Theory of efficient pricing of electricity transmission services

A REPORT PREPARED FOR THE NEW ZEALAND ELECTRICITY COMMISSION

July 2009

Theory of efficient pricing of electricity transmission services

Executive summary

Introduction

This report has been prepared by Frontier Economics (Frontier) for the New Zealand Electricity Commission (the Commission) as a contribution to the Commission's *Transmission Pricing Review: High Level Options Investigation* (the Review).

The purpose of this report is to provide the Commission with:

- The theoretical basis for the pricing of natural monopoly services, as applied to electricity transmission, including the recovery of sunk costs and the efficient signalling of new investment; and
- A review of relevant economic literature.

Objectives framework

As is noted by Green (1997), an effective transmission pricing regime should: (i) promote efficient operation of the wholesale electricity market; (ii) signal efficient investment in generation, load and transmission projects; (iii) compensate owners of existing transmission assets; (iv) be simple and transparent; and (v) be politically implementable.

The need for transmission pricing arises from the 'natural monopoly problem' associated with large scale infrastructure networks such as electricity transmission. In particular, the presence of large economies of scale in network augmentation means that pricing transmission services at marginal cost in accordance with first-best pricing principles tends not to recover sufficient revenue to fund the total costs of the network. This indicates the need for a charging framework that recovers remaining transmission costs in a manner that minimises distortions to the use of the existing network while providing efficient signals for longer-term investment decisions in load and generation projects.

Efficient use of the existing network

In a perfectly-functioning energy-only nodal market, where differences between nodal prices reflect the (efficient) cost of losses and congestion in the network, participants face incentives to make efficient operating decisions and, by implication, also face incentives to make efficient use of the existing network.

In this context, transmission pricing reduces to an exercise of least-distortionary sunk cost recovery. Required revenue in excess of loss and congestion rentals should be recovered in a manner that minimises the impact on participant operating decisions. This suggests that some form of Ramsey-Boiteux pricing, in which network charges are levied on users in accordance with each user's 'willingness to pay', is likely to be most appropriate.

Efficient investment decisions

If undistorted nodal prices can be relied upon to encourage efficient participant *operating* decisions, reliance on nodal prices would appear to be a reasonable starting point for encouraging efficient participant *investment* decisions as well. The literature suggests that in a perfectly-functioning energy-only nodal market, and under strict assumptions, including:

- no economies of scale in generation, load or transmission; and
- perfect augmentation of the network,

nodal pricing will indeed yield *both* efficient short-run (operational) and long-run (investment) signals. In practice, market imperfections both external to and internal to the network planning and investment regime can undermine the real-world relevance of this result.

Market imperfections *external* to network planning and investment

In addition to the standard assumptions (perfect competition, nodal prices fully reflective of the marginal value of energy at each node in the network, etc) the following conditions need to be met in order for nodal pricing to deliver efficient participant investment outcomes:

- no economies of scale in generation or load;
- no uncertainty on the behalf of market participants; and
- perfect expansion of the transmission network.

In practice, these conditions are unlikely to be met. However, transmission pricing has little to no role in correcting these market imperfections – they are either institutional failings or real-world realities.

Market imperfections *internal* to network planning and investment

As noted above, perfect expansion of the network is required in order for nodal pricing to deliver efficient participant investment outcomes. This requires that the network be incrementally expanded to the point where the nodal price difference between any two points on the grid equals the least-cost expansion option between those two points. In practice this does not occur due to:

- economies of scale and/or scope in transmission investment;
- inaccurate pricing of supply security/reliability; and/or
- over-cautiousness on behalf of network planners/regulators.

Each of these factors leads to a muting of nodal prices and a blunting of efficient investment signals. To the extent necessary, transmission pricing can be designed to augment or "boost" muted nodal pricing signals in an attempt to restore efficient investment incentives. In doing so, it is important that transmission prices do not distort participants' incentives to make efficient use of the existing transmission network (see above).

Other considerations

Costs that are unambiguously attributable to one party (such as connection costs) should be paid for by that party in accordance with a 'causer pays' approach to cost recovery. Standardised cost-allocation rules can help overcome inefficient co-ordination failures when multiple connecting parties seek to connect at one point on the grid.

There is theoretical and anecdotal evidence (from international experience) that suggests a broad trade-off exists between locational energy pricing versus locational transmission pricing in terms of incentivising efficient participant investment decisions. Locational energy pricing markets (such as New Zealand) have less theoretical need for, and appear less likely in practice to implement, locational transmission pricing regimes, and vice versa.

1 Introduction

1.1 Background

This report has been prepared by Frontier Economics (Frontier) for the New Zealand Electricity Commission (the Commission) as a contribution to the Commission's *Transmission Pricing Review: High Level Options Investigation* (the Review).

The purpose of this report is to provide the Commission with:

- The theoretical basis for the pricing of natural monopoly services, as applied to electricity transmission, including the recovery of sunk costs and the efficient signalling of new investment; and
- A review of relevant economic literature.

Frontier notes that this report has been prepared on the following basis:

- Other (non-efficiency-related) objectives of the transmission pricing regime, such as equity are not considered except in so far as they influence the achievement of economic efficiency; and
- Other features of the New Zealand electricity industry and markets – including the regulatory arrangements for transmission planning, investment and cost-recovery – are both:
    - Relevant to the role of transmission pricing in promoting economic efficiency; but
    - Broadly assumed to remain in their current form.

This means that this report will focus on how transmission pricing can promote economic efficiency within the context of these existing features of the New Zealand electricity market arrangements.

1.2 Structure of the report

This remainder of this report is structured as follows:

- Section 2 outlines the theoretical framework behind the efficient pricing of natural monopoly services as applied to electricity transmission;
- Section 3 discusses pricing methodologies to promote efficient use of the existing transmission network;
- Section 4 discusses pricing methodologies to promote efficient investment by market participants in generation and load projects; and
- Section 5 outlines some additional considerations pertinent to the design and implementation of a transmission pricing methodology.

2 Objectives framework

Key concepts

This section outlines the high-level principals of an effective transmission pricing methodology, and discusses the rationale for transmission pricing more generally.

As is noted by Green (1997), an effective transmission pricing regime should: (i) promote efficient operation of the wholesale electricity market; (ii) signal efficient investment in generation, load and transmission projects; (iii) compensate owners of existing transmission assets; (iv) be simple and transparent; and (v) be politically implementable.

