NERA
ECONOMIC CONSULTING

Compensating Demand Curtailment to Increase Efficiency in the Operation of the NEM

4 May 2012

Project Team

Adrian Kemp

Jonathan Falk

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

The Australian Energy Market Commission (AEMC) is investigating and identifying the market and regulatory arrangements needed across the electricity supply chain to facilitate efficient investment in, operation and use of demand side participation in the National Electricity Market (NEM) – the Power of Choice review. The investigation is being undertaken at the request of the Ministerial Council on Energy, amid concerns about rising electricity prices and the opportunity to make greater use of demand side participation in the NEM to achieve more efficient market outcomes.

Within this context, NERA Economic Consulting has been asked by EnerNOC – a third party supplier of demand side services – to:

- first, investigate whether the current NEM market design is likely to promote efficient demand side participation in the market; and
- second, to the extent that the current market design does not promote efficient demand side participation, consider how a mechanism, such as that used within the PJM market in the United States, could be applied in the NEM to promote more efficient demand side participation.

Most consumers do not face and so do not respond to wholesale market prices

Electricity markets differ from most other markets because the quantity of electricity produced must match demand at each instantaneous point in time and so prices continually change to equate demand and supply. A combination of highly variable demand and significant variation in the costs of different generation technologies means that wholesale market prices can vary considerably at any point in time. The variability in prices creates risks for both generators and consumers.

As a consequence, generators and retailers on behalf of end-use consumers use financial derivatives to lock in the future price of electricity that will be supplied or purchased. This allows retailers to offer consumer's tariffs that are more predictable while creating incentives for investors to build generation capacity in response to the price signals created through the wholesale market.

The result is that almost all consumers do not – in practice – face real-time prices nor have the relevant skills or information to be able to adequately assess whether they might be better off managing demand either directly or through a third party, to lower overall electricity costs. There are real barriers to consumers using demand curtailment to manage electricity costs.

More price responsive demand will lower the total costs of supply…

It is well accepted that allowing demand to respond to wholesale market price signals will improve the efficiency of the market, thereby lowering the cost of electricity supply. The benefits result from generation cost savings both in the short and long-run, which are partly offset by lost consumer benefits from no longer consuming electricity.

… but the current market arrangements prevent consumers from revealing how they would respond to prices

A lack of information, systems, widespread adoption of metering and other smart technologies, and retailer disincentives to offer real-time prices or other more innovative tariff offerings means that there is little opportunity for consumers to respond to prices, even if they would benefit individually from doing so.

There are potentially significant market inefficiencies if customers that would respond to wholesale market prices are unable to reveal their willingness to reduce demand. The current market design creates barriers for consumers taking advantage of demand curtailment to manage electricity costs. This results in consumers being given limited options to express their responsiveness of demand to changes in price, particularly during generation and network peak periods.

While allowing consumers to directly face real-time prices would address this problem, there are a number of reasons why retailers do not offer consumers real-time prices, including:

- retailers essentially make profits by managing wholesale price risks on behalf of consumers and so if retailers no longer managed those risks then this would have implications for the margin charged to consumers for this service;
- the costs of installing metering and associated transaction and administrative costs which affects the competitiveness of tariff offerings allowing consumers to respond to wholesale price signals;
- the lack of consumer understanding and information about wholesale risks/opportunities and so the increased complexity new tariff products would create;
- preferences by consumers to prefer flatter pricing structures, given the costs and risks that consumers would incur managing half-hourly price variations; and
- there is limited reliable information on the likely responsiveness of demand and so on the wholesale market cost savings that could be derived if consumers were given the opportunity and information to respond to wholesale market prices.

In effect, retailers have little incentive to provide opportunities for consumers to respond to wholesale price signals by curtailing demand. This results in consumers having limited opportunity to reveal preferences for demand response versus increased electricity generation. By implication this means that the current market is unlikely to provide efficient levels of demand response and optimise new generation investment against demand response activity.

Even if more consumers could respond to wholesale market price signals, demand response would be inefficient

Reductions in demand by price responsive consumers also deliver potentially significant system wide benefits that are not captured in the benefits they would receive by simply responding to wholesale prices. These system wide benefits relate to:

- reductions in unserved energy, delivering benefits to those customers who may otherwise have been indiscriminately forced to use less electricity through rolling blackouts; and

- reductions in network costs through the avoidance of otherwise required network capacity augmentations.

Facilitating greater participation of demand curtailment in response to wholesale price signals provides the opportunity for those consumers that value load less to consume less during these periods. This will reduce the level of unserved energy delivering benefits to all consumers.

The benefits from reduced network costs over time arise both organically as extreme peak demand is reduced, and depending on the mechanism used to facilitate demand response, improving the capacity for demand to be contracted to address identified network constraints.

This positive externality for an individual engaged in demand reductions represents a market failure that improving the price responsiveness of demand cannot directly solve.

Incorporating demand side participation directly in the wholesale market will achieve more efficient wholesale market outcomes

More efficient wholesale market outcomes can be achieved by incorporating demand side participation directly in the wholesale market bidding and dispatch process. This would provide an opportunity for consumers either directly or through third party aggregators to reveal preferences for demand reduction. We believe that the current market barriers means that there are likely to be lost opportunities to use cost effective demand reductions to achieve overall wholesale market efficiencies.

A demand side wholesale market compensation mechanism can be designed to provide the same incentive as real-time prices to customers at the margin, but without exposing those customers to needing to manage all of the price volatility risks in the market. It essentially creates an option to respond to wholesale market prices where those prices are higher than the value from using electricity, while simultaneously allowing the consumer to choose a retail tariff product to manage price volatility risks more generally.

