

Options for Electricity Transmission Regulation in Australia^{*}

by

Joshua S. Gans and Stephen P. King^{**}
University of Melbourne

10 September, 1999

The pricing of access to electricity transmission networks in Australia is currently under review. Several options have been proposed including those based on nodal pricing and the assignment of transmission rights contracts. As most of the marginal costs of transmission are recovered through wholesale electricity prices we focus on the key issue of regulation and investment incentives. We find that current options are unlikely to be adequate in terms of encouraging socially optimal levels and timing of new transmission investment. As an alternative, we propose a regulatory scheme, based on a related idea by Sappington and Sibley, that can overcome this problem. Our scheme can potentially generate first best results and is readily applicable given the current institutional structure of electricity markets in Australia. *Journal of Economic Literature* Classification Numbers: L51, L94.

Keywords. electricity transmission, regulation, investment, nodal pricing, incremental surplus subsidy scheme.

^{*} We thank James Bushnell, Danny Price, Frank Wolak and an anonymous referee for helpful discussions. Responsibility for all views expressed lie with us.

^{**} Melbourne Business School and the Department of Economics, respectively. All correspondence to Joshua Gans -- E-mail: J.Gans@mbs.unimelb.edu.au. The latest version of this paper is available at <http://www.mbs.unimelb.edu.au/home/jgans>.

1. Introduction

There are three key stages in the supply of electricity. These are generation, transport or transmission and retailing.¹ Generation can be a competitive activity. Retailing has the potential to be competitive and the possibilities for innovation in retailing are reflected in moves by some companies to provide ‘whole of energy’ service to final customers. But transmission offers no real scope for competition.

In the Australian National Electricity Market (NEM), transmission prices are set by regulation under the *National Electricity Code*.² These prices are based on regions and can differ between regions to reflect different costs but are uniform within each region. Prices are designed to allow owners of transmission assets to receive a reasonable return from the use of these assets. The system operator of the NEM coordinates the actual production and transmission of electricity.

Transmission pricing influences the decisions of both generators and consumers and affects the development of the electricity industry over time. For example, the transmission of electricity involves line losses. Electricity dissipates during transmission and the level of loss depends on distance and the load on the transmission network. In the NEM, line losses are taken into account in the dispatch algorithm. That algorithm uses bid information from generators, demand information and technical estimates of line losses to construct an industry supply curve. The prices paid by customers and generators are augmented by the magnitude of line losses and if the estimates of line losses used in the dispatch algorithm are correct, then this process of augmenting prices will internalise the cost of line losses.

Other elements of the transmission pricing regulations currently used in the NEM, however, may not send the correct economic signals. For example, ongoing operation and maintenance costs are ‘postage stamped’ with users paying for these

¹ Transport is often divided into two classes, transmission and distribution, but the division is somewhat arbitrary. We consider transmission as transport from generation to ‘nodes’ where power is withdrawn. A node could be a single customer. If a node is a self-contained distribution system to a group of customers then any regulation of distribution is beyond the scope of this paper.

² The National Electricity Market is not, in fact, national but includes the states of New South Wales, Victoria, Queensland and South Australia. Reforms to electricity have differed between states. For example, see Independent Pricing and Regulatory Tribunal of NSW (1998) for a discussion of the New South Wales reforms.

costs based on the intensity of their use of the network. If this usage charge reflects short-run marginal costs, then it is an appropriate component of network pricing. If, however, operations and maintenance costs are unrelated to system usage, these prices will distort consumer and generator decisions. Further, given the costs that low maintenance can impose on system users, for example through outages, efficient maintenance charges should involve an incentive scheme that would penalise transmission network owners for outages not related to system constraints.³

The limitations of the current NEM transmission pricing rules have led to considerable debate among those involved in the electricity industry.⁴ But regulators face considerable difficulty when designing electricity transmission regulation that provides appropriate signals to customers, generators and transmission owners. In this paper we review some of these difficulties and solutions that have been suggested both in Australia and overseas. In particular, we consider the role that ‘nodal pricing’ might play in transmission regulation. While nodal pricing might be a useful way to augment the NEM we show that it is unlikely to provide appropriate incentives to facilitate ongoing investment in the transmission network. We consider an alternative approach to reward transmission investment that may better facilitate system augmentation.

2. Regulation and the problem of electricity transmission

Electricity transmission requires regulation because it embodies a natural monopoly technology.⁵ It is efficient to have a single system to provide transmission, so long as the owner of that facility can be prevented from exploiting monopoly power. Regulation is often used to constrain the owners of natural monopoly facilities and many standard solutions, including price-caps, rate-of-return regulation and

³ For a discussion of such incentive contracts see Laffont and Tirole (1992).

⁴ See for example Outhred (1998). In early 1999, both the Independent Pricing and Regulatory Tribunal of NSW and the National Electricity Code Administrator reviewed electricity regulation under the NEM, including transmission pricing.

⁵ See Panzar (1989) for a technical discussion of firm costs and natural monopoly technology. King and Maddock (1996) provides a less formal discussion in the context of infrastructure regulation.

public ownership, are used around the world to deal with the conflict between efficient production and abuse of market power.

To understand the problems of electricity transmission, we need to understand the source of the natural monopoly technology. Electricity transmission, for example, differs from gas transmission. In the gas industry, where competition can be introduced into transmission by allocating pipe capacity or having competition between different pipelines that service a city from geographically separate gas fields, there need not be a long term natural monopoly problem and the need for regulation might be transitory. This suggests that as the NEM develops, we might see competition in Sydney between transmission lines from both the Hunter region in the north and from Victoria and the Snowy hydroelectric scheme in the south. Such competition in electricity transmission, however, may not be feasible or desirable. Having an integrated, ubiquitous transmission system is often the most efficient design and improves the integrity of supply as a failure on one set of lines can be offset by flows through other parts of the integrated system. But unless the relevant systems are completely separate, there will be flows between the systems that are difficult to monitor and control.

In this sense, an electricity transmission system appears more like an integrated local telephone network. The standard regulatory solution in telephony involves the sale of point-to-point switched access. For example, if Optus wishes to terminate a long-distance call on the Telstra local telephone network then it can buy access to that network. Telstra transports the call from the point of interconnection with Optus' network to the relevant destination. Similarly, in electricity, the owner of the transmission system could be required to supply point-to-point access for the physical flow of electricity.

