Residual network costs and the economic efficiency of regulated cost allocation

Derek E. H. Olmstead,[1] Brian Rivard,[2] and Mark Zanewick[3]

July 2022

Comments on this working paper are welcome

1. Introduction

This paper considers the causes of residual transmission network costs and sets out a framework through which the economic efficiency of the regulated rates set to recover these costs can be analyzed.

Electricity transmission networks are a classic example of a natural monopoly: the cost of building a network that connects generators to consumers is minimized by having a single network. Since the single provider would by definition have monopoly pricing power, public utilities regulators are charged with setting rates for network access. It is well accepted that prices equal to short-run marginal cost promote the efficient use of a commodity, including transmission services. However, rates set at marginal cost do not allow the natural monopoly to recover all its sunk and fixed capital and operating cost. To ensure full recovery of these residual transmission network costs, an alternative regulated rate design is required. The fundamental challenge in electricity rate design is the recovery of these residual network costs.

The legislative frameworks and regulatory approaches applied by regulators in setting transmission rates vary across jurisdictions. These frameworks often contain multiple “rate design principles and objectives,” which can include the sufficiency of recovery of prudently incurred costs, equity and fairness, cost causality, justness and reasonableness, rate stability and predictability, environmental outcomes, and economic efficiency. Electricity rate design has a long history, both in theory and practice, that generally involves a normative balancing of these rate design principles and objectives. Regulators have not settled on a single ideal or optimal rate design.

Historically, regulators have tended to use consumption-based charges to recover residual network costs, including “energy charges” levied on each kilowatt-hour of consumption and “demand charges” levied on consumption in coincident or non coincident peak periods. Fixed connection charges are also used to a lesser extent. Each approach involves the regulator making a trade-off between economic efficiency and equity or some other objective. The recovery of

[1] Administrator, Market Surveillance Administrator, Adjunct Assistant Professor, Department of Economics, University of Calgary, and Adjunct Research Professor, Department of Economics, Carleton University. The views expressed herein do not necessarily reflect the views of the Market Surveillance Administrator.

[2] Adjunct Research Professor, Business, Economics and Public Policy, Ivey Business School, Western University and member of Ontario’s Market Surveillance Panel. The views expressed herein do not necessarily reflect the views of the Ivey Business School or the Market Surveillance Panel.

[3] Senior Economist, Market Surveillance Administrator. The views expressed herein do not necessarily reflect the views of the Market Surveillance Administrator.

residual network costs through consumption-based charges — either energy charges or demand charges — induces inefficiencies because doing so distorts the price signals sent to consumers of the social costs that are avoided by reducing consumption. Fixed charges may be deemed unfair or inequitable, if some consumers, particularly low-income, low-volume consumers pay proportionately more for electricity than other consumers.

In approving rates, regulators must ultimately choose an approach that best balances the trade-offs as appropriate for the specific rate design objectives and circumstances of their respective jurisdictions. Many regulators focus primarily on the principle of cost causality when approving rates to recover residual network costs as it has become conventional wisdom that rates based on cost causation provide appropriate price signals that are aligned with enhancing economic efficiency, are fair, objective, equitable, and minimize inter-customer subsidies.[4] As a result, evidence included in regulated rate applications may rely mainly on cost causality studies as justification for proposed rates, rather than provide detailed economic efficiency analyses.

This paper makes a case for greater use of economic efficiency analysis in electricity rate design. To this end, section 2 sets out a simple theoretical model of efficient transmission investment and shows in general how transmission policy objectives, such as a "no congestion" transmission policy contribute to the creation of residual network costs and the need to design regulated rates for utility cost recovery.

Section 3 sets out a detailed numeric example that compares the economic efficiency of a range of common rate designs and establishes three key findings. First, determining the rate design which maximizes economic efficiency is fundamentally an empirical matter. Second, to the extent that residual network costs are caused by transmission policy objectives, analytical methods that claim to be based on principles of cost causation are flawed in principle. Third, an "embedded cost" tariff design[5] does not provide price signals that promote economically efficient short-run or long-run consumption or transmission investment. Embedded cost approaches look at existing costs making up the existing revenue requirement and often focus on the cost of service and usage patterns in a test year. Generally, embedded cost approaches involve functionalization, classification, and allocation of costs. Functionalization identifies the purpose served by each cost (or the underlying equipment or activity), classification identifies the general category of factors that drive the need for the cost, and allocation selects the parameter to be used in allocating the cost among classes (Lazar, Chernick, and Marcus, 2020).

Section 4 summarizes the findings of sections 2 and 3, and section 5 concludes with a discussion of a number of potential applications of this analysis.

---

[4] This is the position taken in Alberta by recent decisions of the Alberta Utilities Commission, as summarized in AESO (2021) at paragraph 2.

[5] This paper examines an embedded tariff design as described in AESO (2021) and NERA (2021). The details of the embedded tariff design are provided in section 3.

2. A model of efficient market-driven transmission investment

This section develops the theoretical benchmark for efficient transmission investment in a competitive wholesale market. The competitive benchmark model provides a point of reference for assessing the implications of a no congestion transmission policy on transmission cost recovery and the relative efficiency of various transmission tariff designs, including the embedded tariff design.

There are two basic approaches to transmission investment within the economic literature (Hogan et al., 2010). The first approach uses a competitive or "merchant" transmission investment framework that relies on market-driven investment to increase transmission network capacity. The second approach relies on regulatory mechanisms whereby transmission owners are incentivized through benchmark or price regulation to invest in transmission capacity to reduce congestion and to provide reliability.

This section adopts the merchant framework using a two-node model based in Joskow and Tirole (2005). The subsequent section builds on this framework and uses a numeric parable to elucidate the efficiency considerations of consumption-based transmission tariffs, including demand charges, energy charges, and the embedded tariff design as described by AESO (2021) and NERA (2021).

i. A simple model of transmission investment

Assume there are two regions, A and B.[6] Supply (S) and demand (D) in each region is defined as S_i and D_i, for $i = A, B$, which are functions of the market price. It will be useful to define excess demand in region i as $ED_i = D_i - S_i$ and excess supply in region i as $ES_i = -ED_i = S_i - D_i$. The excess demand and supply curves reflect the willingness of a region to trade at a given market price.

Assume that the unit cost of transmission capacity (per mega-watt (MW) of transfer capability) is λ and that there is no marginal cost of using this capacity, i.e., assume that transmission losses are zero. Trade will flow from the high-price region to the low-price region but will be limited by the available transmission capacity. The model abstracts away from how trade is organized but assumes (without loss of generality) that "traders" organize trade to take advantage of any available opportunities to trade profitably. Within the control area of an electricity system operator (SO), these trades are organized by the SO through security-constrained economic dispatch, while third parties organize trade between different SOs.[7]

---

[6] The two-node electricity system is a styled representation of an electricity system where wind generation has located in one planning region, and the output of the wind generators flows over the bulk transmission system to populated city centers in the other planning region.

[7] Security-constrained economic dispatch is an area-wide optimization process designed to meet electricity demand at the lowest cost, given the operational and reliability limitations of the area's generation fleet and transmission system (generation offers need not be made cost for this concept to be applicable). Alternatively, there could be a single SO that organizes trade in both regions and between regions through security-constrained economic dispatch.

The analysis begins with autarky, which, in the classic literature on the economics of international trade, is a state where trade does not occur. The analysis then considers the pattern of trade when investment in transmission occurs and trade is allowed, beginning with the case where transmission capacity is free ($\lambda = 0$) and then is costly ($\lambda > 0$). An important distinction between efficient investment in transmission capacity and efficient use of existing transmission capacity is drawn.

ii. Autarky

To consider the implications of trade between the regions, begin by assuming that no transmission capacity connects regions A and B. As a result, the electricity market in each region will clear independently, and the market prices (P_i^*) and consumption (Q_i^*) will be determined based on the intersection of the region's supply and demand curves. Alternatively, the market will clear where excess demand or excess supply is zero. The equilibrium in each region is illustrated in Figure 1.

