Transmission Investments: Who really benefits from "beneficiaries pay?"[¤]

James Bushnell

Department of Economics, University of California, Davis

June, 2014

Abstract

This paper discusses the merits and implications of allocating the capital costs associated with transmission investment in a manner that is proportionate to the estimated benefits the investment provides. This approach has been referred to as "beneficiaries pay" in the United States. I argue that in most restructured markets, a beneficiaries pay approach, while a perfectly reasonable way to allocate costs on the grounds of fairness, is unlikely to contribute much to the efficiency of either generation or transmission investment. In markets where transmission access is priced efficiently according to congestion pricing principles, and transmission planning is not unduly influenced by the location decisions of individual firms and plants, congestion pricing provides an appropriate signal for the location of generation assets. Any further locational prices signals provided by a cost allocation mechanism could therefore distort investment efficiency. Finally, an assignment of costs based upon precise ex-post measures of benefits carries a great risk of distorting behavior and the revelation of information. It is these distortions that pose the greatest potential risks to the efficient construction and usage of network assets.

I. Introduction

There have been tremendous advances over the last two decades in the efficient pricing of electricity transmission usage and congestion. What were once considered intractable problems with unpacking the sources and causes of network congestion have now, in much of the world, become routine components of half-hourly energy transactions. At the same time, the allocation of the capital costs of construction of electricity transmission assets remains a vexing, and in many ways more fundamentally difficult challenge.

There has been a continual tension between the appealing simplicity of allocating costs pro-rata amongst consumers (e.g. "socializing") through postage-stamp rates and the attractive equity of allocating the costs of projects in a manner that aligns with the perceived benefits they provide. The success of congestion pricing methods, such as locational marginal pricing (LMP), which have reversed the congestion problems created by charging for *usage*

[¤] An ealier version of this paper was sponsored by Trustpower Limited. It was commissioned in response to the working paper on beneficiaries-pay methods for pricing transmission published by the New Zealand Electricity Authority on 21 January 2014 (Available at http://www.ea.govt.nz/dmsdocument/17482). I am particularly indebted to James Tipping for helpful comments and input.

based upon postage-stamp methods, has helped advance the notion that postage-stamp rates are inefficient in all contexts. The United States Federal Energy Regulatory Commission (FERC), in its Order 1000, has contributed to this perception by promoting a principle of "beneficiaries pay" as a basis for capital cost recovery.

In this paper, I will argue that beneficiaries pay, while a perfectly reasonable principle for framing the *fairness* of allocating costs, offers little real benefit in terms of the *efficiency* of investment incentives in markets with sophisticated congestion pricing of transmission. Congestion pricing provides the correct signal for location decisions as long as those location decisions do not by themselves endogenously trigger new transmission investment.

The efficiency of generation location decisions becomes problematic when a market does not price short-term congestion efficiently. Markets such as California during the 2000s and the UK today have had problems with too much investment in generation pockets. These investments have triggered either costly payments to generators or expensive transmission upgrades, or both. Under these circumstances it can be argued that a beneficiaries pay approach could have forced such generators to internalize the costs of their location decisions.

In markets that have already adopted efficient congestion pricing in the form of LMP, and that have a transmission planning process that is not distorted by undue influence of the decisions of specific generation firms or large customers, the local prices themselves provide the correct incentives for generation investment.[1] Ironically, the application of efficient congestion pricing, by providing appropriate long-term signals for generation location, mitigates any long-run distortions of a postage-stamp based approach for capital-cost recovery. Even in the United States, beneficiaries pay has proven to be more of a rhetorical device than an economic reality, even after FERC's Order 1000.

This paper builds on similar examinations of these issues in more academic (Joskow and Tirole, 2005) and policy-based (Baldick, et al, Hogan, 2011) settings.[2] Most of these papers focus on broader issues than cost allocation. While most of these papers do not endorse beneficiaries pay principles, they do not reject them either. Hogan (2011) points to an efficiency effect, but this is focused on the prospect of rate-based investment crowding out "voluntary" (e.g. merchant) investment by direct beneficiaries in "mixed" systems.

In policy settings, beneficiaries pay has a strong appeal to a sense of fairness in cost allocation. An allocation of costs that is viewed as being roughly aligned with benefits, for example on a regional basis, can mitigate discontent with the transmission planning process and thereby reduce delays in investment. In the United States, this enhancement of the process has been the primary "efficiency" claimed by supporters of benefits based allocation. In addition, in markets lacking an effective means to ration transmission capacity (via LMP, for example), or with planning processes that can be unduly influenced by the locational decisions of individual generators, the allocation of the costs of transmission to those who benefit can be a means to offset other incentive problems.

---

[1] This type of transmission planning has been characterized as "transmission leads generation" as opposed to "generation leads transmission." See Ross Baldick, James Bushnell, Benjamin Hobbs, and Frank Wolak. "Optimal Charging Arrangements for Energy Transmission." Report prepared for the Great Britain Office of Gas & Electricity Market. 2011.

[2] Paul Joskow and Jean Tirole. "Merchant Transmission Investment," *The Journal of Industrial Economics*, 5(2): 233-264.

William Hogan, "Transmission Benefits and Cost Allocation." Harvard Electricity Policy Group paper. 2011.

However, the relatively modest steps taken to-date toward the allocation of costs to beneficiaries have utilized *ex-ante* estimates of those benefits. While dependent upon potentially heroic assumptions, an ex-ante determination of benefits at least minimizes the risk of distortions to market behavior in the pursuit of manipulating the measurement of benefits. This added risk arises when benefits are measured *ex-post*. By distorting market behavior, precise measures of ex-post benefits threaten to overwhelm any of the relatively modest efficiencies a benefits-based allocation might provide.