To include such a mechanism in the NEM requires consideration to be given to:

- determining the relevant baseline against which demand curtailment would be measured, for each electricity consumer participating in the scheme;
- allowing registered demand curtailment participants to 'bid' load curtailment offers for consideration in the market dispatch process;
- designing mechanisms to ensure that any bid demand curtailment is actually delivered;
- designing mechanisms to ensure that consumers engaged in demand curtailment pay retail tariffs as though demand curtailment had not occurred, ie based on baseline demand rather than actual demand;
- retailers settling in the wholesale market based on baseline demand, rather than actual demand; and
- ensuring that delivered demand curtailment is paid the NEM settlement price for demand curtailed within the relevant 30 minute period, and so is treated equivalently to generation dispatched.

Under this mechanism the settlement price will be determined by comparing baseline demand against the supply offers inclusive of demand response offers. The actual generation dispatched would be reflective of the baseline demand less the dispatched demand curtailment.

Allowing demand side curtailment to be incorporated in the electricity wholesale market bidding and dispatch process is being introduced into wholesale markets throughout the United States, following a decision by the Federal Energy Regulatory Commission. Relevantly, the approach being adopted in the US is equally valid for an energy-only market like the NEM, or a wholesale market with a separate capacity market.

Compensating demand reductions directly in the wholesale market has many desirable features over alternative approaches

Introducing demand side bidding into the NEM would provide similar incentives as real-time prices to customers at the margin, but without exposing those customers to the price volatility risks in the market. It essentially creates an option to respond to wholesale market prices where those prices are higher than the value from using electricity, while simultaneously allowing the consumer to choose a retail tariff product to manage price volatility risks for other periods.

Adopting a mechanism to compensate demand curtailment directly in the wholesale market has many desirable features, namely:

- it creates the same incentives as though a customer faces real-time wholesale prices and so will improve the overall efficiency of the market;
- it allows third party aggregators to directly compete with retailers to manage wholesale risks, thereby effectively increasing the options available to consumers to manage electricity costs;
- it provides an additional option to facing real-time wholesale prices directly for consumers wanting to obtain the benefits from demand response – if real time tariff products start being offered to consumers and are preferred to the wholesale compensation mechanism, then the demand for compensation through the wholesale market for demand response will be reduced;
- it reduces the complexity and so costs to consumers of responding to real-time wholesale price signals;
- it promotes efficient demand response, where the costs of obtaining the response do not outweigh the benefits to the system as a whole;
- it will reveal the opportunities available from demand response in the market, and in so doing provide insight on the materiality of the current market barriers to consumer demand response; and
- it does not require any additional funding as the demand reductions are compensated through the wholesale market settlement.

Finally, if once such a mechanism is introduced in the market there is significant increase in demand response activity, then this would emphasise the benefits that would have otherwise been lost if the mechanism had not been introduced. The uptake of demand response

following the introduction of similar mechanisms elsewhere suggests that the size of the lost opportunity is likely to be significant.

Finally, and most importantly from the perspective of the AEMC's review, allowing demand curtailment to be compensated through the wholesale market has no obvious downside at all, apart from the costs involved in implementation. If consumers choose to not make use of the opportunities that they are currently being denied, then this is at least information about the extent of the problem. However, if the denial of the opportunity is as a consequence of bad incentives or market failures affecting retail offerings, then this mechanism will directly address those failures. Essentially introducing compensation for demand curtailment at worst will do nothing but at best will achieve significant benefits with the truth lying somewhere in between. Regardless of the outcome, it will augment the choices available to those consumers who may currently be denied that choice.

1. Introduction

Electricity prices in the National Electricity Market (NEM) continue to rise as more generation and network investment is needed to satisfy growing peak demand. This has led to questions about whether greater demand side participation in the market could help to avoid these investments and so lower electricity supply costs and prices.

It is within this context that the Australian Energy Market Commission (AEMC) has been directed by the Ministerial Council on Energy (MCE) to investigate and identify the market and regulatory arrangements needed across the electricity supply chain to facilitate efficient investment in, operation and use of demand side participation in the NEM – the Power of Choice review.

As the AEMC has described:[1]

> We consider that the objective of this review is to identify opportunities for consumers to make informed choices about the way they use electricity, and provide incentives for network operators, retailers and other parties to invest efficiently so that there is increased confidence that demand and supply side options are given equal weight in satisfying the community's demand for electricity services.

In principle, an efficient level of demand side participation should arise in the market so long as consumers face retail prices reflective of the cost of supply. However, the absence of retail prices that reflect the variability in both wholesale and network supply costs, particularly during peak periods, results in current levels of demand side participation being sub-optimal.

This means that many consumers currently use electricity during those few peak days or hours through the year when, if faced with the marginal costs involved in supplying electricity during those days, they might choose to not use so much electricity. To put it another way, demand curtailment is likely to be a cheaper substitute for peak generation and network investment for many consumers individually, as well as a more cost-effective resource for the market as a whole.

However, simply charging consumers the marginal cost of supply during each supply period is itself challenging and complex. Indeed this is one reason why retailers exist – to manage wholesale price volatility for electricity consumers. Evidence and experience from around the world suggests that retailers can and have effectively insulated the large majority of customers from exposure to super-peak prices.

To achieve the demand signalling benefits and promote efficient demand curtailment therefore requires consideration of whether alternative market design mechanisms can be used to provide the same price signalling incentives as though all consumers faced the true cost of electricity supply.

NERA Economic Consulting (NERA) has been asked by EnerNOC – a third party supplier of demand side services – to:

---

[1] See page i, AEMC, (2011), *Power of choice – giving consumers options in the way they use electricity,* Issues Paper, 15 July, Sydney.

- first, investigate whether the current NEM market design is likely to promote efficient demand side participation in the market; and
- second, to the extent that the current market design does not promote efficient demand side participation, consider how a mechanism, such as that used within the PJM market, could be applied in the NEM to promote more efficient demand side participation.

