THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume LIII June 2005 No. 2

MERCHANT TRANSMISSION INVESTMENT*

PAUL JOSKOW†
JEAN TIROLE‡

We examine the performance attributes of a merchant transmission investment framework that relies on 'market driven' investment to increase transmission network capacity needed to support competitive wholesale markets for electricity. Under a stringent set of assumptions, the merchant investment model has a remarkable set of attributes that appears to solve the natural monopoly problem and the associated need for regulating electric transmission companies. We expand the merchant model to incorporate several attributes of wholesale power markets and transmission networks that the merchant model ignores. These include market power in wholesale electricity markets, lumpiness in transmission investment opportunities, stochastic attributes of transmission networks and associated property rights definition issues, strategic behavior by potential merchant transmission investors and issues related to the coordination of transmission system operators and merchant transmission owners. Incorporating these more realistic attributes of transmission networks and the behavior of transmission owners and system operators leads to the conclusion that several potentially significant inefficiencies may result from reliance on the merchant transmission investment framework. Accordingly, it is inappropriate for policymakers to assume that they can avoid dealing with the many challenges associated with stimulating efficient levels of investment in electric transmission networks by adopting the merchant model.

I. INTRODUCTION

THE ECONOMIC ANALYSIS OF RESTRUCTURED COMPETITIVE WHOLESALE ELECTRICITY MARKETS has mostly focused on the organization and functioning of spot markets for electric energy. This theoretical and empirical research has studied, among other things, the organization of day-ahead and real time (balancing) energy markets and associated auction rules, the role of bilateral

*We would like to thank Richard Green, Shmuel Oren, Masheed Rosenqvist, Steven Stoft, Frank Wolak, the editor and three anonymous referees for helpful comments on earlier drafts. Joskow gratefully acknowledges research support from the MIT Center for Energy and Environmental Policy Research and the Cambridge-MIT Institute.

†Authors' affiliations: MIT Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02142-1347, U.S.A.
e-mail: *pjoskow@mit.edu*

‡Institut d' Economie Industrielle, Université des Sciences Sociales de Toulouse, Place Anatole France, 31042 Toulouse, France.
e-mail: *tirole@cict.fr*

contracts, congestion management, nodal pricing, physical and financial transmission contracts, and associated market power issues. This work typically takes the transmission network as given. However, in the medium and long term as demand grows and new generating capacity is added to replace older, less efficient, capacity or to meet growing demand efficiently, investments in transmission capacity are likely to be necessary to minimize the overall costs of wholesale electricity supplies, to maintain reliability, to mitigate locational market power, and to improve the performance of competitive wholesale and retail markets.

Despite growing problems associated with stimulating transmission investment in many restructured electricity markets, there has been surprisingly little research on the institutions governing transmission investment. Early formulations of the structure for competitive wholesale markets envisioned the creation of independent regulated regional transmission and system operating entities (Transcos, or regulated Transmission Companies) that would be responsible for building, owning and operating transmission facilities and would be subject to economic regulation (Joskow and Schmalensee [1983]). More recent research has explored the attributes of incentive regulatory mechanisms that could be applied to such regulated transmission monopolies (e.g., Celebi, Nasser [1997], Léautier [2000], Vogelsang [2001]) to integrate energy price (congestion) signals with transmission investment. The institutional arrangements governing transmission operation and investment in England and Wales reflect this basic institutional approach. The regulated Transco model is necessarily subject to the classical challenges of regulated monopoly, namely how to specify and apply regulatory mechanisms that provide good performance incentives to the regulated firm while minimizing the economic rents that the regulated firm can derive from its superior information.

An alternative (or complement) to the regulated Transco model is the merchant investment model. It relies on competition, free entry and decentralized property-rights based institutions, and market-based pricing of transmission service to govern transmission investment (Hogan [1992], Bushnell and Stoft [1996,1997], Chao and Peck [1996]). In return for investment in additional transmission capacity, merchant investors receive property rights that allow them to collect congestion revenues equal to the difference in nodal energy prices associated with the incremental point-to-point transmission capacity their investments create. The value of these rights to receive congestion revenues represents the revenues merchant investors receive to cover the capital and operating costs of these investments and provides the financial incentives that guide 'market based' transmission investment. Previous theoretical research has demonstrated that under a fairly stringent set of assumptions, all profitable investments are efficient (efficient investments, on the other hand, are not all profitable, due to the lumpiness problem discussed in Section 5). This result undermines the traditional assumption that transmission networks are 'natural monopolies' that must be subject to regulation. If it were correct, in practice, it would lead to the remarkable conclusion that both the generation of electricity and the transmission of electricity can be largely deregulated.

Research on this model has focused almost entirely on simple cases where transmission investments are characterized by no increasing returns to scale, there are no sunk cost or asset specificity issues, nodal energy prices fully reflect consumers' willingness to pay for energy and reliability, all network externalities are internalised in nodal prices, transmission network constraints and associated point-to-point capacity is non-stochastic, there is no market power, markets are always cleared by prices, there is a full set of futures markets, and the transmission system operator (SO) has no discretion to affect the effective transmission capacity and nodal prices over time. That is, the analysis has proceeded under assumptions equivalent to those of a simple model of perfect competition. Unfortunately, these assumptions do not reflect the attributes of the transmission investment opportunities that are likely to be most conducive to merchant investment, the stochastic properties of real transmission networks, or widely documented imperfections in wholesale power markets.

We begin with a background discussion of the institutional environment in which the merchant investment model has been developed, the key assumptions upon which analyses of its properties have been based and the primary results of this work derived (Sections 2 and 3). The paper then examines how these results are affected when we relax several assumptions to more accurately reflect the physical and economic attributes of electric transmission networks.[1]

Section 4 shows that when there are imperfections in the competitive wholesale electricity markets that lead nodal spot electricity prices to depart from their efficient levels, investment incentives will be distorted. For example, when unregulated generators have market power, nodal energy prices will be distorted from their efficient levels. These distortions may lead to over-investment or under-investment depending upon where on the network electricity generators have market power. Imperfect government interventions to control market power in competitive wholesale electricity markets may also distort investment incentives.

Network expansion investments that are most conducive to supply by competitive entrants are also likely to be characterized by economies of scale or 'lumpiness.' Section 5 shows how economies of scale will lead to under-

[1] Similar issues are likely to arise with merchant investment in other network industries such as natural gas pipelines, railroad infrastructure and highways. However, we have not explored these issues in the context of other network industries.

investment, to monopoly pre-emption of competitive generation or transmission investments, and distort the timing of investments. We argue that these problems are unlikely to be resolved by relying on bilateral or multilateral negotiations among the market participants who benefit from these negotiations (the 'Coase Theorem'). Indeed, opportunities for multilateral bargaining are likely further to distort investment incentives. We also demonstrate how strategic interactions between merchant transmission owners who are in a vertical relationship to one another can lead to underinvestment in related transmission links.

Section 6 shows that the types of non-contingent transmission property rights that have been assumed to be awarded to investors in transmission capacity are poorly matched to the stochastic characteristics of transmission networks and investments in them. Instead, transmission property rights that are contingent on exogenous variations in transmission capacity and reflect the diversification attributes of new investments would be necessary to properly align investment incentives with the stochastic attributes of transmission networks. Defining and allocating property rights that provide efficient investment incentives is also likely to be inconsistent with the development of liquid competitive markets for these rights or derivatives on them.

Under the merchant transmission model, the ownership and maintenance of transmission facilities are to be separated from decisions regarding security-constrained, bid-based dispatching of generators and price-responsive demand on the network and managing reliability criteria and constraints. The latter decisions are to be made by a monopoly independent system operator (ISO) that is unaffiliated with generators, energy marketers or transmission owners. Section 7 shows that investment and maintenance decisions made by transmission owners are interdependent with dispatch and network reliability decisions made by the ISO. Separating these decisions leads potentially to moral hazard problems and associated inefficiencies. Transmission owners and system operators must have a compatible set of incentives to avoid these inefficiencies, but designing these incentives is challenging.[2]

II. MERCHANT TRANSMISSION INVESTMENT: BACKGROUND

Transmission investment institutions cannot be considered independently of the institutions that govern the determination of energy prices, operating reserves, contingency constraints, congestion management and the measurement of transmission capacity and increments to it. In what follows, we

[2] Vertical integration between transmission ownership and system operations is likely to reduce these incentive problems and, if transmission owners are also independent of generation and marketing of power, may be a superior organizational structure.

shall assume that a nodal pricing system for electric energy is in place with attributes similar to those being proposed by the U.S. Federal Energy Regulatory Commission (FERC) in its SMD proposals and what is in operation in New York, New England and PJM in the U.S. This is the most conducive framework for merchant investment because nodal prices provide a measure of locational scarcity that is necessary to make this framework a plausible option.