II. Categorizing Efficiency

There are several aspects to economic efficiency. The *term allocative efficiency* is used by regulators, such as the New Zealand Electricity Authority, to describe the utilization of the transmission network. In other words, in order to meet demand, is the responsibility for the production of electricity allocated in the most efficient manner? The transmission network clearly plays a role in this concept, as sometimes it might be of least cost to supply all of San Francisco's demand with hydro-power from the Pacific northwest if the transmission network allowed. But frequently transmission constraints make such an allocation impossible. Therefore allocative efficiency seeks to find the least cost solution to meet demand subject to the limits of the network.[3]

In the transmission pricing context, allocative inefficiencies can arise when transmission prices differ substantially from the marginal costs of providing transmission services. Here networks such as telecommunications, transportation, or electricity offer extreme examples, as the *marginal cost*, or cost of transmitting one more kWh, or one more phone call, is usually very low if the network is uncongested, but can be quite high if the network is congested.

A classic example of allocative inefficiency is when a low-cost generator reduces output because it can't make enough revenue from sales to cover both its own costs *and* the transmission charge, even when the network is uncongested. Despite the fact that it would be costless for the network to carry more power, the capacity is unused because prices are well above this low marginal cost level. Anecdotal stories of this kind of inefficiency are disturbingly common. In New Zealand, the EA describes the under-production by South Island generators even during times when the HVDC link from the South Island is unconstrained.[4] In the United States, Calpine is considering retiring its Sutter Power plant in part because of high transmission charges related to investments made by the Western Area Power Administration.

I will use the term *investment efficiency,* to describe the long-term, or investment, decisions of key market participants. In the transmission context there are two related concerns. The first is whether the "right" transmission projects are being developed and implemented. The second is whether power plants and, to some extent, customers, choose the "right" technologies and places to locate given the current and future network. The challenge for

[3] Economists also frequently concern themselves with consumptive efficiency, which relates to whether customers are consuming the right amount of product, and whether the value they place on that product is greater than the cost of producing it. Given that most electricity demand is quite inelastic and not subject to very dynamic prices, this is usually a secondary concern in electricity markets relative to the other forms of efficiency. The discussion is simplified by focusing on production inefficiencies, but the general conclusions of this paper would not change if demand were somewhat elastic.

[4] See section 4.3 of the Electricity Authority's *Transmission Pricing Methodology: Issues and Proposal* Consultation Paper, October 2012. Available at http://www.ea.govt.nz/dmsdocument/13799.

long-term (or dynamic) efficiency in the transmission context stems from the fact that transmission assets take a long time to site and build, are very long-lived (lasting many decades) and that they are physically very large, and economically disruptive, investments.

A classic example of an investment inefficiency would be a situation where an expensive line were built out to a resource rich area (for example an area with promising wind characteristics), yet no generation resources locate there because of the large costs they would incur to pay for the transmission line that had just been built. This would result in an expensive asset going unused. This is a classic concern with charging methods that allocate sunk costs to price-elastic uses. For example, the current New Zealand cost-allocation method appears to have discouraged otherwise economic generation from being developed in the South Island.[5] Another example of an investment inefficiency arises when a generator locates in a remote region with low-cost fuel (such as wind or hydro) and by doing so *causes* expensive network investments to be made. If the network investment was inevitable, because transmission access to the resource-rich region was needed to meet system goals or the region was the obvious source of additional low-cost production, there would be little efficiency consequence. However if the resulting transmission costs *exceed* the value of the potential generation, then investments taken as a whole would be inefficient.

These examples highlight the importance of the role of causality when developing policies to mitigate inefficiency. A distinction needs to be made between investments that made sense ex-ante, but turned out not to be so, and decisions that are expected to be socially inefficient even at the time they were made, but were undertaken because the charging regime distorted an investor's incentives. Take for example, a wind producer's decision to add capacity in Scotland in anticipation of expanded transmission capacity there. It could be the case that this location decision is only profitable if the transmission costs were socialized to load. Even so, the generation investment is inefficient only if it were the trigger that *caused* the transmission investment to be undertaken. If the transmission expansion were inevitable, the expansion of wind capacity to take advantage of that new transmission capacity is the desired outcome. Otherwise, the expensive new capacity would sit unused, decreasing benefits to everyone.

It is the latter category of decisions that we should focus on when developing a cost-recovery policy – those inefficient decisions made specifically in response to the incentives provided by the cost recovery mechanism.

i. Defining Efficient Grid Investment

A grid planning process, when done correctly, tries to develop a network that maximizes the social benefits of its use. To a large extent this means developing a network that can reliably satisfy demand at the lowest possible cost, while also accommodating other social goals such as a reduction in greenhouse gas emissions. While quantifying these benefits over a long time horizon can in practice be very challenging, the concept is simple: if the increased benefits (marginal benefits) exceed the costs of network investment (marginal cost), then the investment should go forward.

In markets where investments can be very small and incremental, market prices solve this calculation in a simple way. The marginal benefit of delivering one more kWh of power is captured by what the last consumer is currently willing to pay for power (the current price). When the investment is small, the current price is virtually the same as the price *after* the investment and so a decision-maker can simply compare prices before the investment to the cost of the investment.

[5] Ibid.

