Chapter 4
Transmission Investment In a Deregulated Power Market
by Steven Stoft

From:

Competitive Electricity Markets and Sustainability
Edited by François Lévêque

Published by
Edward Elgar Publishing Limited
Edward Elgar Publishing, Inc.

Chapter 4 Outline
Transmission Investment In a Deregulated Power Market
Steven Stoft

Chapter 2
Transmission Investment In a Deregulated Power Market

1. Introduction

From the earliest days of commercial power production, transmission has grown steadily in importance. New wholesale power markets have sparked interest in distributed generation, but trade between these markets has only increased the need for transmission investment. As with generation, a market for the *use* of existing assets is not difficult to imagine, but a market to supply these assets is more problematic. In fact, the discrepancies between the properties of transmission costs and benefits and the assumptions of competitive economic theory are so substantial that a market solution is probably not desirable. Even incentive regulation may prove so difficult to design and so inaccurate that a planning solution may be preferable, at least until wholesale power markets are functioning efficiently and the generation-investment problem has been solved.

Transmission, an exceptionally inhomogeneous product, can be both a substitute for and complement to generation, and suffers from returns to scale and lumpiness,[1] as well as major externalities, both positive and negative. This chapter investigates the extent to which these problems can be overcome. To simplify, this chapter ignores transmission losses because they play a relatively small role in the investment problem and one similar to the role of congestion, which is considered. After introducing some properties of congestion prices and transmission costs, three basic approaches to transmission investment are explored.

1.1. Three Approaches to Investment

Three approaches to investment stand out as relatively distinct, although many mixtures of these are possible. A **planning approach** refers to a system that does not include any incentives specifically tailored to the long-run transmission investment problem. Such an approach would be carried out by a group of engineers and economists under instructions to build an efficient system. In practice, it would be backed by rate-of-return regulation with a requirement that investments be “used and useful.” A **merchant approach** would allow any private company to modify the transmission system, subject to certain restrictions, and would reward (or punish) such modifications by allocating transmission rights to investors. A **performance-based-regulation approach** would induce investment by a for-profit owner of the transmission system (a Transco) by adjusting its profit level on the basis of the cost and performance of the system.

1.2. Congestion: the Opportunity Cost of Using the Grid

All three approaches will be assumed to exist within the framework of a wholesale power market based on publicly-known nodal prices (Hogan, 1992; Harvey, Hogan and Pope, 1996). That is, at each relevant point (node) in the network, a price is established, and these prices together clear the market. They may be purely competitive prices or they may be distorted by market power, but in any case there is one price at each node and all energy transactions at a given node take place at that price. These prices are adjusted each time there is a change in supply or demand.

---

[1] Returns to scale imply transmission investment costs are non-convex. Lumpiness refers to having to buy an integer number of transmission lines selected from a small set of available capacities, but it is better understood as simply referring to a cost function with a fluctuating slope.

Such a pricing system automatically prices the use of the transmission system even though it applies directly only to energy transactions. If the price at node A is $20/MWh and at node B is $30/MWh, then the price to transmit energy from A to B is $10/MWh, while the price to transmit it from B to A is negative $10/MWh. Although nodal prices have some peculiar properties, it is important to understand that they are simply the result of the normal forces of supply and demand constrained by the physical limits of the transmission system. When supply and demand are both competitive, nodal prices are simply the standard competitive market prices and have all the properties expected of such prices. Except when prices are determined somewhat arbitrarily because a vertical supply coincides with a vertical demand curves, competitive nodal prices are unique. Although they are often calculated from bids in a centralized auction, they are not the product of any special rules of calculation but are the prices at which a well-arbitraged bilateral market would arrive if the transmission constraints were enforced.

Three distinct costs associated with congestion are often confused, congestion rent (C_R), congestion cost (C_E), and the cost of congestion to load (C_L). Economists focus on the first two, while consumers react to the third. Consider a load pocket with 1000 MW of load and a 300 MW line into the load pocket from a large system that could supply 800 MW at $20/MWh and much more at $40/MWh. These costs are represented as a “Remote Supply” function for the load pocket in question. Suppose that local load is fixed at 1000 MW and that the pocket contains 600 MW of $30/MW generation and 200 MW of $50/MWh generation. The “Net Local Demand” curve shows the demand for imported power net of what would be purchased locally. For example, at a local price of $40/MWh, the load pocket would consume 1000 MW and supply 600 MW, leaving it with a net demand of 400 MW. In other words, Net Local Demand accounts for local supply as well as local demand, and it is local supply that provides the price sensitivity of this “demand curve” and not actual demand responsiveness.

The **congestion cost** is also called the re-dispatch cost because it is the extra cost of dispatching more expense generators than would be needed if the transmission system had ample capacity and did not constrain power transfers. In the present example, 100 MW of local $50 generation and 400 MW of local $30 generation must be used in place of 500 MW of remote $20 generation that would have been used had there been no transmission constraint. Congestion cost is a dead-weight loss, not a transfer payment.

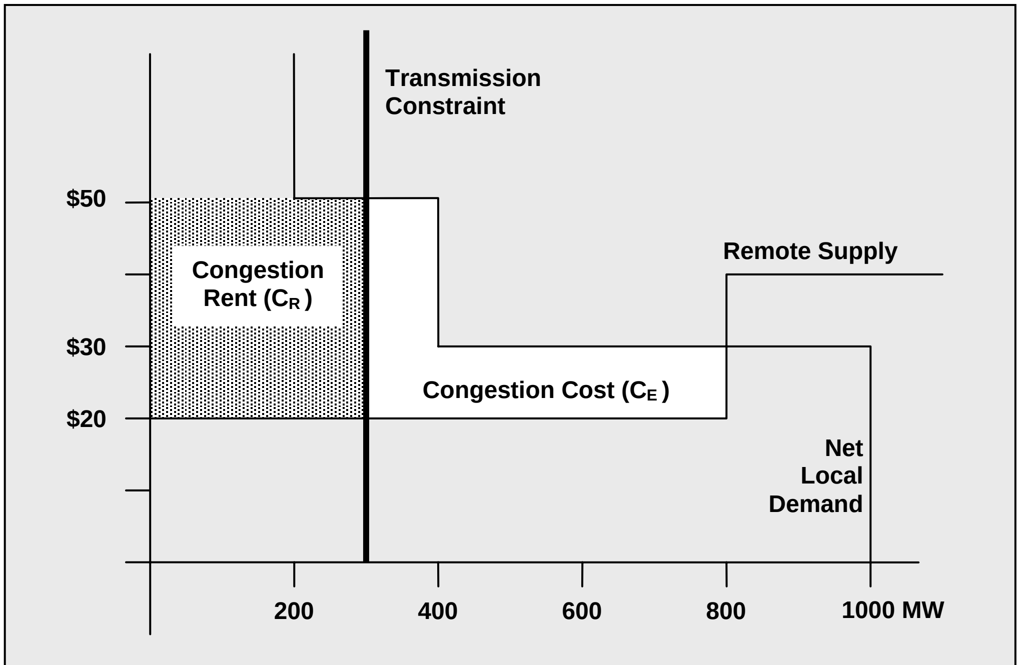

Figure 1. Defining Congestion Rent and Congestion Cost

Congestion rent is the amount collected by the owners of the rights to the transmission line. In a one-line network these rights would typically pay the owners an amount equal to the line's capacity times the difference between the prices at the two ends of the line. In the case of a load pocket, this is the difference between the internal price and the external price. Congestion rent is a transfer payment from line user to line owner, as using the line has no actual cost.

Finally, there is the **cost of congestion to load** (C_L) (see Figure 2). With ample transmission, the load in the pocket would import 800 MW of power and use 200 MW from internal generators. The price would be set by the intersection of supply and demand at $30/MWh, so the total cost of power to load would be $30/MWh × 1000 MW, or $30,000/h. Because of the congestion, the price in the pocket is $50/MW and so load must pay $50/MWh × 1000 MW, or $50,000/h, which makes the cost of congestion to load $20,000/h. The three costs are shown in Table 1.

Table 1. Three Views of Congestion

Congestion (re-dispatch) cost	C_E	$7,000/h
Congestion rent	C_R	$9,000/h
Cost of congestion to load	C_L	$20,000/h

As can be seen there is no particular relationship between these three concepts. Frequently consumers find it unfair that congestion can cost them far more than the "congestion cost." It does not help that it can also cost them more than congestion cost and rent combined with the excess revenue accruing to generators that seem to benefit from the constraint without reason. Although the matter is beyond the scope of this chapter, it should be noted that if the transmission and generation markets satisfy the axioms of perfect competition, nodal prices will just cover the long-run costs of the efficient mix of generation and transmission. In other words there is nothing inherently unfair or inefficient about the distribution of revenues under nodal pricing in a congested system. Of course this does not indicate that either transmission or generation is, or can be supplied by, a competitive market, only that the problems with these costs and prices arise from non-competitive features of the markets and not simply from the method of nodal pricing or the effects of transmission congestion on prices.

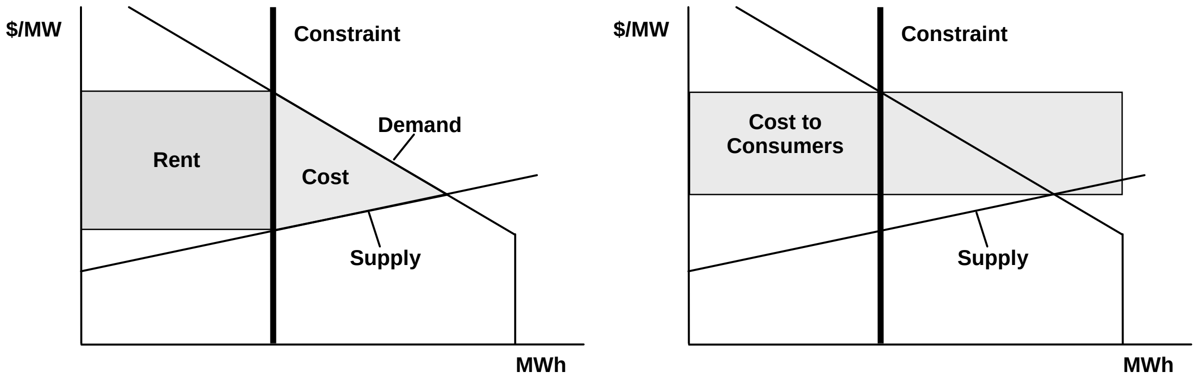

Figure 2. Cost to Consumers Compared with Congestion Cost and Rent

1.3. The Zero-Congestion Fallacy

Because transmission congestion imposes costs, one recurrent view holds that it should simply be eliminated. This is now the policy of the Alberta government (see box 1). Although examples can be manufactured for which this is the least-cost solution, in real power markets such situations never exist. If there is one hour per year in which a remote generator is $10/MWh cheaper than the most expensive local generator in use, and if 1 MW of that generator's output cannot reach local load, then the line is constrained during that hour and the cost of the constraint is $10/year. Adding 1 MW of capacity to that transmission path would cost far more than $10/year. Eliminating all congestion—allowing every last megawatt of trade—is simply not efficient. When transportation is expensive, it is often cheaper to consume local product than to transport slightly cheaper product from a distance.

The fallacy of eliminating all congestion may arise from confusion between congestion and unreliability. Unreliability is the result of having too little local generation to meet local demand net of imports. This can result from congestion,

ALBERTA REGULATION #174/2004, Electric Utilities Act,

Alberta's Electric Utilities Act took effect on January 1, 1996. From the beginning, the Electric Utilities Act has imposed uniform prices (forbidden competitive prices) throughout the Province. With its new requirement (shown below) to over-build the grid, uniform prices will become competitive prices.

From Section 8(1) of the Electric Utilities Act:
(e) taking into consideration the characteristics and expected availability of generating units, plan a transmission system that

(i) is sufficiently robust to allow for transmission of 100% of anticipated in-merit electric energy referred to in section 17(c) of the Act when all transmission facilities are in service, and

(ii) is adequate to allow for transmission, on an annual basis, of at least 95% of all anticipated in merit electric energy referred to in section 17(c) of the Act when operating under abnormal operating conditions;

(f) make arrangements for the expansion or enhancement of the transmission system so that, under normal operating conditions, all anticipated in merit electric energy referred to in clause (e)(i) and (ii) can be dispatched without constraint;

but in almost all cases, congestion is not associated with unreliability. It is simply the result of having more than enough local generation, but at a cost higher than the cost of remote generation that could be accessed with a larger transmission line.