Unfortunately, access based on point-to-point flows of power will often be either impossible or irrelevant for electricity transmission if *any* part of the transmission grid is subject to congestion constraints. This is because the electricity transmission grid, unlike a telephone network, is not a switched system. Electricity will not flow along a simple well-defined path but will 'divide' and flow according to the laws of physics. These 'loop flows' create both positive and negative externalities between access seekers. The ability of one company to transmit electricity between two points in the system will depend on the activities of all other companies using the

system. As a result, point-to-point transmission rights can be meaningless. It may be physically impossible for one company to exercise its transmission rights if another company is also exercising its rights. Conversely, it may only be possible for one company to exercise its rights if another company is also exercising its rights.

To see why loop flows create problems when defining rights to transmission access, consider Figure One.⁶ The simple network illustrated in the figure involves three links. There are generators located at points 1 and 2. There is a customer located at point 3. The customer's load is 900 MW. The generator at 1 can produce more than enough power at a constant marginal cost of \$20 per MWh. The generator at point 2 can also produce more than enough power but at a constant marginal cost of \$40 per MWh. Each of the links is identical in terms of resistance and we ignore line losses up to the (thermal) capacity of each link. However, the capacity differs between the links. Both links 1-3 and 2-3 have capacity of 1000 MW. But link 1-2 has a capacity of only 100 MW.

Power flows on the simple network must lead to the same difference in potential between each pair of points regardless of the route between those points. So, the difference in potential between point 1 and point 3 must equal the difference in potential between points 1 and 2 plus the difference between points 2 and 3. The potential difference between two points equals the resistance of the link times the current. Given our assumption of equal resistance, this means that the flow of current on link 1-3 must equal the sum of the flow on links 1-2 and 2-3. But link 1-2 has a capacity constraint of 100 MW, so if the required flow on link 1-2 exceeds 100 MW then this link, and the transmission system, will fail.

⁶ This is a modified version of a diagram presented in Stoft (1996).

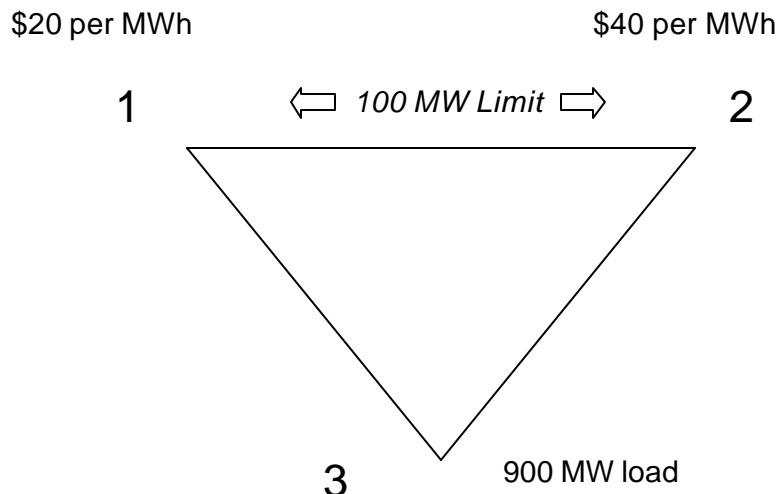


Figure One

The optimal or economic dispatch of the generators to meet the load at point 3 is the least cost method of generating sufficient power to satisfy the load at point 3 without violating the system constraints on power flows. Initially, it might appear that it is optimal to simply dispatch the generator at point 1 to meet the load. This generator is the lowest cost producer. However, if this occurred, then of the 900 MW produced, 600 MW would have to flow directly along link 1-3 and 300 MW would have to flow along links 1-2 and 2-3. But this flow violates our capacity constraint on link 1-2 and is not feasible. In fact, the most power that could be dispatched from point 1 to 3 in the absence of generation at point 2 would be 300 MW. With this dispatch, 100 MW would flow along links 1-2 then 2-3. 200 MW would flow directly from 1-3.

One way to reduce this problem would be to also dispatch the generator at point 2. This will increase the flow on link 2-3 and so allow greater current flow on link 1-3 without increasing the flow on the constrained link 1-2.

Suppose we dispatch 90 MW from the generator at point 2 and dispatch 300 MW from point 1. This was the maximum feasible dispatch from point 1 in the absence of dispatch from point 2. In this case, and ignoring line losses, 390 MW would reach point 3. The potential difference between points 1 and 3 must be the same regardless of the 'route' between the points so the flows must be 230MW on

link 1-3, 70 on link 1-2 and 160 on link 2-3.⁷ These flows are feasible as no link exceeds capacity.

There are three effects from the dispatch of the generator at point 2. First, it raises the total amount of power delivered to point 3. Second, it allows the generator at point 1 to transport more power along the unconstrained link 1-3. Third, the generation at point 2 has reduced the effect of the constraint on link 1-2. In fact, it is no longer constrained. These effects mean that the generation at point 2 directly benefits the customer, by providing more power. The generation also indirectly benefits the customer and the generator at point 1 by increasing the amount of power that the low cost generator at point 1 can ship.

In this example, the optimal dispatch to meet the 900 MW load would be 600 MW from the generator at point 1 and 300 MW from the generator at point 2. The flows would be 500 MW on link 1-3, 100 MW on link 1-2 so this link is operating at capacity, and 400 MW on link 2-3. These are the lowest cost feasible flows to meet the load at point 3.

To see how access based on physical flows would feed into this transmission system, suppose the generator at point 1 wanted to buy the right to transmit 900 MW across link 1-3. This is within the link's capacity constraint. But, by itself, the access right would not be feasible. The access rights available to the generator at point 1 depend on both the load at point 3 and on the activity of the generator at point 2. Loop flows mean that a point-to-point access contract is meaningless in a capacity constrained, ubiquitous electricity transmission grid.⁸

Loop flows are important if there are any binding constraints on the transmission system. If the system never faces any binding constraint, then access can be sustained as if electricity flows directly from point-to-point, even though the actual flows will be more complex. If there are binding constraints at any point in the network then this creates externalities that affect access and efficient transmission.

⁷ Again, as each link has equal resistance, the flow on 1-3 equals the sum of the flows on 1-2 and 2-3.