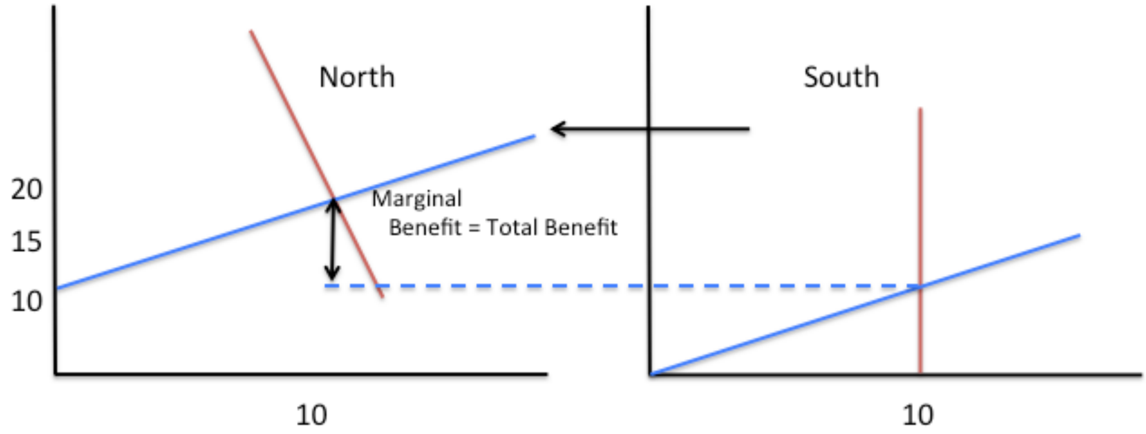

Figure 1

To illustrate this discussion, I will draw upon a simple example of supply (e.g. generation) in two locations, as illustrated in Figure 1.[6] At the current levels of demand, the market clearing price for power is $20/MWh in the North and $10/MWh in the South. The marginal benefit of a very small connection between these markets would be $10. Each extra MW of transfer capacity would increase MC in the South by about $1 and decrease it by $1 in the North. If small network investments were possible, the market signal for transmission investment provided by these two markets would expand capacity until the price differential in the two markets equaled the marginal cost of that transmission capacity. For example if transmission capacity cost about $6 a MW for each MW constructed, transmission would expand until the price differential were about $6 and then stop. In this case this would imply a little more than 2 MW of transmission capacity added.[7]

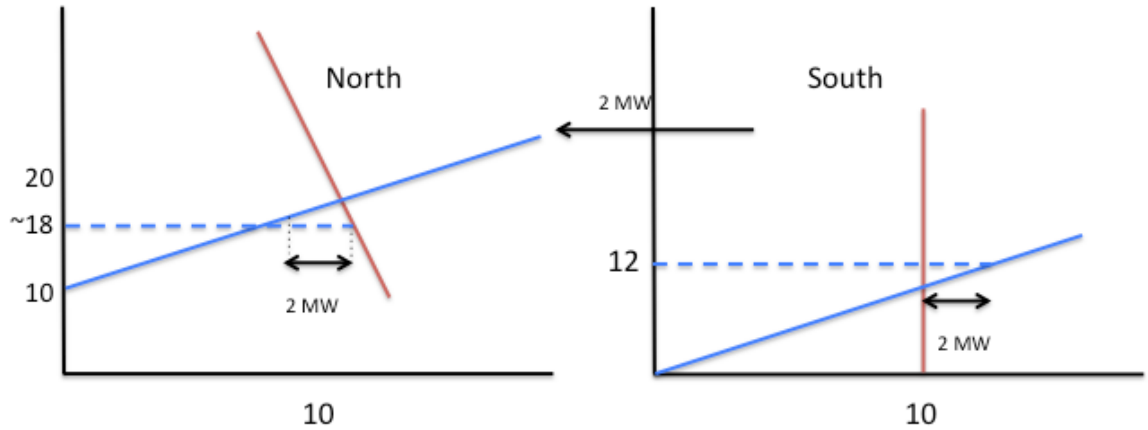

Figure 2

Because grid investments are lumpy and disruptive, prices alone don't provide the answer. Return to our previous example and assume that small investments are infeasible or impractical, so that a 10 MW line is the smallest practical interconnector to construct. This means that power will flow from South to North as long as the interconnection is unconstrained and costs are lower in the south. In fact somewhat over 5 MW (more than 5

[6] The details of the example are not necessary, but for completeness sake we can assume that supply in the Northern market is MC = 10 +q and in the South it is simply MC = q. Thus the south has a cost advantage as long as the supply in the south is less than 10 greater than in the North.

[7] If demand were completely inelastic, this would result in exactly 2 MW of transfer capability, because demand increases when price declines in the North, generation in the North would decrease by a little less than 2 MW, leaving prices a little above $18/MWh. We will ignore those differences for the moment for the sake of simplifying the discussion.

since demand has grown in the North) is enough to raise costs in the South to about \$15/MWh and lower them in the North to the same.

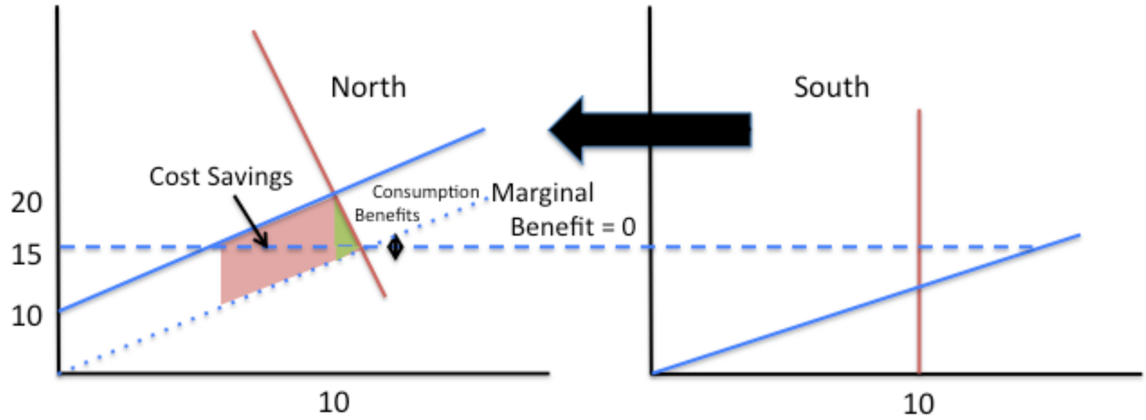

Figure 3

Consider that after a relatively large investment such as this, the marginal cost of delivering that power would be zero (no congestion), meaning the *price* of delivering that power may have gone from \$10 to nothing because of the investment. Although the last couple MW of capacity may have no current value (thus the zero price for its use), the many MWs of capacity that came with it did create much value.