The need for transmission pricing arises due to the 'natural monopoly problem' associated with large scale infrastructure networks such as electricity transmission. In particular, the presence of large economies of scale in network augmentation means that pricing transmission services at marginal cost in accordance with first-best pricing principles tends not to recover sufficient revenue to fund the total costs of the network. This indicates the need for a charging framework that recovers remaining costs in a manner that minimises distortions to the use of the existing network while providing efficient signals for longer-term investment decisions in load and generation projects.

2.1 Guiding principles

As noted above, this report has been prepared to provide a theoretical framework for the Commission's Review of transmission pricing. While the subject of the report is efficient pricing theory, the expression 'efficiency' has a number of dimensions that are worth exploring:

- **Productive efficiency** is concerned with minimising the quantity of inputs used to produce a given output;
- **Allocative efficiency** is concerned with allocating production inputs and outputs to their highest-valued uses, taking productive efficiency and the structure of relative prices as given; and
- **Dynamic efficiency** is concerned with promoting the allocation of resources to their highest-valued uses over time, particularly in terms of encouraging the optimal location and timing of investment.

In some cases, the requirements of various dimensions of efficiency can conflict with one another. For example, achieving allocative efficiency at a point in time may not be fully compatible with promoting dynamic efficiency.

This suggests that the development of a framework for considering efficient transmission pricing needs to have regard to the various objectives and potential trade-offs that are relevant to the inquiry.

Green[1] highlights the following six principles that he suggests should be considered when designing electricity transmission prices. These are as follows:

(i) Promote the efficient day-to-day operation of the bulk power market;

(ii) Signal locational advantages for investment in generation and demand;

(iii) Signal the need for investment in the transmission system;

(iv) Compensate the owners of existing transmission assets;

(v) Be simple and transparent; and

(vi) Be politically implementable.

Brunekreeft et al[2] adopt a similar list – ideally, the structure of network charges should encourage:

(i) the efficient short-run use of the network (dispatch order and congestion management);
(ii) efficient investment in expanding the network;
(iii) efficient signals to guide investment decisions by generation and load (where and at what scale to locate and with what choice of technology – baseload, peaking, etc);
(iv) fairness and political feasibility; and
(v) cost recovery.

For the purposes of this report, these collective principles will be taken into account when setting out the requirements for efficient transmission pricing.

2.2 The 'natural monopoly problem'

Biggar notes that[3]:

> At one level, the design of transmission pricing policies can be viewed as merely the application of principles of efficient monopoly pricing. Principles that are familiar from the theory of monopoly pricing in other sectors have direct application to the theory of transmission pricing. In particular, the value of marginal cost pricing, two-part tariffs, price discrimination and Ramsey pricing all have direct parallels in the setting of transmission charges.

Biggar goes on to note that due to the presence of large economies of scale and scope in electricity transmission, marginal cost pricing – meaning nodal pricing – may not itself recover sufficient revenue to cover the total costs of the

---

[1] Green, R. (1997). Electricity transmission pricing: an international comparison, *Utilities Policy*, 6(3), pp. 177-184, available here.

[2] Brunekreeft, G., Neuhoff, K. and Newbery, D. (2005). Electricity transmission: An overview of the current debate, *Utilities Policy*, 13(1), pp. 73-93, available here.

[3] Biggar (2009), p.7.

transmission network.[4] In other words, the value of the surplus accruing from the settlements process within a nodally-priced market is generally insufficient to recover the total costs of the transmission network. This is because, while the value of the settlements surplus should be positive (i.e. generators should face lower average nodal prices than loads) due to:

- the incidence of transmission congestion between generation-rich and load-rich nodes; and
- the fact that marginal transmission losses tend to be approximately double average losses,

it will typically not be sufficient to recover all the costs of a transmission network unless the prevalence of congestion on the network is extremely high. As discussed in more detail below, network planners tend not to allow such high levels of congestion to prevail, as it is usually indicative of unacceptably low levels of supply reliability. Part or all of transmission network costs could alternatively be funded directly by government through subsidies. However, this option is seldom employed in modern electricity markets.

To the extent that marginal cost pricing will not recover total transmission costs, the issue becomes how best to recover the remainder of transmission costs. This is fundamentally the question that transmission pricing methodology seeks to address. This report follows Biggar[5] in separating the consideration of transmission cost recovery into two parts:

- Where the objective is solely to promote efficient use of the existing transmission network; and
- Where the objective is to promote both efficient use of the existing network and to encourage efficient participant investment decisions regarding generation and load projects.

---

[4] This point is dealt with at a general level in Appendix A: Natural monopoly pricing.

[5] Biggar (2009).

3 Efficient use of the existing network

Key concepts

This section discusses the role of transmission pricing in recovering required revenue when the primary consideration is ensuring efficient use of the *existing* transmission network, putting to one side generation, load and transmission investment considerations.

In a perfectly-functioning energy-only nodal market, where differences between nodal prices reflect the (efficient) cost of losses and congestion in the network, participants face incentives to make efficient operating decisions and, by implication, face incentives to make efficient use of the existing network.

In this context, transmission pricing reduces to an exercise of least-distortionary sunk cost recovery. Required revenue in excess of loss and congestion rentals should be recovered in a manner that minimises the impact on participant operating decisions. This suggests that some form of Ramsey-Boiteux pricing, in which network charges are levied on users in accordance with each user's 'willingness to pay', is likely to be most appropriate.

3.1 Optimality of nodal prices

The efficient use of the existing transmission network requires that network users pay and receive prices that reflect the marginal cost of electricity at different points on the network.[6] This ensures that users implicitly face the marginal cost (or value) of the transmission network at any point in time and at any given location on the network when making consumption or production decisions.

The marginal cost (or value) of electricity at a given point on the network (or 'node') in the context of a bid-based, security-constrained, least-cost dispatch process is the sum of:

- The price of the final (marginal) bid or offer by a participant dispatched to meet the final increment of load at that node; and
- If the marginal bid or offer derives from a participant located at a different node, the marginal value of transmission losses incurred in transporting power from the marginal plant to the relevant node.