Under this model, an independent system operator (ISO) operates a real-time balancing market and allocates scarce transmission capacity using bids to increase or decrease generation or demand at each node. That is, the ISO takes all of the bids (generation and demand) and finds the 'least cost' set of uniform market-clearing price bids to balance supply and demand at each generation and consumption node on the network using a security constrained dispatch model. This establishes day-ahead quantity commitments and nodal prices that reflect both congestion and marginal losses.[3]

The model recognizes that there are incumbent transmission owners (TO) who own the existing transmission assets and requires that the system operator (SO) and TO be separate entities and operate independently.[4] The incumbent TO receives some cost-of-service compensation for the usage of a grid that it no longer controls to compensate it for legacy investments and ongoing maintenance costs. New investments in transmission are anticipated to be made by competing merchant investors whose compensation is based on the value of Congestion Revenue Rights (CRRs)[5] created by their investments. These financial rights represent the right to receive congestion revenues defined as the difference between the nodal prices between the two nodes (point-to-point) covered by the relevant CRR times the quantity of CRRs held.

The separation between ownership (TO) and control (SO) functions in this model is motivated by two considerations. First, a market driven transmission system leads to multiple owners of parts of the grid; while the owners can form a cooperative to operate the grid, their goals are in general

[3] Generators may enter into bilateral contracts with marketers or load serving entities (LSEs) and schedule supplies with the ISO separately from the organized day-ahead market. However, they still have to pay any congestion charges associated with their schedules based on the difference in nodal prices between the injection and receipt points. The day-ahead schedules, nodal prices, and congestion charges are 'commitments.' They can be adjusted in real time by submitting adjustment bids to the real time balancing markets (which again rely on bids, a security constrained dispatch and nodal prices) to allow these schedules to be changed based on real time economic conditions.

[4] Exactly what functions are assigned to the TO and what functions to the SO is a subject of continuing debate and depends in part on whether the TO is 'independent' of generators and marketers that use its facilities to participate in the wholesale market.

[5] Congestion Revenue Rights have also been referred to in the literature as Transmission Congestion Contracts (TCCs) and Financial Transmission Rights (FTRs). These names are interchangeable for our purposes.

antagonistic,[6] and it is well-known that cooperatives of members with heterogeneous interests face complex governance problems.[7] Second, and quite crucially, grid owners face a serious potential conflict of interest when operating a transmission grid if their compensation varies directly with the level of congestion rents.[8] By conservatively 'withdrawing' transmission capacity (under the cover of a safe management of the network), the system operator can substantially raise the congestion rents. Third, in many countries there continues to be vertical integration between generation, power marketing and transmission. The creation of an independent SO is thought to be a way to mitigate the potential problems that may arise from common ownership and control of transmission and generating assets and associated power marketing activities.

It is useful to consider two types of transmission investments that might be governed by a merchant (or regulated) transmission investment framework, network deepening and network expansion investments:[9]

Network deepening investments: These are investments that involve physical upgrades of the facilities on the incumbent's existing network (e.g., adding capacitor banks, phase shifters, reconductoring existing transmission links, new communications and relay equipment spread around the network to increase the speed with which the SO can respond to sudden equipment outages and relax contingency constraints). These investments are physically intertwined with the incumbent TO's facilities. Similar to network deepening investments are network maintenance decisions.

Network deepening and maintenance investments can, as a practical matter, only be implemented efficiently by the owner of the existing lines for several reasons. Defining an efficient 'competitive access to deepening

[6] Incumbent owners of transmission lines that are being compensated based on congestion rents will have incentives to oppose investments by others in generation or transmission that reduce these congestion rents. Generators located in congested areas will have incentives to oppose transmission enhancements that would reduce or eliminate the congestion.

[7] See, e.g., Hansmann [1996].

[8] In practice, due to the lack of market-based penalties for outages, dispatching does not quite correspond to the least-cost optimization used in economic and engineering models; rather, grid operators have substantial discretion over how much outage they are willing to take while dispatching. This discretion in turn potentially provides incentives for system operators to manipulate the congestion rents received by the owners (Glachant and Pignon, [2002]).

[9] We focus on transmission investments that affect congestion on the high voltage network. Regulators in the U.S. often break transmission investments down into additional categories. First, there are local transmission investments 'inside' the demand node. These investments are sometimes called 'reliability' investments. Second, interconnection investments are investments that must be made by an incumbent grid owner to connect new generators with the rest of the network. These are often treated like radial links and are typically paid for by the generators seeking interconnection. However, it is hard to draw bright lines between reliability investments, interconnection investments, and investments that affect network congestion.

We recognize that, in practice, there may not always be a bright line between network deepening and network expansion investments.

investments' policy is likely to be extremely difficult. First, adding third-party facilities that are fully integrated with the existing network from a physical and maintenance perspective creates significant incentive problems with decentralized ownership. The problems of defining a good set of rules for investing in and maintaining facilities of this type with decentralized ownership is further exacerbated by the heterogeneous nature of transmission facilities. While it is theoretically possible to devise contractual arrangements that will solve the incentive problems, including opportunistic behavior of one or more parties, writing and enforcing such contracts will be very difficult as a practical matter. Second and relatedly, one would need to allocate carefully the new capacity of the line between the initial design and maintenance choice of the original owner and the actions of the renters who make deepening investments. This 'moral hazard in teams' problem is a substantial obstacle to the design of an effective third party access policy for this type of transmission investment. Third, the merchant investment model effectively requires that there be free entry into the development of new transmission capacity; entrants, though, are likely to have less information about existing transmission lines than their owners. These considerations suggest that network deepening investments can, as a practical matter, only be implemented efficiently by the incumbent transmission network owner.

Independent network expansion investments: These are investments that involve the construction of separate new links (including parallel links) that are not physically intertwined with the incumbent network except at the point at either end where they are interconnected. These investments can (in principle) be made either by incumbent transmission owners, by stake-holders (generators, load-serving entities), or by a third-party merchant investor. The two operating DC merchant links in Australia appear to fall into this category as do other proposed transmission lines linking different countries or regions and often referred to as 'interconnectors.' While, as in Australia, these links may have effects on power flows on the rest of the network, including on parallel lines, they are physically separable projects from a construction and maintenance perspective.

While the literature underlying the merchant investment model does not distinguish between network deepening and network expansion investments (and some proponents of the merchant investment model appear to conceive it as applying to all network transmission investments) for the reasons discussed above, we think that it is unlikely that a merchant model can or will be applied successfully to network deepening investments. Accordingly, our analysis proceeds under the assumption that the kinds of investments to which the merchant model will apply are investment opportunities with the attributes of network expansion investments (e.g., interconnectors) and we do not attempt to deal with the difficult contractual and incentive issues that would otherwise have to be confronted if network deepening investments were governed by a merchant investment model as well.

III. THE CASE FOR MERCHANT TRANSMISSION INVESTMENT

Let us start from the theoretical case for market driven or 'merchant' transmission investment. This rationale has been developed, *inter alia*, by Hogan [1992] and Wu *et al.* [1996], Chao and Peck [1996] and examined further by Bushnell and Stoft [1996, 1997] for simple cases. The basic argument is conveyed in the two-node framework of figure 1. [This paper focuses on the two-node framework for expositional simplicity. See our discussion paper (Joskow-Tirole [2003]) for an analysis of transmission investment when there are loop flows.]