Relevantly, we do not in this brief paper examine all of the possible options for improving the signals for efficient demand side participation or consider the likely size of the potential efficiency improvement opportunity for the NEM. We leave these matters for future consideration. However, we do recommend that the AEMC provide the time for and consider carefully how best to address the barriers that result in consumers not having the opportunity to respond meaningfully to wholesale price signals.

In addition, while more efficient demand side participation is expected to deliver both wholesale market and network benefits, we have focused only on the benefits from improved wholesale market outcomes. The impediments – if any – to more efficient demand side participation as a substitute for network capacity investment is likely to require a wider consideration of the market rules, which has not been the focus of this paper.

The remainder of this paper is structured as follows:

- section 2 provides a discussion of the efficiency of demand side participation in the NEM, based on a consideration of the theoretical role of prices in markets;
- section 3 describes how a mechanism such as that used in the PJM market, could be applied to the NEM; and
- section 4 provides concluding comments and considerations.

2. The Efficiency of Demand Side Participation in the National Electricity Market

Prior to the market reforms to the electricity industry during the 1990s, the problem of determining the appropriate level of electricity generation investment to meet expected demand was a planning and administrative problem. Studies would be conducted of future electricity demand and decisions made about the appropriate mix of generation plant needed to satisfy the forecast demand, given expectations about fuel costs and the underlying load profile.

The principal purpose of the wholesale market reforms was to improve the efficiency of electricity supply by using the interaction between demand and supply within a market to create price signals for new generation investment. This decentralisation of the generation investment decision making task led to the creation of a competitive wholesale market, and disaggregation of the electricity industry into its network, retail and generation components.

In this chapter we examine whether current levels of demand side participation in the NEM are likely to be efficient, given the current market design. We begin by setting out the relevant characteristics of the NEM's design, before explaining the economic principles for efficient demand participation in a market and the likely market failures arising in the NEM.

2.1. Key characteristics of the NEM design

The wholesale electricity market in the NEM operates as a gross pool energy-only market, across six interconnected regions. In an 'energy-only market', the revenue earned by a generator is a function of the quantity of electricity it sells, not the capacity of its plant. In the absence of a separate capacity market or a capacity payment mechanism, building additional capacity is only profitable if, when the capacity is dispatched, it is dispatched at a price that exceeds its marginal cost. The absence of a capacity payment mechanism in the energy-only market model means that investment in additional generation will be inextricably linked to a generator's expectations about future spot prices and the opportunity to recover the capital investment costs of new generation.

Under the gross pool energy-only market model, registered scheduled generators[2] are required to sell all of their output on the spot market and receive the spot price for each unit of electricity sold into the pool. The spot market is operated and administered by the Australian Energy Market Operator through a centrally co-ordinated real time dispatch process.

---

[2] Base and peak load generators can be classified as:

- *Scheduled generators* – a generator will generally be classified as a scheduled generator if its capacity exceeds 30MW. A scheduled generator is required to schedule its entire output as part of AEMO's dispatch process; or
- *Non-scheduled generators* – a generator will be classified as non-scheduled if it has a capacity less than 30MW or can only offer supply on an intermittent basis. These generators are not required to be scheduled their output as part of AEMO's dispatch process.

The National Electricity Rules, which govern the operation of the market, provide a reliability standard and reliability settings, including a market price cap and cumulative price threshold[3] and market price floor. Collectively these market parameters provide the incentives and bounds for the operation of the wholesale market. The market price cap is currently $12,500/MWh, the cumulative price threshold is $187,500/MWh and the market price floor is -$1,000/MWh.

Given the potential for wholesale market prices to fluctuate, retailers and generators use financial derivatives to lock in the future price of electricity that will be supplied or purchased. A derivative is a financial instrument (eg, a swap, option or futures contract) that derives its value from the trading of rights or obligations relating to an underlying asset, in this case a specific quantity of electricity. All derivatives associated with the NEM are settled on a cash basis.

Retailers therefore manage wholesale price risks on behalf of customers by hedging wholesale market prices through these contractual arrangements. This allows retailers to offer customers retail tariffs from which customers can purchase as much electricity as required regardless of the underlying wholesale cost of electricity.

The effectiveness of retail competition continues to be examined by the Commission and we note that for a number of jurisdictions, retail competition has not been found to be effective.

2.2. Efficient operation of price signals

In any market efficient use and supply of a product or service arises at a price that is determined through the interactions of many buyers and sellers in the market. The price reflects the marginal cost of producing the good or service by the marginal supplier, given existing technologies and techniques. It also represents exactly the value received by the marginal consumer using the good or service, at the price they are willing to pay.

The characteristics of electricity markets – principally that electricity cannot be stored and that there are large daily movements in demand leading to an optimal mix of very different generation facilities – means that for supply and demand to be balanced efficiently the price in every instantaneous period should be equal to the marginal cost of supplying that use, with demand matched to the marginal value obtained from that use.

In practice, electricity markets will typically have processes to generate prices, such that it approximates the marginal value of demand over a defined period (eg, the 30 minute spot settlement price in the NEM is calculated as the average of the market clearing prices for each 5 minute period).

In an energy only market, the return to the investment in generation capacity is paid from prices exceeding short run marginal costs for sufficient periods of time so as to pay for the cost of installing the marginal generator needed to satisfy marginal peak demand.

---

[3] The cumulative price threshold (CPT) provides a mechanism for reducing the dispatch price to the administered price cap (which is currently set at $300/MWh) if the sum of the half-hourly wholesale market spot prices over a rolling seven-day period exceeds the threshold. The CPT is currently set at $187,500/MW-week.