In other cases the desire to eliminate congestion may result from a desire to increase local supply and thereby lower the local market price. For some time, this can save consumers money even if it raises the long-run cost of producing and delivering power. But if expanding transmission lowers consumers costs without lowering total cost, it is a form of monopsony power which essentially expropriates some of the sunk costs of generation in the import-constrained area.

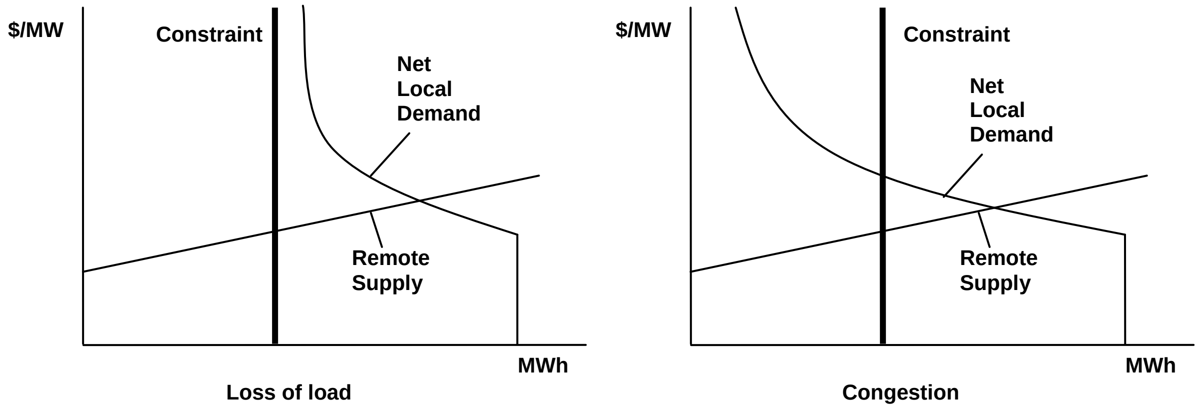

Figure 3. Relationship of Congestion to a Transmission-Cause Reliability Problem

Suppose that generation fixed costs in an import-constrained region are $40,000/MW-year higher than in the surrounding area, and there are 1000 MW of such generation installed in the load pocket. A 200 MW expansion of the import line may virtually eliminate extra fixed (sunk) cost recovery needed by reducing the local price to the external price. This may save consumers $200,000/year per MW of new line. This is probably much more than the cost of the line. This appears to be a savings, largely because it expropriates $40 million per year in generation fixed costs. It also provides a legitimate savings by providing genuinely less expensive external power to the expensive constrained zone.

If capacity produces power in half of all hours and if the new line's capacity is half used, then each MW of line provides a real savings of $40,000/MW, one fifth as much as the initial savings through the expropriation of generation sunk costs. In the long run, internal capacity will retire and eventually the generating capacity in the load pocket will again recover its fixed costs. Perhaps the savings from the line will amount to $240,000/MW for the first five years and $40,000/MW after that. If this has a present value of $1,250,000/MW, then it might be thought that this is the break-even point for building such a line, but that omits the impact of regulatory risk.

Investors take account of the cost of expropriations by factoring that possibility into future investment decisions. This demonstration effect will not be confined to the load pocket in question. It is impossible to predict what the cost of this will be, but the resulting investment risk premiums will affect the rate of return on all capital in the affected load pockets, not just the new investment. This is a consequence of a market-clearing electricity price. Probably the most sensible guess is that all of the money transferred to load from existing generators will be lost to load through higher risk premiums.

In conclusion, it is inefficient to eliminate all congestion, and it is wrong to focus on short-run consumer cost reductions when planning transmission investments. As with any type of productive investment, the goal should be to minimize the total cost of production and delivery. If the market is competitive or the regulation effective, these costs savings will be passed through to consumers.

1.4. Optimal Transmission (Static)

The optimal amount of transmission minimizes the total cost of producing and delivering electricity. It can be determined in simple examples by using the standard first-order condition of calculus. At the optimum, the value of an additional kW of transmission equals the cost of building it. Because realistic transmission cost functions can be quite complex, the optimal design may need to be found by evaluating many options rather than applying calculus.

A simple example will illustrate the main points. Consider a city (load pocket) that can produce power at a cost of \$50/MWh but can buy it for \$30/MWh over a transmission line. Suppose the line can be built for a rental cost of \$6000/h plus \$5/MWh. How large a line should be built? (Please see side bar on defining the rental cost of transmission lines.)

To solve the above problem, the city's load must be specified. Suppose the peak load is 800 MW and its minimum load is 400 MW and it takes on intermediate values according to a uniform probability distribution. The savings from the line will be \$20/MWh for the first 400 MW of line capacity and will then decline linearly to zero for additional capacity up to 800 MW. As long as additional capacity saves more in energy production costs than the cost of the additional capacity, the line should be larger. Since the cost of additional capacity is \$5/MWh, the line should be expanded until the saving falls to \$5/MWh on average. When the capacity is ¾ of the way from 400 MW to 800 MW, load will be great enough to use the last MW of capacity only ¼ of the time. Thus the savings of the last MW will be only ¼ of \$20/MWh or \$5/MWh. Hence 700 MW is the optimal capacity of the line.

This conclusion is actually a bit premature. It is necessary to check that the net benefit from the line is positive when the fixed cost of the line is included. The total benefit is \$(20 × 400 + 12.5 × 300)/h, which is \$10,750/h, while the total cost is \$(6000 + 5 × 700)/h, which is \$9,500/h. The line is worth building, but if the fixed costs had been \$8000/h, it would not have been.

This example is the basis of the conclusion

Optimal Transmission

- Transmission investment should not eliminate congestion.
- Transmission investment should not minimize short-run consumer costs.
- Transmission investment should minimize the total cost of transmission and the <u>production cost</u> of power.

Defining the Rental Cost of Transmission Lines in \$/hour.

The cost of a particular transmission line might be \$500,000,000, but when analyzing power systems it is much more convenient to think of renting capital than buying it outright. Rental cost is naturally expressed in \$/MWh, and these units are particularly convenient for computation. This is not how engineers calculate costs, but it is very convenient for economic calculations.

Given the one-time cost of the line, its capacity, and a savings per MWh of energy transported, it is generally impossible to tell whether building the line saves money or not. This is because one must know how long the line will last, it maintenance costs, and the debt and equity costs associated with the project. All of these are properly taken into account by a rental cost, and so can be ignored once the rental cost is specified.

The one-time cost can be converted to an amortized annual cost, of say, \$50,000,000/year. Adding maintenance of say 2,000,000 year and dividing by 8760 hours/year gives a rental cost of \$5936/h. If the line is a 1000 MW line, then the cost is \$5.94/MWh. This is the cost of renting 1 MW of the line for 1 hour and is independent of whether the line is used or not.

Simple investment problems often involve a transmission cost function such as $C = c + bQ$, where Q is the MW capacity of the line. In this case it will be convenient to specify C, the line's rental cost, in \$/h and c in \$/h, and b in \$/MWh. Sometimes c will be called the fixed cost of the line meaning that it does not depend on the choice of capacity.

that eliminating congestion is almost never the right answer. All transmission lines are used to varying degrees at different times of the day and year. Their maximum potential use will occur for only a few hours. To eliminate congestion it is necessary to build enough capacity to accommodate this maximum, but that means the last megawatt of capacity is used only a tiny fraction of the time and it is almost never economical to build capacity for such infrequent use. It might be argued that the lumpiness of transmission investment will naturally cause a choice between a much-too-small line and a too-large line, and that the economic choice will turn out to be the too-large line and no congestion. Besides the fact that this is unlikely to happen for all lines, there is a deeper problem. As load grows, every line reaches a point where its capacity would be exceeded without re-dispatch for just a few hours per year. To avoid this, a new line, or at least an expansion will be needed, and consumers will have to pay for it. Due to the fixed costs this will come to at least $6000/h in every hour of the year in the above example.

Turn next to the general static optimization problem. This asks what capacity line should be built given the way congestion changes with line capacity. A small line will be congested frequently and congestion rents will be high, while a large-enough line will never be congested. What is the condition for the optimally-sized line?

For simplicity assume the line connects a local region where the total cost of energy production is $C_L(Q)$ to a remote region where it is $C_R(Q)$, and assume power is cheaper in the remote regions so the flow it into the local region. A 1kW expansion of the line will increase production in the remote region and decrease it in the local region by 1kW. If the local marginal cost of power is MC_L and the remote marginal cost, MC_R, then the savings is the difference. In a competitive market, prices equal marginal costs so the savings is (P_L – P_R) times 1kW. The line is worth expanding up to the point where the marginal cost of expansion is equal to the average price differential.

> **Static Optimal Transmission Result (One Line):**
> In an optimal one-line network, the marginal cost of line expansion equals the all-hour average absolute nodal price differential between the ends of the line. In other words, the marginal cost of expansion equals the average congestion rent.

In a network, things are a bit more complex. If we have two lines from A to B and power tends to flow equally on each, but A has much less capacity than B, when line A becomes congested it will limit the flow on line B. This limitation is not a physical restriction, rather the system operator will be forced to limit the total flow on both lines to protect the weaker line. If line A is expanded by 1 MW, this will increase the usefulness of line B by 1 MW. Consequently the value of expanding A is twice what would be computed from the flow on A times the congestion price of A.

> **Static Optimal Transmission Result (Network):**
> Let *d* be the power distribution factor on line A–B calculated as the fraction of power flowing on that line when power is injected at A and withdrawn at B. Then, in an optimal network, the marginal cost of expanding the constrained line equals the average congestion rent divided by *d*.

1.5. Optimal Transmission (dynamic)

Power systems are dynamic. Load grows; generators are built and retire. This dramatically increases the complexity of the optimal transmission-investment problem. To illustrate this, consider the simplest example of a dynamic investment problem. Suppose transmission lines come in two sizes, 600 MW at a rental cost of $5/MWh and 1000 MW at a cost of $4.20/MWh. Suppose that the power transmitted over the path in question starts at zero and grows by 100 MW per year indefinitely and that the savings from transmitting this power is always $6/MWh. Obviously it will be economical to build power lines.

Building the 600-MW line would cost $3000/h, and when load has grown to 500 MW, this would save $3000/h. This is the break-even point, and if the small line is to be built it should be built at this point in time. The alternative is to wait another 2 years until load has grown to 700 MW, at which point the cost and savings of the 1000-MW line would be $4200/h. The smaller lines start saving money sooner, but larger lines will save $1.80/MWh in the long run, while smaller lines will save only $1.00 in the long run. The choice between the two strategies depends on the discount rate, but as can be seen from the graph of savings, it does not require a very low discount rate to make the larger lines the more economic choice.

The conclusion to draw from this dynamic example is that, just because a line is worth building, it should not necessarily be built. It may be better to wait a while and build a larger and more economical project. This is a result of the lumpiness of the investment decision. Note that waiting has no option value in this example because the future is known with certainty.

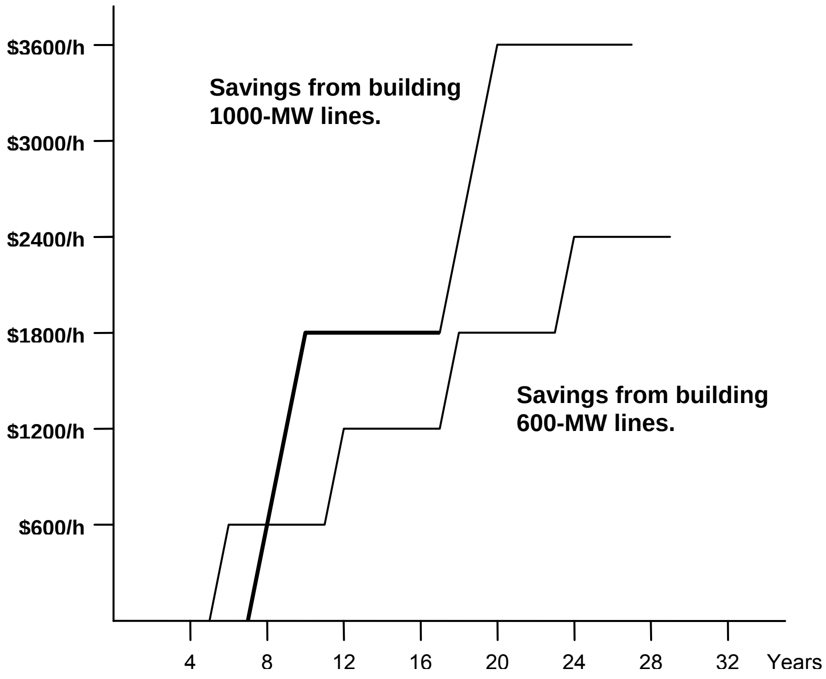

Figure 4. A Positive Present Value Is Not Sufficient.

