to expected spot prices.[98] In the words of Justice French:[99]

> 'Although there are some loose, but not entirely appropriate, analogies between the derivative contract and a form of insurance in my opinion, for present purposes, the

---

[95] Note that it is neither necessary nor appropriate to distinguish between 'peak' and 'off-peak' periods when defining the product market. First, the same basic product – electrical energy – is provided at all times. Second, focussing on peak periods risks encountering the hazards described at length in section 5.3. In particular, it risks misdiagnosing a 'sustained' price increase.

[96] Even OCGT peaking plant, which can be constructed in as little as six months, still requires capital costs in the vicinity of $0.75 to $0.95m per MW and would not constitute supply-side substitution. Rather, any such investment would be understood to be a capacity expansion or market entry, as the case may be.

[97] AEMC Consultation Paper, p.20.

[98] Specifically, the price of hedge contracts is primarily determined by the balance of expectations as to the level and volatility of future wholesale spot market outcomes.

[99] *Australian Gas Light Company v Australian Competition and Consumer Commission (No 3)* [2003] FCA 1525, paragraph 382.

> derivative contracts ought to be regarded as ***an integral part of the pricing and payment arrangements between generators and retailers in relation to the underlying product, which is electrical energy***, and which they deal with 'as if' it had been sold from supplier to retailer.' (emphasis added)

In other words, just as there would be no need to include mortgages in the product market for residential or commercial property, there is also no need to include electricity derivative instruments in the product market for electricity. This implies that the relevant product market for the purposes of assessing the MEU Rule change proposal is likely to comprise electricity energy supplied to the wholesale electricity market.

5.5. Functional Dimension

Defining the relevant functional market requires identification of the vertical stages of production and/or distribution that comprise the relevant arena of competition. This involves consideration of any potential efficiencies from vertical integration and, particularly, whether those complementarities are sufficiently strong for additional stages in the vertical supply chain to be included in the same functional market. The key question is whether the efficiencies of vertical integration between two or more stages of the supply chain are so great that 'market co-ordination between buyers and sellers is superseded by in-house co-ordination'.[100]

In our opinion, it can be presumed that activities take place in separate functional markets unless the transaction costs associated with market procurement or, conversely, the synergies associated with vertical integration, are so *overwhelming* as to *preclude* separate provision (or, expressed in another way, *dictate* that the services be undertaken by a single economic entity). This approach is also consistent with the approach of:

- the ACCC when assessing mergers, ie, a single functional market is defined only when there are 'overwhelming efficiencies of vertical integration between two or more stages in the vertical supply chain';[101]
- the Australian Competition Tribunal (the Tribunal) in *Sydney Airport (No 1)*, in which it concluded that, unless the efficiencies across two related activities were of such a magnitude to 'dictate the services must be performed within the same economic entity', then it should be assumed that separate functional markets can be defined;[102]
- the Tribunal in *Re Services Sydney Pty Limited* [2005] ACompT 7 in which it adopted a test 'which asks whether the complementarities of vertical integration are such as to dictate vertical integration';[103] and
- the Tribunal in the matter of Fortescue Metal Group Limited [2010] ACompT 2, in which it indicated that a relevant question was whether 'in-house provision of a particular good

---

[100] Brunt (1990), p.86.

[101] ACCC, Merger Guidelines, November 2008, p22.

[102] *Re Sydney International Airport* [2000] ACompT 1 at 97.

[103] *Re Services Sydney Pty Ltd* [2005] ACompT 7 at 119.

> or service is always more profitable than if it is purchased from a third party. If the answer is yes, then there are no potential transactions and hence no separate market.'[104]

In the case of electricity generators, the only query is whether the retail function should also be included in the functional dimension of the market, in light of the proliferation of so-called 'gen-tailers'. In our opinion, although there are undoubtedly synergies between the generation and retailing functions (most notably the potential risk management benefits), those efficiencies are not sufficiently strong as to *dictate* that a generator must also incorporate a retail business. This is evidenced quite simply by the fact that many generators are *not* retailers. The relevant functional market is therefore likely to be confined to electricity generation and does not extend to include subsequent vertical stages of production.

5.6. Geographic Dimension

To determine the geographical dimension of a market one should first identify the geographical boundaries to the relevant activities of the enterprise whose market power is at issue. Here, the focus is on individual generators, each of which sells its output in a particular NEM region. The SSNIP test can therefore be applied by considering whether a hypothetical monopolist of all of the generating capacity (ie, the hypothetical monopolist market participant described in section 5.4) in that NEM region could increase the average regional spot price over a one to two year period by 5 per cent above LRMC or, as we explain below, a proxy of LRMC.

The principal constraint upon the hypothetical monopolist would be from generators located in *other* NEM regions that are able to supply electricity to that location via the interconnectors (the cumulative price threshold (CPT) may pose a further constraint[105]). The key empirical question is whether there is a sufficient number of trading intervals during the course of the year during which the hypothetical monopolist was 'pivotal' given the current level of interconnector capacity, and so could therefore engage in strategic withholding.