As a result, prices at various nodes across a network will tend to vary on the basis of:

- **Losses:** Transmission losses experienced in the conveyance of electricity between nodes; and
- **Congestion:** Congestion on transmission elements electrically situated between nodes. Congestion arises when flows on the network reach thermal

---

[6] Biggar (2009), p.7.

or security limits, thereby preventing the lowest-bidding (cost) plant being dispatched to meet load and requiring the re-dispatch of higher-priced, non-constrained plant to meet incremental demand at the relevant node.

In nodal pricing electricity markets, such as New Zealand and many north-eastern United States jurisdictions, wholesale spot prices tend to reflect the marginal cost or value of electricity at each node. So long as there is adequate competition at each node (in some cases, an inappropriate assumption), and prices are not capped below market-clearing levels, participants will have incentives to make efficient operating decisions. By implication, they will also have incentives to make efficient use of the existing network.

For example, where undistorted nodal pricing applies, generators will produce to the extent that their avoidable costs of generation are no greater than the marginal value of electricity at their location. Similarly, (dispatchable) loads will consume up to the point where their willingness to pay for electricity is at least as high as the marginal cost of electricity at their location.

The main 'flaw' with spot market pricing is the existence of non-scheduled loads, such that customers cannot express their real-time willingness to pay for electricity through the energy market.[7] Putting this issue to one side, the result of these incentives is that the transmission network will be utilised by participants in a manner that maximises the net benefits of the (existing) network to the market as a whole.

3.2 Approach to cost recovery

As a result of the salutary efficiency properties of (undistorted) nodal pricing in the energy market, there is no need for transmission prices to be tailored in such a way as to influence generator and load operating decisions if the objective is solely to promote efficient use of the existing transmission network. In fact, participants' use of the existing network will become *less* efficient if transmission prices influence participants' operating decisions away from how they would behave if exposed to nodal prices alone.

Under such conditions, transmission prices should be orientated exclusively towards recovering the costs of the transmission network not recovered through marginal cost pricing in as non-distortionary a manner as possible. That is, remaining transmission costs should be recovered in a manner that minimises their impact on participant operating decisions. In this context, some form of Ramsey-Boiteux pricing, in which network charges are levied on users in accordance with each user's 'willingness to pay',[8] is likely to be most appropriate.[9]

---

[7] Stoft, S. (2002). *Power system economics: Designing markets for electricity*, New York, US: Wiley-Interscience.

[8] We use 'willingness to pay' loosely here. Technically, Ramsey-Boiteux pricing involves setting prices based on each user's inverse of own-price elasticity of demand. Users with more inelastic demand (greater

The appeal of Ramsey-Boiteux pricing is its ability to minimise the 'deadweight loss', or inefficiency, of pricing above marginal cost for transmission services in order to recover total costs.

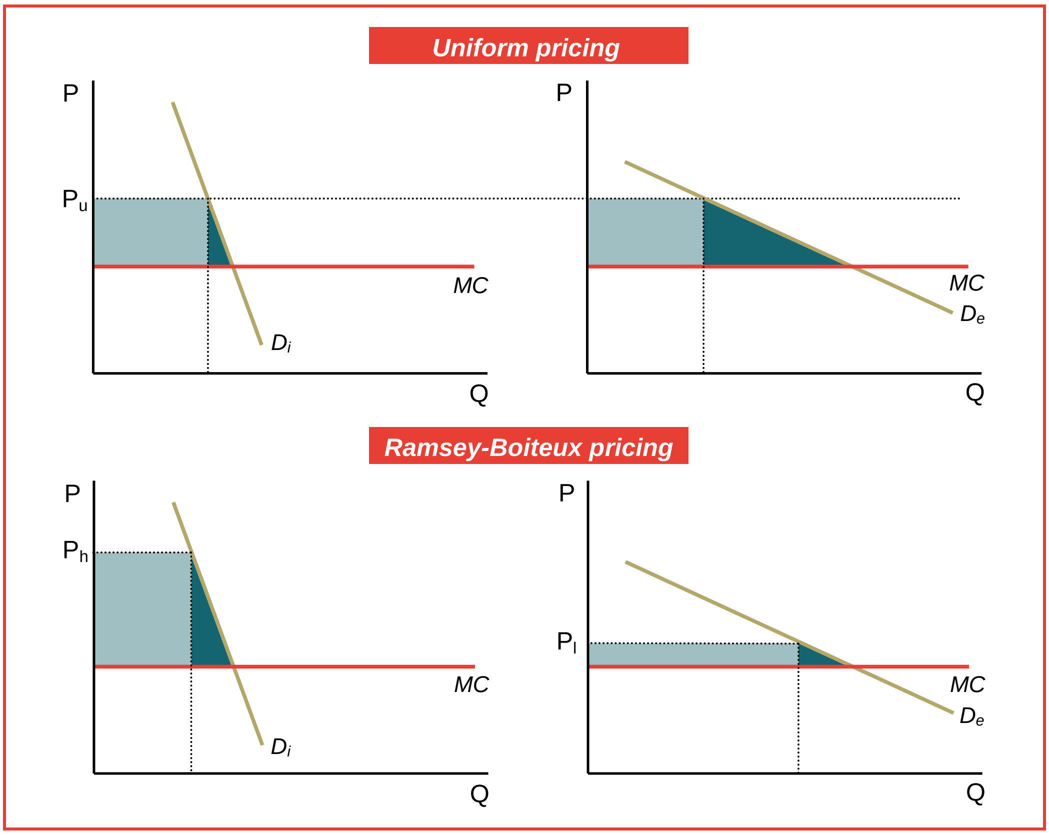

Figure 1 Uniform vs. Ramsey pricing

Source: Frontier Economics

The example outlined in Figure 1 above is taken from Braeutigam[10] and compares the welfare impacts of Ramsay-Boiteux pricing with uniform pricing to all consumers. Under both pricing regimes, a fixed amount of revenue must be raised (the sum of the aqua rectangles).

Under uniform pricing, a single price (P_u) is charged to both users with inelastic (left) and elastic (right) demand. This results in the sufficient amount of revenue being raised (the sum of the aqua rectangles), but also imposes a deadweight loss on society equal to the sum of the dark blue triangles. Under Ramsey-Boiteux

---

'willingness to pay') have demand that is *less* responsive to price changes. When faced with higher prices, these users change their consumption choices by *less* than those with elastic demand.