⁸ The example in Figure One is illustrative and should not be interpreted too literally. In this simple example, the problem of loop flows can be removed by simply closing the link between 1 and 2. In an actual transmission system the loop flows will be far more complex.

If a network has only a few well-defined capacity constraints and the relevant externalities are ‘similar’ within regions, then it may be possible to break the network into zones based on the similarity of the relevant externalities.⁹ Within these zones, point-to-point transmission access may be possible with a more complex form of transmission access operating between zones.

At a minimum, the existence of loop flows is a strong argument for the integrated operation of the electricity system. Externalities between users and the need to plan for contingencies means that a single operator is required to guarantee system integrity and consistency between flows. It may be reasonable for an independent system operator to carry out this function, particularly if there are multiple interconnecting transmission networks that are owned by different parties. If transmission only involves a single owner then this owner could be responsible for operation, subject to regulation that limits any incentives to distort operation to their own financial advantage. Our example above also shows that optimal system operation and dispatch are strongly related so that a transmission system operator would also need to coordinate the dispatch of generators.

3. Alternative approaches to physical access

Because simple point-to-point access contracts cannot coordinate electricity transmission, a number of alternatives have been suggested. Access could involve link-based contracts. A party that wishes to transport electricity would need to consider all possible links that they may need to use to transport power to the relevant customer. They would then need to buy access rights to these links. These rights would depend on the volume of power and may also be contingency based. As the example in Figure One shows, these contingencies might include whether other generators were also dispatching electricity.

Link-based contracts would be complex. If contracts are distinguished by contingencies, then there would be a far larger number of potential link-based contracts than actual links. It is unlikely that individuals could coordinate these

⁹ In terms of the nodal prices referred to below, this means that zones would be regions with similar or only a slow gradient of change in nodal prices. Whether or not breaking the network into zones is a reasonable simplification is a matter for debate. See, for example, Stoft (1996).

contracts through bilateral negotiation. A market for these contracts would probably be thin.¹⁰

Alternatively, access could be based on a contract path that differs from the physical path that is followed by the electricity. A party could buy the right to inject a certain amount of power at a particular point and to withdraw power at another point. These points are referred to as nodes. There is no need for individual contracts to balance or to define the actual flow path of the power. The system operator would check that the purchased rights were consistent with system integrity. Trading of the whole of rights (ie: both the injection and withdrawal of power) could occur without reference to the system operator. Actual system operation would be centrally coordinated and power flows for any individual party might not match their contracted flows. Where a difference arises, the party would be compensated for access rights that went unused and be required to pay for access rights that were used but that they did not hold. Compensation could be on the basis of nodal prices, as discussed below.

Harvey, Hogan and Pope (1997) note that a contract path approach to electricity access is consistent with the guidelines adopted by the Federal Electricity Regulatory Commission in the US. However, they note that these types of contracts cannot cover all power flows. For example, in Figure One, the maximum stand-alone contract right for injection of power at point 1 and withdrawal at point 3 can only cover 300 MW. This is because further flows can only be guaranteed by coordinating flows on link 1-3 with flows on link 2-3. But 300 MW might only be a small fraction of the actual power dispatched by the system operator from point 1 and withdrawn at point 3 under actual system operation. One role of access contracts is to reduce the risk faced by generators and major power consumers. These contracts will only provide partial insurance and as such might be an inadequate approach to access.

Harvey, Hogan and Pope (1997) note that if the aim of access contracts is to limit the risk faced by participants in the electricity market, then financial rather than physical contracts might be a preferred solution. After all, the Independent System Operator controls the actual power flows, so participants need a contract that hedges

¹⁰ Harvey, Hogan and Pope (1997) and Oren, Spiller, Varaiya and Wu (1994) present more complete critiques of link-based access rights.

them against price risk rather than one that relates to the physical operation of the grid. A generator at point 1 in Figure One who has contracted to sell power to a buyer at point 3 is exposed to the risk of relative price fluctuations between points 1 and 3 on the grid. They can insure against this risk by buying a financial instrument whose value depends on the difference in prices (or congestion rents) between nodes in the transmission system. These are called *transmission congestion contracts* (TCCs).

TCCs provide a useful way for participants in the power market to insure against risk. But, as Oren et.al. (1994) note, these contracts are actually derivatives from standard forward contracts that can be written at each node. In other words, so long as there are forward markets for power at each relevant node in the transmission system, then participants can construct the TCCs that suit them by purchasing and selling contracts based on different nodes.

Under the financial contract approach, owners of transmission assets could receive compensation according to the congestion rents on their part of the grid. These are based on the same spot electricity prices that are used for the TCCs, so that ownership of a transmission asset would be the same as ownership of a bundle of TCCs. This raises four questions. First, how many nodes would be defined in the transmission system? In theory, every location where a generator injects power into the grid and every location where a customer withdraws power would need to be defined as a separate node for there to be a complete set of TCCs. With numerous nodes, the forward-contract markets at the vast majority of nodes would be either thin or non-existent. In practice, a nodal approach to transmission requires some aggregation of consumers and generators leading to fewer nodes and fewer TCCs (Outhred, 1998). It is far from clear how this aggregation should be carried out.

Second, would payments from TCCs to the owners of transmission assets cover the cost of providing these assets? In general, unless a transmission system faced considerable constraints, TCC payments are unlikely to cover the costs of building and operating transmission assets. But a system suffering considerable congestion will not be socially optimal so that TCCs by themselves are unlikely to provide adequate compensation to the owners of transmission assets in a socially optimal system. There is also an issue as to whether the TCC prices correctly facilitate new investment in generation and transmission and send correct signals to customers

about the relative benefits of different locations in the grid. We return to this question below.¹¹

Finally, what are the ‘correct’ nodal prices to be used to set the spot price for nodal forward contracts and TCCs?

4. Nodal prices and investment.

For optimal pricing, we want the (spot) price at each node to reflect the short run marginal cost of electricity at that node. These prices can be easily calculated but may appear ‘paradoxical’.

Consider the example in Figure One under optimal dispatch.¹² The relevant prices at points 1 and 2 are obvious. At each of these points there is a generator that has a constant marginal cost, is producing and is not capacity constrained. The prices should reflect the generation costs at the nodes. So at point 1 the nodal price should be \$20 per MWh. At point 2 the nodal price should be \$40 per MWh.