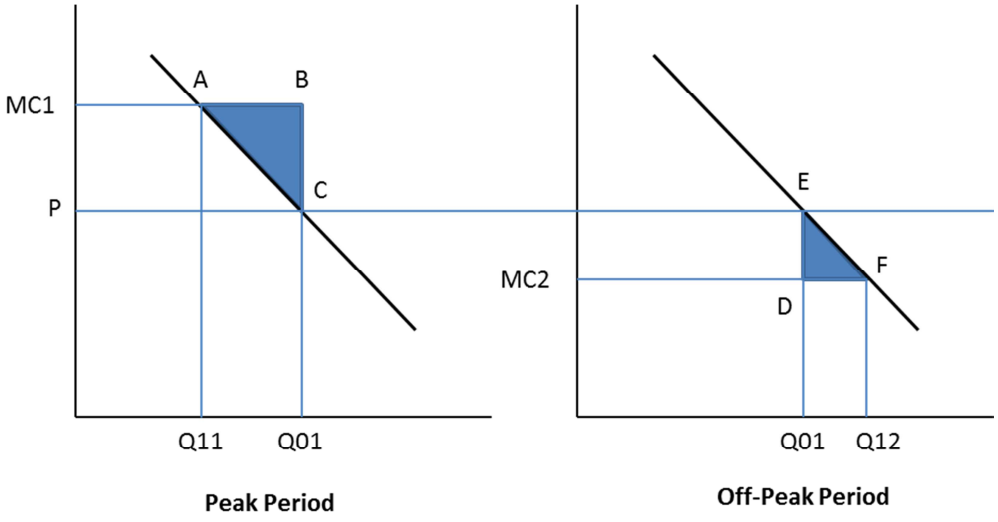
In principle, this market design should result in roughly efficient market outcomes (ie, efficient levels of demand side response and investment in generation). However, this requires all consumers to have the opportunity to face the marginal price reflective of the cost of supply at each time and location, effectively real time pricing (RTP).

However, in practice most customers face a flat tariff because retailers manage the demand volatility risk through forward contracts[4]. As a consequence retailers effectively make decisions on behalf of customers as to likely changes in load shape, absent those customers being capable of signalling the marginal value of demand through choices about how much to consume during each time period in response to the marginal price. This makes the contracting approach more akin to a generation investment planning process for capacity, rather than a market driven process revealing the value of demand to consumers compared against the cost of supply.

As a consequence of a lack of widespread RTP, there are efficiency losses in the market. These losses reflect the additional generation costs and lost value from energy use associated with consumers responding to an average price while the cost of supply is at times either above or below their averaged tariffs. In the former case, inefficiently large amounts are consumed and in the latter amounts that are inefficiently low.

Figure 2.1 provides a highly simplified example of the potential efficiency benefits of moving to retail pricing reflective of the underlying wholesale cost of supply. In this example, we assume that there are only two time periods and the wholesale cost of supply is constant throughout the two time periods – similar to the introduction of time-of-use pricing.

Figure 2.1: Welfare Effects of Time-of-Use Pricing



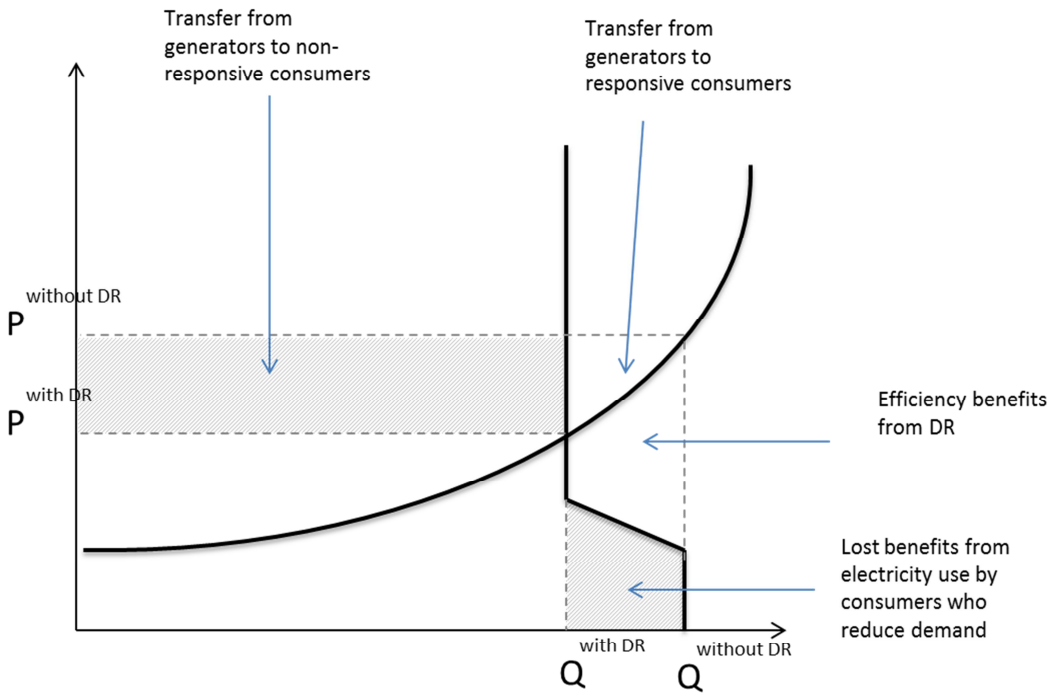
[4] Consumers have demonstrated around the world a strong preference for hedging volatile real-time prices and the Australian market rightly allows them to do so. However, there are obvious limitations in how flexible such hedging schemes need to be to suit the needs and incentives of retailers. This creates the problems that will be discussed in Section 2 below.

By charging consumers the marginal cost during peak periods, demand decreases to Q11 – a movement from C to A along the demand curve. This results in benefits from the generation cost savings, which is partly offset by lost consumer benefits from no longer consuming electricity. The shaded area ABC represents the efficiency benefits.