This kind of problem illustrates why market-driven investment for transmission, where the reward is collecting the value of congestion costs, is so problematic. An investment may create many millions of dollars of benefits, but allocative efficiency dictates that the usage of this asset be set at marginal cost, which means no charge if the investment results in excess capacity. Thus with efficient market pricing, a very valuable asset would create no revenue for its investor. Such investments would therefore have to appeal to benefits beyond the collection of congestion rents.[8] If one individual firm, say a large generator or customer, could enjoy enough benefit by itself that it is willing to take on the investment, then there is still no market failure. Others may benefit from their investment, but that is a *pecuniary externality*, one that does not distort decisions from the efficient ones. Often though, the scale of network investments is large enough that one must combine the benefits of many, many network users to reach a level of benefit that exceeds the cost of the investment. This fact, along with other challenges, conspires to make merchant-based transmission investment an isolated and rare occurrence.[9]

ii. The Timing and Capping of Benefits

The above example, and similar ones used by Hogan (2011) are useful for illustrating the conceptual magnitude and distribution of the benefits to network expansion. They are, however, stylized and static examples. One point to highlight is that the benefits to infrastructure capacity are in many instances highly concentrated in relatively few days or even hours of a year. This is most often the case with any capacity (transmission or generation) that prevents or reduces the probability of load curtailment. The involuntary interruption of load is generally considered very costly in the range of \$3,000/MWh to \$5000/MWh in many planning contexts. and the risks of such interruptions are generally limited to a small number of periods during a year. Thus a transmission line that may

[8] See Bushnell and Stoft, "Improving Private Incentives for Electric Grid Investment." *Resource and Energy Economics* (19):1-2, 1997.

[9] Joskow and Tirole, "Merchant Transmission Investment", *Journal of Industrial Economics*, 2005.

provide a \$10/MWh cost saving during normal times, provides a benefit that is *500 times* greater during periods of scarcity. It is easy to see how a small number of these scarcity hours can overwhelm any savings during other periods.

These specific magnitudes are based upon an artificially precise penalty value used in the operation of the system. They are reflective of the fact that shortages, absent advance load management policies, are extremely costly. This fact, combined with the general lack of storage capability in most power markets, produces price dispersions that would be considered extreme in most markets, but are to be expected in power markets.

One implication of this is that a truncation of the calculated benefits in any given hour can lead to a major adjustment in the underlying magnitude and distribution of those benefits. This would be inconsistent with a market-based outcome, in which such concentrated benefits are the norm, rather than the exception.

That said, extreme spot market prices are generally tolerated only when they can be accompanied by adequate opportunities to hedge the risk of those extreme prices. Financial instruments such as Contracts for Differences and Financial Transmission Rights are necessary cornerstones of restructured power markets. Forward markets are critical for the functioning of power markets, which would otherwise produce high levels of risks for both sides of the market, and suffer from elevated levels of market power.

It is important that a transmission cost allocation scheme provide either some degree of certainty about cost, or provide instruments, such as FTRs, through which this uncertainty could be hedged. These financial instruments do not *truncate* the expected benefits or costs to one party or another. They merely eliminate the uncertainty of the magnitude of such benefits. A cost allocation scheme that seeks to mitigate risks simply by eliminating extreme outcomes becomes a de-facto alternative form of cost allocation.

iii. Winners and Losers from Transmission Investment

The above discussion describes the conditions necessary for market-based investment in transmission assets in the absence of leadership by a monopoly grid company. In general a coalition of users with aggregate total benefits in excess of the costs of investment would have to form and reach mutual agreement on the allocation of costs within the coalition.

Absent from this discussion has been those who are harmed by the network investment. This would include generators in import zones where prices decline due to the transmission, and customers in the export zones whose prices may rise. Their negative benefits would not be part of the calculation undertaken by the investor coalition. As noted in the following section, however, users may easily be made better off by a new investment in some periods, and worse off in others. Parties will therefore make their assessment of whether or not to participate in the coalition based on what I will call their *offsetting* benefits – that is the total benefit they receive from the investment in periods they are made better off, less the disbenefit they receive in periods they are made worse off. I will use the term *net* benefit to refer to the all-in benefits (including charges) to an individual. This is the offsetting benefit less the cost. An individual party will therefore only believe that it is a true *beneficiary* of a project, and decide to join the coalition, if it is made better off by the investment in aggregate, over the lifetime of the investment. The success of the coalition will be based critically on each member determining that its individual offsetting benefits will exceed its individual share of the charges of the investment.

In a market-based setting, those parties made worse off by an investment (i.e. those that are not beneficiaries) are usually not compensated by those undertaking investment. This is not necessarily an admirable phenomenon. In fact economists have identified several cases where market-driven free entry can be inefficient.[10] In the electricity context this problem can be exacerbated by the fact that certain investments can be privately profitable because they create *more* congestion rather than relieve it. Both these points, along with obvious needs to evaluate the reliability impacts of any network additions, imply that any network investment, regulated or merchant, needs to be vetted by a credible central decision making authority.

The standard articulated in the beginning of this section for project approval, that total estimated aggregate offsetting benefits outweigh total estimated costs, will as part of its calculation of benefits net out the transfers from one party to the other. Ignoring the cost of the physical investment itself, the total estimated benefit of the investment over its lifetime will be the sum of the estimated offsetting benefits to all individual parties. This calculation includes those that are made worse off (i.e. those that have a negative offsetting benefit) as well as those made better off. Thus an investment process based upon maximizing social benefits will implicitly account for the losses to some parties, and should reject projects when those losses are not sufficiently outweighed by gains to others.