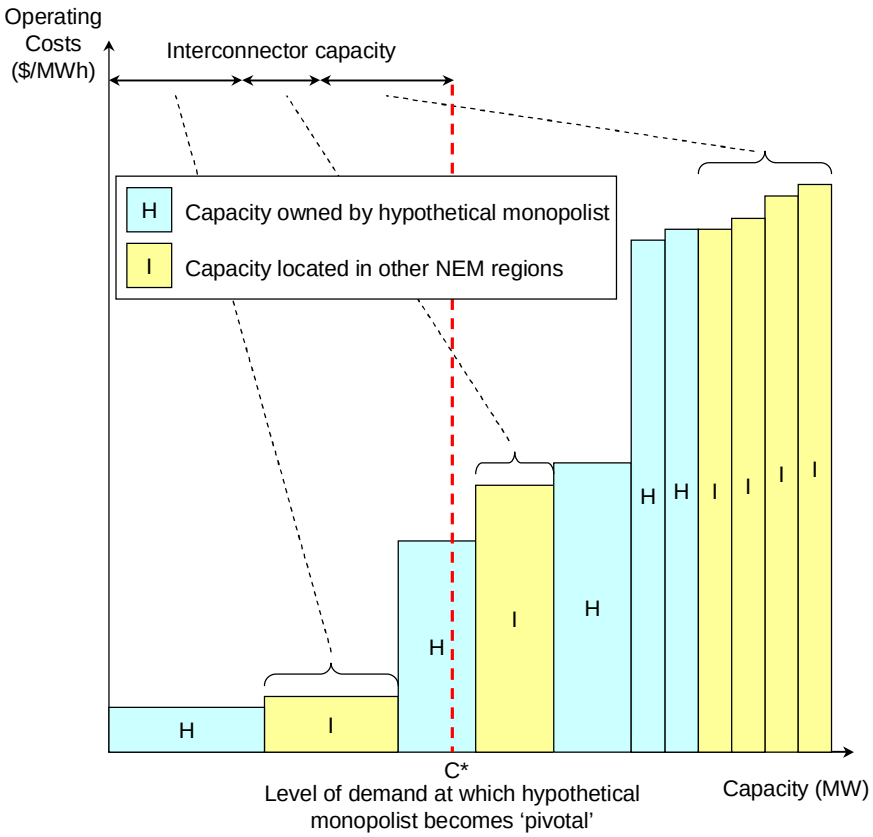
For example, the hypothetical monopolist depicted in Figure 5.1 is pivotal whenever demand exceeds C*, which is the maximum capacity that can be imported from outside the region. In those periods where there was a strong possibility that demand would exceed that level, a hypothetical monopolist may have a strong incentive to withhold some of its capacity so that the interconnector becomes constrained, and the remainder of its capacity is dispatched at a much higher price – potentially the market price cap.

---

[104] In the matter of Fortescue Metals Group Limited [2010] ACompT 2, at 1046.

[105] As noted earlier, an administered price cap (APC) of (typically) $300/MWh is imposed by AEMO whenever the sum of 336 consecutive trading interval prices (ie, 7 days) exceeds the CPT, which is currently $186,000. Once invoked, the APC remains in place until the end of the trading day during which the rolling sum of prices falls below the CPT.

Figure 5.1
'Pivotal' Hypothetical Monopolist



In other words, the first step in the SSNIP analysis would be to estimate what the average regional spot price *would have been* over the course of one or two years if all of the generating units in the region were owned by one firm. In practice, this will inevitably involve a data intensive assessment of historical demand levels in the region in question to determine what the optimal bidding strategy for a hypothetical monopolist would have been over that period (taking into account the potential additional constraint posed by the CPT[106]). Although complex, that modelling exercise is certainly possible provided that all of the relevant information was available, which we expect it is.

The next step would be to establish a benchmark against which to compare that estimated average price. Strictly speaking, the relevant comparator is the LRMC of expanding capacity, for the reasons described in section 4.4 (and elsewhere). Specifically, if the average regional spot price estimated to arise under hypothetical monopoly was:

- more than 5 per cent above the LRMC of expanding capacity, that would establish that a SSNIP could be profitably imposed, and that the relevant geographic market comprised that NEM region; and

[106] Once the APC has been invoked, a generator has no incentive to use its market power to increase the wholesale spot price above the APC (usually $300/MWh). Its bidding incentives may also be influenced if exercising market power would *result* in a breach of the CPT and the application of the APC.

- if it was less than 5 per cent above the LRMC of expanding capacity, that would imply that a SSNIP could not be profitably imposed, and would require the geographic boundaries to be expanded to include other NEM regions, and the test repeated.

In practice, estimating the LRMC of expanding capacity is not straightforward. However, there is a simpler comparison that may obviate the need to undertake that potentially complex assessment. Specifically, it *may* be sufficient to compare the average regional spot price estimated to arise under hypothetical monopoly to the average price that *actually* transpired over the period.

In particular, if the average regional spot price estimated to arise under hypothetical monopoly was *more than* 5 per cent above the actual average spot price over that period, it is reasonably safe to assume that the relevant geographic market comprised that NEM region. This is because it is reasonable to assume that the actual average spot price over the period will be *equal to or greater than* the LRMC of adding capacity, ie:

- if competition in the market is workable, then the average spot price should be approximately equal to the LRMC of capacity, for the reasons described in section 4.4 (and elsewhere); and
- if substantial market power exists in the market, then the average spot price should be *greater than* the LRMC of adding capacity, as a result of the generator (or generators) in possession of that market power engaging in strategic withholding.