Similarly, charging the marginal cost of supply during off-peak periods increases demand to Q12 – a movement from E to F along the demand curve. In this example consumers are better off because of the additional value created from additional electricity use during off-peak periods which is partly offset by the additional generation costs. The shaded area DEF represents these efficiency benefits.

Relevantly, the wholesale market efficiency benefits of reducing demand arising from avoided capacity investment costs in the long run are captured by consumers that face RTP. However, all generators – at least in the short run – transfer lost revenues to all consumers through lower wholesale market prices. For those generators to remain profitable and assuming no changes to underlying generation technology costs, prices in an energy-only market would be expected to rise sufficiently to ensure that those generators earned sufficient profits to fund the initial generation investment. These transfers would therefore be short-lived – Figure 2.2.[5]

Figure 2.2: Benefits of Real Time Pricing



[5] A benefit of compensating demand response directly in the wholesale market as part of the bid stack alongside scheduled generators, as discussed later, is that these short-run transfers do not arise.

In this stylised example we assume that demand is partially responsive and so real time pricing results in quantity demand decreasing from $Q^{\text{without DR}}$ to $Q^{\text{with DR}}$. This in turn results in the peak wholesale price falling from $P^{\text{without DR}}$ to $P^{\text{with DR}}$. As anticipated, this results in efficiency benefits that are captured by the consumer responding to the price signal, which are greater than the lost benefits from the consumption of electricity. The benefits represent the resource benefits of avoiding the fuel and other costs associated with operating those peak plants and also a transfer from generators to those customers.

Overall aligning retail prices closer to wholesale marginal costs during each time period can be expected to improve the efficiency of the entire market. It follows that – at least in principle – the absence of an opportunity for real time prices to be faced by **all** customers to allow all customers the opportunity to respond to the marginal cost of supply, necessarily results in:

- over investment in generation capacity;
- under investment in demand reduction activity by consumers; and so
- efficiency losses from the current market design.

2.3. Market failures within the current design

The AEMC in its Directions Paper indicates that:[6]

> Efficient DSP does not require all consumers to face time-sensitive tariffs. If consumers are able to respond to the price signals they receive, and have easy access to information about the impacts of their decisions, then the most efficient outcomes result from consumers having the ability to choose a tariff which best suits their individual circumstances and preferences. Efficient outcomes require:
>
> - prices created in the wholesale market to reflect the cost of producing electricity in each half hour;
> - network charges to accurately reflect the cost of building additional capacity; and
> - retailers to have an incentive to offer contracts which respond to their customers' preferences.

The AEMC therefore focuses on whether there are disincentives for retailers to offer tariff products that allow consumers to respond efficiently to wholesale market prices, through selecting tariff products that either pass through those prices or provide some exposure to wholesale price volatility risks.

We believe that there are potentially significant barriers to retailers offering tariff products that provide opportunities for consumers to respond to wholesale price signals. This results in inefficient demand curtailment by customers. That said we also believe that there are additional market failures not highlighted by the AEMC, which would result in an

---

[6] See page 52, AEMC, (2012), *Power of choice – giving consumers options in the way they use electricity*, Directions Paper, 23 March, Sydney.

inefficiently low level of demand side participation, even if all customers faced time sensitive tariffs. The remainder of this section explains the basis of these opinions in greater detail.

2.3.1. Barriers to consumers having the opportunity to respond to wholesale market price signals

Current retail tariff offerings can be characterised as an unlimited quantity option for a given price, with the retailer managing the wholesale price risks through hedging. Essentially, consumers can take as much electricity as desired at any time of the day – irrespective of the underlying supply costs – for the given price, whether it is a flat tariff, time-of-use tariff or other variant commonly offered in the market.

Under the current arrangements, there are strong disincentives for retailers to offer tariffs that vary in line with underlying wholesale prices – so called real-time pricing. These include:

- retailers essentially make profits by managing wholesale price risks on behalf of consumers and so if retailers no longer managed those risks then this would have implications for the margin charged to consumers for this service;
- the costs of installing metering and associated transaction and administrative costs which affects the competitiveness of tariff offerings allowing consumers to respond to wholesale price signals;
- the lack of consumer understanding and information about wholesale risks/opportunities and so the increased complexity new tariff products would create;
- preferences by consumers to prefer flatter pricing structures, given the costs and risks that consumers would incur managing half-hourly price variations; and
- there is limited reliable information on the likely responsiveness of demand and so on the wholesale market cost savings that could be derived if consumers were given the opportunity and information to respond to wholesale market prices.

In principle, in a competitive retail market it should be possible for a retailer to offer a targeted product for less peaky customers with a lower flat tariff, so long as those customers do not suddenly become peakier themselves, thereby driving costs higher for the retailer. However, the inability of retailers to adequately manage an individual's load due to a lack of widespread adoption of smart metering, combined with the complexity and limited retail margins, means that innovative tariff products that would enhance the efficiency of the wholesale market do not deliver sufficient benefits for the retailer alone.

The experience with RTP programmes has been investigated in the US.[7] The research highlighted the relatively small uptake of participation in these programmes at that time. The causes were attributed to:

- a lack of awareness by customers of the potential savings from participation in such a programme; and

[7] Barbose, G., Goldman, C., and Neena, B., (2004), *A Survey of Utility Experience with Real Time Pricing*, Ernest Orlando Lawrence Berkeley National Laboratory, December.

- inability by many customers to adequately manage the price risks through the implementation of technology or other mechanisms.[8]

More recent work in the US has investigated barriers to price responsiveness. Around 22 per cent of survey respondents indicated that there were insufficient resources within the organisation to pay attention to hourly prices, while cost was only identified as a barrier for around 6 per cent of respondents.[9] Addressing these barriers in particular – either directly or through mechanisms that provide a similar price signal but without requiring the entity to directly pay attention to hourly prices – would therefore likely lead to greater responsiveness of demand to wholesale market price signals.