III. The difference between planning for and charging for benefits

While the characteristics described above make markets for transmission investment unlikely, the implications of how such a market would function are at the root of the logic for basing transmission cost recovery upon the benefits received by network users. If one could go to the many thousands of potential beneficiaries and give each a take-it-or-leave-it offer to pay a little less than their net benefit in exchange for network investment, then all should accept that offer and the investment can proceed. If the benefits are not sufficient, then there would not be enough customers taking the deal and the investment would not proceed. Natural gas pipeline investments in the United States follow this general dynamic, with an "open season" during which a builder solicits subscribers willing to pay for a share of the pipeline. In electricity transmission, the benefits can be so diffuse that such a multi-part negotiation can be absurdly impractical. The temptation for one user to "free ride" can be overwhelming. Thus the grid planning process makes the decision instead, ideally following roughly the same decision criterion.

Charging proposals based upon beneficiaries pay principles invoke a vision of a charging scheme that replicates the "take-it-or-leave-it" offer paradigm. If charges could be based upon offsetting benefits, *and* all users knew exactly what those offsetting benefits would be, *and* each of those users had proportional influence on the decision process, then only those projects for which the aggregate offsetting benefits exceed the costs would make it through the approval process.

There are several shortcomings to this logic. First is the assumption that the relevant beneficiaries would have an appropriate level of influence on the decision process. Any disparity, or even non-linearity between the benefits of firms and their influence on the decision making process, severs the clean link between benefits and efficient decision making. Consider a not-unreasonable model of participation whereby firms with over $10 million at stake hire staff or consultants to participate in or lobby a process. Other firms and customers with less than this amount do not find it worth their time and expense to

[10] See Mankiw, G. and M. Whinston. "Free Entry and Social Inefficiency." *RAND Journal of Economics.* (17):1, 1986.

intervene in the process. The smooth, continuous link between benefits and their weight in decision making is replaced by a very blunt delineation between *big* beneficiaries and others.

Hogan (2011) also describes a risk to marginal projects of basing costs upon benefits. If an efficient decision is reached when the largest winners and losers speak the loudest (or have the most votes) with regards to a project, then clawing back those benefits will dilute these incentives. Those most strongly in favor of a good project may become less motivated once a large share of their potential benefits are sure to be taken back in the form of cost allocation.

If everything is measured precisely this shouldn't be a problem, but once the inevitable errors in modeling or forecasts arise, projects that would be net positives may wither from lack of support from those who disagree about the potential benefits such a project would provide specifically for them. It is generally the case that forecasting aggregate benefits, while difficult, is still an easier task than forecasting the benefits of individual firms.

The prospect of compensation for "losers" would add even more potential rancor to this "democratic" project approval dynamic. Firms whose estimate of losses is below the official estimate could be turned into eager supporters of a project that makes them worse off, simply because of the compensation they stand to gain.

One alternative short of compensating losers for periods in which projects reduce their surplus would be to net out the negative periods from the positive ones, but bound the payment at zero. As discussed above, it will be upon the assessment of total offsetting benefits (total positive benefits less total negative benefits across all time periods) that a party self-determines whether or not it is a beneficiary of an investment. For example, say flows are reversed during the day and the night so that a generator loses revenues during the night but gains them during the day. The daytime gains upon which the payment is based would be reduced by the nighttime losses.

While this approach can mitigate some of the more perverse behavior in pursuit of explicit compensation for transmission investments, it is highly vulnerable to choice of time-periods over which netting may occur. In the day-night example above, offsetting benefits every day would balance out payments. However, such trade-offs could easily be seasonal (e.g. win during the spring, lose during the fall) or even spread out over several years (e.g. lose for the first two years, gain as load grows after that). The only obviously appropriate time frame for such offsetting would be the lifetime of the investments. This would allow for charges to most closely align with parties' offsetting benefits of the investment, as would occur in a market-based investment setting. But seeing as those asset lifetimes can easily extend beyond 50 years, this is not a practical option.

In short, the question of negative benefits is a thorny one, and one for which competitive markets do not always provide an answer. Fortunately, traditional well-designed transmission planning processes *do* attempt to incorporate offsetting benefits in the correct manner. The presence of potential large "losers" implies that, rather than further complicate both the calculation of, and incentive to distort, benefits in an attempt to allocate costs in a fashion that perfectly accounts for offsetting benefits, it would be better to limit and simplify the degree to which benefits enter the cost-allocation formula.

iv. Ex-post vs. Ex-ante Benefits and Investment Decisions

An important departure in the EA's proposed SPD method from previous approaches to implementing a cost allocation based upon benefits is the fact that the proposal would base charges upon a measure of *actual* realized benefits and not the *expected* benefits at the time of the investment decision. Replicating an "efficient" coalition-based investment decision would require the latter. There are many reasons why actual benefits could stray considerably from those expected at the time of the investment decision, many of them completely independent to the actions of the participants whose dynamic efficiency the scheme is attempting to incentivize.

Consider a circumstance where an existing hydro-electric supplier in the South Island is expected to benefit from grid investment that increases their ability to sell output into the North Island. Even assuming their benefits are properly reflected in the investment decision making process, it is the ex-ante expected benefits that are relevant to the question of dynamic efficiency. Their willingness to support such an investment would be based upon their expectations of inter-island price differentials and water values over the lifetime of the project's economic relevance, and their expectation that these benefits will exceed any increased charges that they face. Assume the investment appears sufficiently net beneficial that it proceeds.