In other words, if a hypothetical monopolist could have imposed a SSNIP on historical average spot prices it can be assumed to be capable of sustaining a SSNIP of the same or greater magnitude over LRMC. However, it is not reasonable to assume the opposite. Specifically, if the average regional spot price estimated to arise under hypothetical monopoly was *less* than 5 per cent above the actual average spot price over that period, the relevant geographic market should not necessarily be broadened. This is because of the potential effect of the cellophane fallacy.

Specifically, the apparent unprofitability of the SSNIP may simply reflect the fact that historical average spot prices had *already been affected* by the exercise of substantial market power. A hypothetical monopolist may not have been able to increase those prices any further because all opportunities to exercise market power had already been exploited. In other words, the fact that a hypothetical monopolist cannot impose a SSNIP on historical average spot prices does not necessarily mean that it could not profitably sustain a SSNIP over LRMC. In these circumstances, there may be no option but to estimate LRMC and to employ that figure as the relevant benchmark.

In other words, the determination of the geographic dimension of the market requires an empirical exercise to be undertaken. It is therefore not possible to reach a definitive conclusion in this paper on the appropriate geographic boundaries of the market. However, in our opinion, there is a reasonable probability that the modelling exercise described in this section would reveal that, in many cases, a hypothetical monopolist in a NEM region *could* profitably impose a SSNIP, indicating a geographic market that is delineated by a series of NEM regions, or combinations of NEM regions.

Of course, even if the appropriate geographic market for assessing a generator's market power is indeed the NEM region in which it is located, that does not mean that generators located in other regions can be ignored. Indeed, when assessing the existence (or otherwise) of substantial market power in that geographic market, it will be necessary to consider the constraint posed by *all generators* who, operating by means of the relevant interconnectors do (or could) supply that region and affect the regional price, wherever they may be based.

5.7. Summary

The purpose of defining a market is to frame the relevant arena of competition to enable the *real* question of interest to be answered. In this particular case, we are interested in whether particular generators possess substantial market power that may be worth addressing by means of a market intervention, eg, by applying price control. The market definition must therefore be capable of identifying circumstances that may necessitate *ex ante* regulation to address structural concerns or enduring market failures for which *ex post* intervention is impracticable or inappropriate.

This suggests that the relevant timeframe for defining the market will need to span *at least* one year and possibly two. A shorter timeframe risks overlooking relevant economic forces and, in particular, mistaking temporary pricing power for substantial market power. A further advantage of adopting such a timeframe is that it allows the entire 'demand cycle' to be accounted for in the comparison of prices and LRMC – a comparison that cannot be meaningfully implemented over, say, a three month period.

The relevant product market for the purposes of assessing the MEU Rule change proposal is likely to comprise electricity energy supplied to the wholesale electricity market. There is no need to extend that definition to include electricity derivatives, since these instruments are simply another way of expressing the price for the same underlying product. There is also no need to expand the functional dimension of the market to include electricity retailing. Although there are potential complementarities between the generation and retailing functions, those efficiencies are not sufficiently strong as to *dictate* that a generator must also incorporate a retail business. Indeed, many generators do not.

It is not possible to reach a definitive conclusion on the appropriate geographic dimension of the market, since this requires the completion of an empirical exercise. However, in our opinion, there is a reasonable high probability that the modelling exercise described in section 5.6 would reveal that, in many cases, the relevant geographic market was limited to a NEM region, or combinations of NEM regions. However, assuming such an analysis did conclude that the market is delineated by a series of NEM regions or combinations of NEM regions, this does not mean that that generators located in other regions can then be ignored in the subsequent assessment of substantial market power. Rather, the constraining effect of those competitors operating by means of relevant interconnectors would be a critical consideration.

6. Conclusion

This report has considered the economic concepts of 'competition' and 'market power', their implications for the application of regulation and how those concepts apply to wholesale electricity generation markets such as the NEM. A workably competitive wholesale electricity spot market functions no differently from most other workably competitive markets. Specifically, with certain limited exceptions, if prices are significantly and persistently *above LRMC* or *below LRAC* this should, given time, prompt a supply-side response that restores prices to those levels.

Market power is the antithesis of workable competition. For this reason, it is particularly important to distinguish the exercise of substantial market power from *temporary pricing power*, which is a common feature of workably competitive markets. A generator (or group of generators) can be considered to possess a substantial degree of market power when it is not constrained by the forces of workable competition, ie, when:

- it has the ability to increase average spot prices to such an extent and with sufficient frequency that they exceed the LRMC of adding capacity, including a return on capital and accounting for risk;[107] and
- it is insulated from the forces competition by significant barriers to entry and expansion (as opposed to, say, minor differences in product attributes) that enable it to sustain average prices at that level.[108]

It follows that any assessment of whether a generator has a substantial degree of market power consequently requires:

- a focus on genuine and enduring barriers to entry and expansion, as the fundamental source the substantial market power, noting that this might also include 'strategic' barriers to entry and expansion; and
- the undertaking of long-term price cost tests, as evidence of the exercise of market power, eg, comparisons of average spot prices to the LRMC of adding capacity, rather than comparisons of spot prices to SRMC at particular points in time.

To be sure, periods of high prices that appear to have arisen from strategic bidding conduct are certainly relevant to that assessment, but only to the extent that they have had a sustained effect on average spot prices that is likely to persist over the long term. It is also unlikely to be necessary to consider the price of hedge contracts since a generator's market power must ultimately be signalled to customers through the spot market.