In the absence of addressing these barriers, consumers are unlikely to demand and retailers are unlikely to offer, tariff products that allow consumers to offer to reduce load even though it might improve the efficiency of the operation of the wholesale market (ie, where the benefits outweigh the underlying costs). This means that consumers will have limited opportunity to reveal preferences for demand response versus increased electricity generation. By implication this means that the current market is unlikely to provide efficient levels of demand response and optimise new generation investment against demand response activity.

2.3.2. Positive benefits to all consumers from reduced unserved energy

An individual engaged in load curtailment when faced with real time prices delivers benefits that are captured by the individual (through avoided wholesale costs) and system wide benefits from reduced unserved energy and avoided network capacity augmentation expenditures that the individual cannot capture. These positive benefits create a market failure because even if consumers faced real time prices, there would be underinvestment in demand side participation.

The NEM reliability standards require that unserved energy should not exceed 0.002 per cent of the total energy consumed in a region in a year. In practice, while uncommon, there are circumstances in the NEM when energy is not served. The economic cost of unserved energy can be substantial, and if it results in rolling blackouts there is little opportunity to discriminate between those customers that value electricity more highly during those periods than others.

Real-time pricing when faced by all customers has the effect of reducing the demand for those customers that value electricity least during periods of insufficient supply. As a consequence there is a commensurate reduction in unserved energy, which is a benefit captured by all consumers through the avoidance of blackouts.

The intuition for this result is as follows. Assume that the entire electricity market consists of two consumers – an individual residential household and a large manufacturing company.

---

[8] As we discuss later, compensating demand reductions through the wholesale market provides incentives for third parties to manage these risks on behalf of consumers, and only if it is efficient given the underlying market benefits and costs of implementation.

[9] See Table 25, KEMA, (2012), *Mandatory Hourly Pricing Program Evaluation Report*, prepared for Consolidated Edison Company of New York, 1 May, New York, page 4-12,

The value of electricity is significantly higher for the manufacturing company compared to the individual household. Now assume that both the manufacturing company and the individual household are charged a flat tariff. During peak periods if supply is insufficient to meet demand, then the household and the manufacturing company would be required to reduce demand through rolling blackouts. However, if they faced the marginal wholesale cost, as supply became constrained relative to demand prices would rise until the residential household decided to switch off. In this circumstance the manufacturing company would benefit from not having to switch off to ensure that demand is met by total supply.

Involuntary lost load is a potentially large cost to the economy and real time pricing provides the opportunity to allow load to be reduced in the order of value to consumers, thereby minimising both probability and size of lost load. The benefit of avoiding these costs is therefore potentially substantial and is a positive externality from individuals responding to prices reflective of the underlying marginal cost of supply.

To give an illustrative example, it would not be at all unusual for some consumers to value lost load at $25,000/MWh for some uses, ie double the market cap. Further, such a customer might have been paying only $100/MWh for service. So, every lost MWh carries a social loss of, on average $24,900. Even at the expected loss rate of 0.002 per cent of hours, this is a loss of 50 cents per year per customer. Once we recognise that these losses are concentrated in peakier hours, the losses are much higher. And there might have been lots of customers who would have had uses valued at more than $100, but far less than the market cap, who would have been willing to forego those losses but there is, without demand bidding, no incentive for them to do so.

The Reliability Panel has the responsibility to periodically review the market parameters including the reliability standard. This typically involves the type of assessment set out above, where the expected benefits from improved reliability is weighed against the additional generation and network costs required. This requires judgements to be made about an acceptable cost to those affected by possible blackouts and the costs to all energy users. Greater opportunities for demand curtailment from those that value demand less than others effectively would allow the reliability standard to be potentially tightened at no cost to other consumers through an associated need for network expansion.

Finally, while we have not examined in detail in this paper, there may be opportunities from a general increase in demand response to lower network capacity augmentation over time.

2.4. Summary

By allowing demand to respond to prices that reflect the marginal cost of supply at each point in time and location, the market can be expected to deliver efficient investment in supply capacity. Introducing the option for real time pricing to all consumers is therefore a first-best approach to ensuring that there are efficient levels of demand response in the electricity market.

However, if real time pricing is only introduced to a subset of consumers or indeed to no consumers is likely to result in inefficiently low levels of demand curtailment – particularly if there are barriers to retailers providing such tariff offerings to consumers.

This analysis highlights that there are 'in-principle' benefits from addressing these market failures either by introducing real-time prices to all consumers, or alternatively, developing a mechanism so that all consumers can respond to changes in prices but without the need to mandate all customers onto real-time tariffs. Indeed the focus of the next section is precisely on one mechanism that can achieve this outcome.

3. Mechanisms to Improve the Efficiency of the Wholesale Electricity Market

Allowing customers to respond efficiently to real time wholesale market prices in the absence of all customers facing real-time wholesale prices, is a particularly challenging task. The mechanism needs to provide incentives as though a customer faces real-time wholesale prices, but ideally without the price volatility risks that facing real-time wholesale prices create.

Allowing consumers to be paid directly, as opposed to needing to avoid higher wholesale prices, for reducing demand is one approach that can be used to achieve this outcome. It relies on the equivalence between charging and providing a rebate of equal value to create the same incentives as if a customer faced real-time wholesale market prices. However, it avoids the price volatility risks by allowing customers to continue to pay flat tariffs – or some other preferred tariff structure – for that portion of customers' load that is not reduced in response to the marginal benefits from reduction that could be earned.[10]

To allow consumers to be paid for demand reductions during peak periods, and to ensure that this does not lead to inefficient reduction in demand, requires a mechanism that:

- provides an objective baseline against which the amount of demand curtailed can be assessed; and
- transfers the value of demand curtailment at each point in time to the consumer engaged in demand reducing activity.