[9] See, for example: Braeutigam (1989); Biggar (2009); Brunekreeft et al (2005); and Laffont, J.J. and Tirole, J. (1994). Access pricing and competition, *European Economic Review*, 38(10), pp. 1673-1710, available here.

[10] Braeutigam, R. R . *Optimal Policies for Natural Monopolies*, in: Schmalensee, R. and Willig, R. (1989). *Handbook of Industrial Organization: Volume II*, 5th edition, Amsterdam, Netherlands: Elsevier, pp. 1320-1323.

pricing, the same amount of revenue is raised in aggregate. However, the user with inelastic demand contributes a relatively larger share (and faces a higher price, P_h) than the user with elastic demand (who faces price P_l). As can be seen, price discrimination of this form results in a lower deadweight loss to society.

The implementation of Ramsey-Boiteux pricing requires significant levels of information regarding network users' own-price elasticity of demand with respect to network services. Such information tends to be difficult to estimate accurately, but various proxies may be available. However, it is unlikely to be possible to recover remaining transmission costs in a manner that avoids influencing participants' operating decisions *at all*. Thus, some degree of inefficiency resulting from the recovery of remaining transmission costs is likely to be unavoidable.

4 Efficient investment decisions

Key concepts

This section discusses the requirements for nodal pricing to deliver both efficient short-run (operational) and long-run (investment) decisions. Market failures both internal and external to the network planning and investment regime are considered, as is the role of transmission pricing in restoring potentially muted investment signals to efficient levels.

In a perfectly-functioning energy-only nodal market, and under strict assumptions (namely: (i) no economies of scale in generation, load or transmission; and (ii) perfect augmentation of the network), nodal pricing will yield both efficient short-run (operational) and long-run (investment) signals. In practice, market imperfections both external and internal to the network planning and investment regime can violate this result.

Imperfections *external* to network planning and investment – in addition to standard assumptions (perfect competition, nodal prices fully reflecting the marginal value of energy at each node, etc) the following conditions need to be met in order for nodal pricing to deliver efficient participant investment outcomes: (i) no economies of scale in generation or load; (ii) no uncertainty; and (iii) perfect expansion of the network. In practice, these conditions are unlikely to be met. However, transmission pricing has little to no role in correcting these imperfections – they are either institutional failings or real-world realities.

Imperfections *internal* to network planning and investment – perfect expansion of the network is required in order for nodal pricing to deliver efficient participant investment outcomes. This requires that the network be incrementally expanded to the point where the nodal price difference between any two points on the grid equals the least-cost expansion option between those two points. In practice this does not occur due to: (i) economies of scale/scope in transmission investment; (ii) inaccurate pricing of supply security/reliability; and/or (iii) over-cautiousness on behalf of network planners/regulators.

Each of these factors leads to a muting of nodal prices and a blunting of efficient investment signals. To the extent necessary, transmission pricing can be designed to augment or “boost” these muted signals in an attempt to restore efficient investment incentives. In doing so, it is important that transmission prices do not distort participants’ incentives to make efficient use of the existing transmission network.

4.1 Introduction

The previous section discussed the appropriate characteristics of transmission prices where the objective was solely to promote efficient use of the existing transmission network. The next issue to consider is how transmission prices should be developed to promote efficient participant investment decisions (i.e. dynamic efficiency).

This encompasses the following aspects of participant decisions:

- **Type of new investment** – how to encourage the appropriate technology of generation plant and load. For example, how could the transmission pricing regime encourage the right mix of peaking versus baseload generation?
- **Timing of new investment** – how to encourage the appropriate timing of generation plant and load. For example, how could the transmission pricing regime encourage baseload plant to be commissioned at the right time?
- **Location of new investment** – how to encourage the appropriate location of generation plant and load. For example, how could the transmission pricing regime encourage generation to be commissioned in the right area?

4.2 Efficient participant investment: theory

4.2.1 The role of nodal prices

If undistorted nodal prices can be relied upon to encourage efficient participant *operating* decisions, reliance on nodal prices would appear to be a reasonable starting point for encouraging efficient participant *investment* decisions as well. In this context, it is relevant to recall the rationale for the design of 'energy-only' markets, such as those implemented in New Zealand, Australia and elsewhere. Energy-only markets seek to encourage the optimal type and timing of new generation through marginal cost pricing alone – they do not incorporate separate capacity markets to assist generation investors to recover their fixed costs.

Stoft shows that in a bid-based energy-only market, generators can (ultimately) recover their total costs even if they always bid at their short-run marginal cost of production (SRMC).[11] He even refers to the view that marginal cost pricing will not recover fixed costs as a 'fallacy'.[12] Furthermore, Stoft argues that marginal cost pricing in the energy market will also tend to encourage the appropriate timing and type of new generation, as investors make predictions regarding the future path of spot prices and respond accordingly.

By extension, it seems logical to posit that an undistorted nodal pricing market will also encourage the appropriate *location* of new generation and load investment. Rational investors will respond not only to the expected path of spot prices at a particular node, but to their expectations of nodal prices across the market.

---

[11] Stoft (2002), pp.121-132.

[12] Stoft (2002), p.121.

4.2.2 Conditions necessary for efficient short- and long-run signals

Biggar notes that under certain strict conditions, nodal pricing will yield both efficient short-run (operational) and long-run (investment) signals. These conditions are:[13]

- No economies of scale and scope in generation and transmission;
- Effective competition between generators at all locations in the network; and
- Transmission is augmented continuously so that the average congestion cost on a transmission constraint is equal to the marginal cost of augmenting the constraint.

Biggar goes on to say:[14]

> In the absence of economies of scale and scope, full nodal pricing, when coupled with this simple rule for efficient transmission investment, will ensure the fully-efficient electricity market outcome. That is, full nodal pricing will ensure both efficient short-run operational decisions, and efficient long-run investment/location decisions in both transmission and generation. Full nodal pricing ensures that generators continually face the short-run marginal cost (SRMC) of use of the transmission network, while the transmission augmentation rule ensures that the transmission network is augmented to the point when the long-run marginal cost (LRMC) of transmission expansion is equal to the average SRMC arising from generator re-dispatch. No further investment/location signals are required.

Similarly, Brunekreeft et al states that:[15]

> The question [of how well nodal pricing works in giving investment signals to new load and generation] is closely related to the question whether LMPs recover all the costs of the network. If the LMPs recover all network costs then the LMPs unambiguously set the efficient investment signals for generation and load. In the long run, with optimal investment, the difference between LMPs would reflect marginal network expansion costs.