This report has also considered the appropriate market definition for the purposes of considering the Rule change that has been proposed by the MEU. That Rule change contemplates a significant intervention into the operation of the NEM, which would have lasting effects. The market definition must therefore be capable of encompassing structural

---

[107] Note again that this may involve engaging in strategies such as predatory pricing, albeit for the same purpose.

[108] As general rule, entry and expansion will provide an effective competitive constraint if barriers to entry and expansion can be overcome in an appropriate time to deter or defeat any non-transitory exercise of substantial market power.

concerns or any other forms of enduring market failure that would warrant such an intervention.

This suggests that relevant timeframe for defining the market will need to span *at least* one year and possibly two. The adoption of a shorter timeframe would risk mistaking temporary pricing power for substantial market power. The relevant product market is likely to comprise electricity energy supplied to the wholesale electricity market. The relevant functional dimension of the market is likely to be limited to electricity generation.

It is not possible to reach a definitive conclusion on the appropriate geographic dimension of the market, since this requires the completion of an empirical exercise. However, in our opinion, there is a reasonable probability that the modelling exercise described in section 5.6 would reveal that, in many cases, the relevant geographic market was delineated by a series of NEM regions or combinations of NEM regions.

Appendix A. Marginal Cost Concepts

This appendix provides a more detailed overview of the concept of marginal cost, and some of the challenges that arise in its estimation over both the short and long term.

A.1. Short Run Marginal Cost

Section 2.1 explained that SRMC can be defined as the cost of an incremental change in demand, holding capacity constant. Importantly, its estimation takes account of the potential costs of shortages faced by customers. In the event supply cannot expand to match demand, SRMC rises to whatever price level is necessary to curtail demand to match available supply. Its application in the context of decisions affecting the future therefore relies as much on probability and expectation as on fact.

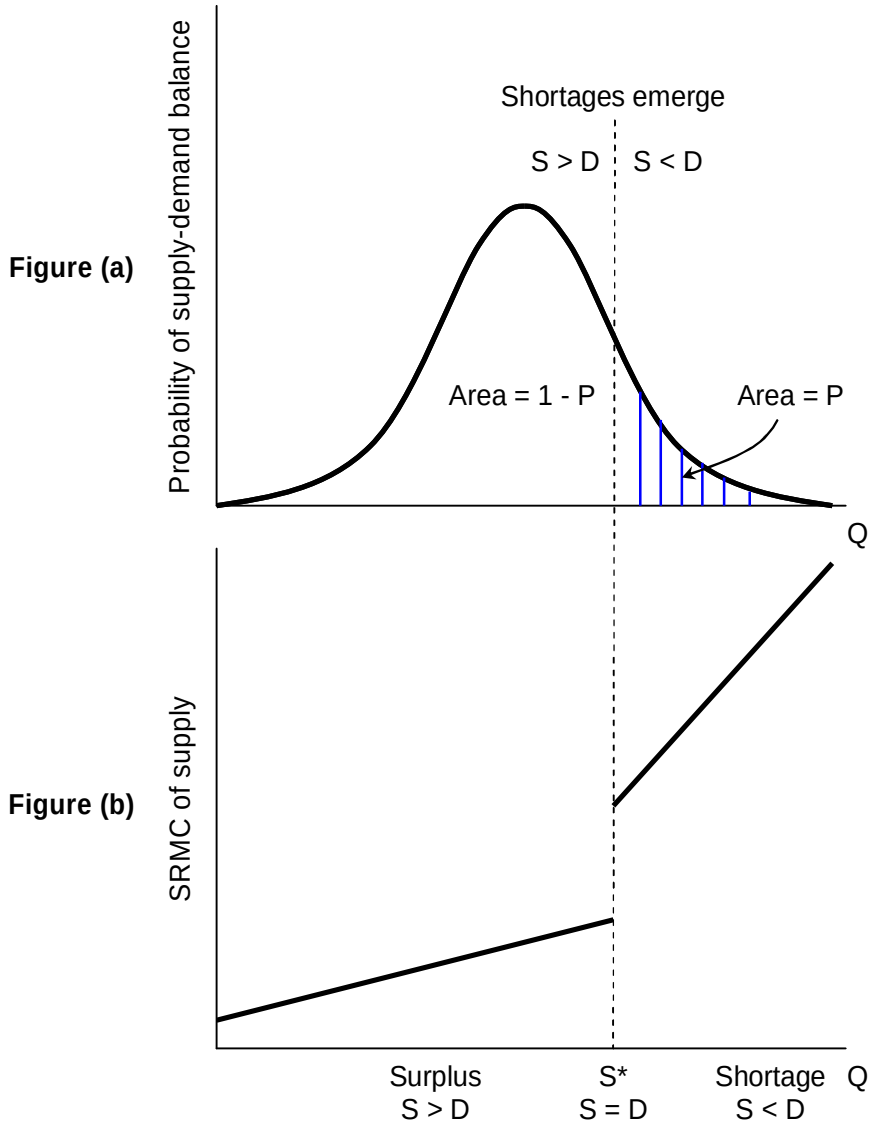
Its estimation for the purposes of decisions that are relevant beyond the immediate point in time involves a probabilistic assessment of possible future outcomes and the costs they entail. Specifically, a forward-looking SRMC is the sum of the various additional costs arising under different scenarios (holding capacity constant), multiplied by the probabilities of these scenarios occurring. Formally, the expected SRMC is given by:

- the SRMC when supply exceeds demand (ie, operating and maintenance costs), multiplied by the probability that supply exceeds demand; *plus*
- the SRMC when supplies are less than demand (ie, *including* the costs of shortages) multiplied by the probability that supply is less than demand.