Such a mechanism has potentially further benefits by allowing third parties to aggregate demand savings and so lower the overall cost of delivering demand response activity to the market. This is not possible when the consumer's only benefit from curtailing their demand is avoiding paying a high price.

Importantly, such a mechanism should result in demand reductions or generation supply being treated equivalently – as would be the case if consumers faced real-time prices. So long as demand reduction is cheaper (in terms of reduced consumer surplus plus the cost of enabling the demand response) than capacity expansion to generation supply, then such a mechanism will promote more efficient market outcomes.

This section provides a brief overview of one mechanism that could be used to achieve these outcomes.

3.1. FERC approach to compensating demand response

In 2011 the Federal Energy Regulatory Commission (FERC) in the United States issued a final rule to allow a demand response resource to participate in an organised wholesale

[10] Relevantly, mechanisms can be designed so that consumers have the flexibility to respond or not at any given time given the particular circumstances faced. This reflects the realities that consumers might not be equally capable of responding to demand signals at all times and so would prefer an option to respond if it is beneficial at a particular point in time.

energy market.[11] The rule is based on the premise that demand response should be compensated equally to generation at the locational market price (LMP), which varies according to geographic location and time – equivalent to the NEM's regional wholesale spot prices. The effect of this rule is to ensure that demand response is paid the equivalent of the marginal wholesale market supply cost. This is equivalent to a consumer facing that price and reducing demand to avoid the incurrence of the marginal wholesale market supply cost.

In summary, the FERC decision:

- implements a net benefits test as a threshold to ensure that any reductions in demand do not result in an increase in the unit cost of electricity to remaining load because of the smaller amount of load paying wholesale costs; and
- requires demand response resources to be compensated at the LMP when dispatch of the demand resource is cost-effective as determined by the net benefits test.

Each independent system operator is considering how to implement the FERC decision in its wholesale market design. Relevantly, many US wholesale electricity markets (eg, the PJM – see Box 3.1) involve both day ahead and real time markets, plus a separate capacity market. Consideration is therefore being given to how to estimate a baseline demand against which customer demand response can be assessed for both the day ahead and real time markets.

While the focus in the US has been on compensating demand response in the context of US wholesale market designs, there is nothing in principle preventing a similar scheme being incorporated in an energy-only market like the NEM. Indeed, as explained further below there are opportunities to include compensation for demand response in the NEM dispatch and settlement system without affecting the prices and incentives of most generators and retailers, while still delivering efficiency improvements to the market as a whole.

In practice, the PJM demand response program allows:

- registered consumers – or third party aggregators on their behalf – to submit bids on both quantity of demand curtailment and price into the day ahead market;
- if demand response is 'dispatched' as part of the day ahead market then the consumer is paid the LMP for the market and receives an obligation to curtail load;
- a registered consumer to indicate an intention to curtail demand up to 1 hour ahead of real market dispatch;
- any real-time dispatch is paid the real time LMP, however, there is no obligation to curtail in the real-time market.

The baseline against which demand curtailment is assessed in the PJM is based on averages of historical information, which is based on actual customer usage data for identified periods of time, (eg, average weekday, average Saturday, average Sunday etc). In addition there are

[11] FERC, (2011), *Demand Response Compensation in Organized Wholesale Energy Markets*, Docket No. RM10-17-000, 15 March.

weather sensitivity adjustments according to the season of the proposed demand curtailment.[12]

Box 3.1: Overview of the wholesale market operation of the PJM

The PJM wholesale energy market consists of two inter-related markets – a day-ahead market (or short-term forward market); and a real time market.

The day-ahead market allows market participants to sell or purchase electricity at binding prices one day ahead of the required dispatch. This allows load – ie large electricity users or retailers on behalf of smaller consumers – to submit forward hourly demand schedules, which also provide an indication of the price sensitivity of demand. Generators can also submit generation offer curves specifying generation capacity and prices for each hour over the next 24 hour period. The PJM market operator calculates the day-ahead dispatch schedule based on the least cost combination of generation to satisfy load.

The real-time market dispatch energy based on generation offers and load for every 5 minute dispatch period, given system reliability requirements and actual network constraints. The LMP for generators that were selected for dispatch in the day-ahead market, reflect those used in the day-ahead market. For those generators that were not dispatched in the day-ahead market, there is an opportunity to rebid LMP offers for the real time market. Actual dispatch reflects the least cost combination of generation given the real time LMP offers.

3.2. Applying a demand response compensation mechanism in the NEM

A mechanism to compensate demand curtailment within the NEM energy-only market design can be, at least in principle, no different to that being adopted by US market operators in response to the FERC decision.[13] The compensation for demand response participating in markets such as PJM is essentially split between the receipt of capacity payments, and payment for demand response operation through the day-ahead and real-time markets. The NEM only has an equivalent to the real-time market.

The inclusion of such a mechanism in the NEM requires consideration to be given to:

- determining the relevant baseline against which demand curtailment would be measured, for each electricity consumer participating in the scheme;
- allowing registered demand curtailment participants to 'offer' load curtailment for consideration in the market dispatch process alongside generation offers;

[12] The methodology used to estimate customer baselines is described in section 10.4.2, PJM, (2012), *PRM Manual 11: Energy & Ancillary Services Market Operations*, Revision: 50, 3 April.

[13] There is a prominent difference between the mechanism described here and the FERC mechanism. The FERC mechanism pays the market price without requiring the load offering supply back to the market to 'buy the baseline'. While we believe that a mechanism that compensates at LMP without rebating the retail rate back to the retailer is superior in terms of efficiency, we propose the mechanism here because it will be much simpler to implement, while clearly being better than the status quo in the NEM. It also avoids the need to introduce an uplift scheme.