However, both Brunekreeft et al and Biggar note that in reality certain market failures result in nodal markets **not** sending appropriate investment signals to market participants.

The following two sections discuss the nature and implications of the market failures that arise:

- Outside the network planning and investment regime (section 4.3); and
- Within the network planning and investment regime (section 4.4).

---

[13] Biggar (2009), p.14.

[14] Ibid.

[15] Brunekreeft, et al (2005), p.75.

Section 4.5 discusses how transmission charges should ideally be structured given the objective(s) such charges are designed to achieve.

4.3 Participant investment and perfectly efficient network expansion

In this section we consider the implications for the efficiency properties of nodal pricing due to imperfections arising outside the transmission sector.

To do this, we initially retain the assumption necessary to ensure that the network is efficiently expanded through time – namely that the transmission network is exogenously augmented continuously, so that the average congestion cost of a transmission constraint is equal to the marginal cost of augmenting that constraint. This assumption is further explained below and is ultimately relaxed in section 4.4.

4.3.1 The assumption of perfect network expansion

The assumption of perfect network expansion implies that incremental network investment can and does occur in such a way that the price difference between any two nodes on the network is always equal to the cost of the least-cost expansion option between those two nodes. This, in turn, requires that transmission investment is infinitely divisible and exhibits no economies of scale or scope. While unrealistic, these conditions ensure that the pattern of nodal prices in a market will not be distorted by any imperfections in the transmission sector.

4.3.2 Pre-conditions for efficient participant investment

In order for nodal prices to provide efficient investment signals, the same pre-conditions apply as those required for efficient use of the network. These are:

- The market must be sufficiently competitive so as to preclude the ability of participants to make offers or bids into the market that deviate from their marginal costs of generation or marginal value of load; and
- The market must be allowed to fully reflect the marginal value of energy at each node – that is, imposed price caps must not artificially restrict nodal prices to levels below the value that participants place on an increment of energy.

However, in addition to these requirements, efficient participant investment also requires:

- **No economies of scale in generation or load**. If generation or load exhibit ‘lumpiness’, participants’ investment signals can be distorted. For example, if a generation investment is of such a size (in MW) that its location at a particular node dramatically lowers prices at that node, prospective generation investors can face incentives to under-invest, inappropriately locate and/or delay investment altogether;[16] and

- **No uncertainty or lack of perfect foresight**. The volatility and unpredictability of nodal prices can result in participants making inefficient investment decisions due to errors in forecasting future prices over the time horizon of the investment. This problem is exacerbated by the lumpiness of transmission investment (discussed below) which can result in nodal prices deviating from their long-run equilibrium levels.[17]

Nevertheless, if all of these assumptions hold, then, as noted by Brunekreeft et al:[18]

> The standard theories of competitive pricing imply that a complete set of prices (for each location at each time and date) will give efficient investment signals for generation (and load)… Investors will make their location and plant choices based on the present value of selling at current and future nodal prices.

4.3.3 Appropriateness of transmission pricing as a tool for ensuring efficient participant investment

To the extent that the assumptions set out above do not hold, reliance on nodal prices will fail to ensure efficient investment in generation and load projects. However, the key question for present purposes is the scope for a particular transmission pricing methodology to *compensate* for these market failures.

In our view, the ability of transmission pricing methodology to compensate for any of the above potential forms of market failure is questionable. For example, if generators at certain locations have the ability to push up prices above their costs, this suggests that transmission prices for service to those nodes be increased to deter over-investment in new generation at those nodes. The greater the exercise of market power, the greater that transmission prices would need to be to discourage new investment at that node. This potentially sets up perverse incentives as it would effectively reward a generator for exercising market power by imposing higher and higher barriers to new investment at that node the more power it exercises.

An even more problematic example is where market prices are capped at levels below consumers' true valuation of energy. In these circumstances, it is not clear how transmission pricing methodology could be used to address this distortion.

Another issue raised above is the presence of uncertainty and the 'bounded rationality' of investors, such that the prediction of future price paths is

---

[16] Brunekreeft et al (2005), p.76.

[17] Ibid.

[18] Ibid.

necessarily uncertain and incomplete. Once again, it is not clear how one might go about designing a tariff structure to overcome these imperfections.

4.4 Participant investment and imperfectly efficient network expansion

As was noted in the previous section, even where transmission augmentation occurs efficiently, nodal prices may not ensure efficient participant investment decisions.

This section takes the next step and considers the efficiency properties of nodal prices and the potential role of transmission pricing methodology in circumstances where transmission expansion does not follow a perfectively efficient path. This may occur – even in markets where network planning is the responsibility of centralised planning bodies, such as New Zealand and Australia[19] – for a number of reasons, including:

- Economies of scale or scope in transmission investment (as distinct from economies of scale in generation and/or load as discussed above);
- Inaccurate pricing of supply security/reliability; and/or
- Over-caution on the behalf of network planners/regulators;

As further explained below, these distortions tend to result in over-building of the network compared to a perfectly efficient network expansion path, and thus the muting of nodal price differences. This, in turn, will tend to undermine the efficiency properties of the investment signals provided by nodal prices. Under these conditions, transmission pricing methodology may have a valuable role to play in encouraging efficient participant investment decision-making.

4.4.1 Imperfectly efficient network expansion

As noted above, centrally-planned network investment may lead to imperfectly efficient network expansion through time for several reasons. These are discussed further below.

Economies of scale in transmission

The technical characteristics of transmission networks are such that economies of scale in network augmentation exist over particular ranges of capacity upgrades. This means that it is rare for 'marginal' augmentations to be undertaken in a manner that preserves the equivalence between nodal price

---

[19] Core grid investment in New Zealand is required to satisfy the Grid Investment Test, while shared network investment in Australia is required to satisfy the Regulatory Test. Network investment may be <u>even less efficient</u> if left to private investors, and therefore, the issue of 'merchant transmission investment' is not considered further in this paper.

differences and the costs of the least-cost augmentation available between the relevant two nodes. In fact, as noted by Biggar and others,[20] the presence of economies of scale means that the cost of overbuilding the network beyond what is strictly efficient tends to be relatively small. On the other hand, the benefits of overbuilding can be quite large in terms of additional reliability (see also below on the 'over caution' of network planners).