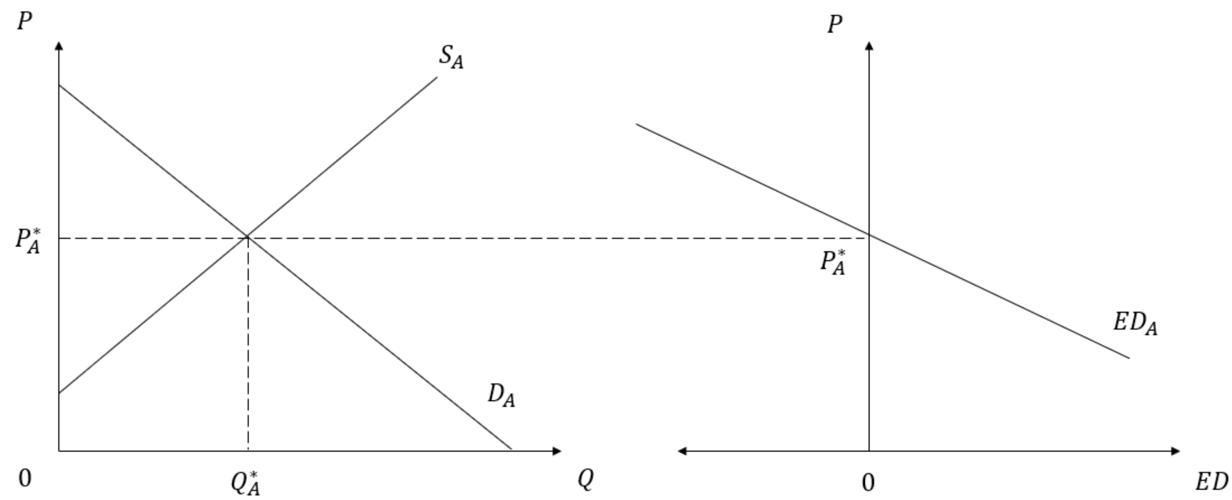
The welfare from trade in autarky occurs separately within each region. Welfare (also called total surplus) is the aggregate gains from trade for all economic agents and is equal to the sum of consumers' and producers' surplus. Consumers' surplus is the area below the electricity demand curve above the equilibrium price (P_i^*) for quantities from zero to the equilibrium level of consumption (Q_i^*). Producers' surplus is the area above the supply curve (marginal cost curve) and below the equilibrium price (P_i^*) for quantities from zero to the equilibrium level of consumption (Q_i^*). When costly transmission capacity is added, welfare also includes the transmission owners' surplus, which is defined as the difference between the revenue they receive for providing transmission services (i.e., tariff revenue) less the cost to build and maintain the transmission capacity.

A market outcome is said to be economically efficient if it is associated with the maximum possible level of welfare or total surplus. In this sense, efficiency reflects an optimal market outcome and any departure from this is inefficient.[8] As illustrated in Figure 1, there is no other combination of equilibrium prices and quantities in each region that could yield greater welfare than P_i^* and Q_i^*. Therefore, these market outcomes are welfare maximizing and economically efficient.

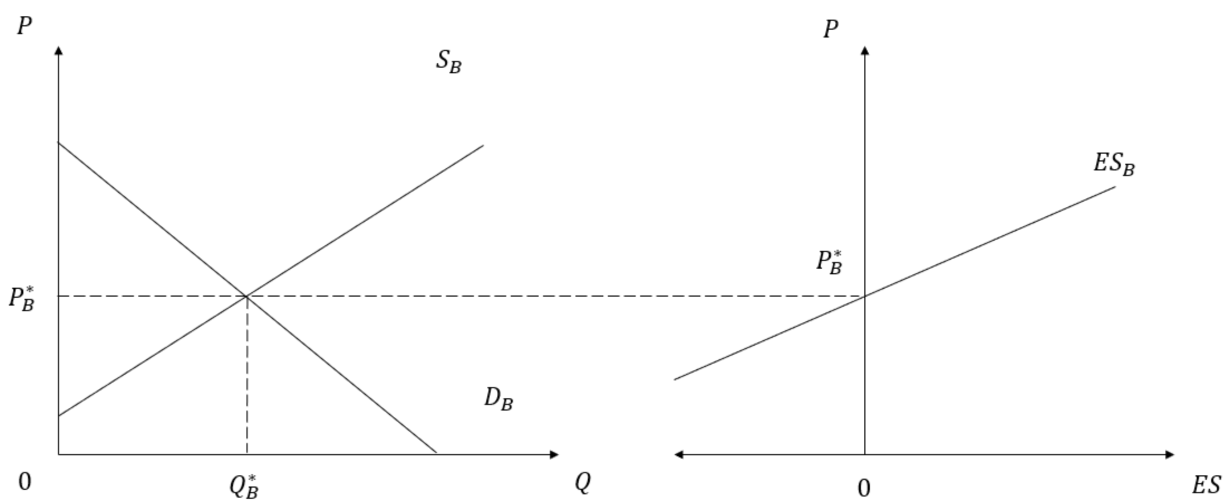
[8] In the context of consumer responses to residual network cost allocation, the inefficiency may manifest itself in the form of either (i) foregone electricity usage that is valued in excess of the marginal cost of production and delivery or (ii) substitution of electricity services from the market to higher cost behind-the-meter generation that is preferred because consuming it would avoid paying for residual network costs.

Figure 1: Autarky equilibrium

Panel A: Region A



Panel B: Region B

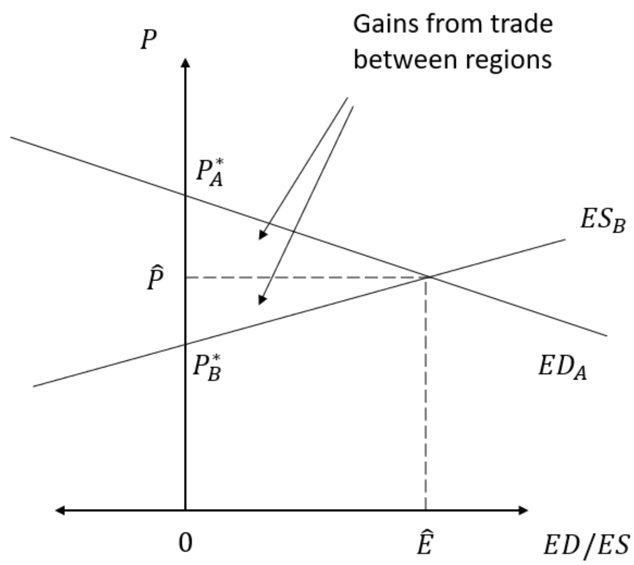


iii. Trade with free transmission capacity

Once trade is allowed, there will be gains from trade if the autarky prices in the two regions are not equal. Assume that $P_A^* > P_B^*$. Also, assume that transmission capacity is free ($\lambda = 0$). With free transmission capacity, any difference between regional prices will induce investment in new transmission to allow trade.

If at least one of the excess demand and supply curves has a non-zero slope, then the trading equilibrium price ($\hat{P}$) will be determined by the intersection of the ED_A (because region A will import) and ES_B (because region B will export) curves. Assume both the excess demand and supply curves have non-zero slopes. In equilibrium, $\hat{E}$ units of trade flow from region B to region A and therefore at least $\hat{E}$ units of transmission capacity must be built. Since transmission capacity is free in this case, it does not matter if unutilized transmission is built but for the sake of simplicity it is assumed that only $\hat{E}$ units of transmission capacity will be built. The trading equilibrium price will be between the two autarky prices, i.e., $P_A^* > \hat{P} > P_B^*$.[9] As a result of trade, the price in region A will decline below its autarky level and the price in region B will rise above its autarky level. This equilibrium is illustrated in Figure 2.

Figure 2: Trading equilibrium trade with free transmission



The net welfare gains from trade compared to autarky are equal to the area above ES_B and below ED_A, between 0 and $\hat{E}$ units of trade. Since this area is positive by definition and the cost of transmission is zero, trade provides a positive overall gain in welfare. It is useful to note however these are net welfare gains as trade does not benefit all economic agents (i.e., trade does not result in a Pareto improvement). Since the price in region A has declined, consumers there are better off while producers there are worse off, and conversely for region B.

iv. Trade with costly transmission capacity: The competitive benchmark

To extend the model, now assume that the cost of transmission capacity is strictly positive ($\lambda > 0$). When the cost to invest in transmission is positive, a difference between regional prices no

---

[9] If one of the excess demand or supply curves is flat, then the autarky price in that region will be the equilibrium trading price. The region with the flat curve would correspond to the world economy in the international trade literature while the region with the sloped curve would be the small open economy and take the world price as given. In practice, one region having a flat excess demand or supply curve means that it is large relative to the other. Since two regions cannot both be large relative to each other, both curves being flat is not a practically relevant case as long as there is some degree of demand or supply elasticity in each region.

longer guarantees that transmission investment will be welfare enhancing in the long run.[10] Instead, welfare-improving investment in transmission capacity will occur only if the difference in autarky prices in the two regions is larger than λ.