1.6. Optimal Transmission (option value)

In the above dynamic example, two sets of projects were considered; (1) build small lines at optimal intervals, and (2) build large lines at optimal intervals. Even if there had been uncertainty in this example, each set of projects could have been evaluated to find its expected net present cost of transmission and generation. If all reasonable sets of projects are evaluated in this manner and the one with the lowest cost is chosen, this constitutes a complete dynamic analysis. Unfortunately the selected project may not be the best choice. This is not because expected net present cost is the wrong criterion, but because the set of choices was unnecessarily restricted.

Besides sets of projects, there are also investment strategies. One such strategy might be to build a small line now and then wait until load increased by 600 MW and build another small line if the wait was 8 years or more and build a large line if it was 8 years or less. A strategy is different from a specific “set of projects” because it waits for more information and then chooses one project or another. Quite often, some strategy will be more cost effective than any specific plan.

For example consider a system with no transmission, one city and a possible remote coal plant with a 50% chance of being built. If it is built, the most efficient transmission project would start now and have a net present value of $200 million. If the coal plant is not built, this project would have a net present value of minus \$100 million. Because there is a 50% chance the coal plant will be built, building the line would have an expected net present value of \$50 million ((200 – 100)/2).

If the line were started a year later it would have a (current) net present value \$180 or minus \$90 million depending on whether the coal plant is or is not built. Waiting a year and then building the line thus has an expected net present value of \$45. Other projects could be considered that delay the line for different amounts of time or build lines of different capacity. But it is quite plausible, and this example will assume, that of all specific projects, building the line now is most valuable. In spite of this, there may be a more beneficial strategy for selecting projects.

If in one year, we will know whether or not the coal plant will be built, the strategy of waiting a year and then selecting the best project is more valuable than any particular of project. The expected present value of this strategy is 50% of \$180 million (if the coal plant is built) plus 50% of \$0, if the coal plant is not built. This strategy has an expected present value of \$90 million which is \$40 million greater than the value of building the transmission line now, the most valuable project given today's information. This forty million dollars is called the option value of waiting a year to decide.

There is often a cost of delaying the start of projects with a positive expected net present value, but there is also generally an option value to delaying the decision to go forward with a project. It is only when this cost of delaying is greater than the option value of waiting to decide that a commitment should be made. Option value is difficult to estimate and it should be estimated for various waiting periods. In short, considering option value greatly expands the number of possibilities that must be considered.

2. Cost Recovery for Optimal Transmission Investments

The previous section demonstrated that the optimal grid will suffer congestion which will result in the collection of congestion rents and that these rents are related to the cost of the grid. This raises the question of to what extent congestion rents on the optimal network will cover the cost of that network. To consider this question it is useful to expand the notion of congestion rent. The congestion rent collected on a one line network is $C_R(L_{AB}) = Q\cdot(P_B - P_A)$, where Q is the power flow from A to B on line L_{AB}. This is the trading surplus collected if Q is sold at A and purchased at B. Expanding this concept to the entire grid results in defining the congestion rent for the grid to be the revenue from selling all energy injections at their nodal prices and purchasing all energy withdrawals at their nodal prices. If W_i is the net energy withdrawal at Node i and P_i is the price at Node i, then the congestion rent for the network G is $C_R(G) = \Sigma W_i \cdot P_i$.

Considering only lossless networks, it is possible to decompose the set of net energy withdrawals into a set of bilateral trades each with one injection (negative withdrawal) and one withdrawal of equal magnitude. Each bilateral trade from node A to node B, which can be any two nodes on the network, has associated with it a congestion rent, $C_R(B_{AB}) = Q\cdot(P_B - P_A)$, where Q is now the magnitude of the injection and withdrawal of bilateral trade B.

Notice that one possible decomposition of the net nodal energy withdrawals into bilateral trades corresponds exactly to the power flows on the individual lines. Doing so associates with each line a congestion rent equal to the line's flow times the price at the withdrawal node minus the price at the injection node.

> **The Lossless Congestion-Revenue Result:**
> If the set of all bilateral trades, B, sums to the total net energy withdrawals from the network, then the total congestion rent is the same whether computed by node for the

entire grid, G, as the sum of congestion rents on all lines, L, or as the some of congestion rents on all trades.

$$\mathrm{C_R(G)} = \sum_{\text{All Lines}} \mathrm{C_R(L)} = \sum_{\text{All Trades}} \mathrm{C_R(B)}$$

Proof: Since the injections of bilateral trades are paid the nodal price and withdrawal are charged the nodal price, the net revenue collected at a node is the nodal price times the sum of withdrawals minus the sum of injections. Since these two sums add up to the net withdrawal, the net revenue collected from bilateral trades at node i is just $W_i \cdot P_i$, and over all nodes bilateral trading revenues sum to the congestion rent calculated for the entire network. That congestion rents on lines sum to the same value depends on these power flows summing to the net withdrawals on the network. This follows from conservation of energy in a lossless network. The net power flows on lines into a node must sum to the net withdrawal from the grid at that node.

This result demonstrates that the trading revenue that is collected from buying and selling power in a congested network with nodal prices (even if these prices are not the competitive prices) will exactly cover the congested rents calculated on a line-by-line basis. This assumes there are no power loss, so that the power that flows out of one end of a line equals the power that flowed into the other end. (In reality losses are typically well under 5% on a high voltage transmission system.)

It would be desirable if the trading surplus in an optimally built network were to cover the cost of the transmission network. The above result makes it reasonable to investigate this question by looking at the cost-recovery properties of congestion rent from a single line.

2.1. Congestion Rents Recover Linear Line Costs

Suppose the cost of a transmission line is strictly proportional to the megawatt capacity of the line, so that cost is given by $\mathsf{C} = c \cdot Q$, where c is in \$/MWh and Q is in MW. In this case, according to the Static Optimal Transmission Result (1-Line) given in Section 1.4, the marginal cost of the line, c, should equal the average congestion rent per MWh, $\overline{P}$. As a result, the revenue from the line, $\overline{P}Q$, will equal the cost of the line $c \cdot Q$. In the optimal 1-line (lossless) network with linear costs, congestion rents will cover the cost of transmission capacity exactly.

Not surprisingly this result extends to networks. If the power distribution factor on the Q-MW line from A to B is d, then when line A—B is congested, it controls the power flow of Q/d. If the congestion price from A to B is P, then the congestion rent associated with this constraint is $P \cdot Q/d$ and the benefit of increasing the lines capacity is $P \cdot \Delta Q/d$. In this way the cost-recovery of constraints in a network can be seen to be analogous to the cost-recovery of a single line in a one-line network. An optimally constructed network with linear cost functions and no losses, will recover its fixed costs through marginal-cost (competitive) pricing.

2.2. Congestion Rents and “Fixed” Cost Recovery

When investing in transmission, some costs are roughly independent of the capacity of the line and some are roughly proportional. Those that are independent will be called “fixed” costs in this chapter. The linear cost function just considered has no fixed-cost component, so to add realism, consider one of the form $\mathsf{C} = f + c \cdot Q$. In this case, the same first order conditions will determine the optimal transmission system, so congestion rents will recover $c \cdot Q$, but not f. In other words all costs that are proportional to capacity will be recovered, but none of those that are independent of the line’s capacity will be recovered from congestion rent in an optimally sized system.

The cost function just considered is one example of returns to scale. Another model of returns to scale has a cost of line capacity that is proportional to the square root of capacity, $C = a\, K^{1/2}$. In this case the marginal cost of capacity is $MC = (a/2)\, K^{-1/2}$. If the line is built to the socially optimal level, MC will equal the average congestion rent, which is paid on the whole line capacity, so revenue is $R = (a/2)\, K^{1/2} = C/2$. In other words, at the socially optimal level of investment the line provides a congestion rent of exactly half of what it costs. This is true regardless of how large or small the optimal line is.

Lumpy technology also exhibits fixed costs and, at least, limited returns to scale. But the two concepts, returns to scale and lumpy technology, can be usefully distinguished. Figure 5 below provides the basic intuition.

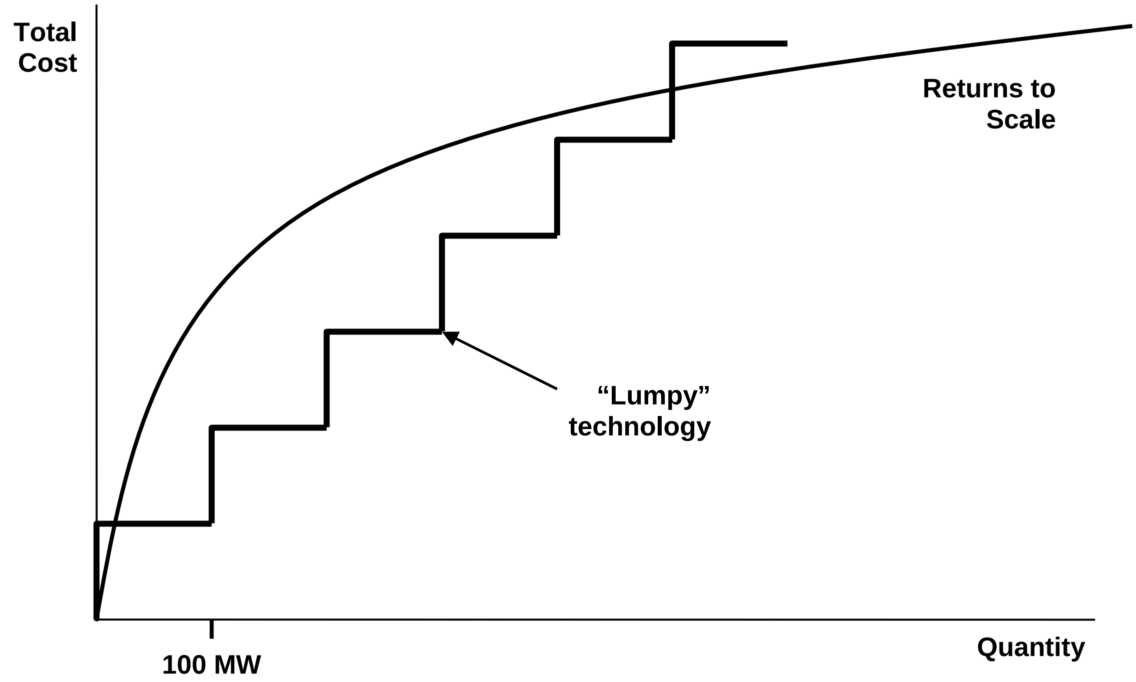

Figure 5. Lumpy Technology May Not Exhibit Returns to Scale in the Long Run

Both types of technology violate the "non-convexity" assumption required for perfect competition. Both types have a fixed-cost component. The lumpy technology shown above requires a fixed cost to provide its first megawatt of capacity, but the marginal cost of capacity is then zero up to 100 MW. In this sense, the cost of this technology is all fixed. On a small scale, up to 100 MW, there are returns to scale; over large changes in capacity, there are no returns to scale.

3. Transmission Planning with Strategic Generation Investment

Under a planning approach, no performance-based incentive mechanisms are applied to the problem of deciding on long-lived transmission investment, mainly lines and transformers. The planning might be carried out by the ISO or by a Transco, and payments for the cost of investment will be carried out under rate-of-return regulation. To simplify discussion, it will be assumed that the planning is done by a Transco that owns and operates the grid, but that there is an ISO which operates the wholesale energy market. Because PBR is not used for long-run investments does not mean it would not be used for the day to day operation of the grid, but the problems of efficient grid operation will not be discussed in this chapter (Joskow, 2004).