It is important to note that overbuilding is not a result of irrationality on the part of network planners *per se* (although it might be if the costs, or more likely the benefits, of overbuilding are inappropriately considered). As noted by Perez-Arriaga et al,[21] power system expansion in the real world is a discrete rather than a continuous decision variable. This almost necessarily implies that the network planner cannot augment the network to the theoretically perfectly efficient size. It is generally the case that a decision must be made as to whether under- or over-invest relative to the perfectly efficient augmentation size and this often (but not always) results in 'over-investment'. When faced with a set of discrete augmentation options, choosing an option that represents such 'over-investment' may be perfectly efficient *within the context of the available augmentation options*. Thus, while nodal prices resulting from such 'over-investment' in the network may not be theoretically perfect, they may still provide reasonable investment signals for participants. More profound distortions to nodal prices caused by network planning decisions are discussed immediately below.

Inaccurate pricing of supply security

In some jurisdictions, networks are augmented according to a centralised process using either a cost-effectiveness or cost-benefit framework. The intent of these frameworks is to minimise the cost, or maximise the net benefit, to the market of augmenting the network.

The cost-effectiveness approach is generally adopted in conjunction with deterministic reliability standards, such as 'N-1' redundancy. A cost-effectiveness approach ignores the potential benefits of network augmentations falling short of or going beyond the required standard. As a consequence, augmentations targeted at meeting those standards are assessed purely on the basis of which augmentation (or alternative) option is least-cost. This means that to the extent that the costs of augmentations required to meet deterministic standards exceed the benefits of maintaining those standards, the application of deterministic

---

[20] Biggar (2009), p.16.

[21] Perez-Arriaga, I.J., Rubio, F.J., Puerta, J.F., Arceluz, J. and Marin, J. (1995). Marginal pricing of transmission services: An analysis of cost recovery, *IEEE transactions on Power Systems*, 10(1), pp. 546-553, available here.

planning standards will result in the network being overbuilt compared to planning purely on the basis of maximising net economic benefits.[22,]

A somewhat different approach is adopted in New Zealand, where deterministic reliability standards currently apply, but a cost-benefit framework is used to determine the appropriate augmentation (or alternative) option for satisfying the standard.[23] This means that reliability and other benefits are quantified and incorporated into the analysis, thereby allowing the selection of a more efficient option than yielded by a pure cost-effectiveness approach. However, while an improvement, this approach does not *guarantee* that all transmission augmentations undertaken will actually be net beneficial. Having said that, a paper by Gleadow, Todd and Smith found that a number of recent transmission investments undertaken to meet deterministic standards were in fact consistent with the maximisation of net market benefits.[24]

The likelihood of network over-building is even less likely under a regime of 'probabilistic' reliability standards, where both the likely costs and benefits of an augmentation are considered and only augmentations that produce both positive and maximal net market benefits can proceed. However, inefficient over- or under-building can occur if the assumed value of supply reliability to customers is incorrectly calibrated. In such cases, whether the network is over- or under-built will depend on whether customer valuations of reliability are over- or under-estimated, respectively.

Over-caution of network planners

A final reason as to why efficient network expansion through time may not occur is the risk aversion of network planners. As noted by Biggar[25] and Brunekreeft et al, network planners tend to explicitly or implicitly 'err on the side of caution' when it comes to planning and augmenting the network. In many cases, this is justifiable given the uncertainty involved in estimating power flows, future generation and demand growth and location, and other stochastic variables involved in the augmentation decision-making process. There is also a commonly perceived 'asymmetry in costs' between network over- and under-investment, given the unpopularity of load curtailment during times of supply scarcity.

---

[22] Biggar (2009), p.15; Brunekreeft et al (2005), p.78.

[23] *Electricity Governance Rules, Schedule F4 – Grid Investment Test*, clause 4.1.1.

[24] Gleadow, J.C., S. Todd and B.C. Smith (2009). *Development and use of advanced grid reliability standards in New Zealand*, paper presented at the 2009 Electricity Engineers Association Conference, Christchurch 19-20 June 2009.

[25] Biggar (2009), p.15; Brunekreeft et al (2005), p.78.

4.4.2 Role of transmission pricing in augmenting muted signals

All of the above reasons as to why transmission networks might not be augmented in a perfectly efficient manner lead to a tendency for networks to be over-built. An over-built network, in turn, produces a systemic muting of nodal price divergences and thus a distortion to participants' investment decisions.

As was discussed in section 4.3.3, the key question for present purposes is the scope for transmission pricing methodology to drive efficient investment decisions by compensating for distorted nodal prices. This suggests that transmission pricing methodology should seek to *augment or amplify the signals emanating from nodal prices* to the extent that nodal prices are influenced by transmission investment proceeding in advance of when it is efficient.

One implication of this approach is that transmission charges to generators should be highest for those generators that benefit most from the network being over-built (i.e. through nodal prices being higher than would otherwise be the case). At the same time, network charges to loads in the same areas should be relatively low, as those loads are effectively penalised by higher nodal prices caused by network overbuilding. On the other hand, charges to generators should be relatively low (or negative) in areas where generators are worse off due to network overbuilding (e.g. generators located in load-rich areas, who experience lower nodal prices than would otherwise be the case). Likewise, transmission charges to loads in these areas should be relatively high.

Another implication of this approach is that transmission charges to different participants *at any given location* should be different depending on the timing of their consumption or production.

Biggar[26] makes this point with the aid of a simple example involving a base-load and peaking generator located at the same node. Both generators are connected to a load node via an (overbuilt) network line that never reaches the stage of becoming congested. We are then asked to assume that in the absence of network over-building, this line would be congested during peak-load times only (say, for one hour a day) and that the peaking plant only generates during this one-hour peak period while the base-load plant generates continuously.

In this stylised example, the generator that only produces at times congestion would be expected in the absence of over-building (in this case, the peaking generator) should face higher transmission prices than the generator that produces relatively less at peak times (in this case, the base-load generator). This is necessary to replicate the signals that would emanate from a perfectly-functioning energy-only nodal market.

---

[26] Biggar (2009), p.20.

Thus, Biggar contends that in order to restore the efficient investment signals that are likely to have been muted due to network overbuild, even different parties *at the same node* may need to face different charges, in addition to facing different charges to those applying at other nodes.