To consider the case where welfare-improving investment is feasible, assume that $P_A^* - P_B^* > \lambda$. One implication of this assumption is immediately apparent: efficiency in the presence of strictly positive transmission costs requires the possibility that market prices can vary according to market conditions in the various regions. In electricity markets, such prices are called locational marginal prices (LMP).[11]

When the cost of transmission capacity is positive, it is no longer welfare maximizing to invest in it to the point that prices are the same in each region, i.e., $\hat{E}$ units of transmission is inefficiently high. Instead, the efficient amount of investment in transmission capacity, $\hat{E}^o (< \hat{E})$, is such that the equilibrium price in the exporting region ($\hat{P}^o$) is lower than the equilibrium price in the importing region ($\hat{P}^o + \lambda$) by exactly λ, the marginal transmission cost, and there are $\hat{E}^o$ units of trade between region B and region A. Indeed, if greater than $\hat{E}^o$ units of transmission capacity were installed, the marginal gains from trade would be less than the marginal cost of transmission capacity expansion, creating deadweight loss.[12] In equilibrium, the net welfare gains from trade are equal to the area above ES_B and below ED_A, between 0 and $\hat{E}^o$ units of trade, less the cost of transmission, $\lambda\hat{E}^o$.[13] Since this is the efficient, welfare maximizing outcome with costly transmission, this is referred to as the competitive benchmark.[14] This equilibrium, including

[10] In this context, "long run" means the net effect of infrastructure investments and system use whereas "short run" means only system use and treats infrastructure investment as fixed and sunk.

[11] There is a substantial academic and regulatory literature that discusses the necessity of LMP for economic efficiency. For a selection of academic perspectives, see, e.g., Schweppe et al. (1988), Stoft (2002), and Biggar and Hesamzedah (2014). For a regulatory perspective focused on Alberta, see Church et al. (2009).

[12] It is important to note that even if it is not efficient to invest in marginal transmission capacity in the long run, any available transmission capacity should be used in the short run if it is available to maximize welfare. The existence of fixed costs does not change this because sunk costs are not relevant to short-run marginal decision-making. This is analogous to the energy market itself where market participants, in the absence of market power, have the incentive to offer incremental supply at short run marginal cost and to produce output at this or any higher price, even if this is less than the firm's long run average cost.

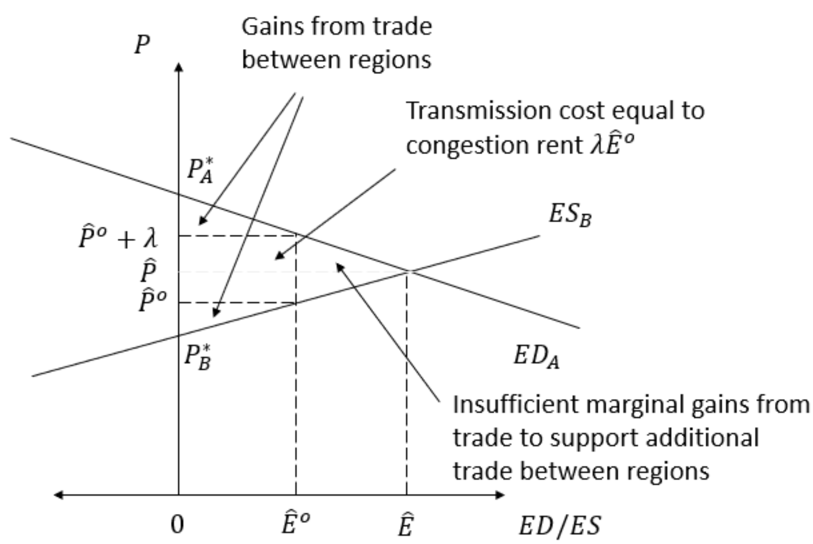
[13] Again, there is a net gain in welfare between autarky and trade with $\hat{E}^o$ units of transmission capacity. However, trade does not lead to a Pareto improvement whereby all participants are made no worse off and at least one better off. Consumers in region A will be better off than under autarky, while producers in region A will be worse off and conversely for region B off. Traders will just break even with congestion rents equal to the cost to build transmission.

[14] This competitive benchmark is derived in Joskow and Tirole (2005). They examine investment by a merchant transmission company that is rewarded through financial transmission rights that pay a dividend equal to the congestion rents (discussed further below) and show that under several strong assumptions and conditions, all investments that are profitable to a merchant transmission company are economically efficient. The appealing nature of this result is that it allows unfettered competition to govern investment in new transmission capacity rather than central planning and regulatory oversight. It also allows investment in generation capacity in the constrained regions to compete with investment in transmission capacity. Unfortunately, the strong assumptions and conditions required for unfettered competition and efficient investment are likely to be inconsistent with the actual attributes of transmission investment and the operation of wholesale markets in practice.

equilibrium regional prices (LMP), pattern of trade, and optimal amount of investment, is illustrated in Figure 3.

To consider the implications of transmission being costly, note that the equilibrium price in the exporting region (B) is above its autarky price but below the free transmission price, with the converse for the importing region (A). That is, assuming that both the excess demand and supply curves have non-zero slopes, then $P_B^* < \hat{P}^o < \hat{P} < (\hat{P}^o + \lambda) < P_A^*$. The flow from region B to region A is limited by the amount of transmission capacity (i.e., $\hat{E}^o$) and the transmission network is said to be "congested." Traders earn the difference between the region A price (what they are paid for the power) and the region B price (what they pay for the power) per unit of trade. This amount, $\lambda\hat{E}^o$, is referred to as "congestion rent."[15]

Figure 3: Trading equilibrium trade with costly transmission



In the competitive benchmark, the congestion rent is just equal to the total cost of transmission investment, $\lambda\hat{E}^o$. In effect, this means that consumers in the importing region pay for the transmission capacity through the congestion rent (i.e., congestion rent represents a market-determined "transmission fee" that consumers in region A willingly pay to traders to cover the traders' cost to construct the $\hat{E}^o$ units of transmission capacity required to allow the level of trade that benefits consumers in region A).

Furthermore, under the competitive benchmark, there are no residual transmission costs to be recovered; these costs are incorporated in the energy prices that consumers in region A pay. In other words, the price paid by consumers in region A is equal to the marginal cost of obtaining

[15] In an international trade context, this surplus is known as the merchandizing surplus. Note also that if there were no traders and instead, there was a single SO organizing the flow across the transmission line, the SO would collect the congestion rent by charging the consumers in region A the region A price for the electricity imported to region A and paying the generators in region B the region B price for producing the electricity exported from region B.

one more unit of energy from (i) producers in region A or (ii) the market in region B including the cost of building one more unit of transmission capacity, which are equal in equilibrium. Meanwhile, the price paid by consumers in region B is equal to (i) the marginal cost of obtaining one more unit of energy from producers in region B and (ii) the price in region A less the cost of building one more unit of transmission capacity, which are equal in equilibrium.[16] Therefore, given regional price differences (LMP), there is no need under the competitive benchmark for a regulator to impose a regulatory mechanism to allocate transmission costs.

If greater than $\hat{E}^o$ units of transmission capacity were installed, then the congestion rent would be insufficient to cover the total cost of transmission investment and a regulatory cost recovery mechanism would be required to ensure full transmission cost recovery.