For node 3, consider the marginal cost of increasing the load by one MW. Note that this extra MW could not simply be dispatched from the generator at point 1 because of the constraint on link 1-2. Rather, to maintain system balance, every time an extra MW is dispatched from the generator at point 1 it must be matched by an extra MW from the generator at point 2. In other words, to satisfy an extra MW of load at point 3 requires that an extra half MW is produced at each of point 1 and 2. The marginal cost of extra load at point 3 is given by the (weighted average) generation cost, \$30 per MWh. So, the correct nodal price at point 3 is \$30 per MWh.

Figure Two below reproduces Figure One but includes the nodal prices under optimal dispatch. Note that the price at node 2 is higher than the price at node 3. In other words, there is a negative price gradient between node 2 and node 3 with the price “higher at the sending end of the line than at the receiving end” (Oran, et. al., 1994, p.3). Also note that the nodal prices differ between the ends of unconstrained links, such as link 1-3 and link 2-3.

¹¹ Hogan (1992) presents the case for using TCCs to coordinate transmission investment.

¹² From Section 2, this involved dispatch of 600 MW from node 1 and 300 MW from node 2.

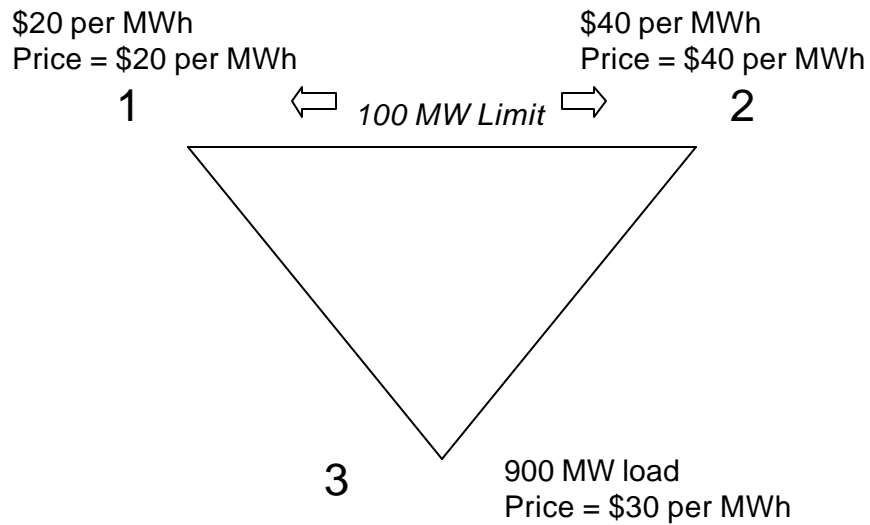


Figure Two

The Independent System Operator, using the same type of computer programs that are used to coordinate dispatch, can calculate nodal prices and it would be relatively easy to make nodal prices available under the NEM. Whether relevant financial markets would develop so market participants can use these prices to hedge against risk, is uncertain.

Nodal pricing can be used to provide information about the optimal location of new infrastructure investment. For example, suppose an individual has up to 200 MW of new generation capacity at a marginal cost of \$35 per MWh. This might be power output to be fed into the grid from a co-generation project. Intuitively, we might believe that it is optimal to locate the new capacity at point 3, near the customer. However, this is incorrect because it ignores the positive externalities created by locating generation capacity at point 2.

If the new generation capacity is located at point 3 and supplies 200 MW of power to this node then the remaining 700 MW must be supplied by the generators at points 1 and 2. Optimal dispatch of this 700 MW requires 500 MW from the generator at point one and 200 MW from the generator at point 2.¹³ But this raises the total cost

¹³ So the current flow on link 13 (400MW) equals the sum of the flow on 12 (100 MW) and 2-3 (300MW) and the capacity of link 1-2 is not exceeded.

of supply. The 200 MW located at point 3 crowds out 100 MW of generation from point 2 and 100 MW of generation from point 1. The average cost of this ‘crowded out’ generation is only \$30 per MWh, below the \$35 per MWh of the co-generation plant. Overall, having the new generation facility locate at point 3 and dispatch to capacity has raised, not lowered, the cost of supplying the load at point 3.

If the new generator located at point 3, optimal dispatch would keep that generator idle even though the higher cost generator at point 2 would be dispatched. Note that this is reflected in the nodal prices. The marginal cost of power from the grid at point 3 is \$30 per MWh. If consumers at point 3 faced this price then they would not want to buy from a co-generation facility with a minimum price of \$35 per MWh.

This does not mean that the extra generation capacity is not needed. Rather, nodal prices signal that it should be located at point 2, where the marginal cost of electricity is \$40 per MWh, rather than at point 3. If the co-generation operation was located at point 2 then it would optimally be fully dispatched, crowding out 200 MW of power from the existing generator at that point. This would lower the total cost of power supplied to point 3 by \$1000 but, in this example, would not alter the nodal prices as this generation would all be ‘inframarginal’.¹⁴

This simple example shows how nodal pricing can be used to coordinate generation and consumer decisions. If a consumer wants to take power off the system then it will optimally choose to do this at the node with the lowest price. If a generator wishes to put power into the network then they will locate where the nodal price is highest.

5. Transmission investment and nodal prices

In section 4, nodal prices correctly guided participants’ location and investment decisions because the participants behaved like ‘price takers’. However, if participants do not act as nodal price takers then their decisions need not be optimal. This is most likely to arise in two situations. First, with ‘thin’ nodal markets large

participants might be able to alter nodal prices by their market behaviour. Large firms might exploit their market power and distort the electricity market. Secondly, transmission investments are often ‘lumpy.’ But a large lumpy investment will generally alter the subsequent nodal prices. Investors might not be able to pursue profitable investments suggested by nodal prices because the act of investing will change these prices and remove the profits.

To see these problems, first suppose that the transmission network is owned by a monopoly and only this monopoly can augment the system. In this case, the network owner’s market power means that congestion rents based on nodal prices will send the exact wrong signals for investment. To the degree that these rents can be expanded by degrading the system or by not investing in new capacity, the network owner will have an incentive to increase these rents.¹⁵ This is a standard problem in utility regulation. If a monopoly facility owner gains congestion rents through pricing (such as peak load pricing) that sends efficient signals to the users of the facility, then the monopolist will have incentives to increase congestion (King and Maddock 1996).