As illustrated in the previous section, planning transmission investments is a complex problem, and is in fact far more complex than these illustrations suggest because of the complexity of the network. This is widely recognized. An additional complication is also recognized. Because generation is not planned, transmission planners facing a wholesale power market must forecast the location of generation and load many years in advance.[2]

These problems are common to all three approaches mentioned in the introduction (planning, merchant, and PBR). Under each approach, the dynamics of the combined transmission/generation system and the option value of waiting must be taken into account. Under each approach the investor will not have control or direct knowledge of future generation investment, and under each, the complex cost structure and network externalities cause additional difficulties. This section does not focus on the common problems but on strategic issues peculiar to the planning process when generation investment is deregulated.

3.1. Strategic Manipulation of the Zero-Congestion Policy

The possibility of strategic manipulation is particularly acute in the case of a zero-congestion policy, such as Alberta's. In the example for static transmission optimization, the optimal line size was 700 MW, and if an 800 MW line had been built there would never have been any congestion. This makes it appear that building for zero-congestion would not be too expensive. In this case it would require only a 14% increase in transmission capacity. But consider the case of a wind farm that can be located at various distances from load. If it is close to load, the cost of transmission may be only $1/MWh while if it is in the most remote location it might be $20/MWh. Suppose that the most remote location is the windiest and it is no more expensive to build wind generators there than closer in. What is the effect of a zero congestion policy on the location of this wind farm?

Clearly, it is most profitably for wind generators to locate in the most remote location whether or not the benefit of more wind comes close to offsetting the extra cost of transmission. Moreover, since wind power has nearly zero marginal cost, it is always "in merit," and consequently transmission capacity must be built to accommodate the windiest hour of the year. The last megawatt of such transmission capacity will have almost no value. (One can be sure a zero-congestion regulation will be violated in practice simply because adding capacity to a remote location to capture one hour of supply makes so little sense.) This illustrates the fundamental point of the Strategic-Generation-Investment problem.

> **Strategic-Generation-Investment Problem:**
> The policy of the transmission planners influences the distribution of installed generation. Hence, planning policies cannot be appropriately compared on a static model of generation and transmission but must be compared using a game-theoretic approach that recognizes generation-investment strategies.

This problem complicates transmission planning in a market environment because the problems of predicting generation investment and planning transmission for that investment can no longer be separated. For example, one might think that planning could proceed by having two groups of planners, one of which forecasts generation investment, and the other of which takes the forecast and plans the "best power lines" for that predicted investment. The only communication necessary between the two groups would be the transfer of forecasts from the first group to the second. Unfortunately, the Strategic-Generation-Investment Problem shows that this may not work.

---

[2] This is not as bad as it sounds, because with growth, almost any line that saves more than its rental cost at the time it goes into service will continue to be economic in the long-run.

First, the strategic-investment problem tells us that the forecaster will need to know what policy the transmission planners will follow. This is necessary because generation investors will react to this policy and the forecasters need to understand this reaction to make good forecasts. Second, the planners cannot simply build the best lines for the predicted generation. They must choose an investment policy that induces good generation investment, so they must understand how generation investment will respond to their policy.

3.2. The Difficulty of Implementing Simultaneous Optimization

There is one obvious hope for resolving this chicken-and-egg problem. If the planners choose the "truly best" transmission investment policy—call this the Ideal Transmission Planning Policy, this rule should induce generators to make the best possible investment decisions. In this case, transmission investment policy should be optimal for the implied generation investment incentives, and these should be optimal for the transmission policy. But this doubly optimal system deserves a closer look as it involves more than technical difficulties.

At the opposite end of the spectrum from the zero-congestion policy, is the Ideal Planning Policy. This requires the planner to estimate future load growth and then plan both generation and transmission simultaneously to minimize the total expected present cost of delivered power.[3] Ideal planning is extremely difficult, but, besides the technical difficulty, another problem blocks our path.

> **The Ideal Transmission Planning Policy:**
> Plan transmission and generation together and optimize both for expected load growth, then build that transmission and hope the market induces suppliers to invest in optimal generation.

In principle, this policy works because, given optimal transmission, investors will find the co-optimized generation to be their optimal strategy. But, both joint optimization and any real markets for generation are subject to error. Consequently, there will be times when ideal optimization directs that generation should be built in location X, and a corresponding transmission line be built to serve that generation. If such a line is built and the market decides not to build generation at location X, the planners will be severely embarrassed by their line to nowhere.

The problem to focus on is not the error, but the "embarrassment." Errors are taken into account by our theory of minimizing expected cost. The "embarrassment" causes a more fundamental problem. It prevents planners from adopting the Ideal Transmission Planning Policy.

Planners will not undertake a project that can lead to such an embarrassing situation. Instead, they will simply attempt to optimize their transmission for generation that has been built, is being planned, or at the least, appears to be an obvious extension of an existing trend. They will not predict optimal generation and build transmission for that prediction. It might seem that this will make no fundamental difference, but it does. The Strategic Generation Problem tells us that as soon as the generators realize the actual planning policy is no longer the Ideal Planning Policy, they will no longer invest optimally for that Policy. Instead they will invest optimally for the new policy, and that is an entirely different matter.

Transmission planners will not build for theoretically determined future generation investments but will instead build for generation as determined by generation investors. This means generation investors can manipulate the transmission planners by their selection of generation investment sites. They will learn what policy the planners are using and game that policy. This is the Strategic-Generation-Investment Problem.

---

[3] In fact it requires even more. Load is to a small extent determined endogenously by the price of power, so it should be determined simultaneously, but this complication will be ignored.

3.3. Strategic Manipulation of Optimal Transmission Planning

Having given up on the Ideal Transmission Planning Policy, the question becomes, what is the best realistic planning policy? There are many choices, but only a few can be stated simply. One of these, the zero-congestion policy has already been ruled out. Can we find a better one? The next-best practical alternative will be considered. This specifies that the planner should optimize transmission taking generation and anticipated generation as given. This will be called the 'Practical Planning Policy."

The Practical Planning Policy:
Build the transmission system that is optimal given actual and anticipated generation.

If planners follow the Practical Planning Policy, and generation decides mistakenly or perversely to locate in a remote region where fuel is cheap but transmission is so expensive that the combination is uneconomical, the planners may well have to build accommodating transmission.[4] Although the result may not be optimal, it will be far more sensible than building to the point of zero congestion. To analyze the planner's choice, consider the net social benefit of a transmission and generation project.

This project consists of remote generation which "exports" power over a transmission line to the central market. Net Social Benefit consists of (1) the net benefit to consumers, (2) generation profits and (3) transmission profits. To simplify the calculation, assume the entire project is small compared with the central market and that long-run supply in the central market is very elastic. This implies that the remote generation project will not change the price paid by central consumers, and as a consequence it will be of no net benefit to them. The project, if efficient, will displace central production with cheaper remote production and delivery, but the savings will be entirely captured by the investors. As a consequence Net Social Benefit reduces to generation profits plus transmission profits.

Net Social Benefit = Generation Profits + Transmission Profits

Assume that remote generators are paid one price and central consumers are charged a higher price which creates congestion rent. From this, the cost of the transmission investment can be subtracted to find Transmission Profits, which can be negative or positive.

Net Social Benefit = Generation Profits + (Congestion Rents – Line Costs)

In effect, remote generation is paying the congestion rents because the central price is determined by the central-market cost of supply. If remote generation is required to pay the full cost of the transmission line through congestion rents, then Net Social Benefit equals Generation Profits, and profit maximization by generators will maximize Net Social Benefit. In this case suppliers will invest optimally in generation and if the planners follow the Practical Planning Policy, the combined project will be optimal.

For linear (proportional) transmission costs, congestion rents do cover line costs for the optimal line. Consequently if transmission costs are linear, the Practical Planning Policy will induce optimal generation investment and the optimal transmission investment for generation. The combined system will be optimal.

The corollary of this result is that when transmission costs are not linear, investment is unlikely to be optimal. For example, consider again a wind farm. Under the Zero-Congestion Policy, the wind farm could locate as far as it liked from the central market with complete impunity. Under the Practical Planning Policy, the planner will only build the optimal transmission line and the further from the central market that the wind farm locates, the smaller will be that line. The marginal cost per MW of line capacity increases roughly in proportion to the length of the line for any line capacity. Since the optimal

---

[4] This section concerns strategic manipulation by generation, but it assumes that the wholesale generation market is in every respect competitive. Manipulation is not the result of market power, but of generation's first-mover advantage in the optimization game.

line capacity occurs at the point where the marginal cost of line capacity equals the average congestion rent, the longer the optimal line, the greater the average congestion rent. Since the wind farm will pay these rents, it will care how long the line is.

This is good news. The Practical Planning Policy should not only build more economic lines for existing generation, but it should also induce the existence of a more efficient spatial distribution of generation. The combined saving should be significant and perhaps enormous. But this does not imply that the induced generation investment will be optimal under the Practical Planning Policy.

Suppose the cost of transmission is $c\sqrt{K}$, which implies that for an optimal line, congestion rent is exactly half the total cost of the line. The wind-farm investors will realize that they must pay only half the cost of the line since, under the example's assumptions, the only transmission cost they pay is the congestion rent. When they differentiate profit with respect to distance, x, from the central market in order to optimize their location, their value for the dC_R/dx will be half what it should be and this will cause them to locate too far from the central market. The planners will then be forced to build a line that is optimal for generation located too far away, but the result will be a combined generation-transmission project that is sub-optimal.

The obvious remedy for this problem is to charge generation the difference between the congestion rent and the cost of the line. This implies a charge that varies by location and some measure of a generator's size, but that is quite different from a congestion charge. The need for such charges has long been recognized (Vogelsang, 2004; Brunekreeft, Neuhoff and Newbery, 2004) and has been the focus of many economic schemes, few of which have been informed by legitimate economic analysis. Many of these go under the name of "MW-mile" charges, though the most elaborate one was an Alberta scheme called SERP. It based locational charges on an analysis of line-by-line power flows that would be caused by a short-circuit at the location in question.

Without such a scheme, the fallback position is to charge every MW delivered to the system a fixed (independent of location) charge per MWh. This has the advantage of preserving short-run efficiency by leaving the dispatch unaffected. It will, however, leave the Practical Planning Policy sending inefficient long-run signals for generation location.

These insights define a fundamental problem for transmission planning in the context of wholesale generation market that includes market-driven investment.

> **The Fundamental Transmission-Planning Problem:**
> Is there a mechanism for collecting transmission fixed costs which, when coupled with the Practical Planning Policy and nodal pricing based on competitive energy prices, will improve the total efficiency of the power system? Both short-run dispatch efficiency and long-run incentives for generation investment must be considered. The standard of comparison is a per-MWh charge on all supplied energy independent of location.

Although it seems unlikely that any charging mechanism can be found that will be both short-run and long-run optimal, it does seem likely that some improvement is possible at least for realistic networks. Because so many erroneous collection mechanisms have already wasted time and money, none should be seriously considered until some proof is given that they will lead to improved efficiency on some well specified collection of power systems. These test cases can be gross simplifications of real power system, but they must include the possibility of locational choices for generation investment.

4. Merchant Transmission Investment

If transmission is not built by a public or regulated entity, it will be built by private investors. This happened before the power industry was regulated and it happens today. But the question of interest is how much will be built and how efficiently it will be built. Certainly there are no easy assurances from theoretical economics that an unregulated market will perform well, and to date there is no empirical evidence for this proposition. There are, however, several theoretical reasons for concern.

Returns to scale suggest that investors will need market power to recover their fixed costs, but market power, in this industry as in all others, leads to underinvestment. Externalities generated by the use of the grid for trade suggest there will be free-rider problems, which will exacerbate the under-investment problem. Transmission investments provides two other externalities, the value of which private investors will not easily capture. First it reduces market power in the generation market (Stoft, 1997; Borenstein *et al*. 2000), and second it provides reliability. Generation market power is a particularly knotty problem because energy suppliers are known to lobby government bodies in an attempt to block transmission investment that is not in their interest. Any attempt to reduce market power with transmission is likely to be the target of supplier lobbying.

In spite of these difficulties, transmission investment is not always as difficult to finance as many assume. Often congestion rents are viewed as the sole source of remuneration to merchant transmission. In fact, lines may be built without any assumption of congestion income. If there is a cheap but remote area for generation investment, the suppliers that locate there may build lines simply to bring their power to market. They will still have problems with free-rider problems and the like, but they will be motivated by other than future income from congestion rents. Similarly a city may find local supply too expensive and may build transmission out to the larger network simply to access cheaper power with no thought of future congestion rents. Alternately all three motivations may coincide.