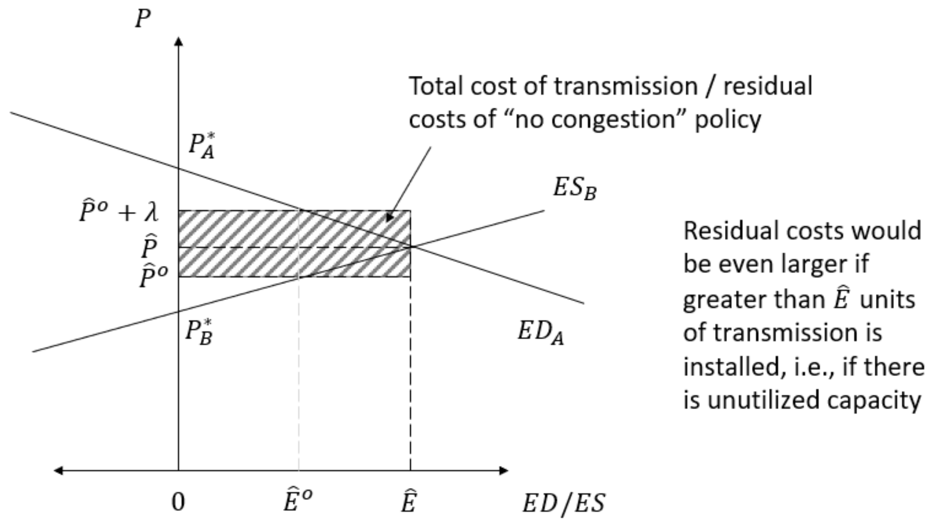
v. No congestion policy with non-distortionary residual cost recovery charges

Continuing to assume that transmission investment is costly ($\lambda > 0$), assume that a policy decision is made to invest in transmission capacity so that the transmission system is not congested.[17] This means that there must be sufficient transmission capacity such that the prices will be the same in both regions. In addition, assume that any residual costs are recovered in a manner that has no distortionary effects on any economic agents in the market (i.e., through fixed charges based on willingness to pay). The trading equilibrium price ($\hat{P}$) is determined by the intersection of the ED_A and ES_B curves as in the case when transmission was free. In equilibrium, $\hat{E}$ units of electricity flows from region B to region A. This equilibrium is illustrated in Figure 4.

[16] It may be argued that the competitive benchmark outcome could be replicated by a regulatory direction to build $\hat{E}^o$ units of transmission capacity, set the energy price in both regions equal to $\hat{P}^o$, and then impose a regulated energy charge at the rate λ only on consumers in region A. This will not, however, lead to an equivalent market outcome. The reason for this is that the regulatory charge λ that is imposed on consumers in region A will introduce a wedge between the price that consumers in region A pay and what producers in region A receive. In particular, because producers in region A would not receive λ, they would be paid less than under the competitive benchmark and will therefore reduce their production. As a result, this regulatory alternative would result in a lower price and reduced consumption (thereby inducing a reduction in welfare) in region A.

[17] This is the policy in place in Alberta. Specifically, the *Transmission Regulation* requires the AESO to, "tak[e] into consideration the characteristics and expected availability of generating units, plan a transmission system that: is sufficiently robust so that 100% of the time, transmission of all anticipated in-merit electric energy…can occur when all transmission facilities are in service, and is adequate so that, on an annual basis, and at least 95% of the time, transmission of all anticipated in-merit electric energy…can occur when operating under abnormal operating conditions" (*Transmission Regulation*, section 15(e)) and "make arrangements for the expansion or enhancement of the transmission system to that, under normal operating conditions, all anticipated in-merit electricity…can be dispatched without constraint" (*Transmission Regulation*, section 15(f)). This policy decision, including its rationale, is discussed in Government of Alberta (2003). A fundamental motivation of the "no congestion" policy decision is the claim that it is necessary to facilitate Alberta's competitive wholesale market. The merits of this policy and the claims made in support of it are beyond the scope of this paper; for a more detailed overview and critique of it, see Church et al. (2009).

Figure 4: Trading equilibrium under a "no congestion" transmission policy, costly transmission capacity, and non-distortionary residual network cost recovery



It is clear from Figure 4 that the amount of transmission required for the "no congestion" policy to be implemented is greater than the amount under the competitive benchmark, i.e., $\hat{E}^o < \hat{E}$. The total cost of this transmission will be $\lambda\hat{E}$. Since the prices in both regions are the same, there is no congestion rent and therefore <u>all</u> transmission costs are residual costs.

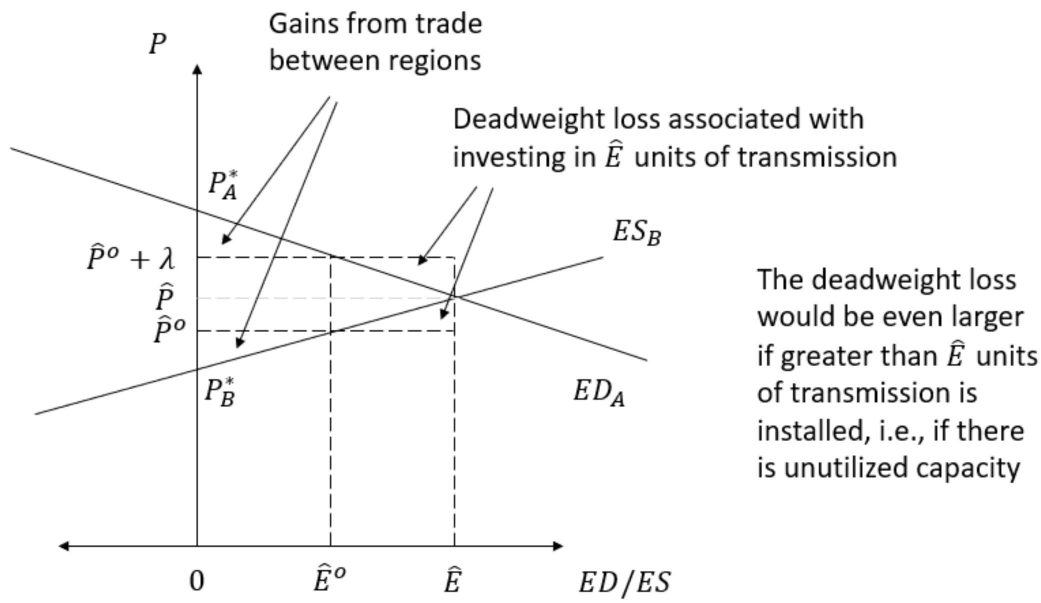
The total welfare implication of the "no congestion" policy is illustrated in Figure 5. The increase in transmission costs of the "no congestion" policy above that of the competitive benchmark are greater than the additional gains from increasing the level of trade between regions. This level of investment creates deadweight loss. Notably, a no congestion policy unambiguously reduces total welfare relative to the competitive benchmark.[18] This does not imply that there is a net reduction in welfare compared to there being no transmission capacity and autarky. It is an empirical question whether the gains from trading up to the level of the competitive benchmark are greater than or less than this deadweight loss.[19]

---

[18] This does not mean that no economic agent benefits from the transmission that is built in excess of the competitive benchmark. Rather, it means that the costs to obtain these benefits are larger than the benefits themselves, hence the deadweight loss, and that some economic agents benefit from inefficient outcomes.

[19] To repeat a previously made point, if transmission capacity exists then its costs are sunk and it is efficient to use it in the short run even if it was not efficient to build it in the long run.

Figure 5: Welfare implications of "no congestion" transmission policy, costly transmission capacity, and non-distortionary residual network cost recovery



vi. No congestion policy with distortionary consumption-based charges

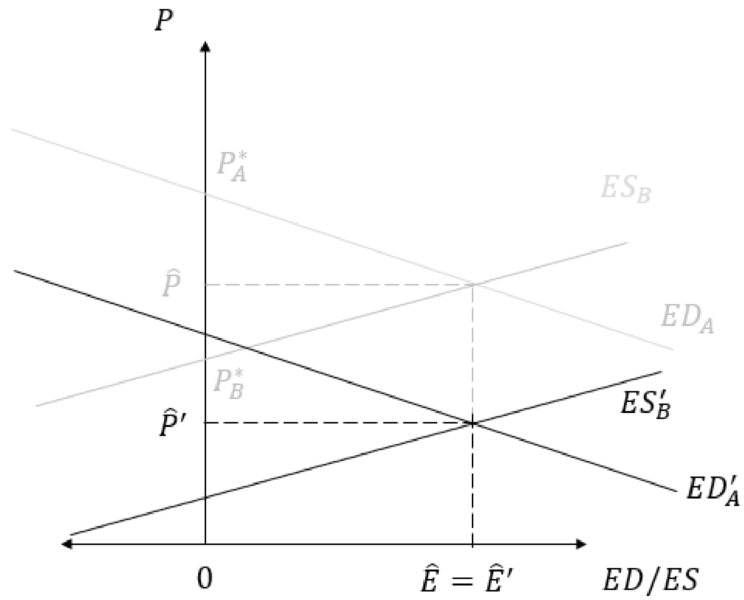
The previous situation assumed that transmission cost recovery occurred through an unspecified non-distortionary charge (e.g., a charge that does not affect consumers' demand for electricity in the short-term). Building on the previous situation, now assume that the residual network costs are recovered through a consumption-based charge levied equally on consumers in both region A and B. In a single-period model, this amounts to an energy charge rather than a demand charge as the on/off peak distinction is not meaningful.