After the investment comes on line, there is a drought in the South Island and water values climb to the point that it makes no sense to send power to the North Island. The ex-post benefits of the project are zero, but the decision to invest was still correct given the information available at the time it was constructed. Conversely, if power prices soar in the North Island, ex-post benefits rise substantially above ex ante estimates, yet these extra benefits did not – or should not have – influenced the decision to build the project.

The point of this example is not that there is no justification for revisiting the calculation of benefits after the construction of a transmission investment, but rather that such updated calculations by definition would have nothing to do with the improvement of dynamic efficiency. It was the ex-ante offsetting benefits (net of any increased charges) that should have influenced the generator's incentives to support (or not) the network investment. Once the network investment has been made, LMPs on their own provide the correct dynamic incentive for future generation and load investments as well as operational decisions. Charging those incoming generation plants and loads for a portion of their network benefits can only reduce dynamic efficiency.

For the incumbent generators, who would have been willing to pay an amount approaching their expected offsetting benefits to support the line, charges based upon their actual ex-post efficiency threaten to reduce the efficient utilization of that network asset. Anticipation of the reduced efficiency in the utilization of the network could in turn impact investment decisions and possibly have a negative impact on dynamic efficiency. I will explore this possibility more in the discussion of allocative efficiency.

v. Transmission Planning with Imperfect Information

The above discussion illustrates several cases in which dynamic efficiency is not in any way enhanced by an ex-post beneficiaries pay approach to cost allocation. In most locations major transmission infrastructure plans evolve over a long time-frame and are based upon an expectation of wide-spread distributed benefits that may nonetheless vary considerably from year to year or even decade to decade. As long as this variation is random – that is *independent* of – the behavior of market participants, the charging scheme would not improve dynamic efficiency.

One argument for a beneficiaries pay approach on the grounds of *ex-post* investment efficiency arises in cases where the transmission planning process is strongly impacted by asymmetric information about the future needs of the network. This problem can arise in restructured markets where the entities responsible for transmission investment are not privy to the internal plans of generation companies and large end-users.

If this is the case, then in contrast to the previous discussion, the planner's best current estimates of benefits may *not* be the best measure of expected total benefits, even at the time the decision is made. A single large generator or user may possess information about the future that could swing the social benefits of a transmission project from red to black, or vice-versa.

However, when assessing the impacts of asymmetric information it is useful to consider which cases are most likely to create problems, and what a firm with private information might do to take advantage of that information. In many cases, large firms have an incentive to reveal their future benefits as these are likely to be critical to getting projects approved. Consider a firm planning a large expansion of wind generation in the South Island. If the benefits of the wind generation depend upon adequate transmission capacity, this firm would want to highlight, not conceal, those benefits in the context of a transmission planning process. Similar logic would apply to a consumer that may be considering expansion of its load and would benefit from increased import capacity.

The incentive to conceal any knowledge of future benefits would arise from a hope to avoid the allocation of the costs of a project. This is where the allocation of costs based upon *ex-ante* estimates of benefits can risk dynamic inefficiencies. Once a firm judges that enough benefits to other firms have been identified for a project it supports to go forward, it can hope to "free-ride" on the future expansion by concealing its plans in the hopes of avoiding a cost allocation. If the project is in fact built, there is no efficiency loss, but firms could easily misjudge the approval process and, by keeping quiet about their benefits, inadvertently cause the demise of projects they support.

In the United States, any "inefficiency" caused by ex-ante measurement of benefits largely stems from the technically complicated and frequently litigious nature of the measurement process. Disagreement over forecasted benefits (which themselves depend on models and parameters) is obviously one reason why allocating charges based on ex-post benefits would appear attractive[11]. When there are drawn-out fights over whether a network project should go forward, and those fights are linked to disputes over the allocation of costs, this can create serious economic consequences through the delay or even cancellation of projects that benefit all but those who would have had to pay for them.

It is worth noting that with asymmetric information, allocation based upon *ex-post* benefits creates the opposite problem. It becomes almost costless to proclaim ambitious expansion plans in regions where a firm *may* benefit from additional capacity. If those plans change, or were simply exaggerated at the time of a transmission project's approval, the *ex-post* benefits and therefore the cost share of this hypothetical firm would be relatively modest. Under the EA's proposal, a large proportion of these costs would not be able to be allocated via beneficiaries-pay, and would therefore have to be spread more widely.

[11] Even ex-post evaluation of benefits requires making potential controversial assumptions. In New Zealand, the EA's 2014 proposal illustrates the number of design choices and assumptions required to be made in order to assess benefits ex post. Their analysis demonstrates the substantial impact those design choices can have on the distribution of charges amongst market participants.

IV. Dynamic Efficiency and the Grid Planning Process

To explore this aspect further, we need to consider exactly which investment (generation or transmission) is being influenced by which charges. First consider a case where a single generation plant is making a location decision on where to build, given its knowledge of the current and future status of the transmission network.

For any generation plant that lacks the ability to unilaterally alter the network investment plan, the expected local price provides the fundamental building block for investment decisions. The locational marginal prices will place a different value on energy at different locations, and the incentive to build generation at a particular location will therefore automatically internalize the expectation of future energy prices in that location. No further locational incentives would be necessary. In fact additional locational charges on top of the LMPs would reduce dynamic efficiency.

It is when the investment in generation is seen as *causing* costly investments in transmission that locational prices can be insufficient to provide the proper incentive to locate generation plants, or load. For instance, if a generator anticipates that there would be a reasonable chance of transmission expansion soon after its generation investment, this network investment could diminish or eliminate congestion costs between the generation plant site and the rest of the system. To the generator facing these ex-post low LMPs, its location may appear to have no higher costs than other locations, even though large sums may be needed for the transmission expansion that eliminated the congestion.