Figure A.1 shows the SRMC under conditions of surplus and shortage. Figure (a) shows a probability density function of different supply and demand balances. The maximum capacity that the system can supply is shown as S*. Below S*, supply exceeds demand (S>D), and there is no shortage. Above S*, supply is less than demand (S<D), and shortages occur. The shaded area under the probability density function shows the probability of shortages (P). The probability that there are no shortages is given by the remaining area under the curve (1-P).

Figure (b) shows the cost conditions associated with surplus and shortage. In the absence of shortages, SRMC is low, but increase when supplies become less ample relative to demand in order to ration the available capacity (which is fixed in the short run). This reflects the fact that the cheapest means of supply will be used first, followed by increasingly expensive supplies as the supply-demand surplus falls, as discussed above. Beyond the capacity constraint, S*, shortages occur, resulting in a sharp jump in SRMC costs, reflecting the cost to customers of the marginal unit of shortage.

Figure A.1
Short Run Marginal Cost



The SRMC of a product will also be affected by changes in the demand/supply balance. The tighter is that balance, the more likely it is that SRMC will need to increase to curtail demand, since the more susceptible is the market to temporary disruptions, ie:

- when supply is plentiful, there is little probability of shortages and SRMC is relatively low, ie, the probability-weighted cost of curtailing demand it will low or zero; and
- when supply becomes scarce, the probability of shortages increases and SRMC will rise, ie, the probability-weighted cost of curtailing demand will start to increase.

By way of example, a natural disaster that destroys a large proportion of the world's annual sugarcane crop can be expected to have a greater expected effect on the SRMC of sugar when it exacerbates an already critically tight demand/supply balance.

To summarise, SRMC can be defined as the cost of an incremental change in demand, holding capacity constant. Importantly, its estimation takes account of the potential costs of shortages faced by customers. In the event supply cannot expand to match demand, SRMC rises to whatever price level is necessary to curtail demand to match available supply.

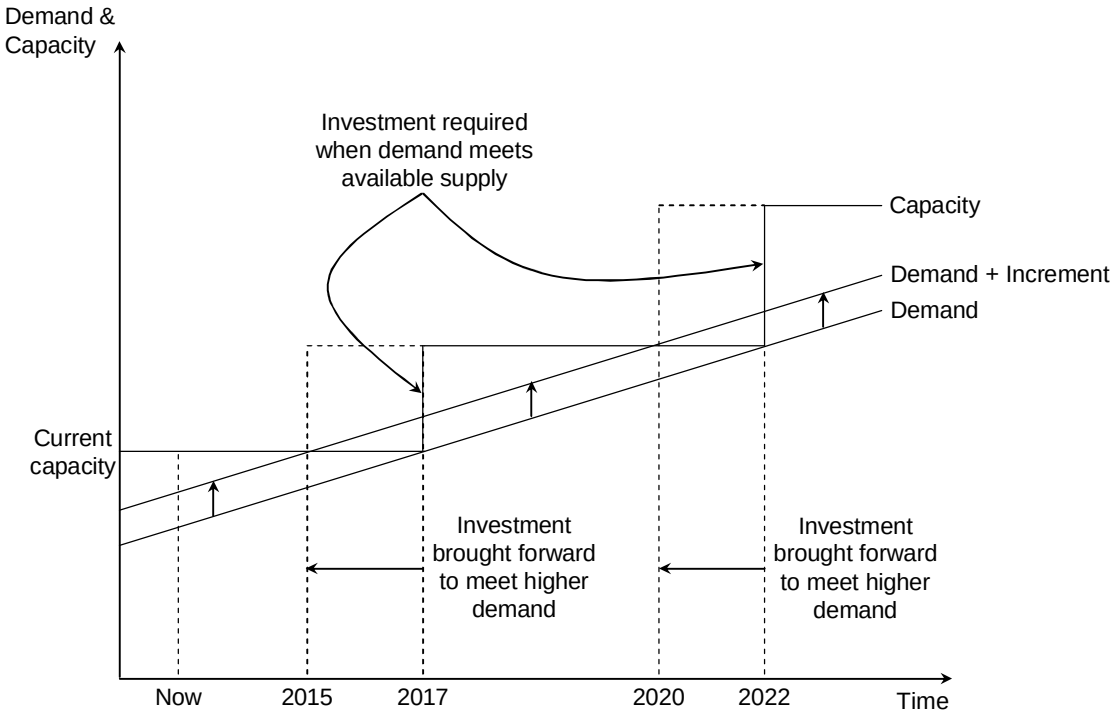
A.2. Long Run Marginal Cost

Section 2.2 explained that LRMC reflects the cost of serving an incremental change in demand in a market, assuming all factors of production can be varied. Importantly, because LRMC is a long run concept, it accounts for the fact that firms have the option of *expanding their capacity* in order to meet an incremental increase in demand. Measuring LRMC involves estimating the costs involved with undertaking a capacity expansion *sooner than would otherwise be the case* in response to that change in demand.