Figure 1 depicts a simple situation in which load serving entities (distribution companies or marketers buying on behalf of retail consumers, or large industrial customers buying directly) in the South (say, a large city) buy their power from cheap generation sources in the North and, possibly, in addition, more expensive sources in the South. Alas, the capacity of the line from North to South is limited to K and, faced with net demand/supply curves in the North and the South, the system operator is forced to dispatch 'out of merit'. For example, the system operator calls on expensive generators in the South while generators in the North would be willing to supply this amount at a lower price if more transmission capacity were available. The rationing of the scarce North-South capacity is implemented by setting two nodal prices, p_S and p_N that clear the markets in the South and the North, respectively. The difference, $\eta = p_S - p_N$, is the shadow price of the transmission capacity constraint. The area ηK is the *congestion rent* and the

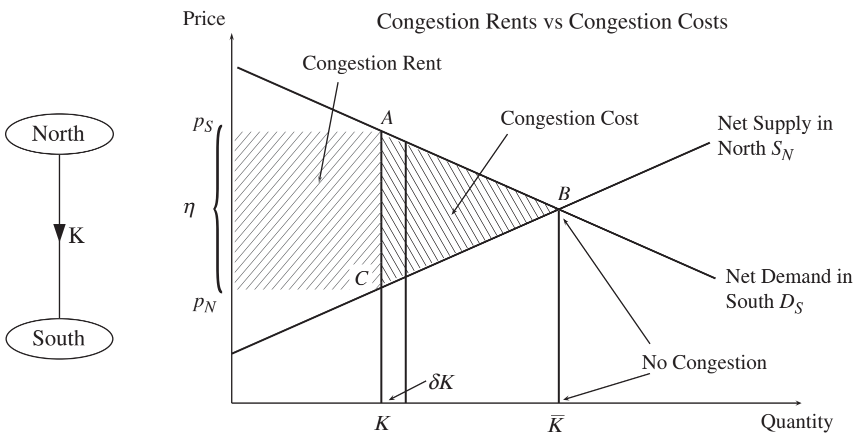

Figure 1

triangle ABC is the *congestion* or *redispatch cost*. The latter represents the cost of running more costly generation in the South because less costly imports from the North are limited by transmission congestion. The congestion *rent* and the congestion *cost* are sometimes confused and it is important to recognize their appropriate definitions and meaning.

Now consider a *marginal* (unit) increase in transmission capacity (δK). This unit increase allows one more kWh to flow from North to South, replacing a marginal generator in the South with cost p_S by a cheaper generator in the North producing at cost p_N. That is, the social value of the investment is given by the reduction in the area ABC in Figure 1.

Assume that the builder of this marginal capacity, whether it is a new entrant or the incumbent TO, is rewarded through a financial transmission right that pays a dividend equal to the shadow price of the transmission constraint. A non-incumbent merchant company will enter to build this extra capacity as long as η exceeds the cost of building it. By contrast, *if* an incumbent grid owner is compensated through the payment of congestion rents, it may not want to make this marginal investment as it must compare the extra revenue η net of the cost of expanding the capacity with the reduction in the congestion rent on its inframarginal transmission units $(-Kd\eta/dK)$. It is only when the incumbent grid owner's capacity has been rated at some level K^* not too different from actual capacity, and that the corresponding rights, with value ηK^*, have been auctioned off, that the monopoly distortion vanishes. The incremental capacity then yields $\eta + (K - K^*)\frac{d\eta}{dK}$ close to η. As in the case of Contracts for Differences,[10] forward sales restore proper incentives for a player with market power.

Hogan [1992] and Bushnell and Stoft [1996, 1997] show that under certain conditions (besides nodal pricing, no increasing returns to scale, congestion rights satisfying feasibility constraints, no market power in the wholesale market, well defined property rights, a complete set of competitive liquid forward markets to provide sufficient information on long run demand and supply conditions and risk management, etc.), all profitable transmission investments are efficient. This powerful result appears to transform the transmission investment problem from one that appears to be almost intractable to one that requires a simple implementation of a property-rights based market system.

Accordingly, merchant investment's appeal is that it allows unfettered competition to govern investment in new transmission capacity, placing the risks of investment inefficiencies and cost overruns on investors rather than consumers, and bypassing planning and regulatory issues associated with a structure that relies on regulated monopoly transmission companies. In addition, in theory, it allows investment in new generating capacity in the

[10] See Green (1999).

constrained area to 'compete' with new transmission investment that reduces the import constraint. In this way, market driven transmission investment is an economist's dream, solving the problems associated with imperfect regulation of a 'natural monopoly' transmission company and aligning competitive transmission investments with the newly developed competition in the generation segment. Unfortunately, the optimality of the market driven approach depends on a number of strong assumptions and conditions that are likely to be inconsistent with the actual attributes of transmission investments and the operation of wholesale markets in practice.[11]

We turn now to a discussion of what we view as the most important assumptions underlying the case for the merchant model and the implications of relaxing these assumptions.

IV. IMPERFECTIONS IN ENERGY MARKETS

The reasoning above assumes that the prices that clear the markets in the North and the South reflect the marginal costs of production (and the marginal willingnesses to pay[12]) at each location, so that the congestion rent perceived by merchant investors does reflect the social savings brought about by the investment. That is, potential investors in new transmission capacity see the correct locational price signals in the wholesale markets. Factors that may distort prices include market power, regulatory interventions like price caps, the absence of a complete representation of consumer demand in the wholesale market, and discretionary behavior by system operators when the network is constrained.

Suppose for example that there is a generator with market power in the South, and that the latter region is import constrained. The generator exercises market power by withdrawing capacity and driving the price in the South up. Hence

$$p_S > c_S,$$

where c_S denotes the marginal cost of production in the South. The measured congestion rent then overestimates the cost savings associated with the replacement of one unit of power generated in the South by one unit of power generated in the North, suggesting an *over-incentive* to reinforce the link, ignoring other market imperfections.[13] On the other hand, the increased transmission capacity does not replace production in the South one-for-one; it also leads to an increase in total energy consumption in the

[11] Of course, these are attributes with which any alternative investment framework, including any regulated investment framework, must contend.

[12] We will present the argument in terms of cost savings; because what matters is net supply at each node, the same argument would apply to the demand side.

[13] For example, as we discuss below, lumpiness in transmission investment leads to *under-investment*.

South, which yields a social benefit equal to $(p_S - c_S)$ times the expansion in consumption. The box below shows that, under a weak assumption, the first effect dominates, and therefore market power results in an over-incentive to invest in merchant transmission. The assumption is that the Southern monopolist's reaction curve be downward sloping in a Cournot game. Intuitively, the transmission line creates a Cournot 'duopoly' in the South, in which the Southern firm faces a fixed output from its (transmission) rival. A downward sloping reaction curve means that the Southern firm curtails its output as the transmission capacity expands. This implies that the energy consumption increase effect is smaller than the inflated signal effect (the two effects would cancel if the output in the South were invariant to an increase in imports from the North).[14]

The impact of locational market power on merchant investment incentives

- Consider a monopoly supplier in the South producing at marginal cost c_S and facing demand function $D(p_S)$. Let $P(\cdot)$ denote the inverse demand function and $S^g(\cdot)$ the gross surplus. Provided that the transmission capacity K between North and South is fully utilized, the monopolist solves:

$$\max_{p_S}\{(p_S - c_S)[D(p_S) - K]\};$$

equivalently, this monopolist selects a consumption q_S in the South so as to solve :

$$\max_{q_S}\{[P(q_S) - c_S](q_S - K)\}.$$

Letting c_N denote the marginal cost in the North and neglecting consumption in the North, social surplus is

$$W = S^g(q_S) - c_N K - c_S[q_S - K].$$

The marginal gross surplus, dS^g/dq_S is equal to price p_S in the South, and so when the line's capacity is increased by dK, resulting in a consumption change dq_S, welfare changes by

$$dW = (p_S - c_S)dq_S + (c_S - c_N)dK.$$

Note that, with *perfect competition*, $p_S = c_S$ (and $p_N = c_N$) and so $dW = \eta dK$. With monopoly power in the South, though,

$$p_S - c_S > 0$$

[14] Similarly, and to the extent that reinforcing the line is akin to adding production capacity in the South, this suggests that entrants in generation have too much of an incentive to invest in the South as well. The box verifies that this is indeed the case.

and

$$c_S - c_N < \eta.$$

There is an over-incentive to invest if and only if

$$dW < \eta dK,$$

i.e.,

$$(p_S - c_S)dq_S + (c_S - c_N)dK < (p_S - c_N)dK,$$

that is if and only if

$$dq_S < dK,$$

For there to be an over-incentive to invest, the monopolist must 'absorb' some of the increase in imports from the North. To know whether this is the case, differentiate the first-order condition for profit maximization:

$$\frac{dq_S}{dK} = \frac{P'}{2P' + (q_S - K)P''}.$$

Thus, there is an over-incentive to invest under merchant investment if and only if

$$P' + (q_S - K)P'' < 0.$$

A sufficient condition for this is that the demand curve be concave. More generally, this condition is the standard condition for quantities to be strategic substitutes.

- The same reasoning can be applied to generation investments in the South. Indeed, K could alternatively denote the amount of power produced by a competitive fringe in the South in the profit maximization exercise. And so (regardless of the fringe's marginal cost)

$$dq_S < dK$$

as long as

$$P' + (q_S - K)P'' < 0.$$

Hence, there is in general an over-incentive to invest in generation in the South as well.

Conversely, a generator with market power in the North may (while still making full use of the link) be able to raise price p_N by withdrawing production capacity — perhaps to the level of p_S if it faces no competition in the North (Oren [1997], Stoft [1999], Joskow and Tirole [2000]). In this case, the congestion rent underestimates the gain from expanding the line's capacity, resulting in an *under-investment* by merchant transmission investors. At the same time, it could lead to inefficient entry of generating capacity in the North in response to the short run monopoly rents created there.