Importantly, for there to be an incentive problem creating the dynamic inefficiency, the transmission investment decision needs not only to come after the generation investment decision, it needs to be *directly caused* by the generation investment. If a long-planned transmission expansion happens to go into effect a couple years after generators locate in a region that benefits from it, there is no dynamic inefficiency. In fact, by taking advantage of the access the new transmission provides, the generator is increasing dynamic efficiency. It would be inefficient for a generator to delay or avoid its investment because it was trying to avoid charges for the fixed costs of that investment.

Most countries with a restructured electricity sector try to maintain some distinction between transmission investments that are directly triggered by specific generation projects and those made in response to more widespread network needs. New Zealand's approach is consistent with this philosophy. The former category of network investment, usually called "connection," is recognized as being caused by a specific entity and the costs associated with that investment are assigned to that entity. Note that there is usually no attempt to measure the benefits of such investments. It is understood that the entity paying for them would not have started the process if it did not expect its benefits to exceed those costs.

More of a gray area exists when examining investments that may coincide with connection and "interconnection" costs, but could arguably be seen as doing more than the minimum necessary to accommodate a new asset. The planned construction of a modest amount of wind capacity in the relatively remote Tehachapi region of California provides a good example of this. Clearly this initial new capacity required, and therefore "caused", additional network investments to be made. However, given the economies of scale of such network investments, it usually makes sense to build far more capacity than would be needed simply by the first project. This is particularly true in cases of renewable investment where it is highly likely that additional future generation investments would also take advantage of the new network capacity. Without the new capacity, no investment would (yet) be needed in the network, but if the *entire* network upgrade were assigned to this single generation project, neither would be built.

In California, this tension was resolved by assigning a capacity-weighted portion of the infrastructure costs to the initial generation projects. The remaining costs were assigned to load via a postage-stamp charge pending future generation investments. Those future investments would be responsible for a similar capacity-weighted share of remaining outstanding costs upon their entry to that location. [12]

Similar tensions arise when the interconnection of specific generation plant proves to be the spark that triggers investments in other areas of the network. In cases where congestion costs or reliability concerns become sufficiently large with the new generation, additional network investment can become socially beneficial. In a sense the new generation "causes" the network expansion, yet the benefits would be widespread and the actual trigger for this investment could plausibly have been one of many pending projects.

In fact, in most cases in most countries, major network infrastructure can be thought of as a long-term plan consisting of major projects that are evolving over a very large time scale. When there is a sound, well-informed body that oversees the selection and implementation of major network investments, specific projects are often part of a long time-horizon plan. A strong case can be made for this type of approach. When this is the dominant paradigm, the costs of any specific project may coincide with the timing of specific generation plants, but may also be thought of as arising independent of the actions of any specific plant. In those cases, the plans of generation should inform the network planning process but charges based upon those plans would not enhance dynamic efficiency.

It is in the markets that lack efficient market-based processes for dealing with transmission congestion where the generation "tail" can wag the transmission "dog". In markets with zonal pricing regimes, generation is usually deemed to have the "right" to deliver power at the zonal price; when intra-zonal constraints prevent that from happening, the generator is compensated for its curtailment.[13] This dynamic creates a perverse incentive for generation to locate in regions where it can maximize its payments for its own curtailment rather than produce the most energy profitably. In response, transmission planners have had to build out capacity in order to reduce the payments to generators who located in places where they would not have if they had been financially responsible for the congestion consequences of the location of their plant. [14]

---

[12] See Wolak, Bushnell, and Hobbs. "Opinion on Alternative Treatment of New Transmission for Interconnection of Renewable Generation." Market Surveillance Committee of the California Independent System Operator. October, 2007. Available at https://www.caiso.com/informed/Pages/BoardCommittees/BoardCommitteesArchive/MarketSurveillanceCommitteeArchive.aspx

[13] The United Kingdom power market features a single zonal price. The processes used to deal with the resulting congestion both in the short and long run have contributed to inefficient operations and most likely promoted inefficient investment in both transmission and generation.

See Baldick, Bushnell, Hobbs, and Wolak. "Optimal Charging Arrangements for Energy Transmission." December 2011. Available at https://www.ofgem.gov.uk/publications-and-updates/optimal-charging-arrangements-energy-transmission-final-report-united-states-academic-team.

[14] Until recent reforms, California's market suffered from an indirect form of this problem. Although California has priced transmission usage using LMP since 2009, generation assets are also compensated for their capacity under a resource adequacy (RA) framework. These RA payments account for transmission constraints only in crude ways. As a consequence there were aspects of generation leading transmission in the planning process. This process was reformed in 2012 to adopt a more holistic planning paradigm that should not be distorted by the location decisions of individual plants.

See Bushnell, James B., Harvey, Scott M., and Hobbs, Benjamin F., "Opinion on the Integration of Transmission Planning and Generator Interconnection Procedures. March, 2012. Available at

Under this kind of situation, where generation decisions are truly leading transmission investments, it can be appropriate and more efficient to place the costs of such transmission onto the plants that have caused it. This can be viewed, however, as a very crude way of implementing the spirit of LMP by imposing a constant level of "average" transmission costs onto plants rather than just charging them for their time-varying congestion costs in the first place. In a way, this mirrors the argument that generators should be compensated for their capacity through capacity payments to make up for the suppression of time-varying peak energy prices in some markets. While beneficiaries pay can be appropriate under these conditions, it is best to avoid these conditions in the first place.

The dynamically efficient system relies upon the long-run plan that is well informed, not easily swayed by individual generation location decisions, and allocates costs in a way that minimizes any distortions to either investment or usage of the network. I will discuss such allocations in a later section, but a frequently-used option is to apply postage stamp pricing to load.