Although congestion rent may not be the primary motivation for building lines, rights to new lines should be given to investors to encourage such investment and internalize the line’s benefits to the extent possible. Transmission rights can protect an investor from congestion cost that would have otherwise been paid if the investor used the line and can provide income to the extent others use it when it is congested. Another benefit of transmission rights is to cause negative externalities to be internalized. Before turning to the difficulties of merchant investment, it is worthwhile understanding how transmission rights should be granted in return for investment and what useful role they can play.

4.1. Transmission Rights

Because alternating current (AC) transmission lines are so thoroughly integrated with the power grid, their owners generally cannot be given physical control or the right to charge anyone who uses the line. For example, almost any power flow from one point to another on a power system causes some power to flow on most transmission lines in the grid, although the amount that flows on remote lines is too small to matter. AC power lines are in some ways like a set of connected water pipes without valves between them. Pushing water from one point to another affects the flow in almost every pipe. Because of such physical complexities, the standard proposal is to reward the investor in a transmission line with a set of financial rights, not physical rights.

The standard financial transmission right is a Congestion Revenue Right (CRR) which is defined by a quantity, Q, source and sink, A and B, and a set of time intervals, T. At any point in time during T, the CRR pays $(P_B - P_A)Q$, which is the congestion price from A to B times the megawatt quantity of the right. The payment has nothing to do with actual power flows associated with the owner of the right.

Although CRRs can, in principle, be privately issued, they are generally issued by the ISO and that will be assumed throughout this discussion. Consequently, at any point in time, there is a well defined set of CRRs, $\Re$, that have been issued. An important property of $\Re$ is its feasibility, which is defined as follows. Corresponding to every CRR there is an imaginary power flow of *Q* MW from A to B, during time intervals T. This imaginary power flow has nothing to do with actual flows on the grid. Since every CRR in this set corresponds to an imaginary power flow, we define $\Re$ to be a feasibly set of rights if the corresponding set of imaginary power flows could take place on the system without violating any reliability constraint. This has nothing to do with load or generation and concerns only the transmission system.

The following procedure can be used, at least in principle, to reward investors in transmission upgrades. First, sell a set of CRRs, in an auction that does not withhold any feasible CRR. This should leave no valuable CRR unallocated. When a transmission upgrade is completed the system should be able to accommodate more power flow reliably and this should expand the feasible set of CRRs. The investor is allowed to claim any set of CRRs which, combined with the existing set, forms a feasible set in the upgraded system. This allows the investor a certain amount of choice and it accounts, to some extent, for positive external affects of the upgrade. There is one more rule. If the "upgrade" has actually reduced the feasible set of CRRs the investor must take counter-flow CRRs such that the new allocated set is feasible. These will have negative financial value and will properly discourage any system downgrades, provided the initial set of CRRs matched the actual flows on the system (Bushnell and Stoft, 1996 and 1997). Unifying these two rules, the reward for a modification of the transmission system is this:

Feasible CRR Allocation Rule for Rewarding Transmission Investment:
The modifier of a transmission system must take from the ISO a set of CRRs such that together with the pre-existing CRRs the new complete set is a feasible set on the modified system.

This approach to rewarding investment has several advantages. First, if the pre-existing set of CRRs matches the flows on the system, it makes it unprofitable to damage the system. Second, it gives the investor the maximum possible congestion rent while treating others fairly. Third, for feasible sets of rights, the cost to the ISO of paying CRRs is never more than the congestion rent collected.[5] Unfortunately there are also a number of drawbacks. First, the awarded CRRs do not adequately compensate investors. Second, to live up to its potential, the set of CRRs that the ISO makes available needs to be quite complex. For example, the investor may want north-south rights at some times and south-north rights at other times. Third the results assume investors have no market power in the energy market.

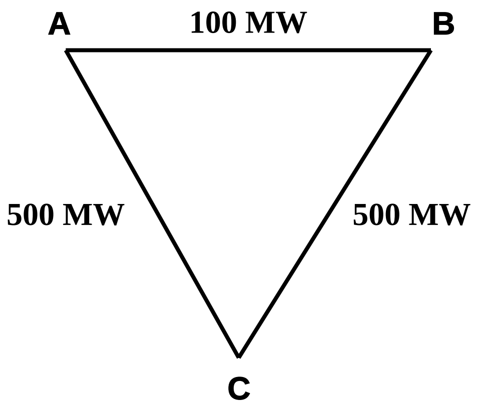

Figure 6. Option Rights Reduce the Feasible Set of Rights.

Other Styles of Rights. There are number of other styles of transmission rights (Hogan, 2002; Gribik *et al*., 2005; Baldick, 2005). PJM uses Financial Transmission Rights, FTRs, which in their original (obligation) form are identical to CRRs except that the revenue associated with the entire set of rights is adjusted to equal the congestion rent collected from the entire transmission system. This correction is generally small. The FTR option (as opposed to obligation), introduced into PJM in 2003, is a fundamentally different from of transmission right. At any point in time

---

[5] This assumes the set of feasible transmission flows is convex, which is not quite true, and that the feasible set of right is based on average feasible power flows and not the power flows that are possible only under ideal conditions.

an FTR option from A to B pays its holder the maximum of what a similar FTR obligation pays, or zero. Unlike FTR obligations, FTR options, like other options, have only non-negative values.

Using FTR options instead of obligations results in a smaller feasible sets of rights because cancellation of rights in opposite directions is not allowed under the normal meaning of “feasible.” For example consider a three node triangular network with generation at A and B and with load at C, and with a 100 MW line from A to B and two 500 MW lines to C. With the standard, obligation-style, CRRs, two 500 MW FTRs, one from A to C and the other from B to C, are simultaneously feasible. With FTR options, only two 300 MW FTRs are feasible because if one option were not exercised, the other would correspond to a flow that would load the A—B line to its limit.

Now, the power flows used to compute the feasibility of a set of FTRs are purely imaginary, so the same feasibility rule could be used with options as with FTR obligations. But in this case, two 5,000 MW *options*, one from A to C and the other from C to A would be feasible (because these flows cancel). The option in the congested direction, from A to C, would earn a payment of 5,000 MW times the congestion price while the ISO would collect only a congestion payment of, at most, 600 MW from A to C. Since the counter-flow option would not have a negative value and would pay the ISO nothing, the ISO would find itself short 4400 MW of congestion rent. For this reason PJM computes feasibility of its options rights as follows:

> The FTR Auctions maximize the quote based bid value of a set of simultaneous feasible FTRs awarded in the auction. To ensure feasibility, counterflow created by an FTR Option bid must be ignored when FTRs bids are tested for feasibility.[6]

There may be no economic justification for the ISO to create option rights. The system operator is in a good position to create standard CRRs because it can back them with congestion revenue without risk and consequently does not need to charge any risk premium for them. But because adding options to the mix reduces the total amount of hedging available, it may be better to let a private derivatives market supply CRR options if there is a demand for them.

The motivation for these alternative styles of rights, and for that matter the motivation for transmission rights in general, is mainly to allow the hedging of energy transactions. For example the main FTR page of the PJM web site states that “Financial Transmission Rights (FTRs) [are provided] to assist market participants in hedging price risk when delivering energy on the grid. … The FTRs provide a hedging mechanism that can be traded separately from transmission service. This gives all market participants the ability to gain price certainty when delivering energy across PJM.”[7] No mention is made of transmission investment.

Market power is another area of concern with transmission rights (Joskow and Tirole, 2000). This also is not closely related to transmission investment. If an energy supplier has market power in a load pocket, it can enhance its power by purchasing transmission rights into the pocket. These rights will pay more when it raises the local price of energy, which makes its exercise of market power more profitable.

Presently, the main use of financial transmission rights has been as a substitute for prior rights held by transmission owners. This has been quite useful because of the compatibility between CRRs and nodal pricing. This substitution and the more general use of transmission rights has also provided a useful hedging mechanism for nodal price differences, i.e. congestion rents. The value of CRRs in this regard is

---

[6] From PJM’s document FTR Market Frequently Asked Questions, Updated Feb. 1, 2005, question 49.

[7] Found on www.pjm.com/markets/ftr/ftr.html, the web site of PJM Interconnection, a regional transmission organization (RTO) that coordinates wholesale power trading in a region of the U.S. from Michigan to Washington DC, to Kentucky. It includes 44 million customers and 135,000 MW of generating capacity.

still not well documented, but they seem to be well accepted in this role. Although their use as a partial incentive for transmission investment has long been advocated, and at least PJM and NYISO have rules in place to this effect, there does not seem to be any documented instance of CRRs playing a significant role in any merchant transmission project.

4.2. The Paradox of Transmission Rights

The appeal of rewarding transmission investment with CRRs comes in part from their properties in an idealized world of perfect competition. In this economic model, marginal investments are always possible and their cost is linear. Consequently a line may be upgraded by one megawatt for 1/100 the cost of a 100 megawatt upgrade. Moreover, any investor can upgrade any line; there is no ownership of the transmission path. This brings perfect competition to each line in the system. In such a system the congestion rent that would be earned by a marginal upgrade of a transmission path would exactly equal the value of the upgrade in reducing the redispatch cost caused by congestion. For example, if a line is congested for 1000 hours per year with a price differential of $10/MWh, then a 1 kW (1/1000 of a MW) upgrade of the line would save $10 × 1/1000 × 1000, or $10 per year in redispatch cost by allowing cheaper generation to substitute for more expensive generation. Similarly, if the investor is granted a CRR for 1/kW in the direction of the congestion, the investor will earn $10 per year, exactly what the line is worth.

Investment on every path will proceed as long as the marginal investment costs less than the congestion rent earned by that investment. Because this equals the value of line expansion, investment will proceed exactly to the point where is no longer pays to continue investing. Supposes a 100 MW line needs a 50 MW expansion because of a new load. Some investor may build 30 MW of the expansion, but at that point further investment may become unprofitable because it lowers the congestion rent on the 30 MW of transmission rights received for the initial investment. In the real world this would most likely stop investment before the optimal transmission capacity is achieved, but in the idealized world of perfect competition, some other investor will continue the investment, perhaps for another 10 MW. Then as this investor’s stake in high congestion rents discourages further investment, yet another investor will take over the project. In this way every last kilowatt of economic investment will be made.

Under the heroic assumptions of perfect competition, rewarding investors with all of the congestion rents provides the ideal incentive for investment (provided the allocation of rents is also ideal). Under realistic assumptions, which include market power, paying investor more when there is more congestion on their line results in withholding of investment and too much congestion. As will be discussed in Section 5, the exact opposite payment scheme has merit when the investor is a monopoly Transco. In this case charging the investor the amount of the congestion rent instead of paying the amount of the congestion rent results in an ideal investment incentive.

4.3. Returns to Scale and “Lumpiness”

Returns to scale, as discussed above, means that optimal transmission investments will simply not generate enough congestion rent to pay for themselves. Obviously, this means merchant investors will build less than the socially optimal level of investment (Joskow, 2003; Joskow and Tirole 2004).

The same does not hold for lumpy technology. Consider Figure 7, which shows transmission investment coming in lumps. If the rental cost of a lump of transmission is $8/MWh, it would not be worth building the second lump because the re-dispatch “Cost” triangle that would be eliminated averages less than $8/MWh.[8] Thus the social optimum is to build only one lump and the rent on this lump will be

---

[8] At the left is the cost is $16/MWh, but it falls linearly to zero before reaching the right end of the lump.

$16/MWh which pays for the line twice over. This shows that with lumpy technology, socially optimal investment may produce more than enough congestion rent to pay the cost of that investment.

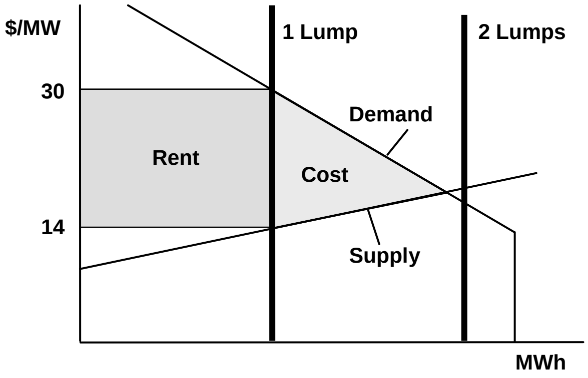

Figure 7. Optimal Investment in Lumpy Technology May Be Profitable.