The consumption-based charge, like a consumption tax, has the effect of reducing consumers' willingness to pay in each region, which shifts downward the market demand curves in Figure 1 (the autarky figure). This will cause the amount of desirable trade to decline in each region, i.e., both of the ED_A and ES_B curves will shift downward.

In the special case where both of the ED_A and ES_B curves shift down by the same amount (i.e., the elasticities of supply and demand are the same in each region, at least in the vicinity of the market equilibrium), then the (single) trading price will fall because of the consumption-based charge but the level of trade under the "no congestion" policy will not change. This situation is illustrated in Figure 6.

However, under more general conditions that allow for the shapes of the ED_A and ES_B curves to be different (because the underlying elasticities of supply and demand in the regions are different) — which is ultimately an empirical question — consumption-based charges may increase or decrease the amount of trade between the regions under the "no congestion" policy. In the case that the amount of trade between regions increases, the implication of the "no congestion" policy is not only to create residual costs that must be allocated by the regulator but also to increase the amount of transmission further, and raise residual costs further, above the case of the "no congestion" policy with non-distortionary cost recovery. Whether this situation occurs or not is, again, an empirical question.

Figure 6: Trading equilibrium under a "no congestion" transmission policy, costly transmission capacity, and a special case of distortionary residual network cost recovery



In all cases, cost recovery through a consumption-based charge results in lower overall consumption across the two regions, and welfare declines relative to the competitive benchmark and relative to the no-congestion policy, non-distortionary charge situation.

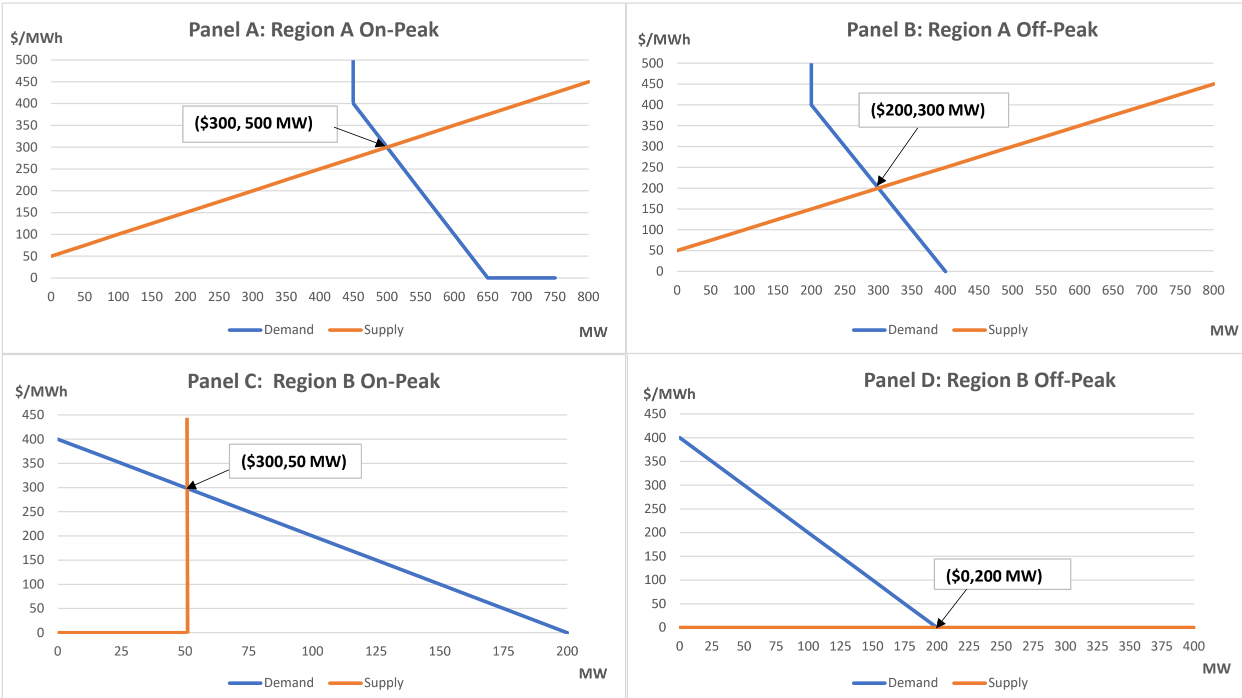
3. Efficiency attributes of different transmission rate designs

This section analyses the efficiency implications of different transmission cost recovery mechanisms, namely an energy charge, a demand charge and the embedded tariff design. The analysis is presented through a parable.. For comparison purposes, the parable first derives the autarky, competitive benchmark and no congestion policy, non-distortionary charge equilibria described generically in section 3. This is followed by the derivation of the equilibria with a no congestion policy and the application of a demand charge, an energy charge, and an embedded cost-based tariff. The mathematical derivations of each equilibrium presented in this section are provided in Appendix A.

i. A numeric parable

Consider a country with a two-node electricity system that is divided into two regions (A and B) and that operates in two periods (on-peak and off-peak) as depicted in Figure 7.

Figure 7: Two-node, two-period example with no transmission network



The natural conditions in region B are conducive to the deployment of low-cost wind generation that can produce 50 MW of electricity in the peak period, and 400 MW in the off-peak period at a cost of $0/MWh. Higher cost generation resources must be deployed in region A with an aggregate marginal cost curve defined as MC(Q) = 50 + 0.5Q in both periods.

Most consumers live in region A and electricity demand is greater in region A than in region B. In region B, there is a single price responsive industrial customer that consumes according to the linear inverse demand curve P(Q) = 400 - 2Q in both periods. Electricity demand in region A consists of residential and industrial demand. Residential customers do not respond to electricity prices. Residential demand is equal to 450 MW in the peak period and 200 MW in the off-peak period as represented by the vertical portion of the on-peak and off-peak demand curves in Figure 7, Panel A and Panel B. There is a single price responsive industrial customer in region A that consumes according to the linear inverse demand curve P(Q) = 200 - Q in both periods (the downward sloping portion of the demand curves in Figure 7, Panel A and Panel B).[20]

ii. Autarky

The parable begins with the two regions operating independently (autarky) as there is no transmission network connecting region A and region B. A single SO directs the flow of electricity

[20] The two-node electricity system describe in this parable was purpose-built to reflect key features of the Alberta electricity system, where the majority of wind generation has located in the south and central planning regions of the province, and the output of the wind generators flows over the bulk transmission system to the populated city centers of the Calgary and Edmonton planning regions.

within each region through security constrained economic dispatch and determines an equilibrium quantity and price for each region at the intersection point of the respective supply and demand curves.[21] Consumers pay the equilibrium price per unit of electricity consumed to the SO, and the SO pays the equilibrium price per unit produced to generators.