As a second example, in the case of market power in the South (this is the situation that will generate very high prices for consumers in the South), the regulator may be tempted to impose a price cap.[15] While the price cap improves economic efficiency if it really is about constraining market power, it may also distort price signals if high prices are at least sometimes due to tight competitive supply and demand conditions rather than market power. A tight price cap in the South then reduces the congestion rents during those hours that are very important because they produce the bulk of the rents to support investment, yielding *under-investment* in transmission.[16]

Third, prices may not clear supply and demand in real time because market clearing processes are not fast enough to respond to rapid changes in supply and demand conditions while maintaining physical requirements for frequency, voltage, and stability on the network. To maintain physical network parameters, administrative rationing is then substituted for prices to balance supply and demand as a consequence of what is effectively a problem of incomplete markets (Wilson [2002]). Moreover, the system operator has discretion in determining the exact nature of the responses to operating reserve deficiencies or 'scarcity.' For example, Patton [2002] shows that during tight supply conditions, the system operator takes actions that tend to depress market-clearing prices. Whether it is administrative rationing in response to incomplete markets or price controls motivated by efforts to constrain market power or price distortions caused by market power or discretionary decisions by the system operator, actual prices will depart from the efficient prices required to give the efficient signals for new investment. These imperfections are potentially important with regard to transmission (and generation) investment because the prices that create significant congestion rents tend to occur in a relatively small number of hours and these hours also happen to be the hours when these types of price distortions are most likely to occur.

[15] Or a *de facto* price cap as when the system operator curtails load administratively when prices don't clear the market.

[16] Capacity payments cum capacity obligations have been introduced or proposed in some jurisdictions to respond to the distortions caused by price caps on spot markets. We show elsewhere (Joskow and Tirole [2004]) that under certain conditions capacity payments may compensate for the distortions caused by spot market price caps. If capacity payments are not properly reflected in the prices seen by potential merchant transmission investors, their investment decisions will be distorted.

V. LUMPY TRANSMISSION INVESTMENTS

V (i). *Pre-emption*

The analysis of the effects of market power on investment incentives in the previous section, as well as most of the literature upon which the merchant investment model relies, assumes that transmission investment is not characterized by economies of scale or 'lumpiness.' However, *network expansion investments* are likely to be lumpy. That is, the average cost of a new link declines as its capacity increases, other things equal (Baldick and Kahn [1992], Perez-Arriaga *et al.*, [1999]). The impact of lumpiness is illustrated in figure 2. The initial capacity is K_0 and social efficiency would command that it be brought to a level K_1 in a single, discrete investment.

Assuming that the energy market participants are perfectly competitive, so that net demand/supply curves represent the true marginal costs/ willingnesses to pay, the surplus created (or congestion cost reduced) by the expansion of capacity from K_0 to K_1 is depicted by the shaded area in figure 2. The value, $\eta_1(K_1 - K_0)$ of the transmission rights (equal to the ex post congestion *rents*) granted to the merchant investors building this capacity expansion understates the social surplus S_1 it creates by reducing congestion *costs*. Lumpiness thus results in an *underincentive* to reinforce the system for the same reason that an incumbent grid owner rewarded by congestion rents has suboptimal incentives to remove these congestion rents.[17] What is needed to create proper incentives for investment is that the merchant investor be granted the entire shaded area in figure 2. The corresponding incentive mechanism requires knowledge of net supply and demand curves, though, and does not fit directly with the merchant investment paradigm.

Another source of lumpiness for *network expansion* investments arises because there may be a *scarcity of rights of way*, for example a unique corridor between a cheap and an expensive area. The difficulties that new transmission corridors face in obtaining siting authority suggests that the available corridors for new lines through many areas will be limited in the sense that (for example) one additional corridor may be available through the Pyrenees between France and Spain, and it may accommodate *one* new link that could be of any size between 100 MW and 1000 MW. This scarcity is particularly problematic as demand grows. Merchant investment is then likely to end up in a 'pre-emption and monopoly' situation. In a system with growing demand, pre-emption leads to an investment at the first date at which the discounted value of the financial rights on the additional capacity is equal to the investment cost. The box below also contains an analysis of the incentive to get a toehold by sinking a small investment to pre-empt additional entry and produce monopoly rents for the merchant transmission

[17] The underincentive effect associated with lumpiness is also discussed by Gans and King [2000].

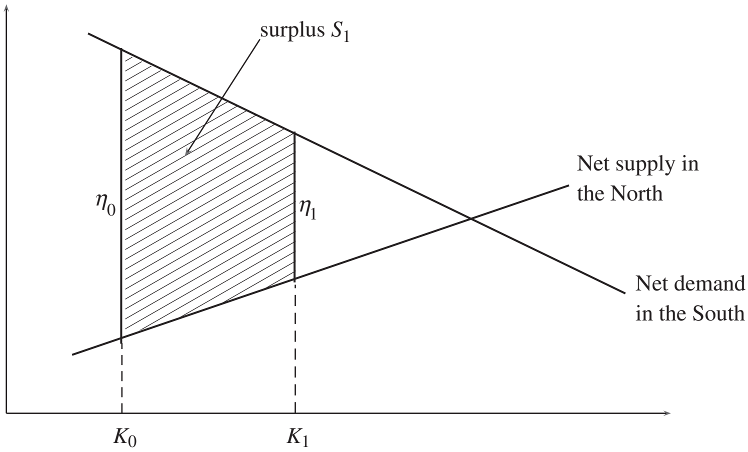

Figure 2

investor and under-investment in transmission capacity generally. A merchant will install a small capacity on the corridor to gain a toehold and will later expand this capacity (presumably, the merchant will underinvest in this expansion as we have seen), to the extent that the expansions are now deepening investments.

Lumpy investments: pre-emption and toeholds

- Suppose that at some future date T, the net demand in the South jumps up (forever) to a new level (the dotted line in figure 3); the post-reinforcement shadow price jumps from η_1 to $\eta_2 > \eta_1$. Letting r denote the interest rate and I the investment cost, suppose that

$$\eta_1(K_1 - K_0) < rI < \eta_2(K_1 - K_0).$$

Then, under free entry into merchant transmission investment, investment occurs at date $\tau < T$ such that

$$\left[\frac{1 - e^{-r(T-\tau)}}{r}\eta_1 + \frac{e^{-r(T-\tau)}}{r}\eta_2\right](K_1 - K_0) = I$$

(see Fudenberg and Tirole [1985]). Note that this pre-emption is actually socially beneficial if the surplus S_1 brought about by the expansion before the increase in demand (the shaded area) exceeds the interest on the investment cost, i.e.:

$$S_1 > rI.$$

Otherwise, pre-emption is socially wasteful.

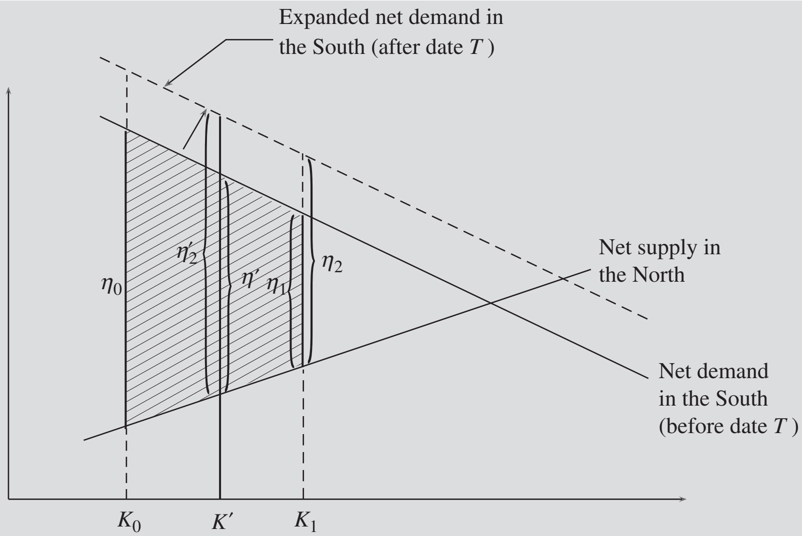

Figure 3

And the point about underinvestment remains: Letting S_2 denote the surplus after demand has grown, if

$$\eta_2(K_1 - K_0) < rI < S_2,$$

then no merchant investment ever takes place even though it is socially desirable.