- designing mechanisms to ensure that any demand curtailment accepted and scheduled is actually delivered;
- designing mechanisms to ensure that consumers engaged in demand curtailment pay retail tariffs as though demand curtailment had not occurred, ie based on baseline demand rather than actual demand;
- retailers settling in the wholesale market based on baseline demand, rather than actual demand; and
- ensuring that delivered demand curtailment is paid the NEM settlement price for demand curtailed within the relevant 30 minute period, and so is treated equivalently to generation dispatch.

Under this mechanism the settlement price will be determined by comparing baseline demand against the supply offers inclusive of demand response offers. The actual generation dispatched would be reflective of the baseline demand less the dispatched demand curtailment.

The implications of this mechanism for the operation of the market is as follows:

- during a consumer's demand response dispatch, their retailer would bill them on the basis of an AEMO-provided baseline rather than for their actual electricity use;
- where demand response is a cheaper cost option to generation during the relevant time period then less generation will be dispatched; and
- the efficiency of the wholesale market will be improved as reduced generation dispatch and so revenue to generation results in delaying the need for new generation investment over time.

Importantly, this mechanism resolves the 'missing money' problem because the consumer continues to pay retail tariffs and so the retailer will settle in the wholesale market on the basis of the baseline load. In this circumstance, the consumer receives the benefit of the time-specific wholesale market price less any cost of implemented demand response, less any lost benefits from the use of electricity during that period. This amount should be higher than the payment of the retail electricity cost, reflective of the difference between the marginal cost of supply and the actual retail tariff during that time period.

3.3. Further considerations

Introducing such a mechanism into the wholesale market design would involve costs, and so consideration would need to be given to whether these costs are outweighed by the potential benefits. Relevantly, apart from the costs of modifying the market rules and the processes and procedures of the AEMO to facilitate the bidding of demand response, the additional costs of facilitating demand response would be borne by those consumers that would benefit from the use of such a mechanism.

While all consumers bear the costs involved with any changes to the wholesale market design, they also receive the benefits of lower capacity investment requirements as demand responds. These benefits are substantial and so would likely exceed the implementation costs incurred.

Once a demand compensation mechanism is introduced to the market, consumers would be willing to participate where the cost of curtailment during peak periods is less than the wholesale market price in those periods. The cost of curtailment will equal the sum of:

- the lost benefits from reduced electricity consumption during peak periods;
- any costs associated with third parties providing demand response services, including administrative and metering costs.

This means that demand response activity will exactly equal the efficient level as though the consumer faced the real-time price and third parties charged consumers directly for services to facilitate management of those price risks.

Finally, given the metering and associated monitoring costs required to allow demand response to be compensated in the market, it is possible that only some – potentially larger consumers – would participate in demand curtailment programmes.

3.4. Summary

The opportunity to avoid future generation capacity investments to address peak demand by allowing consumers to respond by charging electricity use decisions based on the real-time value of demand curtailment has the potential to improve the efficiency of the wholesale market in supplying electricity.

Compensating demand reductions directly through the wholesale market is one mechanism that encompasses the existing positive externalities associated with demand response within the current NEM design. It allows consumers to receive the total system wide market benefit of demand curtailment, which is greater than the retail price paid and so will result in efficient levels of demand response without subjecting customers to price volatility risks.

In addition, this approach enables third parties to compete to procure demand response from consumers. This creates strong incentives for the costs of facilitating demand reductions to be reduced over time.

Allowing demand response to be compensated directly through the wholesale market in the NEM is feasible. Consideration would need to be given to how customer baselines would be determined, as well as the practical processes and mechanisms for dispatching demand response as a substitute for dispatching generation. The US examples provide significant confidence that these matters can be addressed.

Finally, and most importantly from the perspective of the AEMC's review, allowing demand curtailment to be compensated through the wholesale market has no obvious downside at all.[14] If consumers choose to not make use of the opportunities that they are currently being denied, then this is at least information about the extent of the problem. However, if the denial of the opportunity is as a consequence of bad incentives or market failures affecting retail offerings, then this mechanism will directly address those failures. Essentially introducing compensation for demand curtailment at worst will do nothing but at best will

[14] We acknowledge that implementation will involve costs, but expect these to be relatively small against the potential benefits.

achieve significant benefits with the truth lying somewhere in between. Regardless of the outcome, it will augment the choices available to those consumers who may currently be denied that choice.

4. Conclusions

Ensuring that demand side participation in the NEM is efficient requires consideration to be given to the relative cost of delivering effective and reliable demand reductions compared against the cost of building more peak generation capacity. Ideally, the efficient level of demand response would result from the interaction of electricity suppliers and consumers responding to market prices. However, in practice the volatility of wholesale market prices and consumers' desire to manage these risks through retailers means that most consumers face relatively flat retail tariffs. As we have demonstrated, this in turn reduces the efficiency of the market as:

- too much electricity is consumed during extreme peak periods when the retail price is lower than the marginal cost of supplying peak electricity; and
- too little electricity is consumed during other periods when the cost of supplying electricity is lower than the retail price.

There are considerable barriers to consumers being given the opportunity to respond to wholesale price signals, even if it would result in lower electricity costs for the individual. A compensation mechanism for demand response funded through the wholesale market settlement process will significantly lower these barriers for consumers and so will promote more efficient levels of demand response.

In addition, there is a market failure in the electricity market as reductions in demand curtailment by a group or an individual delivers system wide benefits from reduced unserved energy. This means that, even if consumers faced real-time prices, there is likely to be sub-optimal investment in demand curtailment compared against the system wide benefits, warranting a market mechanism to improve the efficiency of the market outcomes.

Compensating demand response directly through the wholesale market dispatch and settlement system provides one means by which consumers can face the same incentives as real-time prices and allows the identified market barriers to be addressed. The experience from applying similar mechanisms in the US suggests that such mechanisms are entirely feasible.

NERA
ECONOMIC CONSULTING

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

NERA Australia Pty Ltd, ABN 34 092 959 665