In summary, this suggests that in a market with an overbuilt network, an efficient transmission pricing regime may need to impose relatively high charges on:

- new (or expanded) loads in areas of the network *and* at times during the day and year when drawing power from the network is expected to contribute to the case for future network augmentation; and
- new (or expanded) generators in areas of the network *and* at times during the day and year when injecting electricity is expected to contribute to the case for future network augmentation.

Clearly, the informational and predictive requirements of setting such charges are considerable. These difficulties need to be weighed up against the benefits of imposing such differentiated transmission charges, which in turn will depend on the extent to which the transmission network is overbuilt by comparison to strict economic efficiency criteria.

Biggar also makes a number of other points on the formulation of transmission charges that are relevant to the New Zealand context:[27]

- First, the impact of charges depends on the spatial differentiation in the charges, rather than their absolute level. It is the spatial differentiation that gives rise to the locational signals, not the amount recovered from the charge;
- Second, and following from the first point, the net allocation of costs to generators versus loads is essentially arbitrary. This does not mean that it is not necessary to impose charges on both generators and loads. Rather, so long as the extent of locational differentiation of charges is sufficient, the net amount of transmission costs recovered from generators *vis-à-vis* loads is not important for economic efficiency. However, this proposition is based on the assumption that the wholesale market is reasonably competitive, such that consumers will ultimately pay for all transmission costs in the long run; and
- Third, transmission charges should be as predictable and stable as possible to enable investors to make robust decisions. Unlike nodal price divergences, which in principle should be able to be hedged through financial transmission rights or nodal forward markets, changes in fixed transmission charges cannot as readily be hedged. Investors should be able to agree long-term transmission charges in advance of making investment decisions that are invariant to other network users' entry and exit decisions.

---

[27] Biggar (2009), pp.23-24.

4.5 Structure of 'fixed' transmission prices

If transmission charges are to be imposed in a way that diverges from least-distortionary sunk cost recovery, it is important that transmission charges are not based on usage of the transmission network, in terms of MWh injected or withdrawn from the grid.[28] Usage-based charges operate as a tax on usage, deterring the utilisation of sunk assets. Dynamic efficiency requires that charges influence participants' generation and load investment decisions but minimise their impact on operational decisions, such as electricity consumption and generator bidding/dispatch. Better options could include:

- Charges based on the rated capacity of the relevant generation or load facility; or
- Charges based on independent or coincident peak demand or injections, with the peaks determined on a basis unlikely to interfere with day-to-day operational decisions.

The second of these options has the advantage that it focuses the allocation of charges on loads and generators based on the times at which loads and generators are likely to experience more favourable nodal prices (i.e. lower for loads and higher for generators) due to premature network augmentation, as discussed above.

---

[28] Biggar p.17.

5 Other considerations

Key concepts

This section discusses two additional considerations when designing an appropriate transmission pricing methodology – (i) the treatment and classification of connection costs and (ii) the role of transmission pricing in non-LMP markets.

Costs that are unambiguously attributable to one party (such as connection costs) should be paid for by that party in accordance with a 'causer pays' approach to cost recovery. Standardised cost-allocation rules can help overcome inefficient co-ordination failures when multiple connecting parties seek to connect at one point on the grid.

There is theoretical and anecdotal evidence (from international experience) that suggests a broad trade-off exists between locational energy pricing versus locational transmission pricing in terms of incentivising efficient participant investment decisions. Locational energy pricing markets have less theoretical need for, and appear less likely in practice to implement, locational transmission pricing regimes, and *vice versa*.

5.1 Costs of connection

The discussion in section 4 refers to pricing for the recovery of transmission costs that are incurred as a result of a centralised planning process orientated towards maximising net economic benefit or meeting reliability standards at least cost.

However, certain transmission costs may be incurred due to a specific request by an actual or prospective network user. For example, the costs of connecting a new load or generation participant to the existing shared network could fall within this category.[29]

In principle, such costs should be recovered directly from the party making the request.[30] But this may give rise to complications where the party making the request cannot prevent a later-connecting party from utilising the connection assets paid for by the first party. In such cases, it may be appropriate to implement a cost-sharing methodology to avoid 'free-riding' by the subsequent connecting parties on the investment made, and costs incurred, by the first connecting party.

Free-riding may harm efficiency where it leads to a lower provision or consumption of a good or service compared to where there is no free-riding. Inefficient delays could occur if each prospective connecting party finds it

---

[29] This category of costs may also reasonably include costs that are not requested by the connecting party but are unavoidable if that party's request for connection is to be fulfilled.

[30] See, for example, Biggar (2009) and Brunekreeft et al (2005).

worthwhile to delay connecting if they think the other party will connect (but not otherwise) and vice versa. This could create a type of 'Prisoners' Dilemma' amongst intending connecting parties.

Given the scope for negotiation between the parties, it is possible that the parties could agree to an appropriate cost allocation and avoid inefficient delays. On the other hand, a mandated cost allocation could potentially avoid the transactions costs of negotiation and the risks of stalemate.

The appropriate cost allocation would ideally be based on what the parties would have agreed if they were to enter into a joint venture at the outset, prior to the upgrade being made. In such a joint venture, the parties would agree on a cost allocation designed to maximise the total economic surplus of the venture. This would imply each should connect at the socially optimal date. The cost allocation methodology should therefore be designed to encourage connections as close as possible to the same socially optimal dates. A cost allocation rule based on relative profitability would be one way of securing this outcome.[31]

5.2 Role of transmission pricing in non-nodal markets

Both the theoretical literature (and to a lesser extent the practical experience with transmission pricing internationally) indicates that there is somewhat of a trade-off between (i) differentiated energy prices; and (ii) differentiated transmission prices.