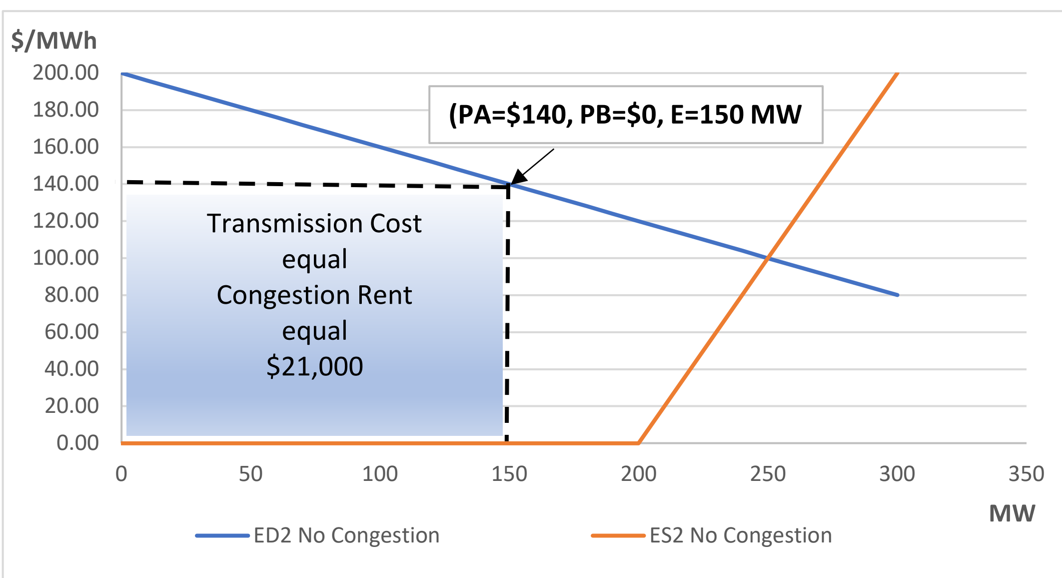
In the peak period, the equilibrium price of electricity is $300/MWh in both regions. In the off-peak period, given the abundance of low-cost supply in region B, the equilibrium price is $0/MWh. At this price, the industrial customer in region B consumes 200 MW. In region A, the equilibrium price is $200/MWh. Generators produce 300 MW with residential customers consuming 200 MW and the industrial customer consuming 100 MW.

iii. Competitive benchmark

The off-peak price difference between the regions presents an opportunity for investment in transmission capacity to allow trade from region B to region A. Assume the SO is charged with the regulatory obligation to facilitate *efficient* transmission investment. The SO facilitates transmission investment through a competitive procurement process to identify a transmission company that can construct transmission at the lowest cost per unit of transmission capacity. Assume the winning bid to construct transmission is $140/MW of transfer capacity.

The SO determines the optimal amount of transmission capacity at the point where the marginal benefit of adding new transmission capacity (the marginal welfare gain from trade) is equal to the marginal cost of adding new transmission capacity. This is illustrated in Figure 8.

Figure 8: Competitive benchmark, off-peak trading equilibrium



[21] In this parable, generators do not exercise market power so that the electricity is supplied at marginal cost.

The optimal level of transmission capacity is equal to K=150 MW. The addition of E=150 MW of transmission capacity allows 150 MW of $0/MWh electricity to be imported to region A from region B. With the addition of 150 MW of imports, the price in region A falls to $140/MWh while the price in the region B remains at $0/MWh. The marginal benefit of the 150th unit of transmission is equal to the cost savings realized by replacing a unit of output from a generator in region A with a marginal cost of $140/MWh, with a unit of output from a generator in region B with a marginal cost of $0/MWh; a cost savings of $140/MWh. The marginal benefit of the 150th unit of transmission is just equal to the marginal cost of the 150th unit of transmission, which is $140/MW.

The total cost to build and maintain K=150 MW of transmission transfer capacity is equal to $21,000 ($140/MWh x 150 MW). Transmission cost recovery is addressed through the SO's collection of congestion rents. Congestion rents are the difference between what consumers in region A pay to the SO for the electricity imported from region B and what the SO pays the generators in region B for the electricity exported to region A multiplied by the amount of trade (i.e., ($140/MWh - $0/MWh) x 150 MW = $21,000). The total congestion rents collected by the SO equals the total cost of transmission. The SO pays the congestion rent to the transmission operator and there is no residual transmission cost.

iv. Effects of a "no-congestion" transmission policy

Assume now that the country implements a no-congestion transmission policy, such that the SO is required to facilitate investment in transmission capacity to the point where trade results in prices being the same in regions A and B in both periods (assuming no transmission losses). Upon review the SO determines that a common off-peak price of $100/MWh in both regions is achieved with E=250 MW of trade from region B to region A, which would require at least K=250 MW of transmission capacity. This outcome is illustrated in Figure 9.

Figure 9: No transmission congestion policy, off-peak trading equilibrium

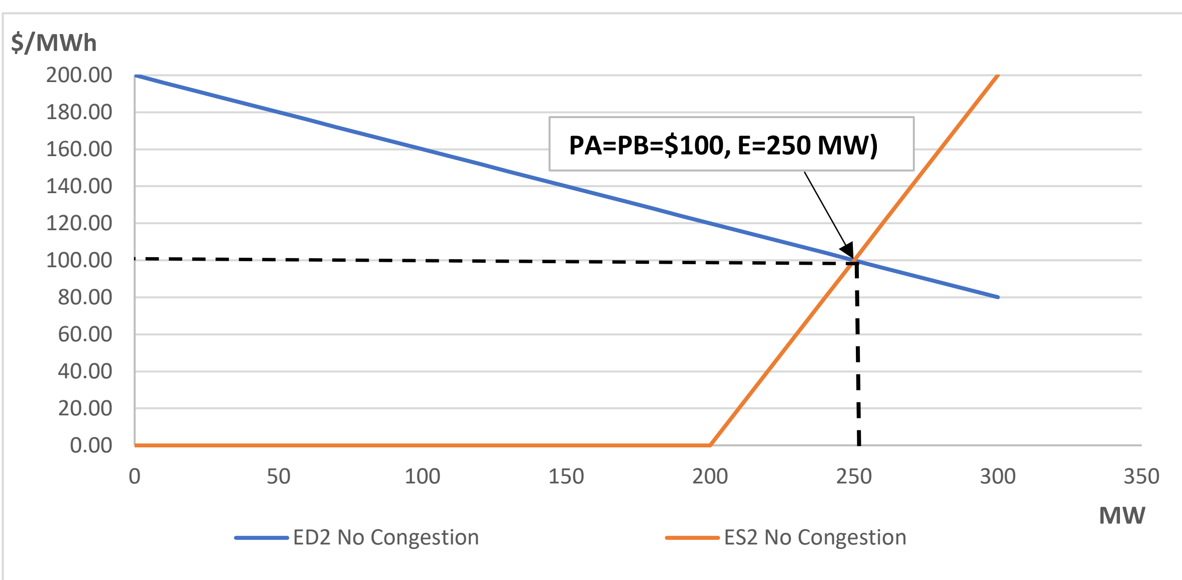


Table 1 provides a comparison of the welfare effects of the autarky, competitive benchmark and "no congestion" policy outcomes (a comparison of the transmission capacity investment options). In this parable, investment in 150 MW of transmission capacity (the competitive benchmark) improves total welfare by 1.5% relative to autarky. Note that investment efficient transmission capacity leads to a welfare improvement for consumers in region A but a welfare loss for producers in region A. The no congestion policy leads to an investment of 250 MW of transmission capacity and a lower level of total welfare relative to the competitive benchmark. When compared to autarky, the no congestion policy improves the welfare of the consumers in region A and producers in region B, but reduces the welfare of the consumer in region B and producers in region A.

Table 1: Welfare effects of different transmission investment options

Scenario	Transmission Capacity (MW)	Consumer' Surplus Region A	Producers' Surplus Region A	Consumer' Surplus Region B	Producers' Surplus Region B	Transmission Cost	Congestion Rent/Tariff Revenue	Total Surplus	Efficiency Gain from Autarky
Autarky	0	$162,500	$85,000	$42,500	$15,000	$0	$0	$305,000	0.0%
Competitive Benchmark	150	$181,400	$70,600	$42,500	$15,000	$21,000	$21,000	$309,500	1.5%
No congestion policy	250	$195,000	$65,000	$25,000	$55,000	$35,000	$0	$305,000	0.0%

However, under a no-congestion policy, with K=250 MW and prices equal in both regions, the SO no longer collects a congestion rent to cover the transmission owner's cost. The SO must therefore identify an alternative solution, one that requires a regulatory mechanism to ensure transmission cost recovery. Assume that the SO has a regulatory obligation to propose an approach for the recovery of the cost of transmission that is consistent with the principles of efficiency (appropriate price signals), fairness, objectivity, and equity.[22] The SO is further obligated to design a tariff that is levied on consumers only, and on all consumers equally (i.e., "postage stamp rates"). The SO considers three tariff designs: a demand charge, an energy charge, and a tariff based on the embedded cost approach.