If we call the lumpy technology in Figure above, linear-lumpy technology, indicating it has constant returns to scale over a range of lumps, we can ask the question: Would merchant investors under-invest in linear lumpy technology? Figure 7 shows that they might invest optimally and earn well above a normal rate of return. If demand were a bit greater in Figure 7, so that the intersection of supply and demand were to the right of "2 Lumps," the social optimum would be two lumps. However, merchant investors would fail to build the second lump because it would earn almost no congestion rent. The result would be under-investment and excess profits. Of the three possibilities, under-investment, optimal investment and over-investment, they will avoid the third. A static analysis would conclude that this should lead to too little investment on average but to excess profits. The disconcerting aspect of this conclusion is that it appears they will on average, in fact almost always, earn above-market rates of return. This seems unlikely.

In a dynamic market this result appears even more suspicious. Investors will anticipate the possibility of high rates of returns and will invest a little early, which will increase their average investment and lower their rate of return. Considering this dynamic effect, there seems little reason to suspect that lumpy technology will lead to systematic under investment, or even that investment will be wrong on average. This is an optimistic result, but one more effect needs consideration.

Lumpy investments pay least when they are first made and most just before the next investment is completed. In fact, many optimal transmission investments lead to a protracted period with virtually no congestion and hence no congestion rent when they are first completed. This effect may be dramatic. When a lumpy upgrade is first made it is common for the line to be almost completely uncongested and this situation may last for years. The island of Nantucket off the coast of Massachusetts is served by a single DC cable which is about to become congested for a few hours in August. This might cause a very partial blackout during the tourist season which is unacceptable, so a second identical cable will soon be added. This one will not be congested for perhaps another twenty years. Even when a lumpy transmission investment can expect full recovery of its costs from congestion rents over the long run, the cost recovery may well not begin for years and will be very slow when it starts. This "back-end loading" of the revenue stream creates grave risks for the investor. What if a new technology, such as cheaper high-voltage DC lines or Aluminum-Zirconium wires, comes on the market before high levels of congestion kick in? What if load growth is slower than anticipated? What if gas pipelines are built to fuel new generation that competes with power imported on the transmission line (Barthold, 2003)?

This investor's payment stream does not mirror the stream of social benefit which results from the elimination of, or reduction in, previous congestion rents. Because that benefit stream starts out at the rental cost of the line, it is far less risky than the stream of congestion rent. Risk is costly, so merchant investment based on collecting congestion rents from CRRs issued in return for the investment will be much more costly then a socially sponsored investment in the same project. Low-risk investing is cheaper than high-risk investing.

A simple example may help explain the relationship between the social benefit stream and the congestion rent stream on a lumpy transmission investment. Suppose load in a load pocket takes on values between X and X + 200 MW with a uniform probability distribution. Suppose the price difference between supply from the load pocket and external supply is \$20/MWh. Suppose additional transmission costs \$5/MWh and comes in 100 MW lumps. When should transmission be built?

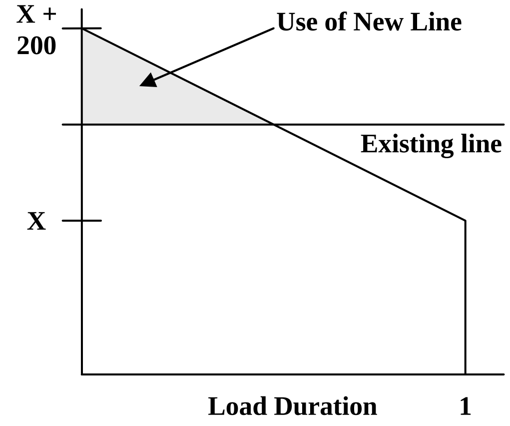

Figure 8. Optimal Investment Eliminates Congestion

Only when the present line is congested would a new line add value. When the line is congested half the time, as shown in Figure 8, the value added will range from zero when it is just barely congested to \$20/MWh of new-line capacity when the load is at its peak value and the new line is fully utilized. On average, under these conditions the new line would provide \$10/MWh of line capacity in social benefit while congested and \$0/MWh while not congested. Consequently when the existing line is congested half of the time, it will provide \$5/MWh of benefit on average. Since this is what the line costs to build, this is the break even point. When the line is congested less often, investment is not warranted.

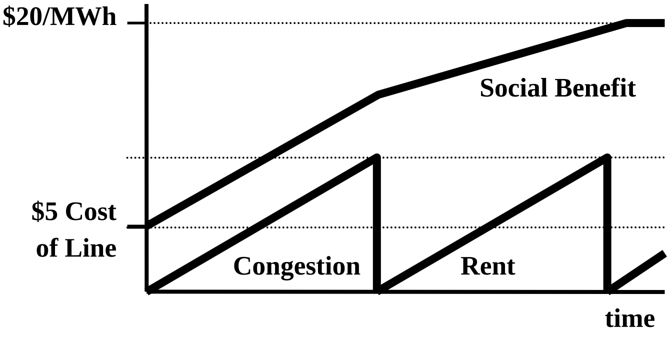

Figure 9. Investors Should Not Capture Full Social Benefit.

If new lumps of transmission are added at the socially optimal times, congestion rent just before the new addition will be \$10/MWh on average, and just after the new addition it will fall to zero. When the new line is first put in place the line will earn nothing, but its earnings will grow until the next lump of transmission is built. If the growth in X is linear, then, on average, the line will earn \$5/MWh, exactly enough to cover its cost. (This example assumes a zero discount rate.) At least in this case of lumpy investment, optimal investing will be rewarded with exactly the right level of congestion rents.

Notice the difference between the social benefit from a transmission upgrade which starts out covering the rental cost of the line on day one and the stream of congestion rents which flow to the merchant investor. These start out at zero and only reach the break-even point half way to the point in time when the next investment will be made and rent will again drop to zero. Also notice that the social benefit of building the line is much greater than its cost. It is normal to find such “consumer surpluses” in a well functioning market. Obviously, this is a very narrow result, but it disproves a common view which holds that lumpy investments inherently underpay investors similar to the way in which increasing returns to scale (a declining marginal cost curve) underpays optimal investing. The main problem with lumpy investments is that they pay off merchant investors very late, which makes them extremely risky for a merchant investor even though risk in social benefit is low. This can greatly increase the cost of merchant lines relative to their cost if built under rate of return regulation.

4.4. Free Riders?

When large merchant transmission projects are contemplated it is often noted that many will benefit from such a project in the initial years but all will attempt to avoid paying for it. As soon as the line is completed certain loads will find their prices reduced and certain suppliers will find they can sell at a higher price. Both will want the line to be built, but all will want others to pay for it. Figure 9 may illustrate such a situation.

In the first year after the line is built, market participants will benefit by $5/MWh of line capacity, yet the investors will be paid next to nothing in congestion rent. This observation usually leads directly to the conclusion that all those who benefit without paying are free riding. But, in this example, that is not the case. The investor will be paid in full and the gap between social benefit and congestion rent is not the result of free riding but is simply normal consumer surplus.

Of course, transmission projects are likely to suffer from the effects of returns to scale as well as lumpiness, so it is likely that investors will be underpaid if they depend on congestion rents. In this case free riding is a correct diagnosis, but it will be extremely difficult to asses the extent of the free riding. In particular it is wrong to believe that investors should capture the entire consumer surplus even at the beginning.

An investor with market power may be able to capture some of the value that would otherwise accrue to free riders. Similarly, an investor with market power may be able to capture part of the normal consumer surplus that would be provided by optimal investing and complete fixed-cost recovery. Both reasons probably help to explain the many proposals to allow the exercise of market power by merchant transmission investors. Some market power would help cover investment costs, so there is some legitimacy to the suggestion. But once started down this path, merchant investors may well ask for more market power than needed to break even. Proposals to base the fundamentals of cost recovery on market power fail to provide a rational for a market that must rely largely on market power to cover costs. The investment efficiency of such a market is unclear at best.

The principle argument for efficiency in such a market is that suppliers with market power still minimize costs, given their output. But this argument also holds for suppliers who do not decide what transmission is needed but simply respond to an auction held by the system operator for the provision of certain transmission capacity. Until some means can be found of tailoring the exercise of market power to provide the right level of fixed cost recovery only on efficient transmission investments, the argument for the deliberate introduction of market power as a method of inducing investment is weak to non-existent.

4.5. Mixing Planned and Merchant Transmission

The possibility of planned transmission both discourages and threatens merchant transmission. Before a merchant line is built, potential subscribers to the project would prefer to induce the planners to build the line and spread the cost over the broader market. This discourages participation in the project by those who should buy a long-run contract for the use of the line. Essentially, this exacerbates the free rider problem.

Once a merchant line has been built, those who have not pre-paid for its use will still wish to encourage the planners to build a competing line, as actually happened in Australia (Firecone, 2003; Littlechild, 2004). The merchant line, being a direct-current (DC) line could directly charge those who used it (Brunekreeft, 2004b), but had it been an AC line those wishing to use the path would still have reason to lobby the authorities to over-invest by building a second line, thus driving down congestion rents and providing a cheaper alternative than use of the merchant line. This possibility is a threat to merchant investment.

Because merchant transmission has so far proven itself entirely inadequate, some have suggested a mixture of merchant and planned transmission (Hogan, 1998; Rotger and Felder, 2001; Chandley and Hogan, 2002). For this to succeed, the level of discouragement and threat must be reduced. This can be accomplished if the role of planning can be defined with enough clarity. If merchant investors know which lines will and will not be built by the planners, then planning should not discourage investment in lines that will be needed for merchant purposes, but which the planners will not build.

One suggestion for the bright line between planned lines and merchant lines is that planners will build only lines which would require the cooperation of “many parties” because in this case the free rider problem is thought to be particularly severe. While this would give merchants clear guidance at the two extremes, there would inevitably be an important middle ground of ambiguity, perhaps encompassing most transmission expenditures.

Another proposal for separation was made several years ago in the context of the Alberta market (Stoft 2002). That market, to a greater extent than most, fails to send adequate locational signals for generation investment and, like all power markets, lacks proper real-time demand elasticity. Consequently, transmission investment is occasionally required for reliability purposes. It was proposed that the planners build only for reliability, and that when such a project is undertaken, merchant investment be allowed to expand the project for the incremental cost of the expansion, thus avoiding significant fixed costs. To further facilitate merchant investment, it was proposed that the Transmission Administrator (planner) also facilitate joint investments by joining merchant projects under certain circumstances when lumpiness is a problem. The Transmission Administrator would buy a part of the line and keep it out of use until a new party decided to purchase it. This proposal was not viewed as ideal, but only as a better alternative than the rule Alberta eventually did implement, requiring that congestion be completely eliminated. It also has the advantage of not depending on or blessing the exercise of market power.

5. Performance Based Regulation for Transmission Monopolies (Transcos)

The planning process provides non-directive and generally weak incentives. Planners know that if they do a demonstrably poor job, they may find themselves out of work. This provides an incentive and most engineers are actually quite motivated by this and by professional pride and a desire for professional recognition. Consequently, it is a mistake to believe the planning approach lacks good incentives. However, these incentives may differ from a pure incentive to minimize the total cost of delivered power and may put too much weight or not enough weight on reducing complaints about the occasional outage or about congestion that inhibits trade. Consequently, it may be better to design explicit formulas that determine monetary rewards for minimizing total cost. These rewards cannot be easily applied to individuals, so the standard approach is to apply them to the profits of a regulated monopoly, a Transco.

Any regulation of a monopoly provides financial incentives, but often these have not been explicitly designed or even considered. The incentives of cost of service regulation are usually poorly thought out and derive mainly from unintentional lags in rate setting and the subjective application of rules such as the requirement that investments must be “used and useful.” When financial incentives are explicitly designed, the regulation is called “performance based regulation” (PBR), or “incentive regulation.”

5.1. A Direct Approach to PBR for Transmission Investment

If it is assumed that power is available to all customers at the competitive price, then the objective of transmission investment is the minimization of the total cost of delivered power. Without the assumption of availability, cost could be minimized by maximizing blackouts, and in the limit, delivering no power at all. Because of occasional blackouts (load shedding events), the cost-minimization framework must be maintained by assigning a cost to "un-served load." Assuming this assignment of cost can be accomplished, the goal of transmission investment is total cost minimization.