The lumpy nature of investment may lead to perverse incentives for a monopoly owner of transmission. Let us modify our example in Figure Two so that the links between 1-3 and between 2-3 both have a capacity of 700 MW. Note that these constraints are not binding under optimal dispatch and do not alter the nodal prices. Also note that the nodal price at 2 is below that at 3. In other words, if transmission payments are based on nodal prices then the owner of the transmission capacity between points 2 and 3 would have to pay a per MWh charge and would have an incentive to close the link. But, if they did close the link, then this would change the optimal dispatch and the nodal prices. In particular, optimal dispatch would involve 700 MW dispatched from the generator at point 1 and nothing from the generator at point 2. The nodal price at point 1 would still be \$20 per MWh. The same

¹⁴ When dispatching 300 MW from point 2, the system operator will first fully dispatch the new low cost generator, then, for the final 100 MW, dispatch the older high cost generator. So the marginal cost of electricity at point 2 is still the marginal cost of the original generator, \$40 per MWh.

¹⁵ As Oren, et.al. (1994, p.12) note, “because a higher difference in nodal prices will usually translate in a higher link-specific merchandising surplus, owners of the link will have perverse investment incentives. In particular, they will have an incentive to degrade the link. This is not different than the problem created by the monopolist’s incentive to restrict supply, and it was so recognized in Chile and Argentina. Thus, in both countries, third parties can expand or request an expansion of a particular link.”

nodal price would hold at point 2. But the nodal price at point 3, in this simple example, would be infinite. The load of 900 MW cannot be satisfied. Clearly it is socially undesirable to close the link 2-3. Of course, now that it is closed, the owners have an incentive to open it again – but this will again alter the nodal prices and leave the owners with a transmission debt.¹⁶

Allowing third parties to make transmission network investments may reduce some of the concerns about manipulation of nodal prices but still fails to create correct investment incentives. Suppose that, in Figure Two, an entrepreneur notes the constraint on link 1-2 and decides to invest in an extra 201 MW of transmission capacity. Assume that the cost of this extra capacity is rather low so that, given the existing nodal prices, it is both privately profitable and socially desirable to build the extra transmission capacity. Once built, however, the extra capacity will alter the optimal dispatch and will change the nodal prices. In fact, link 1-2 will no longer be capacity constrained after augmentation and the nodal prices at points 1 and 2 will be identical. The arbitrage opportunity will have vanished and the investor will get no return. Knowing this, however, the investor will not undertake the socially optimal investment in the first place.

The existence of forward contracts based on nodal prices may partially overcome this problem by allowing the investor to ‘lock in’ current nodal prices. But, if the relevant financial markets are efficient, then they will take the hedging activity of a potential investor into account and the contract prices will alter even as the investor tries to lock in any gains. If the market were sure that the investment was going to go ahead, the forward contracts would reflect the expected equality of nodal prices in the future.

The inability to lock in gains will not necessarily dissuade all investment. But it will effect the level of investment. While an augmentation of 201 MW is socially desirable, the investor may wish to augment the system by less to maintain a differential in nodal prices. The investor, in this imperfectly competitive market, will wish to manipulate nodal prices and congestion rents to their own advantage.

¹⁶ This example is clearly stylised for simplicity. Oren, et.al. (1994) present a more sophisticated example.

As Bushnell and Stoft (1995) note, if a sufficiently large coalition of network users can form, then it will be in the interest of this coalition to undertake socially optimal investment. This simply reflects the standard economic argument that a social dead weight loss is Pareto suboptimal and there is an alternative outcome that can benefit everyone. However, forming such a coalition might be problematic. For example, suppose the load at point 3 represents a large number of small consumers. It might be difficult to get these consumers to agree to fund a link between points 1 and 2 even though this will lower their electricity prices. In particular, each consumer will have an incentive to free ride on the investment made by others and gain the benefit of the cheaper power without paying for the augmented link. A similar situation can arise if there are multiple generators at point 1 or if the same company owns the generation capacity at points 1 and 2.

Finally, even if nodal prices and TCCs could provide optimal incentives for an individual investor, the complexity of the externalities in an electricity system means that this might not lead to the correct pattern and timing of total investment. One investment decision impacts on others. It is likely that an optimal configuration requires coordination (Bushnell and Stoft, 1995).

6. Does nodal pricing pay for transmission?

Nodal pricing is unlikely to create the correct incentives for transmission investment. However, it has a superficial appeal because it appears to provide a relatively ‘hands off’ way to regulate and remunerate transmission companies. Unfortunately, this appearance is misleading. Congestion rents associated with nodal prices and TCCs will not in general recover the costs of transmission investment and operation. Standard regulatory intervention is often needed to evaluate the revenue requirements of transmission companies and to meet these requirements from system users (Wu et.al., 1994).

To see this, consider again the example in Figure Two. Suppose the marginal cost of operating the transmission system is $\$C$ per hour and that the sunk capital cost of building the network is $\$K$ per hour. Denote the maximum willingness-to-pay for the 900 MW load by consumers at point 3 by $\$V$ per hour. The cost minimising, or

‘optimal’ dispatch to meet this load has a total cost of \$24,000 per hour.¹⁷ So, it is socially desirable to build and operate the transmission network to meet the load at point 3 if and only if $V \geq C + K + 24,000$.

Suppose the only remuneration received by the network owner was the congestion rents. These are \$3,000 per hour. The operator ‘pays’ a total of \$24,000 for electricity and ‘sells’ it at the nodal price for \$27,000. These payments might be reflected in transmission congestion contracts and passed onto the owner of the transmission assets. The network owner is only fully compensated for the cost of the network if $C + K \leq \$3,000$. But if $V > \$27,000$ there will be situations where it is socially desirable to build and operate the network, but an investor who was only remunerated through congestion rents derived from nodal pricing would be unwilling to build this network. For example, if $V = \$40,000$ and $C + K = \$9,000$, it is socially desirable to build the network but nodal pricing offers an inadequate return.¹⁸ In such circumstances, the evaluation and allocation of a payment to make up for the short-fall of congestion rents will involve a regulator who faces a number of alternatives.