- Similarly, we can show that pre-emption may encourage inefficiently small investments. Suppose that capacity K_1 can either be reached in one stage, at cost I, as discussed above, or in two stages. We assume that the second-stage upgrade can be performed only by the first-stage merchant investor. The first stage costs I' and yields capacity K', $K_0 < K' < K_1$, which can then be upgraded at cost I'' to K_1.

Let $\eta' \in (\eta_1, \eta_2)$ denote the congestion cost for capacity K' before demand in the South jumps up. Similarly, let η_2' denote the shadow price after T when transmission capacity is K'. See figure 3. The first-stage merchant investor has an incentive to upgrade the facility at date T to yield total capacity K_1 if and only if

$$\eta_2 \frac{(K_1 - K_0)}{r} - I'' > \eta_2' \frac{(K' - K_0)}{r},$$

which we will assume. Let us look for an equilibrium in which a merchant investor preempts at date $\tau < T$ by investing a little (I') and then upgrades the line at time T:

$$I' = \frac{1 - e^{-r(T-\tau)}}{r}\eta'(K' - K_0) + e^{-r(T-\tau)}\left[\frac{\eta_2(K_1 - K_0)}{r} - I''\right].$$

For this, it must be the case that preemption at $(\tau - \varepsilon)$ with the full investment does not pay off:

$$I \geqslant \left[\frac{1 - e^{-r(T-\tau)}}{r}\eta_1 + \frac{e^{-r(T-\tau)}}{r}\eta_2\right](K_1 - K_0),$$

or

$$I - \left[I' + e^{-r(T-\tau)}I''\right] \geqslant \frac{1 - e^{-r(T-\tau)}}{r}[\eta_1(K_1 - K_0) - \eta'(K' - K_0)].$$

The right-hand side of this inequality is negative if the total value of the rights (the total congestion rent) decreases with the capacity of the link in the relevant region. The left-hand side represents the cost of building all the links at once rather than in two steps, and may be positive or negative.

Aside from the timing considerations discussed above, note that given an entry at τ, a social planner might want to jump to capacity K_1 directly, as the difference in social surplus for capacities K_1 and K' exceeds the increase (or decrease) in the flow revenue of rights holdings. Thus, there may be too small an investment.

Remark: Merchant transmission projects that increase capacity between an import constrained area with high nodal prices and an export constrained area with low nodal prices are, in a sense, substitutes for generation projects of equivalent capacity inside the import constrained area. This competition between transmission and generation projects itself raises pre-emption issues. Suppose for instance that a merchant investor plans to invest in a new North-South line, whose acquisition of siting permits plus actual construction will take say ten years.[18] If the transmission project is not built, assume that a generator with equivalent capacity would be built in the South and that such a project would take only three years to obtain siting approvals and to be constructed. There is room for only one of these two alternative ways of reducing the price wedge between North and South. The merchant transmission investor is at a strategic disadvantage even if his project is socially more valuable. If few sunk costs of building a new line are incurred within the first two years, then the merchant investor is likely to cancel his project if the new generation plant in the South is built. Knowing this, the generator may well try to use his short-term investment period to pre-empt

[18] Transmission lines do not take very long to build once they have obtained siting permits. However, for major new transmission corridors, the permitting process can be very lengthy.

the transmission project even if the merchant investor has announced his intention and has started (limited) work on this project.[19]

V (ii). *Gaming between merchant investment projects*

Section 5.1 argued that substitute merchant investments give rise to preemption games. By contrast, complementary merchant investments, assuming that they are best undertaken by separate entities, may give rise to wars of attrition. As illustrated in the following example, with complementary investments each owner has an incentive to dimension his project slightly smaller than the other owner's project; the former project then is the bottleneck and receives the entire congestion rent while the latter exhibits excess capacity and generates no rent. Such a situation is conducive to wars of attrition in which no owner wants to move first by fear of being 'underdimensioned' by the rival owner.

Gaming and coordination of merchant investment projects

An important benefit of market mechanisms is the freedom economic agents enjoy in their investment decisions. They don't need to coordinate with other agents. In the case of electric networks, coordination will be needed not only between transmission and generation investments, as we just discussed, but also between transmission investments. To illustrate this, consider the pair of transmission investment projects depicted in figure 4. Complementary investments from North to Middle and from Middle to South will allow cheap power to flow from North to South. While this pair is really a single investment from an economic viewpoint, the investments may be undertaken by different entities for technological reasons (one is an AC line and the other a DC line and the two companies have different expertise) or other reasons (for example, through separate ownerships of rights of way or different political jurisdictions enabling siting).

Suppose the two complementary projects are built (that is, the standard 'coordination failure' does not arise) and that companies choose their capacities K_{NM} and K_{MS}. The value of the point-to-point rights are

$$K_{NM}(p_M - p_N),$$

and

$$K_{MS}(p_S - p_M),$$

respectively. The incentive for gaming comes from the fact that the lower-capacity line grabs the entire rights' value: Suppose, for

[19] Such timing issues are of course not specific to transmission investments. But the latter are particularly vulnerable to pre-emption strategies due to their long lead time.

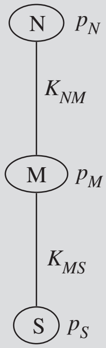

Figure 4

instance, that $K_{MS} < K_{NM}$. Then, $p_M = p_N$ while $p_s > p_{\mathrm{M}}$. This gives rise to a game in which each would like to have a capacity slightly lower than the other. Hence, none dares to move first as the other will then make sure to collect the entire rent.

To be certain, this example is extreme, but it illustrates a general point: Merchant investment is conducive to pre-emption [see section V (i)] and war-of-attrition strategies (Tirole [1988] pp.311–314). These detrimental behaviors can be avoided through a centralized process involving various forms of incentives to promote incentive compatibility. But, by so doing, the market moves closer to a centralized process.

V (iii). *Does the Coase Theorem Apply?*

It is sometimes argued that the problems created by lumpy investments can be resolved through negotiations between the various market participants who will benefit from the investment; that is, that the ‘Coase Theorem’ applies. There are many reasons, exposited in the following example, to believe that negotiations among the affected market participants is unlikely to solve the problems.

Lumpy projects: Will winners get on board and losers be compensated?

As was discussed in section V (i), the price system does not supply the proper signals when the transmission investment has a substantial impact on nodal prices. Merchant investment, to be efficient, then must involve some stakeholders’ process. Its proponents build on the Coase Theorem to argue that winners will participate in the funding of socially desirable projects, while losers will pay enough so to prevent socially undesirable ones.

There are however several issues with this argument, some general and well-known, and some more specific to the context:

- *Transaction costs*: Coasian bargaining involves transaction costs, especially when the number of stakeholders is large.
- *Asymmetric information*: In order to reach an efficient agreement, parties must make reasonable demands. When imperfectly informed as to what others will accept, participants in a bargaining process may end up being too greedy, resulting in an inefficient bargaining breakdown.
- *Absence of future players*: Efficiency requires that all parties be present to negotiate a deal. Future newcomers in generation, transmission and consumption by definition do not sit at the bargaining table and so their interests will not be accounted for by the existing parties.
- *Non-excludability of winners and free riding*: The design of restructured electricity markets imposes uniform prices at a given node: this is mechanistically so in the case of pools, but it applies as well to bilateral-transactions-based systems. Thus, a consumer in the South benefits from the reinforcement of the North-South line and has no incentive to join other consumers in the South and other winners (producers in the North) to contribute to defray the transmission investment if such free-riding does not prevent the investment from being made. The current market design does not call for exclusive rights for the financing parties, and thus encourages free riding in a merchant investment context.
- *Hold-up of potential losers*: To reach an efficient agreement when a socially inefficient transmission investment is being considered, it must be the case that losers be able to 'bribe' the winners not to make this investment. While this bribe enables an efficient outcome 'ex post', when anticipated 'ex ante', it may provide parties with the wrong incentives.

To illustrate the last (hold-up) point, consider a fixed demand q_S in the South (corresponding to gross surplus vq_S — we assume an inelastic demand for computational simplicity). The unit cost of building generation capacity in the South is $I_G \equiv c_0 q_S$, where c_0 is the per-unit cost of investment. This capacity can be built by a monopoly generator (the analysis can be altered to accommodate the case of a competitive set of generators in the South). The variable (ex post) cost of producing q_S is cq_S.

Suppose now that, at cost

$$I_T > I_G,$$

a link with capacity exceeding q_S from another region (North) to South can be built. There is already plenty of competitive generation in the North at the same variable cost c. Thus, the building of the transmission line involves a social loss of $I_T - I_G$ if the generation in the South is not yet built and of I_T if this generation is already in place.