Demand charge

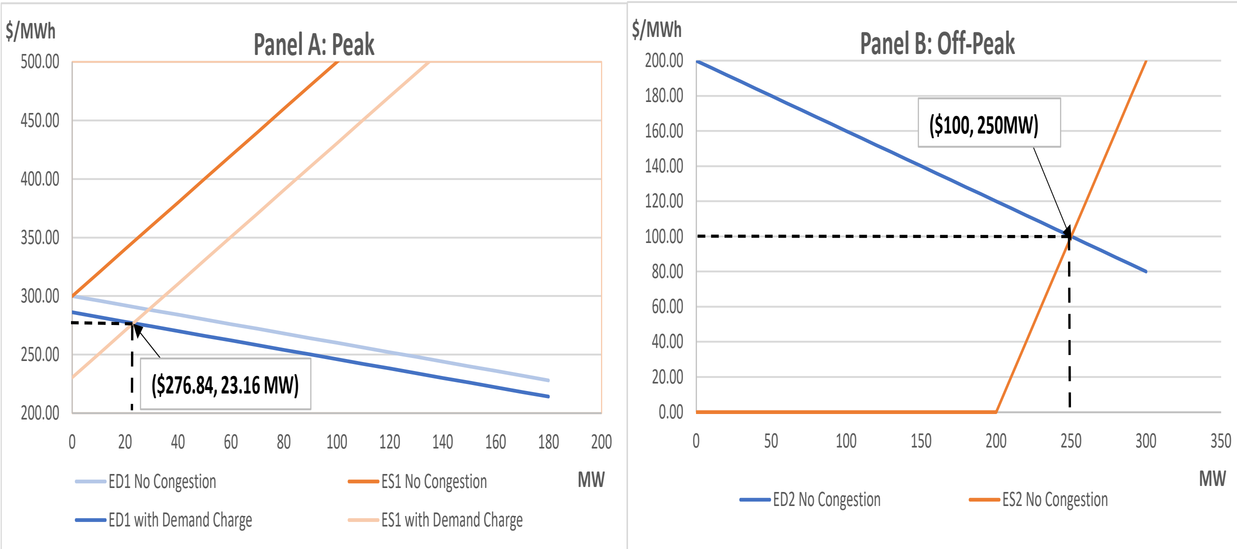
The SO first considers the implications of a demand charge levied in $/MW of coincident peak demand. In the peak period, each consumer is levied a tariff that is equal to their share of total peak period demand in both region A and region B, times the total cost of transmission.

The SO recognizes that a demand charge, which is akin to a peak period consumption tax, would affect the price responsive industrial consumers' willingness to pay for electricity in the peak period. For a given peak period electricity price, industrial customers in region A and region B would be willing to buy a smaller quantity at a given pool price than they would without the demand charge. This would result in a leftward shift of the peak period demand curves in region A and B. Correspondingly, the excess demand curve in region A would shift to the left and the excess supply curve in region B would shift to the right. Consumer demand in the off-peak period would

[22] These are the key rate design principles identified in AESO (2021).

be unaffected by the demand charge. The effect that a demand charge would have on the trading equilibrium is reflected Figure 10.

Figure 10: No transmission congestion policy trading equilibrium with a demand charge



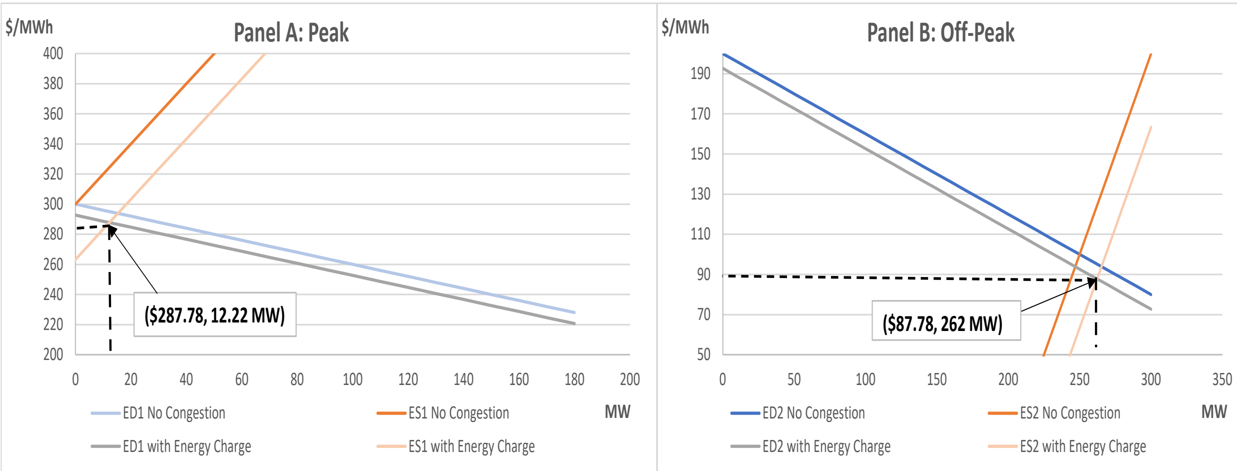
In equilibrium, a demand charge of $69.49/MW would fully recover the residual transmission costs. Note however, that since the supply curve in region B is horizontal at $0/MWh but is upward sloping in the region A, an equal shift in demand in region A and region B would lead to a smaller price reduction in region A than region B. This would induce the export of 23,16 MW from region B to region A in the peak period so that prices in both regions converge to $276.84/MWh (Panel A). The trading equilibrium in the off-peak period would remain the same as the no congestion policy trading equilibrium, with transmission capacity k=250 MW (Panel B).

Energy charge

The SO next considers the implications of an energy charge levied in $/MWh of total consumption. Like a demand charge, an energy charge would affect the price responsive industrial consumers' willingness to pay for electricity. However, an energy charge would affect industrial demand in both periods, not just the peak period. The effect that an energy charge would have on the trading equilibrium is reflected Figure 11.

In equilibrium, an energy charge of $36.67/MWh would fully recover the residual transmission costs. The leftward shift in electricity demand in both regions during the peak period would put downward pressure on the equilibrium price in both regions and induce the export of 12.22 MW from region B to region A in the peak period until prices in both regions would converge to $287.78/MWh (Panel A).

Figure 11: No transmission congestion policy trading equilibrium with an energy charge



Interestingly, the downward pressure in off-peak would induce more electricity export from the region B to region A than would occur in the no congestion with non-distortionary cost recovery situation. This is again due to supply being perfectly elastic in region B and more inelastic in region A. As a result, with an energy charge, the SO would need to increase transmission capacity to K=262 MW to accommodate the additional flow of electricity and avoid congestion (as per the policy). The additional transmission capacity would also lead to an increase in the residual transmission costs that would need to be recovered.

Embedded cost approach

Finally, the SO considers the implications of designing a tariff based on the embedded cost approach to transmission cost recovery.

As described in NERA (2021), "the embedded cost methodology seeks to signal to customers the long-run costs of providing transmission, in a way that identifies which costs have been incurred historically to accommodate (or have been caused by) particular patterns of usage, particular customers, and/or particular services."[23] In this regard, the embedded tariff design is one that focuses on the concept of cost causality.

NERA (2021) outlines the steps required to calculate its recommended tariff design under the embedded approach to include:

[23] NERA (2021) considered two approaches to tariff design: a marginal and embedded cost approach. It determined that the embedded cost approach better fits the legislative framework, namely the no congestion transmission policy, postage stamp transmission rates, and the characteristics of the Alberta transmission system. It describes the marginal cost approach (at paragraph 106) as setting a tariff based on an estimate of how a change in demand from a customer will affect the future costs of a utility. The competitive benchmark is an example of a marginal cost approach as the congestion rent reflects the marginal forward-looking cost of adding a unit of transmission transfer capacity to accommodate an additional unit of demand in region A.