If the network has already been built, then one option would be to simply ignore the short-fall. Unfortunately, this may not be a good way to encourage future network investment. Alternatively, the generators might pay a licence fee. Because the network costs $C + K$ are not related to power flows, an efficient licence fee would involve a fixed charge per generator. This can have two problems. In our simple example, the generators are not making any profit so the licence fee would lead them to exit the market. This reflects the assumption of constant marginal cost. However, it also illustrates the important point that fixed charges will affect the entry and exit decisions of participants. A fixed generator charge will create an entry barrier, particularly for small generators. Secondly, even though the generators might face an efficient ‘lump sum’ licence fee, if the generators cannot themselves price discriminate to their customers, then this fee will be recovered by a higher marginal

¹⁷ This involves 600 MWh dispatched from point 1 at a cost \$20 per MWh, and 300 MWh dispatched from point 2 at a cost of \$40 per MWh.

¹⁸ We are not arguing that nodal pricing will never adequately compensate transmission investment. Rather, we are simply saying that nodal pricing will not always adequately compensate socially desirable investment, and in such circumstances, congestion rent payments must be augmented through standard regulation.

price downstream. Put simply, efficient upstream pricing can be undone by downstream limitations on price discrimination.¹⁹

The customers could pay a fixed fee. For example, the government might levy a per person poll tax on residents at location 3 to fund the network. This result would be efficient, but is probably politically unsaleable.

An alternative would involve the fixed network costs being recovered by a power charge. But this charge will distort the marginal price faced by consumers and will lead to allocative inefficiency. Also, because a 'network tax' would raise the nodal price at point 3, it could distort investment decisions. For example, if the nodal price at point 3 was raised above that at point 2 due to the tax, then a new generation facility could have an incentive to locate at point 3 rather than at point 2. The generator would be by-passing the network and could reasonably claim that electricity produced from its facility should not be taxed as it does not use the network. The regulator would either have to tax all electricity, even that sold without using the network, or inefficient by-pass could be encouraged.

7. A regulatory alternative for transmission

Sections 5 and 6 have shown that nodal prices and TCCs, by themselves, are unlikely to elicit socially optimal investment. In this section, we consider an alternative, regulatory solution to create appropriate incentives for new investment.

A regulatory solution needs to guarantee that the private incentives of the transmission network owner and market participants reflect the social value of their decisions (see Gans and Williams, 1999a; 1999b). Appropriate transmission investment will reduce capacity constraints on the grid. If there are no constraints, then there is no need for investment. If there are constraints however, the key regulatory issue is how to give the network owner the incentive to invest in the grid at the appropriate time.

Sappington and Sibly (1988) show how an *incremental surplus subsidy* (ISS) *scheme* can align the private and social incentives. When a monopolist undertakes a

¹⁹ King (1997) provides an example of this in a general access framework.

relevant regulatory activity, the scheme awards the monopolist the entire social surplus from that activity for one regulatory period but only for one period. In other regulatory periods, the monopolist is regulated on the basis of costs. The intuition behind the ISS scheme is that optimal decisions are often based on marginal gains and losses. So long as the monopolist faces the correct marginal incentives then it will make socially desirable decisions. In their original paper, Sappington and Sibly were concerned about inducing a monopolist to price at marginal cost over time. Here we are concerned with inducing optimal investment and investment timing. Nonetheless, the basic intuition of awarding the monopolist a single period's social surplus can be applied to transmission investment.

To demonstrate this alternative approach to the regulation of transmission investment, consider again the example in Figure Two. Suppose that there is a single firm that owns the grid and this firm can make a once-off capital investment that will eliminate the capacity constraint on link 1-2 forever. Once this investment is made, the independent system operator will be able to meet the entire load at node 3 by dispatching the low cost generator at node 1. This will result in a saving of \$6000 per hour or approximately \$53m per year.²⁰ For simplicity, we take one year as being the relevant regulatory period, and assume that interest rates are ten percent per annum. Suppose that the monopolist has five alternatives. First, it could not invest at all. Alternatively, the monopolist could invest immediately. However, due to technological progress, the cost of augmenting the link 1-2 is not constant but is falling over time. If the monopoly invests immediately, then the capital cost of this investment is \$540m. For simplicity, we will assume that there is no maintenance so that immediate investment involves a per period cost of ten percent of the capital cost or \$54m. The monopolist could also wait before investing. Because of expected changes in technology, waiting one year is likely to reduce the capital cost to \$350m or \$35m per period while waiting for two years is likely to reduce the capital cost to \$280m or \$28m per period; three years, \$270m or \$27m per period; and four years

²⁰ For this example, we will assume that the load is constant forever and allow optimal investment timing to depend on the cost of the investment. In general, the load is likely to also grow over time raising per period benefits from investment. This would make the example more complex but would not alter the basic intuition.

and beyond, \$260m or \$26m per period. Table 1 summarises this information.²¹ The first row represents the total social saving from the investment. The second row is the augmentation cost in the relevant period and the third row is the net present value today of the social gain from investment in the relevant period.

Table One: Relevant Investment Outcomes

Year	Today	1	2	3	4
<i>Social Saving</i>	530	530	530	530	530
<i>Investment Cost (Current)</i>	540	350	280	270	260
<i>Net Present Value</i>	-10	164	206	195	184
<i>TCC Lock-in</i>	-280	-82	-17	-7.5	0
<i>ISS</i>	-1	16	21	20	18

It is clearly undesirable for the network owner to immediately invest in augmenting the link – the current cost exceeds the benefit. If the investment is delayed by one year, there is a congestion cost of \$53m. But this is more than offset by the expected cost savings due to technological change, \$190m. Further, there is a benefit of \$54m in terms of the one year ‘interest saved’. Similarly, if investment is delayed from the first to the second year, the delay results in an extra social cost of congestion of \$53m. But it also leads to a gain in terms of both the lower expected cost of investment, \$70m, and a saving due to delayed expenditure of \$35m. Again, the benefits outweigh the cost and it is socially desirable to wait until the second year to invest rather than invest either immediately or in the first year.

It is not, however, socially desirable to delay until after three years. The social gain from delay from two to three years is \$38m – the \$10m in lower capital

²¹ Gans and King (1998) provide the formal version of this model and the associated regulatory schemes. The numbers picked for this example illustrate the differences between regulatory options but are readily generalisable. That paper also deals with the case where social value might be growing over time due to increasing demand for electricity.

expenditure plus \$28m in interest saved. But this is less than the \$53m congestion costs associated with delay. Socially optimal investment will occur after two years when the capital cost is \$280m.