In Figure A.2 below, an incremental increase in demand would result in capacity expansions that would otherwise have taken place in 2017 and 2022 being moved forward in time to 2015 and 2020, respectively. The capital cost component of LRMC in such an industry can therefore be estimated by taking the difference in the present values of the capacity expansions that will occur at these earlier dates and the present value of the later expansions that would have occurred *without* the incremental increase in demand.[109]

[109] Formally, the LRMC can be calculated as the present value of the difference between predicted expenditure under forecast demand and predicted expenditure with incrementally increased demand, divided by the discounted total incremental increase in demand. *See:* Turvey, R (2000) *What are Marginal Costs and How to Estimate Them?*, Centre for the Study of Regulated Industries (CRI), University of Bath. Note also that the LRMC of adding capacity (and the LRAC associated with reducing capacity) will be determined by the operating and capital costs associated with the optimal investment profile to meet the increment in demand.

Figure A.2
Estimating LRMC



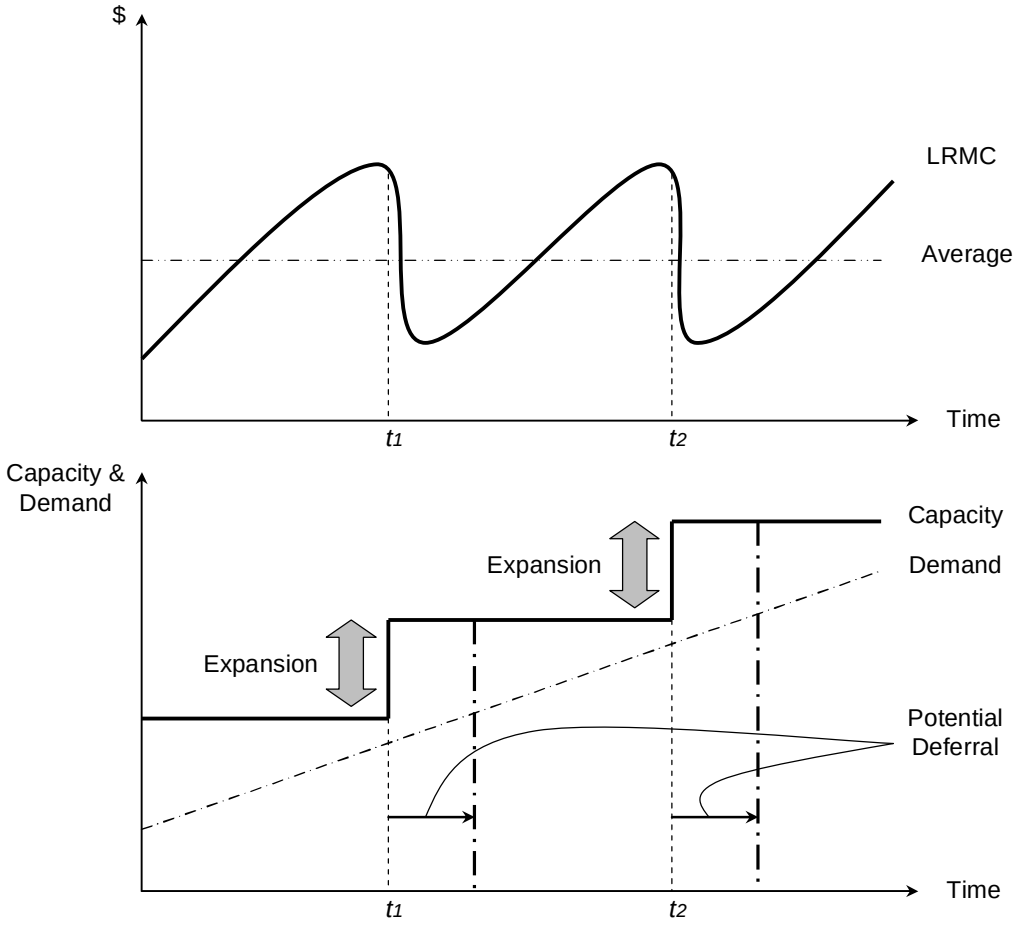
The approach set out in Figure A.2 implies that where capacity must be added in 'lumpy units' (rather than in very small increments), this gives rise to *time-dependent fluctuations* in LRMC. Specifically, the LRMC of supply in such a market will be relatively low when capacity utilisation is low and the next capacity expansion is some distance in the future, but will rise as capacity utilisation increases and the timing of the next expansion is nearer. By way of illustration, Figure A.3 displays the stylised, optimal expansion profile required to serve market demand and the associated LRMC of capacity over time. It shows that:

- in the time period immediately following a capacity expansion (ie, those following t_1 and t_2) the LRMC of the next increment to capacity is low, because the value of any potential deferral of that future capacity requirement is relatively low due to the effect of discounting; and
- as spare capacity declines over time and the need to invest in new capacity approaches (ie, the time periods leading up to t_1 and t_2), the LRMC of the next increment to capacity increases, because the value created through any potential deferral is closer in time and so less (negatively) affected by discounting.