Now suppose that there is investment in generation in the South (the efficient solution). The monopolist charges v per unit, resulting in profit $(v - c)q_S$ (gross of the investment cost I_T). Suppose next that the consumers (or their LSEs) form a coalition and threaten to build the transmission line. If they do so, Bertrand competition brings the price down to c, and so the threat is credible if

$$(v - c)q_S - I_T > 0,$$

which we will assume. Coasian bargaining implies that the monopoly generator will bribe the consumer coalition not to implement the project, where the bribe b lies in an interval:

$$b \in [(v - c)q_S - I_T\ ,\ (v - c)q_S].$$

The lower bound of the interval corresponds to the coalition surplus from bypass, and the upper bound to the quasi-rent enjoyed by the monopolist in the absence of bypass. So, if β denotes the coalition's bargaining power:

$$\begin{aligned} b &= \beta[(v - c)]q_S + (1 - \beta)[(v - c)q_S - I_T] \\ &= (v - c)q_S - (1 - \beta)I_T. \end{aligned}$$

Knowing that it exposes itself to paying a ransom, the generator in the South invests if and only if

$$(1 - \beta)I_T \geqslant I_G,$$

a condition that is violated unless β is small. So, if the condition does not hold, the efficient investment in generation does not occur.

The inefficiency can be prevented by moving Coasian bargaining to an ex ante stage — that is, in effect by having the generator and the consumers engage in long-term contacting. (This solution of course raises other issues, such as the absence ex ante of relevant parties at the bargaining table; or the barrier-to-entry potential of long-term contracts as studied by Aghion and Bolton [1987].)

VI. STATE-CONTINGENT RIGHTS AND DIVERSIFICATION

The existing analyses of merchant investment assume that the capacities K_0 and K_1 are well-defined and non-stochastic, and abstract from some important issues. In practice, even in the two-node model, the actual capacity of the North/South link depends *ex post* on exogenous environmental parameters such as temperature and other exogenous contingencies;[20] furthermore, system operators have substantial discretion in defining and implementing security constraints, affecting the actual power flows on the link in real time. And, of course, even a well-maintained system will have some random outages that cause the available capacity of the link to be reduced.

Moreover, the *ex ante* estimated physical capabilities of a transmission network are defined by relatively crude administrative risk criteria (N-1, N-2) and a set of assumed system 'study' conditions that are discretionary decisions of the system operator that could change in the future.

This all raises the issue of the number of financial rights to be allocated for the existing system and, as a consequence of new investments, how congestion revenue deficiencies or surpluses arising from deviations between the number of rights allocated *ex ante* and the actual capacity of the network *ex post* are handled, and how these allocation and compensation decisions affect investment and the ultimate performance of the system.

We consider this issue in the simple two-node case. Suppose that K varies with the state of nature ω: $K = K(\omega)$, $K'(\omega) > 0$ and ω is distributed between ω^- and ω^+. Let's say that the line is congested for all values of ω, but the value of η will vary with $K(\omega)$. For which value of ω should one compute the number of financial rights? One could be conservative and set the number of financial rights equal to $K(\omega^-)$. One would issue $K(\omega^-)$ financial rights and owe the holders $\eta K(\omega^-)$ in congestion payments. When the realized ω is ω^-, one satisfies the feasibility and revenue adequacy condition. But what happens when $\omega > \omega^-$? The merchandising surplus will exceed what is owed to the rights holders. What does one do with the excess and how does the distribution affect investment incentives? At the other extreme, one could set the number of financial rights to reflect the maximum capacity $K(\omega^+)$. There would be revenue adequacy when $\omega = \omega^+$ but not when $\omega < \omega^+$, which would be most of the time since the system operator would owe $\eta K(\omega^+)$ regardless of the actual realization of ω. Where does the shortfall come from and how does it affect investment incentives? The

[20] The import capacity of Path 15 (connecting Northern and Southern California) when all of the associated AC lines are operational varies between about 2600 MW and 3950 MW depending upon the ambient temperature and remedial action schemes that are in place to respond to unanticipated outages of generating plants and transmission lines. The variance reflects both thermal limits and the application of a set of reliability criteria based on complex system contingency studies. The rated capacity of Path 15 falls by about 600 MW as the ambient temperature rises, other things equal. The rated capacity of Path 15 varies by about 1300 MW depending on the availability of various remedial action schemes to respond to transmission and certain generation outages. California ISO, Operating Procedure T-122A, November 6, 2002. It is also important to recognize that in the U.S. and Europe, there is not a single SO controlling the network, but multiple SOs controlling independent segments of the network. To maintain reliability and avoid free riding, less flexible contingency criteria must be defined than might be the case if there were a single SO operating the network in real time.

answers to these questions necessarily affect the incentives merchant investors will have to make transmission investments.

The impact of a generous [$K(\omega^+)$, say] or conservative [$K(\omega^-)$, say] distribution of rights on investment incentives depends on the way the resulting shortfall or surplus is financed or redistributed. Suppose, first, that one appeals to the taxpayer. Even if taxation were lump-sum, there would still be distortions in investment behavior; generous distributions over-incentivize merchants, while conservative ones under-incentivize them. By contrast, appropriate taxes on users of the transmission network[21] make biased distributions of rights neutral in this radial network, provided that the dispatch is efficient: An efficient dispatch implies that in each state of nature, the cum-tax (or subsidy) price at a given node fully utilizes the link. And so end-users are unaffected by a generous or conservative distribution of rights on the line. Because there is no source or sink of money outside of the industry, rights owners receive the same overall dividend (for example, through a smaller per-unit dividend in the case of a generous distribution). This taxation solution is however complex, since the taxes have to be state contingent.

Rather than dealing with the payment shortfall or surplus problems associated with the kinds of property rights that have been proposed, it seems more natural simply to divide the merchandizing surplus proportionately among the rights' owners based on their relative ownership shares. The next box analyzes the optimal relative allocation rule (in the same way percentage ownership, but not the total number of shares, matters to determine one's proceeds from the distribution of a firm's dividend, the exact number of rights does not matter as long as the merchandizing surplus is distributed proportionately among rights' owners). It shows that the optimal allocation rule derives from standard asset pricing (CAPM) principles in finance. An addition to an existing link is particularly valuable if its actual capacity remains high when the primary link is very congested. Its construction then creates a diversification benefit. These diversification benefits are ignored both in the traditional definition of fixed MW point-to-point property rights payment obligations and with a simple proportional allocation rule.

For example, suppose that the primary North-South AC line exhibits reduced capacity (or breakdowns) during very hot weather. Then an addition along the same path has less social value if it is an aerial AC line than if it is an underground DC line not subject to the same climatic shocks

[21] We here have in mind proportional taxes on electricity in the South (so the price is $p_S + \tau_S$, where p_S is the nodal price in the South) and a tax on exports from the North (where generators receive $p_N - \tau_N$). When actual capacity (K_a) is larger than the number of rights K, this effectively reduces the net dividend paid to rights holders so that there is sufficient revenue to compensate them.

even if average line availability is the same for both. Allocations of rights based only on expected link capacity therefore miss the point that, for a given capacity, some lines may provide better insurance than others. More generally, an allocation of rights solely in proportion to (or equal to) expected capacities provides insufficient incentives to build lines whose availability covaries with the shadow price less (in absolute value) than that of the existing lines.[22]

Non-contingent financial rights under state-contingent capacities

Consider the North-South network (see, e.g., figure 1). The initial expected capacity of the link is K (the actual capacity will in a moment be assumed to be state-contingent). A merchant investor contemplates adding a small amount δK of expected capacity to the line.

Actual dispatching depends on the realization of the state of the world ω. The state of the world encompasses the uncertainty about net demand in the South, $D_S(p_S, \omega)$, that about net supply in the North, $S_N(p_N,\omega)$, and the actual capacity of the lines:

$$[1+\theta(\omega)]K \quad \text{for the existing links(s)}$$

and

$$[1+\mu(\omega)]\delta K \quad \text{for the new facility,}$$

where we can normalize the noises to have zero means:

$$E[\theta(\omega)] = E[\mu(\omega)] = 0.$$

Let us assume that the SO dispatches optimally given the state of nature; and so, in states of nature in which the North-South link is congested:

$$\begin{aligned} D_S(p_S(\omega), \omega) &= S_N(p_N(\omega), \omega) \\ &= [1+\theta(\omega)]K + [1+\mu(\omega)]\delta K. \end{aligned}$$

Let $\eta(\omega) \equiv p_S(\omega) - p_N(\omega)$ denote the (state-contingent) shadow price of the link.