If a regulator had perfect information about these opportunities then it could simply order the monopolist to augment the transmission link after two years. Unfortunately, any regulator is likely to have only poor information about the investment opportunities that are available to the transmission company. Some form of incentive regulation is needed to align the private interests of the transmission company and social interests.

How would a regulated monopoly transmission company respond to these investment opportunities? If the monopolist receives the nodal price congestion rents each period then it will never invest. Recall that the system-wide congestion rents are \$3000 per hour or approximately \$26m per year. If the monopolist invests, then these rents disappear so that the transmission company will never invest as it can never recoup its investment cost.

Alternatively, what if the monopolist receives the current nodal price congestion rents forever once the investment is made? This would be equivalent to the monopolist being able to lock in the TCCs for the entire system at the current nodal prices forever. As these rents are only \$26m per year, the monopolist would choose to invest in the fourth year. Only then would the investment yield the monopolist a non-negative return.²² Clearly, in this example, rules based on nodal price congestion rents fail to align private and social interests.

If the transmission company is regulated by standard rate of return regulation, then this will also not lead to optimal investment timing. If the company is just compensated for its cost of capital on any investment then it is always indifferent between investing and not investing. If the company is over compensated, say by receiving a regulated return of eleven percent, then it will want to invest immediately as this both maximises its regulated capital base and gives it rewards as soon as possible.

²² The relevant present value figures for the monopolist's return are given in row four of table 1. Note that if the monopolist only received the nodal congestion rents at current prices on link 12

As an alternative, suppose the monopolist is allowed to retain the social surplus created by any transmission augmentation for one year. During that year, the company must also pay the costs of this augmentation. Once the year is over, the company is then compensated for the actual costs of their investment in each successive year.

This regulatory scheme perfectly aligns social and private incentives. When the monopolist invests they receive the social benefit of \$53m for one year. But they must pay the capital cost of the investment for that year. The monopolist will not invest immediately because the annual capital cost of \$54m outweighs the payment of \$53m. The monopolist will also not invest after one year. While they make a profit of \$18m if they invest after one year, they make a profit of \$25m if they invest after two years. Discounting this profit back so that it is expressed in 'year one' dollars, the monopolist makes approximately \$21m by investing after two years rather than \$16m by investing after one year.

The monopolist, however, will not wait for three years. Investing after three years makes the monopolist a once off profit of \$26m. But in terms of current dollars this is just under \$20m. The transmission company prefers to invest after two years as this maximises its profit. This is also the socially optimal investment timing.²³

The intuition behind this optimal regulatory scheme is straight forward. The pricing formula for investment gives the monopolist one period of social surplus but requires them to bear the investment costs for one period. The difference between the social surplus generated by the investment and its one-period cost provides a rent for the transmission company. If this rent is increasing (in present value terms) over time, then the monopolist will prefer to wait and defer the investment. The transmission monopolist will maximise this rent by investing when the difference between social surplus and investment cost is at its greatest (taking into account impatience). Hence, its incentives are the same as the social investment incentives.²⁴

forever, then this is only \$2000 per hour or approximately \$17.5m per year and the monopolist would never undertake the investment.

²³ The relevant present value figures are given in row five of Table One.

²⁴ In the above example, a regulatory period was given by one year. This was purely for convenience. In reality, the meaningful period will include considerations such as the time for

The figures used in the above example are, of course, only illustrative. But the logic of the example – that an ISS scheme but not TCCs will align private and social incentives – is more general, as is illustrated in Figure Three. This figure shows the effect on the market for electricity when transmission capacity is augmented. By reducing network constraints, transmission augmentation lowers the social marginal cost of electricity. This is shown by the MC curve shifting to the right from MC_0 to MC_1 . The per period social gain from the additional transmission capacity is given by area B . Suppose that the cost of augmentation today is A but if the monopolist waits for one year then the cost will fall by ΔA . From a social perspective, we prefer the monopolist to invest today if and only if $B \geq \Delta A + rA$. Under the incremental surplus subsidy scheme, the monopolist will invest today if and only if

$$B - rA \geq \frac{(B - r(A - \Delta A))}{1 + r}$$

Simplifying, the monopolist will only invest immediately if $B \geq \Delta A + rA$. But this is the condition for socially optimal investment timing, so the ISS scheme perfectly aligns private and social incentives.

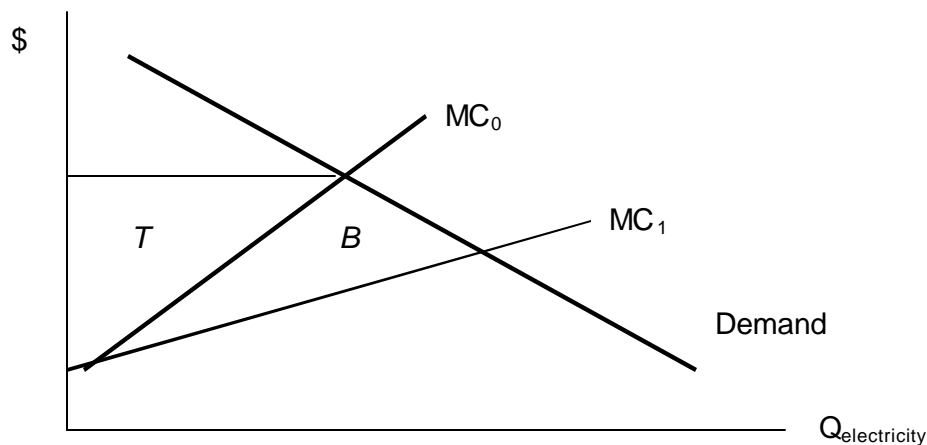


Figure Three

In contrast, the value of TCCs are given by the per period congestion rents. These will be part of the inframarginal rents that accrue under the initial marginal cost

construction and seasonal factors that influence the level of social surplus. This said, for electricity transmission, a period of one year is likely to be reasonable.

curve, which are given by area T in Figure Three. These inframarginal rents, however, are irrelevant from the perspective of timing socially optimal investment. If the part of T that represents the TCCs happens to equal area B then using TCCs will motivate optimal investment. But this would be purely serendipitous. There is no reason to expect that this will be the case. Rather, if the per period value of the TCCs is greater than B then investment will occur too rapidly from a social perspective. If the per period value of the TCCs is less than B , as in the numerical example above, then transmission augmentation will be undesirably delayed from a social perspective.