In other words, LRMC *changes over time* as new capacity is added. The LRMC associated with meeting the incremental demand shown in Figure A.2 would therefore be higher in, say, 2014 than it would be today. This is because the cost today of, say, bringing forward by one year a $1m investment that would otherwise have taken place in 12 months' time is much

greater than the cost today of that same one year rescheduling applied to a $1m investment expected to be made in 10 years' time, because of the time value of money.[110]

Figure A.3
LRMC, Demand and Capacity over Time



In summary, LRMC reflects the cost of serving an incremental change in demand in a market, assuming all factors of production can be varied. Importantly, because LRMC is a long run concept, it accounts for the fact that firms have the option of *expanding their capacity* in order to meet an incremental increase in demand. Measuring LRMC involves estimating the costs involved with undertaking a capacity expansion *sooner than would otherwise be the case* in response to that change in demand.

A.3. Relationship between SRMC and LRMC

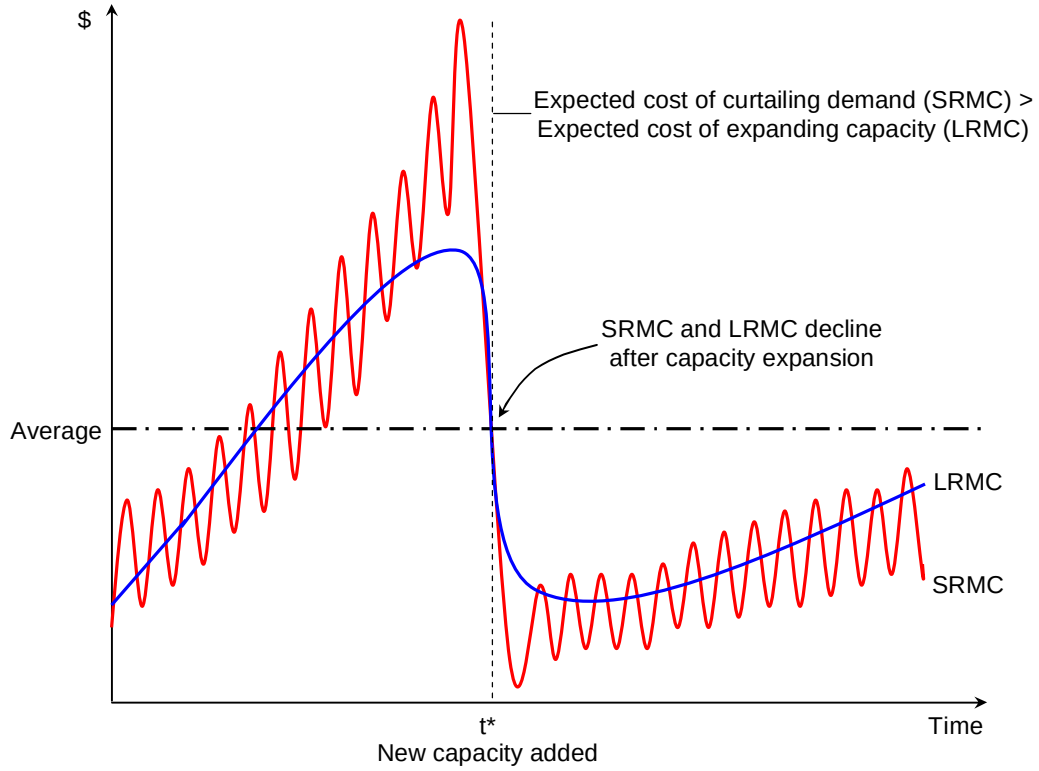
The previous sections explained that SRMC is the cost of an incremental change in demand, holding capacity constant, whereas LRMC reflects the cost of meeting that change in demand

[110] Put another way, the *value* today of *deferring* by one year a $1m investment expected to be made in 12 months' time is much greater than the value today of that same one year deferral applied to a $1m investment expected to be made in 10 years' time.

assuming capacity can vary. Unless assets are highly mobile and capacity can be added in very small increments – conditions that are rarely seen[111] – there is no reason to expect SRMC and LRMC to be the same *at any particular point in time*. However, there is still a strong 'in principle' link between SRMC, LRMC and capacity expansion decisions.

This relationship is illustrated in Figure A.4 below, which depicts the SRMC and LRMC in a market in which demand is increasing over time. In the first instance, medium term demand growth can only be met through increased risk of congestion, or the need for demand curtailment during short run peaks, as reflected in the rising SRMC leading up to t*. However, there eventually comes a 'tipping point' at which the expected SRMC of *curtailing* demand increases beyond the expected LRMC cost of expanding capacity to *meet* that demand. This occurs at t*, at which point new investment takes place.

Figure A.4
SRMC, LRMC and Capacity Expansion



Beyond t* there is significantly more capacity and the probability of shortages emerging that will require demand curtailment is much reduced. SRMC is therefore lower, on average, than during the period leading up to t*. LRMC is also much lower after t* than during the period immediately prior. This is because, beyond t* the LRMC of the *next expansion* is low,