Suppose further that K^* total rights are distributed among all rights owners, including the merchant investor, and that the distribution is

[22] This point applies as well to the case of networks with loop flow. That is, if a transmission investment has a diversification benefit, it must be reflected in the allocation of rights if the rights are relied upon to stimulate efficient transmission investments.

proportional to average capacities; and so the merchant investor receives

$$\frac{\delta K}{K+\delta K}K^* \text{ rights.}$$

The merchandizing surplus is distributed to rights owners. Needless to say, distributions of rights that would not reflect expected capacities could by themselves introduce a bias in merchant investors' incentives. For example, suppose that the incumbents in the past received a very generous rating of the existing lines from North to South, while rating standards are strengthened for new comers. The latter then receive a disproportionately small share of total rights, which penalizes them when the merchandizing surplus is distributed among rights' owners, and thereby gives little incentive to sink new investment. To avoid such obvious biases, we assume an allocation of rights proportional to average capacity. Even so, merchant incentives may be inappropriate, as we will see shortly.

The congestion dividend, $d(\omega)$ paid to the owner of a right is therefore:

$$K^* d(\omega) = [[1+\theta(\omega)]K + [1+\mu(\omega)]\delta K]\eta(\omega).$$

The merchant investor's expected revenue, R, is therefore

$$\begin{aligned} R &= E_\omega\left[\left(\frac{\delta K}{K+\delta K}K^*\right)d(\omega)\right] \\ &= (\delta K)E_\omega\left[\frac{[1+\theta(\omega)]K + [1+\mu(\omega)]\delta K}{K+\delta K}\eta(\omega)\right] \\ &\simeq (\delta K)E_\omega[[1+\theta(\omega)]\eta(\omega)] \end{aligned}$$

or δK small.

By contrast, the increase in social welfare is

$$\delta W \simeq (\delta K)E_\omega[[1+\mu(\omega)]\eta(\omega)],$$

Since in state ω, the merchant investor's expansion delivers $[1+\mu(\omega)]\delta K$ units of transmission capacity which each have value $\eta(\omega)$.

Hence,

$$R \leqslant \delta W \iff \operatorname{cov}(\mu,\eta) \geqslant \operatorname{cov}(\theta,\eta).$$

Let us draw the implications of this simple characterization in specific environments:

Example 1 (diversification effect). Suppose that all uncertainty results from line availability. Existing line(s) may exhibit reduced capacity due to harsh weather (extreme heat) conditions. The merchant investor's line, by contrast is not (or is at least less) subject to these harsh weather conditions (or is better protected against them). For example, the new line could be underground, or cross a climatically distinct area. Then $\theta(\omega)$ and $\eta(\omega)$ are (in a first approximation) perfectly negatively correlated, while η is not perfectly correlated with θ:

$$\mu = k\theta + \varepsilon$$

with

$$k < 1 \quad \text{and} \quad E(\varepsilon|\theta) = 0.$$

Hence:

$$\text{cov}(\mu, \eta) \geqslant \text{cov}(\theta, \eta),$$

implying

$$R < \delta W.$$

Non-contingent rights create an under-incentive to invest. Intuitively, the new line supplies a disproportionately high share of the transmission capacity in those states of nature in which transmission capacity is scarce and therefore very valuable. This contribution however is not reflected in the distribution of dividends which is based on fixed (non state-contingent) shares. It is only when the availabilities of the lines (old and new) are perfectly correlated that the private and social incentives coincide.

The analysis can be generalized to encompass uncertainty about energy market participants' demand and supply curves. Suppose that

$$\begin{aligned} S_N &= a_N + b_N p_N + \varepsilon_N \\ D_S &= a_S - b_S p_S + \varepsilon_S. \end{aligned}$$

So $\omega = (\theta, \mu, \varepsilon_N, \varepsilon_S)$. Under efficient dispatching

$$\eta(\omega) \simeq \left[\frac{a_S}{b_S} + \frac{a_N}{b_N}\right] + \left[\frac{\varepsilon_S}{b_S} + \frac{\varepsilon_N}{b_N}\right] - \left[\frac{1}{b_S} + \frac{1}{b_N}\right] K(1+\theta).$$

This implies that the analysis above generalizes when line availabilities are independent of demand and supply shocks (cov (ε_i, θ) = cov $(\varepsilon_i, \mu) = 0$ for $i = N, S$).

On the other hand, line availability may be related to demand and supply shocks. For example, it may be that a line (old or new) is subject to the same climatic shock as the demand node. Hot weather may simultaneously increase demand and limit the capacity of the line bringing electricity from a cheaper node (precisely when the line is most needed). Such a line obviously has a lower social worth than one whose availability is less negatively correlated with increases in demand at the expensive node.

Example 2 (uncertainty about energy market players only). Suppose that there is no uncertainty about the actual capacities of the lines:

$$\theta(\omega) = \eta(\omega) = 0 \quad \text{for all} \quad \omega.$$

Hence all uncertainty comes from generation and consumption. In this case, the private and social incentives coincide:

$$R = \delta W.$$

VII. NOMINAL AND ACTUAL TRANSMISSION CAPACITY

The difficulty in putting a number in front of a line's 'capacity' (in the two-node case) raises other issues. In a nutshell, the actual capacity of the line depends on discretionary choices, and under a merchant transmission model control is likely to be separated from ownership due to the conflict of interest associated with the measurement of congestion rents. This section considers one such discretionary action: dispatching.

As we have already noted, rewarding merchant investment through congestion rents requires separating ownership and dispatch in order to obtain an unbiased measure of this rent. But this separation of ownership and dispatch raises a moral-hazard-in-teams problem. The electric system's state-contingent output (to simplify, the intensity of power in the absence of outage and the probability and duration of an outage) depends on both the care and the forecasts of the owner (the quality of the line, its maintenance, and its adequacy to consumers' needs) and the quality of the management of the grid by the system operator, as the latter must use her acumen to get lots of power through without creating a high risk of outage. This raises two points: First, one cannot consider incentives given to merchant investors without also specifying those of the system operator. Second, moral hazard in teams reduces accountability. An outage can be claimed to result from poor line maintenance or from imprudent dispatching. Conversely, high power prices may be due to an undue conservatism of the system operator or to a proper dispatching motivated by low line quality.

There is also a potential moral-hazard-in-teams problem among line owners. Recall that merchant investment incentives are better aligned with the public interest when merchants own inframarginal transmission facilities whose congestion rent is to be preserved. The total North-South capacity may then belong to different owners. The same value of a given actual capacity K selected by the independent system operator may correspond to different quality configurations of the various components of the network with multiple owners. The question is then one of allocation of total capacity and congestion rents among the different owners.[23]

Moral hazard in teams : transmission owners and system operator

Consider the North-South network. Let K denote the nominal capacity of the line. In a first step, we assume that this capacity is known to the system operator (for example, the line's maintenance is perfectly observed by the SO). The system operator chooses to allow an amount $\widehat{K}$ to flow through the link. The greater is $\widehat{K}$, the higher is the probability that the link breaks down. We assume that with probability $x(\widehat{K}-K)$ the link breaks down and no power flows through it. With probability $1-x(\widehat{K}-K), \widehat{K}$ flows through. The function x is increasing in $\widehat{K}$. Let $L(\widehat{K})$ denote the out-of-merit dispatch cost when the realized capacity of the line is $\widehat{K}$. [For example, in figure 1, L was equal to the area of the triangle ABC (for capacity K)]. We assume that there is no market power at either node and so L represents the social loss attached to the inability to import power without constraint from the North. Note that

$$L' = -\eta.$$

The socially optimal dispatch solves, for a given K,

$$\min_{\widehat{K}}\left\{x\left(\widehat{K}-K\right)L(0)+\left[1-x\left(\widehat{K}-K\right)\right]L\left(\widehat{K}\right)\right\}.$$

And so $\widehat{K}=\widehat{K}^{*}$ is given by

$$x'\left(\widehat{K}^{*}-K\right)\left[L(0)-L\left(\widehat{K}^{*}\right)\right]+\left[1-x\left(\widehat{K}^{*}-K\right)\right]L'\left(\widehat{K}^{*}\right)=0.$$

The marginal social gain from capacity expansion is then (using the envelope theorem):

$$\frac{dW}{dK}=x'[L(0)-L(K)]=(1-x)\eta.$$

And so, if the marginal investment is rewarded by the congestion cost in the absence of outages, merchant investors face the proper signal for investment.

[23] At an abstract level, one can view transmission owners and the SO as a team (in the sense of Holmström [1982]) jointly delivering an output — state-contingent power— to the final consumers. A general principle is that proper incentives require that each member of the team be made the residual claimant for the team performance $(xL(0)+[1-x]L\left(\widehat{K}\right)$ in the notation of the next example). Making each member residual claimant may however be very costly because of adverse selection and collusion.