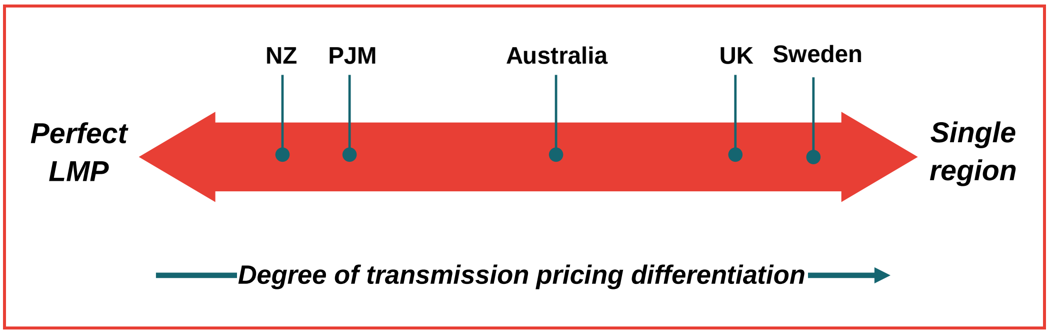

Figure 2 Locational energy vs. transmission pricing

Source: Frontier Economics

At one end of the spectrum is a perfectly-functioning[32] full nodal market, whereby each point on the network has a unique price that perfectly reflects the marginal value of electricity at that node. While in practice no such market exists due to the presence of various market failures (see discussions above), the New Zealand market design reflects such a philosophy.

---

[31] See Gans, J.S. and P.L. Williams, 'Efficient Investment Pricing Rules and Access Regulation', *Australian Business Law Review*, Vol 27, 1999, pp.267-79.

[32] No economies of scale, uncertainty or market power.

At the other end of the spectrum is a single-price (i.e. single region) market in which no differential *energy* pricing exists. In such a market, policy-makers often implement a system of differentiated (spatial, time) transmission charges. The British and Swedish markets most closely reflect this approach. Frontier has prepared a separate paper for the Commission outlining the different approaches adopted by various markets around the world to energy market pricing and transmission pricing.

The representation of the degree of differentiation in energy versus transmission pricing as a spectrum reinforces the point that these two systems of prices are, in theory, substitutable with respect to their role in sending appropriate investment signals to participants. This trade-off is acknowledged by Brunekreeft et al:[33]

> … in the absence of [locational marginal pricing], there is a strong case for a locational element to grid charges, and these should be computed to guide location decisions to minimize the present discounted cost of all [generation] and [transmission] investments required to maintain reliability and security standards.

The extent to which differentiation in transmission pricing may be desirable in the presence of market failures (discussed above) thus largely depends on:

- the level of differentiation (i.e. signalling) in energy market prices (more likely appropriate in a single-region/zonal market than a full nodal market); and
- the severity of market failure (more likely appropriate in cases where market failure(s) are severe, such that nodal price signals (to the extent they exist) are significantly distorted).

---

[33] Brunekreeft et al (2005), p.79.

Appendix A: Natural monopoly pricing

Braeutigam (1989)[34] demonstrates how marginal cost pricing *necessarily* implies the under-recovery of total costs – specifically, a natural monopolist engaging in marginal cost pricing will fail to recover its fixed costs. This can be seen within a simple algebraic framework, whereby a natural monopolist is assumed to possess an affine cost function of the form $C = Q\cdot c + F$, with Q representing output, c marginal cost and F fixed costs. From the textbook definition of profit, namely $\pi = Q\cdot(P-c)-F$, it is evident that marginal cost pricing necessarily implies profits for the natural monopolist of $-F$, and thus the under-recovery of fixed costs[35].

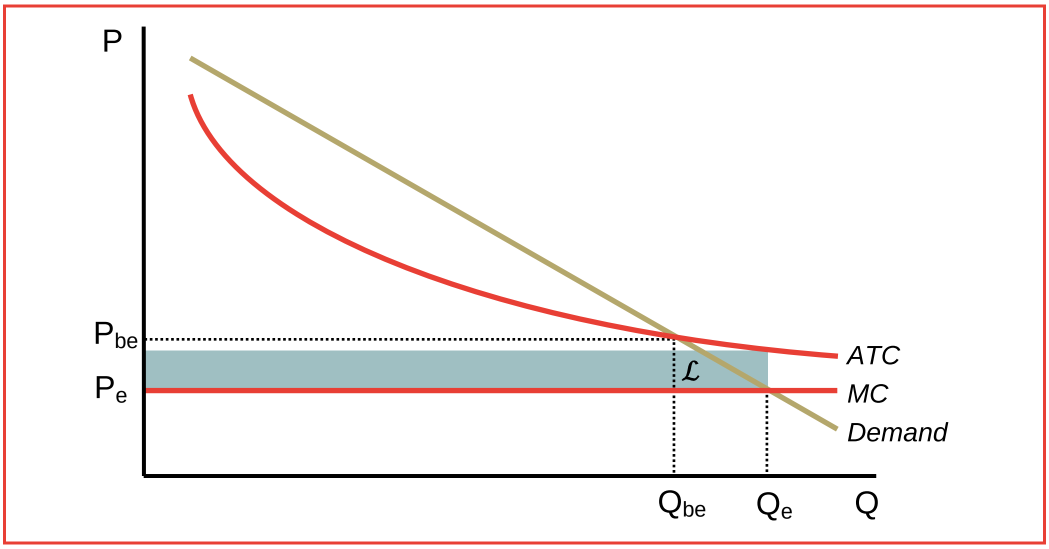

Figure 3 Efficient pricing of natural monopoly services

Source: Frontier Economics

This canonical result can be demonstrated diagrammatically as in Figure 3 above. The welfare-maximising (read 'efficient') level of output is Q_e, where the natural monopolist charges a price equal to its (assumed constant) marginal cost. At this quantity-price choice, the monopolist's profit is $\pi = Q\cdot(P-ATC) < 0$, and thus the monopolist makes a loss equal to the aqua rectangle – this area represents the monopolists fixed costs as outlined above.

The breakeven quantity-price choice of (Q_{be}, P_{be}) represents the point at which the natural monopolist just covers its total costs and earns zero profit. The value $P_{be}-c$ represents the minimum mark-up over marginal cost necessary for the monopolist to breakeven. Note that charging a price P_{be} results in society incurring a deadweight loss equal to the triangular area $\mathcal{L}$.

---

[34] Braeutigam (1989) in Schmalensee & Willig (1989).

[35] Braeutigam (1989) in Schmalensee & Willig (1989), pp. 1300-1301.

FRONTIER ECONOMICS AUSTRALIA
BRISBANE | MELBOURNE | SYDNEY
Frontier Economics Pty Ltd 395 Collins Street Melbourne Victoria 3000
Tel. +61 (0)3 9620 4488 Fax. +61 (0)3 9620 4499 www.frontier-economics.com