This goal is easily translated into a theoretical scheme for incentive regulation (Gans and King, 2000; Léautier, 2000). A monopoly Transco should be paid a fixed but generous sum, R, per megawatt of delivered power less the cost of congestion (C_E, the re-dispatch cost) and less the cost of un-served (lost) load, C_{LL}. In most power systems, \$10/MW hour of delivered power would be more than sufficient for R[9]. The Transco's profits would then be

Profit = $R - C_E - C_{LL} - CT$, where CT = the rental cost of the transmission system.

As with a single transmission line, the rental cost of the system includes the cost of capital as well as maintenance. Note that the Transco does not keep the congestion rent. If the Transco can reduce the sum of C_E and C_{LL} by more than \$1 by investing and thereby raising CT by \$1, it will find it profitable to do so and this will be beneficial to society. Hence this incentive mechanism aligns the Transco's incentives with social welfare. Any reduction in $C_E + C_{LL} + CT$ increases the social surplus by the same amount, and this amount goes into the pocket of Transco.

Notice that this incentive scheme properly rewards "effort" which has a non-monetary cost to the Transco and is consequently not observable by the regulator. If the Transco can increase its monetary profit, as defined above, by \$10, but only by expending \$9 worth of unobservable effort, it will pay it to do so, as it should, since this is socially beneficial. As will be seen shortly, this property is only shared by what are called "high powered" incentives mechanisms.

This scheme presents three difficulties, measuring C_E and C_{LL}, and setting R. Congestion costs, C_E, are the difference between the actual wholesale cost of energy and the lower cost that could be achieved without any transmission limits. The former is known in any system that publishes nodal prices, and the later is usually not very difficult for the system operator to compute. Occasionally there may be some difficulty with knowing how much certain generators could produce were they not constrained by the transmission system, but most systems have on record a realistic estimate of each generator's output capacity and this should serve as an adequate proxy for the true value. (Losses are quite easily estimated.)

The cost of un-served load is far more problematic. The standard error for such an estimate is probably a factor of three. In other words, if it is estimated to be \$15,000/MWh, there is probably only a 68% chance (roughly) that the true value is between \$5,000/MWh and \$45,000/MWh. (This is a purely subjective conjecture.) More problematic is that fact that this cost occurs erratically. Major cascading blackouts may happen about once every 20 years and result in half the load being lost for six hours. The cost of such a blackout would be $6 \times 0.5 \times \$15{,}000$ per MW of load which is equivalent to about \$7/MWh for an entire year. When such a blackout does occur it could cost the Transco a year's revenue and result in bankruptcy.

Perhaps a solution to this is to undervalue lost load by a factor of ten or twenty and use a nominal value such as \$1000/MWh. This would still result in a rather erratic cost stream, but it might be tolerable. Certainly, consumers would pay a significant risk premium to a Transco under such an incentive. Sometimes it will also be difficult to tell whether loss of load is due to generation or transmission problems, and this will lead to litigation costs and other inefficiencies.

The problem of setting R is the most fundamental problem of regulation. The regulator always has less knowledge (information) of the cost minimizing solution than does the regulated firm. Because of this, the regulated firm can extract some "information rent" from the regulator. In general the stronger the incentive, the better the performance of the regulated firm, but the more rent it can extract. The present

---

[9] For completeness the cost of replacing losses should be included in C_E

scheme provides the strongest incentive; the Transco keeps every dollar of cost that it saves. As a consequence it will be able to extract considerable rent. Specifically, the regulator knows that if it sets R below total cost, $C_E + C_{LL} + CT$, the Transco will go out of business, a result that must be avoided, but it does not have a good estimate of total cost. Its only reasonable choice is to set R three or four standard deviations above the cost-minimizing investment level of $C_E + C_{LL} + CT$. Given the level of uncertainty in estimating this total, this is likely to result in an extremely high rate of return for the Transco.

5.2. Two Approaches to Reducing Information rent

The advantage of the PBR scheme just described is that it strongly motivates the supplier to expend cost-saving *effort* that is difficult if not impossible for the regulator to observe. The disadvantage, as we have just seen, is that it requires the payment of high information rents. In the case of transmission investment, these rents could be extremely high and well beyond any acceptable level.

Because of such high information rents, it is desirable to reduce the power of the incentive mechanism. Many types of PBR are forms of price-cap regulation including the mechanism just described, though it is a rather non-standard price cap incentive. In that scheme, R can be thought of as the fixed part of a two-part transmission price, the other part being the congestion price. There are two standard ways of reducing the power of a price-cap incentive. First, the cap, R in this case, can be reset periodically. The more frequently it is reset, the lower the power of the incentive it provides. Second, profits under the mechanism can be shared between the monopolist and the consumers. The smaller the share kept by the monopolist, the lower the power of the incentive. In both cases, lower power will correspond to lower information rents paid to the Transco.

Consider the periodic resetting of the price cap. When the price cap is reset, the objective is to provide the monopolist with a certain allowed rate of return during the next period. To this end, the values of C_E , C_{LL} and CT will be estimated for that period. If the period is short, most of the Transco's costs (CT) will be correctly anticipated and covered by the allowed rate of return. The longer the period, the more cost will be saved or incurred unexpectedly. This will lead to unexpected changes in C_E , C_{LL} which will change revenues. These intra-period changes in expenditure (CT) and revenue ($R – C_E – C_{LL}$) result in profit deviations from the targeted allowed rate of return and this provides some incentives for both cost minimization and beneficial investment. The longer the period between rate cases, the greater the proportion of expenditures for which the price-cap mechanism can provide an incentive. At one extreme lies pure price-cap regulation and at the other pure rate-of-return regulation. In between we find actual rate-of-return regulation in which price-caps are reset roughly every three years.

Timing is the major problem with using periodic price-cap setting to achieve a lower-powered incentive and lower information rents. Investment costs are often incurred over a much shorter period of time than the benefits from the investment. Hence the reset period may be long relative to costs but short relative to benefits. In this case a lower proportion of costs than revenues will be captured in the resetting process. This will tend to discourage efficient investment. This is related to the well-known incentive problems that occur shortly before a rate-case. At this time, it becomes advantageous to make costs appear high and revenues appear low. Also the incentive for investment in diminished shortly before a rate case, because the regulator may view such expenditures as already paid for.

Profit sharing avoids these timing issue because it does not make periodic adjustments, but instead shares profit in some fixed proportion on a continuous basis. Economic profit, for example $\pi = R – C_E – C_{LL} – CT$, accounts a normal rate of return on investment for (in CT), so profit sharing can never prevent a supplier from achieving a normal rate of return, it will only bring it closer to that level.[10] If un-shared profit, π, is maximized by a certain investment strategy, then half of un-shared profit, $\pi/2$, will also be maximized by exactly the same investment strategy. Consequently, if there were no information problem with computing π, profit sharing would leave the Transco's behavior unchanged. This would be ideal. The excess wealth transferred by the high profits of pure price-cap regulation could be reduced by any amount simply setting the sharing parameter appropriately and this would cause no loss of efficiency. But there are information problems, and the fundamental tradeoff of regulation assures that, if the power of the incentive is greatly reduced, efficiency will suffer. That is true whether the incentive's power is reduced through periodic resetting of the cap, as previously described, or through continuous profit sharing, but it is easier to explain the effect in the continuous profit-sharing context.

To glimpse the contradiction inherent in ignoring the information problem, consider, the case in which the dollar-valued economic profit, π, divided by the invested capital needs to equal 50% for the initial determination of R in order to avoid any significant probability of bankruptcy. If the normal rate of return on equity is 15%, then the monopolist will start out making a 65% rate of return on equity.[11] To reduce this, consider a profit sharing ratio of 1% for the Transco and 99% for load. This reduces π to $\pi/100$ and brings the initial return on equity to 15.5%. If the Transco raises π to 100% with a superb effort, it will receive 16% on equity and if it performs terribly, letting π fall to 0%, it will still receive 15% on its equity. Even though it prefers 16% to 15%, this limited reward is not likely to induce the effort level required to raise π divided by invested capital from 0% to +100%.

The important point about effort is that it is a real cost that is not included in *CT* because it is not monetized. It is a cost that does not appear on the books. It will, of course, affect costs that do appear on the books. Effort will reduce these costs for the same level of transmission performance or increase performance for the same level of monetary cost. Lack of effort raises monetary cost relative to performance. Lack of effort, like effort, can take many forms. It can take the form of "gold-plating" offices and equipment or "shirking" by management and workers. Inefficient expenditures can purchase tickets to the San Francisco Giants baseball games as was done by a major California utility. "Graft" is another type of lack of effort; for example, the Transco can subcontract a job to someone who pays a kickback to someone at the Transco. As the rewards for effort (good behavior) fall, all types of unwanted behavior will increase.

Profit sharing on a 50/50 basis takes away half of π, which means it takes away half of revenues minus costs. If effort had a monetary value, then a \$10 effort that produced an \$11 revenue would increase π by \$1 before profit sharing and by only \$0.50 after profit sharing. In either case the effort is worthwhile and will be induced. But because the effort goes unobserved, its cost is not shared by the profit-sharing mechanism, and so the result of the \$10 effort is an after-sharing revenue of \$5.50 which fails to compensate for the \$10 effort. Consequently, with 50/50 profit sharing such efforts will not be undertaken. A \$10 effort that produced a \$30 increase in revenue would still pay off even after profit sharing allowed the Transco to keep only \$15 of the revenue. So profit sharing does not eliminate the incentive to provide unmeasured but costly effort, it only reduces it. As the share of profits kept by the Transco decreases towards zero, the incentive to expend effort decreases toward nonexistent. This explains why profit sharing is limited as a means to control information rents.

---

[10] For example if profit = 1, then net operating revenue, $R - C_E - C_{LL}$, equals $CT + 1$, which means transmission investments are earning more than a normal rate of return. If profit is cut in half, operating revenues still cover $CT + ½$, and investments still earn more than a normal rate of return.

[11] Economic profit is profit above the normal rat e of return on invested capital.

5.3. Difficulties with PBR for Transcos

The price-cap mechanism just described it most likely impractical because it suffers from at least two severe difficulties. First, just as with merchant transmission the Transco's investments will pay off with very long lag times and are consequently very risky (Brunekreeft and McDaniel, 2005). Typically, a large cost must be incurred over a period of several years, then for several more there will be little or no return on the investment and finally, ten or fifteen years after the start of the project, significant payback will begin. This is just one possibility, but a very plausible one. The risk of this delayed payback contrasts sharply with society's risk, which is much less because the societal payback starts immediately upon completion of the line at a rate equal to the rental cost of the line.

Second, transmission investments are tightly linked to reliability (Crew, Kleindorfer and Spiegel, 2004; Sun, Sanford and Powell, 2004), and the reliability part of this mechanism, C_{LL}, introduces severe risks and promises to be extremely controversial if attempted. Most of the major blackouts in the U.S. over at least the last 35 years have been linked more to transmission problems than to generation problems. This has tended to involve operations and investments other than major line and transformer investment, such as tree trimming, computer systems for state estimation, and the setting of line-trip relays. It should be possible to design a separate mechanism for performance in these areas, but that will not completely disentangle reliability from line investment.[12]

The motivation for upgrading a transmission path, which typically involves several lines, is congestion on that path and the congestion cost, C_{E}, that it causes. But this cost can be reduced in two ways, first by a physical upgrade, and second by re-rating the path to a higher capacity. The second approach is far cheaper, in fact is almost free, but it decreases reliability. Unfortunately path ratings are not easily audited as they are somewhat controversial even among engineers. This is because they are not based primarily on hard data, such as the temperature at which a wire melts. Instead, ratings must couple hard data with somewhat subjective data, including probabilities of contingencies such as line and generator outages.[13] Any powerful incentive to upgrade lines will also be a powerful incentive to cut corners on contingency ratings. Consequently if there are strong incentives to upgrade lines, there must also be heavy penalties for cutting corners on path ratings. Since path ratings are too difficult for regulators to monitor, these penalties must instead be applied directly to blackouts which are very costly but occur very rarely. Imposing the cost of lost load, C_{LL}, is such a penalty, but as explained, it introduces severe risks and would be extremely controversial. The danger of degrading reliability appears to create severe difficulties for designing a useful PBR incentive for upgrading lines. [François, I have re-written this paragraph.]

A third difficulty, less severe than the prior two, faced by any Transco proposal is the Fundamental Planning Problem described above. Any realistic Transco incentive will induce the Transco to invest in the lines that are optimal for existing generation, not for optimal generation. The consequence will be that generation investors will build generation based on the Transco's response. For this circularity to produce the efficient outcome, the charges used to supplement the congestion rents and provide the Transco's revenues must be allocated in a way that induces the correct location of generation.