• *Dispatcher with conservative incentives.*
Turn now to the system operator's incentives. Suppose that the SO is penalized more for outages than she is rewarded for increases in the amount of power flowing through the network ; that is, she solves:

$$\min_{\widehat{K}}\left\{x\left(\widehat{K}-K\right)\theta L(0)+\left[1-x\left(\widehat{K}-K\right)\right]L\left(\widehat{K}\right)\right\},$$

where $\theta > 1$. This yields first-order condition:

$$x'\left[\theta L(0)-L\left(\widehat{K}\right)\right]+[1-x]L'\left(\widehat{K}\right)=0.$$

The marginal social gain from capacity expansion is

$$\frac{dW}{dK}=x'\left[L(0)-L\left(\widehat{K}\right)\right]-x'(1-\theta)L(0)\frac{d\widehat{K}}{dK}.$$

And so

$$\frac{dW}{dK}=(1-x)\eta+x'(1-\theta)L(0)\left[1-\frac{d\widehat{K}}{dK}\right],$$

using the SO's first-order condition.

Next, rewrite the SO's optimization program as the choice of a risk factor $\Delta \equiv \widehat{K}-K$:

$$\min_{\widehat{K}}\{x(\Delta)\theta L(0)+[1-x(\Delta)]L(K+\Delta)\}.$$

The cross-partial derivative of the minimand with respect to K and Δ is positive as $L'' > 0$, and so, by a revealed preference argument, Δ is decreasing in K, or:

$$\frac{d\widehat{K}}{dK}<1.$$

In words, the system operator takes less risk as K increases, because the marginal gain from increased throughflow decreases. We therefore conclude that

$$\frac{dW}{dK}<(1-x)\eta,$$

and so congestion rent payments over-incentivize merchant investors. In a sense, the SO's conservative behavior implies that insufficient use will be made of the added capacity and so the shadow price of the link overstates the value of additional capacity. This result shows that one cannot properly analyze merchant investment (or, for that matter, the incentives of a Tranco company not responsible for dispatching) without considering the independent system operator's incentives.

VIII. CONCLUSION

We conclude that the apparently attractive properties of the merchant transmission investment model are seriously undermined when more realistic assumptions about the attributes of transmission networks, transmission investment and unregulated power markets are introduced. Clearly, policymakers cannot proceed under the assumption that they can avoid dealing with the difficult issues associated with stimulating efficient investment in electric transmission networks simply by adopting the merchant investment model. The merchant model ignores too many important attributes of transmission networks and the behavior of transmission owners and system operators. In principle, a regulated Transco model can deal directly with issues associated with lumpy investment, market power in wholesale power markets, gaming behavior of merchant investors and stochastic attributes of transmission capacity, and avoids the need to separate transmission ownership and system operations. However, a regulated Transco model will necessarily confront inefficiencies resulting from asymmetric information and political interference in planning and investment processes and may be less effective than a merchant model in providing the high powered incentives that lead to the identification of innovative transmission investment options, construction costs minimization and efficient tradeoffs between generation and transmission investments. The role of reliability rules driven by the unusual physical attributes of electricity and the need to maintain network frequency, voltage and stability parameters on a continuous basis leads to additional complications that have hardly been addressed in either the merchant investment or regulated Transco models.

We believe that it is unlikely that policymakers will be able to rely primarily on a merchant model to govern transmission investment and the limited amount of merchant investment that has been forthcoming to date in electricity markets where it is permitted and encouraged, is consistent with this view. An important research challenge is to develop good regulatory mechanisms to apply to regulated Transcos that also provide opportunities for merchant investors to develop projects when they are the most efficient options.

REFERENCES

Aghion, P. and Bolton, P., (1987), 'Contracts as a Barrier to Entry,' *American Economic Review*, 77(3), 388–401.

Baldick, R. and Kahn, E., (1992), 'Transmission Planning in the Era of Integrated Resource Planning: A Survey of Recent Cases,' Lawrence Berkeley Laboratory, LBL-32231.

Bohn, R. E.; Caramanis, M. C. and Schweppe, F. C., (1984), 'Optimal Pricing in Electrical Networks over Space and Time,' *Rand Journal of Economics*, 15(3), 360–376.

Bushnell, J., (1999), 'Transmission Rights and Market Power,' *Electricity Journal*, 12(8), 77–85.

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

Bushnell, J. and Stoft, S., (1997), 'Improving Private Incentives for Electric Grid Investment,' *Resource and Energy Economics*, 19, 85–108.

Celebi, M., (undated), 'An Analysis of Incentives to Provide Line Capacity and Reliability in Deregulated Power Networks,' mimeo, Brattle Group.

Chao, H. P. and Peck, S., (1996), 'A Market Mechanism for Electric Power Transmission,' *Journal of Regulatory Economics*, 10(1), 25–59.

Chisari, O., *et al.* (2001), 'High-Tension Electricity Network Expansion in Argentina: Decision Mechanisms and Willingness-to-pay Revelation,' *Energy Economics*, 23, 697–715.

Fudenberg, D. and Tirole, J., (1985), 'Preemption and Rent Equalization in the Adoption of a New Technology,' *Review of Economic Studies*, 52(170), 383–401. Reprinted in *Readings in Industrial Economics*, L. Cabral ed. (2000), Blackwell Publishers.

Gans, J. and King, S., (2000), 'Options for Electricity Transmission Regulation in Australia,' *Australian Economic Review*, 33, 145–160.

Glachant, J-M. and Pignon, V., (2002), 'Nordic Electricity Congestion's Arrangement as a Model for Europe: Physical Constraints and Operators' Opportunism', mimeo.

Green, R., (1999), 'The Electricity Contract Market in England and Wales,' *Journal of Industrial Economics*, 47(1), 107–124.

Green, R., (1997), 'Transmission Pricing in England and Wales,' *Utilities Policy*, 6(3), 185–93.

Hansmann, H., (1996), *The Ownership of Enterprise*, Belknap Harvard.

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

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

Holmström, B., (1982), 'Moral Hazard in Teams,' *Bell Journal of Economics*, 13, 324–340.

Holmström, B. and Tirole, J., (2000), 'Liquidity and Risk Management,' *Journal of Money, Credit and Banking*, 32(3), 295–319.

Joskow, P., (2002), 'Electricity Sector Restructuring and Competition: A Transaction Cost Perspective,' in Eric Brousseau and Jean-Michel Glachant (eds.) *The Economics of Contracts: Theories and Applications*. Cambridge: Cambridge University Press.

Joskow, P. and Schmalensee, R., (1983), *Markets for Power*, MIT Press.

Joskow, P. and Tirole, J., (2000), 'Transmission Rights and Market Power on Electric Power Networks,' *Rand Journal of Economics*, 31(3), 450–487.

Joskow, P. and Tirole, J., (2003), 'Merchant Transmission Investment,' mimeo, MIT and IDEI.

Joskow, P. and Tirole, J., (2004), 'Reliability and Competitive Electricity Markets,' mimeo, MIT and IDEI.

Laffont, J-J. and Tirole, J., (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.

Léautier, T., (2000), 'Regulation of An Electric Power Transmission Company,' *The Energy Journal*, 24(1), 61–92.

Nasser, T. O., (1997), 'Imperfect Markets for Power: Competition and Residual Regulation in the Electricity Industry,' PhD Dissertation, MIT Department of Economics.

Oren, S., (1997), 'Economic Inefficiency and Passive Transmission Rights in Congested Electric Systems with Competitive Generation,' *The Energy Journal*, 18, 63–83.

Oren, S.; Spiller, P.; Varaiya, P. and Wu, F., (1995), 'Nodal Prices and Transmission Rights: A Critical Appraisal,' *The Electricity Journal*, 8(3), 24–35.

Patton, D., (2002), 'Summer 2002 Review of New York Electricity Markets', October 15, 2002, pp.25, 42–64, New York Independent System Operator.

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.

Stoft, S., (1999), 'Financial Transmission Rights Meet Cournot: How TCCs Curb Market Power,' *The Energy Journal*, 20, 1–23.

Tirole, J., (1988), *The Theory of Industrial Organization*, MIT Press.

Vogelsang, I., (2001), 'Price Regulation for Independent Transmission Companies,' *Journal of Regulatory Economics*, 20(2), 141–165.

Wilson, R., (2002), 'Architecture of Power Markets,' *Econometrica*, 70(4), 1,299–1,340.

Wu, F. P.; Varaiya, P.; Spiller, P. and Oren, S., (1996), 'Folk Theorems on Transmission Access: Proofs and Counterexamples,' *Journal of Regulatory Economics*, 10(1), 5–24.