In order for such an approach to be sustainable, and thereby efficient, customers who are subject to these postage-stamp charges need to be assured that the allocation is not unreasonably disconnected from benefits. Often this means that geographically varied stakeholders get reasonable benefits out of the long-run investment plan. Thus rates could vary over broad regions (or zones) according to an estimate of the benefits, which is a reasonable approach in these circumstances.

vi. Cost Recovery in Theory and Practice

The above discussion has focused primarily on *what* the investment in, and allocation of costs for, transmission should be based upon. Equally important for allocative efficiency, and therefore, indirectly, dynamic efficiency, is *how* those costs are recovered. The examples described above illustrate how setting the level of charge based upon a measure of benefits can distort offers into the energy spot market. It is the endogeneity of the cost allocation to market behavior that causes a risk to efficiency.

However, even if the costs were completely exogenous (for example coming from a long-run transmission plan or a modelled forecast of benefits), not all pricing methods are alike when it comes to impacts on efficiency. The general principles of *Ramsey pricing* are well suited to this problem. Ramsey pricing tries to recover costs of investments in a manner that minimizes the deadweight loss (inefficiencies) caused by the cost allocation. This usually means applying proportionately more of the costs to participants who are less likely to change their behavior as a result.

In one dimension this means costs are allocated proportionally to the least responsive, or price-elastic, *users*. In another dimension, for any given set of customers, the charges would ideally be assigned in a way that is less closely tied to their hourly volumetric use.

Along the first dimension, this has usually meant applying transmission investment costs to load, rather than to generation. Generators are widely considered more sensitive to volumetric charges linked to their daily or hourly usage of the transmission network (short-term elasticity) than are most customers, for whom these charges are a smaller portion of their overall budget. There are exceptions here, as some customers (usually large ones) can be quite responsive to hourly prices when motivated to be.

---

https://www.caiso.com/Documents/MSCFinalOpinion-Integration-TransmissionPlanning-GeneratorInterconnectionProcedures.pdf.

Along the second dimension, this would also ideally involve charges being distributed in a way that is least likely to elicit a behavioral change. For example, a fixed monthly fee would have no impact on short-run behavior, compared with a volumetric (per MWh) charge. In this dimension, utility (and system operator) practices have lagged. Most postage-stamp transmission charges are applied to small customers as volumetric charges.

Often large customers (and sometimes generators) are applied charges linked to their peak demand usage, but this can still induce a behavioral change. For example a large firm may limit its production, or install inefficient back-up generation to "shave" its individual peak. In some cases, however, inducing such a behavioral change can be efficient, if the demand charge method was based upon a need to ration periodically limited transmission (or generation) capacity in specific periods.[15] As a generic method for cost recovery, this can induce some inefficiencies, but only during those peak periods. The impacts of peak demand charging can be less severe and less frequent than those created by volumetric fees, which have the potential to impact behavior in every period of the year[16].

For purposes of recovering sunk costs, the method that is least likely to change behavior (and therefore create efficiency problems) would be a fixed charge that is only loosely calibrated to the usage of a customer. This could, for example, consist of a small number of charging "bins" per customer class with customers sorted by the average level of demand (in kW) and perhaps general location (e.g. South or North Island).

V. Conclusions

Coordinating the investment and allocating the costs of large infrastructure investments is a difficult challenge. Lumpy investments cause the implicit (or explicit) costs of using transmission networks to swing from levels of very high to nearly zero marginal costs. Benefits can be widely distributed, and in many cases the transfers from winners to losers can greatly exceed the net total benefits provided by network investments.

Given these conditions, the allocation of the cost burden of these investments in a manner that is at least roughly commensurate with the benefits they provide is natural, common-sense, and fair approach. Such approaches have been discussed for decades. The trick has always been to balance the need for a fair cost distribution with the risk of distorting either the long-term or short-term usage of those very network assets.

While the attraction of beneficiaries pay is hard to dispute on grounds of equity, its benefit in terms of efficiency is much harder to identify. In some environments, such as the U.K., the allocation of investment costs according to benefits can arguably offset distortions created by the inefficient pricing of network usage and transmission congestion. In the United States, the FERC has advanced the notion that allocating costs in a way that is perceived to be related to benefits can reduce disputes and speed up the network investment process.

---

[15] Whether this is appropriate will depend upon what prices the customer already faces. For example a supplier or large industrial customer that is already charged a real-time LMP that reflects all generation and transmission scarcity costs would not require any additional pricing incentives to reduce peak consumption.

[16] The impact of volumetric and/or peak charges on allocative efficiency will obviously depend on the relative price elasticities of the parties being charged, and the structure of the charge (for example whether the charge was based on coincident versus anytime peak demand). The empirical evidence in this area is not conclusive. With the advent of more widespread half-hourly metering, this is an important area for future research.

These kinds of experiences contribute to the perception, advanced by the U.S. FERC, that giving firms "skin in the game" of transmission planning will promote more efficient and expeditious transmission investment. However these firms already have substantial interests at stake, as their business relies upon the ability to deliver and receive electricity. If transmission congestion is properly priced, their "skin" is in the right place before any further adjustments for capital costs. For every example of how a firm may drive a planning process in a better direction because of its cost exposure, it is easy to generate an example where the opposite could happen.

It is better to have a planning process with both a broad and long-term horizon, where no individual firms have undue influence over specific project approvals. Costs can still be assigned based upon rough determinations of benefits to ensure that specific players are not forced to continually pay for a series of projects when none of those projects provide benefits. This helps to prevent the disruption of the planning and investment cycle by firms that justifiably see themselves as perpetual losers in the process.

Beyond this, however, it is difficult to envision cases where an allocation of capital costs linked to precisely calibrated estimates of benefits can significantly improve efficiency. At the same time, such an assignment carries a great risk of distorting behavior and the revelation of information. It is these distortions that create the greatest impediments to the efficient construction and usage of network assets.