A number of PBR alternatives for Transcos have been suggested, including some by Vogelsang based on the price-cap tradition, and are reviewed by Rosellón (2003) and Vogelsang (2004). His most recent proposal is related to these and the social surplus mechanism described above. Unfortunately this proposal is still not able to solve the investment problem taking into account reliability and load growth. His current conclusion is "Long periods mark the limits of regulatory commitment and are still short

---

[12] Wilson (1997), addresses reliability in a franchise Transco context.

[13] Stability ratings, though more firmly based in physics, also involve subjective judgments as to how close is too close to the point of instability.

relative to network investments. As a result, incentives should be further weakened by adjustments based on rate-of-return regulation with a ‘used and useful’ criterion.” PBR for Transcos will be useful for shorter term incentives (Joskow, 2004), but it cannot yet be relied on to solve the long-term investment problems.

6. Conclusion

Currently, wholesale power markets are undergoing a slow and inefficient development process marked by such events as the California meltdown, the over-building of gas-fired generation in large parts of the Eastern U.S., the largest blackout in U.S. history, and the complete redesign of the British market. In particular the generation investment problem seems to remain far from solved, though some reasonable incentive mechanisms seem to be on the drawing board.

Moreover, it should be recognized that transmission investment is crucial to the functioning of the new energy markets. The less congestion, the less market power in wholesale energy markets (Stoft, 1997; Borenstein *et al.*, 2000; Gilbert, Neuhoff and Newbery, 2002). For example, San Francisco and New York both suffer from serious market power because both have limited transmission and must rely for significant portions of their energy on local suppliers. The more transmission into these cities, the more competition in the wholesale energy market. Moreover, in every market in the U.S., there are numerous examples of generation units under “regulatory must run” contracts. These give the market administrator the right to require the plant to run and in return provide regulatory payments which are often substantial. Such contracts exist largely where these generators have extreme market power during some hours of the year because of transmission limitations. Such situations have been numerous and problematic from the beginning and show no signs of disappearing.

Fortunately it is extremely cheap to over-build the transmission system a small amount and thereby reduce market power below the level that would be found under an optimized network. This is because, at optimal investment, the derivative of total system cost with respect to increased capacity is zero. That is the first-order condition for optimality.

Neither a merchant approach nor a PBR approach is conducive to alleviating wholesale-market-power problems by overbuilding the network. Both have biases towards underinvestment, and both are likely to be erratic in their behavior during the decades it will take to tame the likely flaws in their designs. Rate-of-return regulation is more adaptable. Provided the regulator declares in advance that a line will be considered used-and-useful, it should not be difficult to get the Transco to build it.

Neither a merchant approach nor a Transco/PBR approach has yet been developed to the point where it could be considered useable in practice. Both appear to be in rather early stages of theoretical development. Moreover, it appears that any application of these approaches will require a level of understanding and subtlety that is not yet apparent among regulators, at least in the U.S. Given these difficulties and the poor record of the energy-market deregulation process, it appear too early to begin any policy initiatives based on either of these approaches. With transmission investment costs amounting to only 3–8% of the retail costs (Joskow and Tirole, 2002), it is better to rely on a relatively safe approach to transmission investment, even though it is a bit less efficient than results promised by some poorly understood theoretically approaches. This is not to say that merchant transmission investment should be discouraged. It should be allowed and regulated only lightly (Brunekreeft, 2004a), but it cannot be depended on and should not be the focus of those concerned with assuring sufficient transmission capacity.

Rate-of-return regulation applied to a Transco or to wires companies under the direction of an ISO is not without its drawbacks. Besides the Fundamental Planning Problem discussed above, there is the additional problem in the context of a deregulated generation market that both generators and load will constantly lobby regulators for more or less transmission. If the Fundamental Planning Problem is not solved, some generators will lobby for more transmission to get a free ride to market, while others will argue for less to keep their local price up and allow them to exercise more market power. Load will argue for more transmission from cheap regions to expensive ones, not just to save production costs and dampen market power (legitimate reasons), but also to exercise monopsony power against generation in the high cost regions. For years to come, transmission investment appears to be the knottiest problem in the deregulation process.

Bibliography

Barthold, L.O. (2003), “Whither EHV? Distributed generation reverses the trend,” IEEE *Power and Energy Magazine*, Jan-Feb 03 88, 85.

Borenstein, S., Bushnell, J. and Stoft, S. (2000), “The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry,” RAND Journal of Economics, Vol. 31, No. 2, Summer 2000, pp. 294–325.

Brunekreeft, G., (2004a), ‘Regulatory Issues in Merchant Transmission Investment’, Cambridge Working Papers in Economics, CWPE 0422, and Cambridge-MIT Working Paper 38.

Brunekreeft, G. (2004b), “Market-based investment in electricity transmission networks: Controllable flow,” Utilities Policy, 12(4), 269-281.

Brunekreeft, G., T. McDaniel (2005), “Policy Uncertainty And Supply Adequacy In Electric Power Markets,” TILEC Discussion Paper, DP 2005-006. Forthcoming in revised and shortened version: Oxford Review of Economic Policy.

Brunekreeft, G., K. Neuhoff and D. Newbery (2004), “Electricity Transmission: An Overview of the Current Debate,” Cambridge Working Papers in Economics, CWPE 0463, and Cambridge-MIT Working Paper 60.

Bushnell, J., and S. Stoft (1996) “Electric Grid Investment Under a Contract Network Regime," *Journal of Regulatory Economics*, 10: 61–79.

—— (1997) “Improving Private Incentives for Electric Grid Investment,” *Resource and Energy Economics*, 19: 85–108.

Chandley, J.D. and W.W. Hogan (2002), “Independent Transmission Companies In A Regional Transmission Organization,”

Chisari, O.O., P. Dal-Bó and C.A. Romero (2001) “High-Tension Electricity Network Expansion in Argentina: Decision Mechanisms and Willingness-to-pay Revelation,” *Energy Economics*, Volume 23, Issue 6: p.697-715.

Crew, M., P.R. Kleindorfer and M. Spiegel (2004), “Reliability, Regulation and Transmission Investment,” Risk Management and Design Processes Center, Working Paper 04-20-PK.

Firecone Ventures Pty Ltd (2003), “Regulatory and Institutional Framework for Transmission (Australia), Final Report,” November.

Gans, J., and S. King (2000) “Options for Electricity Transmission Regulation in Australia,” *Australian Economic Review*, Vol.33, No.2, June 2000, pp.145-161.

Gilbert, R., K. Neuhoff and D. Newbery (2004), “Allocating Transmission to Mitigate Market Power in Electricity Networks,” (with), *RAND Journal of Economics*, Volume 35, No. 4,Winter 2004, pp. 691-709.

Paul R. Gribik, Dariush Shirmohammadi, Joseph S. Graves, and James G. Kritikson, “Transmission rights and transmission expansions,” To appear in *IEEE Transactions on Power Systems*, 2005.

Hogan, W.W. (1992) “Contract Networks for Electric Power Transmission,” *Journal of Regulatory Economics*, 4: 211–242.

Hogan, W.W. (1998), “Transmission Investment And Competitive Electricity Markets,” mimeo, Harvard University.

Hogan, W.W. (2002), “Financial Transmission Rights Formulations,” mimeo, Harvard University.

Harvey, S., Hogan, W., and S. Pope (1996). "Transmission Capacity Reservations Implemented Through a Spot Market with Transmission Congestion Contracts," mimeo.

Joskow, P.L. (2003), “Remedying Undue Discrimination through Open Access Transmission Service and Standard Electricity Market Design,” AEI-Brookings Joint Center For Regulatory Studies, Regulatory Analysis 03-1, February.

Joskow, P.L. (2004), “Performance Based Regulation,” May (slides).

Joskow, P.L. and J.J. Tirole (2000), “Transmission Rights and Market Power on Electric Power Networks," *RAND Journal of Economics*, 31(3): 450–487.

Joskow, P.L. and J.J. Tirole (2002), “Transmission Investment: Alternative Institutional Frameworks,” mimeo.

Joskow, P.L. and J.J. Tirole (2004a), “Merchant Transmission Investment,” *Journal of Industrial Economics*, 2004.

Léautier, T.-O., 2000, ‘Regulation of an Electric Power Transmission Company’, *Energy Journal*, Vol. 21, No. 4, pp. 61-92.

Littlechild, S., 2004, ‘Regulated and Merchant Interconnectors in Australia: SNI and Murraylink revisited’, Cambridge Working Papers in Economics, CWPE 0410, and Cambridge-MIT Working Paper 37, May.

Littlechild, S.C. and C.J. Skerk (2004), “Regulation of Transmission Expansion in Argentina: Part II – Developments Since the Fourth Line,” Cambridge Working Papers in Economics, CWPE 0465, and Cambridge-MIT Working Paper 63, May.

Perez-Arriaga, I. J. et al (1995) “Marginal Pricing of Transmission Service: An Analysis of Cost Recovery,” *IEEE Transactions on Power Systems*, 10(1): 546–553.

Rosellón, J. (2003) Different Approaches Towards Electricity Transmission Expansion,” *Journal of Network Economics*, Vol.2, Issue 3 – September 2003.

Rotger, J. and F. Felder (2001), “Promoting Efficient Transmission Investment: The Role Of The Market In Expanding Transmission Infrastructure,” sponsored by TransÉnergie U.S. Ltd.

Stoft, S. (1997), “The Effect of the Transmission Grid on Market Power,” Lawrence Berkeley National Laboratory Working Paper, LBNL-40479

Stoft, S. (2002), “A Proposal for Long-Run and Short-Run Congestion Management in Alberta,” Before the Alberta Energy And Utilities Board, File No.: 1803-4, March 4.

Sun, H., M. Sanford and L. Powell (2004) “Justifying Transmission Investment in the Markets,” Electricity Transmission in Deregulated Markets, Dec 15-16, 2004, Carnegie-Mellon University, Pittsburgh, PA.

Vogelsang, Ingo (2001) “Price Regulation for Independent Transmission Companies,” *Journal of Regulatory Economics*, 20(2), September: 141–165.

—— (2004) “Transmission pricing and performance-based regulation,” Prepared for Carnegie Mellon Conference on Electricity Transmission in Deregulated Markets. Pittsburgh, December 15-16, 2004.

Wilson, R. (1997) Implementation of Priority Insurance in Power Exchange Markets," *The Energy Journal*, Vol. 18, No. 1.

Woolf, F. (2003) Global Transmission Expansion: Recipes for Success, Pen Well, Tulsa, Oklahoma.

7. Glossary and Symbol Table

C_E	Congestion cost. Excess cost of energy from dispatching out of merit order because of transmission constraints.
C_R	Congestion rent. Revenue from energy injections at nodal prices less revenue from energy withdrawals at nodal prices.
C_L	Congestion cost to load. Excess cost of energy to load due to transmission constraints.
C_{LL}	Cost of un-served (lost) load
CT	The rental cost of the transmission system paid by a regulated Transco. Includes the annualized costs of capital and maintenance.
CRR	Congestion revenue right which pays $(P_B - P_A)\,Q$, during time periods T.
FTR	PJM's Financial Transmission Right. The obligation variety is the same as a CRR except that the revenues are adjusted for any revenue surplus or insufficiency it total congestion rents. The option variety omits the negative payments possible with an obligation.
G	The transmission grid.
ISO	One of several "Independent System Operators" that run markets in the U.S. They are now switch status to become RTO's, "Regional Transmission Organizations," in keeping with FERC's changing terminology.
K	The capacity in MW of a transmission line.
NYISO	The New York Independent System Operator.
P_A ,P_B	Nodal energy prices at nodes A and B.
PJM	The ISO that now covers Michigan, Pennsylvania, Washington DC, Tennessee and more, with 44 million consumers and 135,000 MW of generating capacity.
Q	The megawatt flow named in a CRR, or on a line.
R	The regulated price paid to a Transco per MWh of delivered power.
RTO	Regional Transmission Organization. The new name for an ISO in the U.S.
S	The share of profit in excess of the target profit level kept by a Transco.
T	The time period covered by a CRR, e.g., peak hours during 2010.
Transco	A regulated monopoly transmission company
W_i	The net power withdrawal at node i.