The incremental surplus subsidy requires the measurement of the increment of social surplus generated by the investment in the relevant year. This requires the calculation of a counterfactual: what would social surplus have been without the investment? In other industries this might be an insurmountable problem requiring approximations based on careful estimation of demand and supply elasticities. Fortunately, the electricity spot market organised under the NEM provides a ready means of calculating these social surplus values. Recall that the dispatch procedure utilises line loss and constraint information as well as generator bids to form a schedule for supplying quantity demanded at every node. That determines the system marginal price paid to generators that period. It would be a relatively simple matter to also calculate what the system marginal prices would have been for the same demand and generator bids if the transmission network was configured without the investment. The difference between the price with and without the investment multiplied by quantity demanded approximates the social value generated by the new investment during the year. Further, charging consumers the ‘without-price’ and paying generators the ‘with-price’ could raise the relevant funds. If the investment has a negative impact at some nodes then this is also captured and the monopolist is penalised.

Note that this method of calculating social surplus is robust to two potential problems. First, it takes into account the effect transmission investment can have in reducing the local market power of generators. Any changes in social surplus that are the result of technical efficiency or an increase in effective competition are measured in the counterfactual experiment proposed here. Secondly, because generators receive the payments they would have without a surplus being paid to the transmission owner, their bids are not distorted by this regulatory scheme.

Nonetheless, this funding scheme potentially involves a distortion to allocative efficiency, as the price faced by users is higher than it ought to be. However, this only lasts one period and given the inelastic nature of short-run demand for electricity we believe that this is unlikely to result in a serious social loss.

8. Future directions in Australian electricity

The current structure of transmission pricing under the NEM involves broad zones with some pricing signals based on the reference node in each zone. In this sense, it has some elements of nodal pricing. But the zones are large so that the potential benefits of a nodal pricing system are not realised under the NEM. Transmission prices are set to partially provide signals to generators and consumers but averaging and inefficient cost allocation means that the relevant signals to guide participants are distorted. Investment in transmission occurs under government fiat and it is far from clear that current investment is socially optimal. As transmission systems are privatised, the problem of regulating transmission investment will become increasingly important. Congestion constraints and loop flows that lead to both positive and negative externalities between participants are poorly handled by the NEM. As the demands placed on the system increase, congestion and loop flow problems will increase and it is not clear that the NEM is equipped to adequately handle these problems. In brief, while the NEM rules appear to be working adequately at present, they are unlikely to be satisfactory in the future.

This paper has considered a number of alternatives that could improve transmission regulation under the NEM. Nodal pricing could easily be further developed. The system operator could calculate the relevant spot prices and the government could encourage spot markets in financial contracts based on these prices to develop. These contracts would help participants to insure against risk and will help guide participants' decisions. This said, nodal pricing is not a complete answer to electricity transmission. The government needs to carefully weigh up the number of nodes in the system with the depth of financial markets that are likely to develop in the price at each node. If there are many nodes then this leads to accurate signals but also increases the risk of market manipulation by large participants. Further, while

nodal pricing aids participants' decisions, major investments that alter nodal prices will not necessarily be well coordinated through nodal pricing.

Transmission will need to be regulated. The regulatory scheme that is proposed in this paper is a variation of the incremental surplus subsidy scheme. It provides a simple, easy-to-implement way to align social and private incentives. This scheme is particularly apt for regulating electricity transmission investments because of the modelling information that is readily available through the operation of the current dispatch and pricing system.

References

- Bushnell, J. and Stoft, S. (1995) "Transmission and generation investment in a competitive electric power industry," University of California Energy Institute, Working Paper PWP-030, Berkeley, CA.
- Gans, J.S. and King, S. (1998) "Efficient investment pricing for electricity transmission", Paper presented at the ACCC/University of Melbourne conference on Electricity Transmission Network Pricing, University of Melbourne, December.
- Gans, J.S. and Williams, P.L. (1999a), "Access Regulation and the Timing of Infrastructure Investment," *Economic Record*, 79 (229), pp.127-138.
- Gans, J.S. and Williams, P.L. (1999b), "Efficient Investment Pricing for Access Regulation," *Australian Business Law Review*, 27 (4), pp.267-279.
- Harvey, S., Hogan, W. and Pope, S. (1997) "Transmission capacity reservations and transmission congestion contracts" Working Paper, Kennedy School of Government, Harvard University, Cambridge, MA.
- Hogan, W. (1992), "Contract Networks for Electric Power Transmission," *Journal of Regulatory Economics*, 4 (3), pp.212-242.
- Independent Pricing and Regulatory Tribunal of NSW (1998) "Pricing for electricity networks and retail supply", Issues Paper, September, Sydney.
- King, S. (1997) "Access pricing under rate-of-return regulation", *Australian Economic Review*, 30, 243-255.
- King, S. and Maddock, R. (1996) *Unlocking the infrastructure: the reform of public utilities in Australia*, Allen and Unwin, Sydney.
- Laffont, J-J. and J. Tirole (1992), *The Theory of Incentives in Procurement and Regulation*, MIT Press: Cambridge (MA).
- Oren, S., Spiller, P., Varaiya, P, and Wu, F. (1994) "Nodal pricing and transmission rights: a critical appraisal", Working Paper, university of California, Berkeley.
- Outhred, H. (1998) "Network pricing – proposals in the National Electricity Code," Paper presented at the ACCC/University of Melbourne conference on Electricity Transmission Network Pricing, The University of Melbourne, December.
- Panzar, J.C. (1989), "Technological Determinants of Firm and Industry Structure," in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*, Vol.1, North-Holland: Amsterdam, pp.3-60.
- Sappington, D. and Sibly, D. (1988), "Regulating Without Cost Information: The Incremental Surplus Subsidy Scheme," *International Economic Review*, 29 (2), pp.297-306.
- Stoft, S. (1996) "Zones: simple or complex," Working Paper, E.O. Lawrence Berkeley National Laboratory.

Wu, F., Varaiya, P., Spiller, P. and Oren, S. (1994), "Folk Theorems on Transmission Access: Proofs and Counter Examples," *University of California Energy Institute Working Paper*, PWP-023