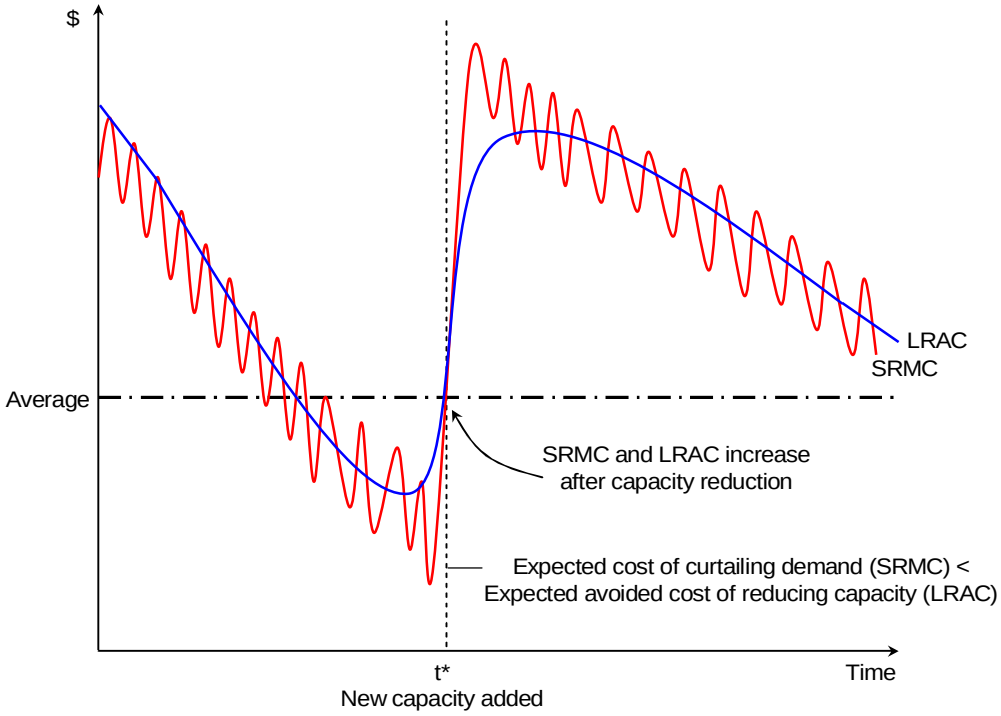
[111] When these conditions are present, *there is no distinction* between SRMC and LRMC since, by definition, there is no difference between the short run and the long run. Any level of demand can be met by quickly adding (or subtracting) capacity and so the need to curtail demand never arises. In these circumstances, SRMC and LRMC are always equivalent, and constant at all times. Of course, industries that exhibit such characteristics are rarely seen.

because the cost associated with bringing forward that future capacity requirement is relatively small due to the effect of discounting.[112]

This is because the costs that would be incurred today by deferring for one year a $1m a capacity expansion that is expected to be made in 12 months' time are much higher than the costs that would be avoided by undertaking that same capacity reduction in 10 years' time. It follows that LRMC must fall immediately following a capacity expansion, since the next expansion is unlikely to be needed for some time and the costs of deferring that investment will be relatively modest for the time being.

Exactly the same principles apply to a market in which demand is *declining* over time. In the first instance, declining demand can be met by firms continuing to supply the market with their existing capacity, as reflected in the declining SRMC leading up to t*. However, there will again be a 'tipping point' at which the long run costs that would be ***avoided*** by reducing or redeploying capacity exceed the SRMC of continuing to supply the product at the current level of capacity. This occurs at t* – at this point, capacity is redeployed to other markets where returns are more attractive.

Figure A.5
SRMC, LRMC and Capacity Reduction



Beyond t* there is less capacity and the probability of shortages emerging that will require demand curtailment is increased. SRMC is therefore higher, on average, than during the

[112] This is because the costs that would be incurred today by deferring by one year a $1m a capacity expansion that is expected to be made in 12 months' time are much higher than the costs that would be avoided by undertaking that same capacity reduction in 10 years' time. It follows that LRMC must fall immediately following a capacity expansion, since the next expansion is, by definition, more distant than prior to the investment.

period leading up to t*. The long-run avoidable cost (LRAC) associated with a decrement in demand[113] is also higher after t* than during the period immediately prior. This is because, following t*, the LRAC of the *next capacity reduction* is higher, because the costs that would be avoided by bringing forward that future capacity reduction are relatively *high*.

This is because the costs that would be *avoided* today by *bringing forward* by one year a $1m a capacity reduction that is expected to be made in 12 months' time are much *lower* than the costs that could be avoided by that same capacity reduction in 10 years' time. This simply reflects the fact that a greater proportion of costs can be avoided over a longer timeframe, ie, costs that cannot be avoided within one year potentially can be avoided within ten years.

Of course, in practice, the 'tipping points' described above will not be as well defined as they are in Figures A.4 and A.5. Indeed, it is often very difficult to time capacity expansions and reductions to coincide perfectly with the emergence of inefficient levels of demand curtailment, ie, when scarcity is either too common or too infrequent. This is particularly the case when capacity must be added and withdrawn in large increments that alter substantially the supply/demand balance. There may therefore be times when:

- SRMC is *above* LRMC for a period as the market waits for new capacity to come on-stream; and
- SRMC is *below* LRMC for a period as the market waits for redundant capacity to be re-deployed elsewhere.

However, such instances of 'misalignment' are neither unexpected, given the imperfections that can affect real world markets, nor a cause for concern, provided that they are transitory. Even accounting for such periods, there is no reason to expect SRMC to differ materially from LRMC, on average, provided they are properly defined and assessed over a sufficiently long timeframe. Equally, although both SRMC and LRMC can fluctuate over time (as Figures 2.4 and 2.5 illustrate), there is no reason to think that either will diverge over the long term.

[113] LRAC is a measure of the cost that would be saved by reducing capacity in response to a small reduction in demand. Recall that SRMC is a measure of the cost that would be incurred by increasing capacity in response to a small increase in demand. The two concepts are synonymous.

NERA
Economic Consulting