(i) ***classification*** of transmission costs between demand and those associated with accommodating flows of in-merit energy;

(ii) ***functionalization*** of the demand related costs in to two categories, bulk system demand costs and regional system demand costs; and

(iii) ***allocation*** of bulk system costs through a 12CP demand charge, regional system costs through a charge on billing capacity, and the costs associated with accommodating flows of in-merit energy through an energy charge.[24]

The key step in the process is the classification of costs between demand and in-merit energy. NERA (2021) uses a "minimum system approach" to classify costs between demand and in-merit energy. The minimum system is defined to reflect the size of the transmission system required to meet peak load. The size of the minimum system defines the proportion of costs classified as demand-related, while the difference between the actual and minimum systems defines the proportion of costs classified to accommodating the flow of in-merit energy.[25]

As a proxy for the minimum system, NERA (2021) uses the maximum hourly metered net load in a regional planning area measured in MW for a given reference period. To estimate the actual system in the regional planning area, NERA (2021) determines if the transmission system required to accommodate flows of in-merit energy in each planning area exceeds the size of the minimum system. The proxy for the transmission system required to accommodate flows of in-merit energy in a regional planning area is the maximum hourly generation measured in MW for a given reference period. Note, the peak demand and peak generation hours in a region may not be the same. If the peak generation exceeds the peak demand in a region, the actual transmission needed to accommodate in-merit energy is greater than the minimum system. Conversely, if peak demand exceeds peak generation the minimum system to meet demand is sufficient to accommodate in-merit energy flow.

Once the minimum and actual system for each region is determined, the individual regional results are aggregated to define the overall minimum system and the actual system. The overall systems are used to allocate costs between demand charges and energy charges. The portion of transmission costs allocated to a demand charge are calculated as the ratio of the overall minimum system and the actual system measured as a percentage. The portion of transmission costs allocated to accommodate flows of in-merit energy are then equal to 100% minus the portion of costs allocated to a demand charge.

Following the approach as set out by NERA (2021), the SO uses the minimum system approach for cost classification and the allocation of costs to a demand charge and an energy charge. Note that the embedded cost approach uses historical flows to allocate costs; in this case, the SO

---

[24] NERA (2021), paragraph 260. Note that our parable only examines bulk transmission and the allocation of bulk transmission costs. As a result, we do not consider step 2, functionalization in the presentation of our parable. This does not change the general findings.

[25] NERA (2021), paragraph 262.

assumes that the consumption and generation amounts of the no congestion policy trading equilibrium without a consumption-based charge reflect historical usages and the causes for transmission capacity investment.[26] The results of this approach are set out in Table 2.

The minimum system of a region is proxied by the region's peak demand. In region B, peak demand occurs in the off-peak period with 150 MW consumed. In region A, peak demand occurs in the peak period with 500 MW consumed. The actual system of a region is proxied by the region's peak generation. In region B, peak generation occurs in the off-peak period at 400 MW while in region A it occurs in the peak period at 500 MW. The actual system in region B exceeds the minimum system indicating that transmission must be built to accommodate the flow of in-merit energy from region B to region A. The minimum system in the region A equals the actual system, indicating that the minimum system in the region A is sufficient to accommodate the flow of in-merit energy.

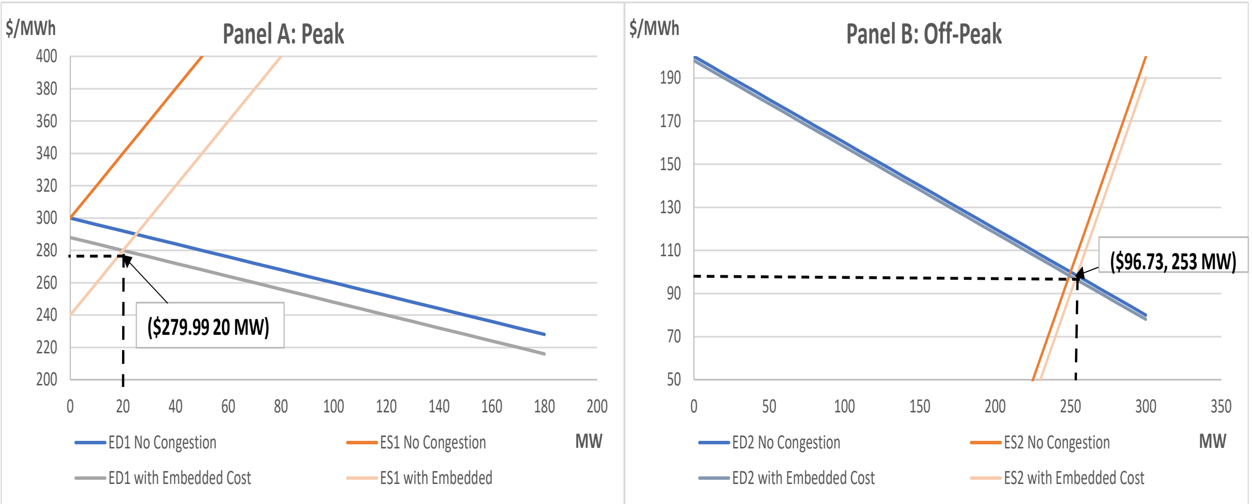
Table 2: Embedded cost approach to tariff design and transmission cost allocation

Classification			
	Region B	**Region A**	**Overall System**
Minimum System (Peak Demand in MW)	150	500	650
Actual System (Peak Generation in MW))	400	500	900
Allocation			
Demand Charge (%)		72.2%	
Energy Charge (%)		27.8%	
Period	**Proxy**	**Region B**	**Region A**
Off-peak	Demand	**150**	350
Peak	Demand	50	**500**
Off-peak	Generation	**400**	100
Peak	Generation	50	**500**

The SO determines the percentage of total transmission cost that would be allocated through a demand charge by first computing the overall minimum and actual systems (the sum of the regional systems) and then multiplying the ratio of the overall minimum system size to the overall actual system size by 100 (i.e., 650 MW/900 MW x 100 = 72.2%). The percentage of total transmission costs that would be allocated through an energy charge is calculated simply as the residual percentage or 27.8%. This results in a demand charge of $50.21/MW and an energy charge of $9.82/MWh. The effect that a tariff based on the embedded cost approach to cost recovery would have on the trading equilibrium is reflected Figure 12.

[26] Note, the act of cost allocation through a consumption-based demand and energy charge may affect consumers consumption decisions, which could modify future flows from historical flows and the potential need for transmission.

Figure 12: No transmission congestion policy trading equilibrium with an embedded cost tariff



As in the case of the energy charge discussed above, levying an energy charge during the off-peak period puts downward pressure on off-peak prices and increase the demand for trade from region B to region A relative to the case with no congestion and non-distortionary cost recovery. This would require an additional 3 units of transmission capacity (K=253 MW) to be built to accommodate the additional trade.

<u>SO's general assessment of tariff options</u>

The SO must select the tariff option that best balances the principles of efficiency (appropriate price signals), fairness, objectivity, and equity. Table 3 provides the welfare effects of each of the different tariff options considered by the SO. The autarky case is presented as a point of reference for defining welfare losses or gains.

Table 3: Welfare effects of tariff options

Regulated Tariff Options	Transmission Capacity (MW)	Consumer' Surplus Region A	Producers' Surplus Region A	Consumer' Surplus Region B	Producers' Surplus Region B	Transmission Cost	Congestion Rent/Tariff Revenue	Total Surplus	Efficiency loss relative Autarky
Autarky	0	$162,500	$85,000	$42,500	$15,000	$0	$0	$305,000	0.0%
Demand charge	250	$203,644	$53,955	$23,220	$53,842	$35,000	$35,000	$299,661	-1.8%
Embedded methodology	253	$202,088	$55,079	$22,429	$52,691	$35,458	$35,458	$296,827	-2.7%
Energy charge	262	$198,355	$57,965	$20,409	$49,499	$36,711	$36,711	$289,517	-5.1%

First note that all the tariff options reduce total welfare relative to the case of autarky in our parable. In general, the relative welfare effect of a given tariff design to the case of autarky (no transmission investment) is an empirical issue. From an overall efficiency perspective, the demand charge provides the highest net welfare gain among the regulated tariff options, followed by the embedded tariff design and the energy charge. From a participant standpoint, consumers in region A and B and producers in region B are better off under a demand charge than under the other tariff options. Producers in region A are worse off under a demand charge than under the other tariff options. Given the requirement that all consumers must pay for transmission, the demand charge may be considered the fairest and most equitable among the tariff options.

A cost causation-based tariff design means customers are charged based on how their use of the system drives transmission costs over the long term.[27] However, under a no congestion policy with an obligation to use postage stamp tariffs levied on all consumers, a true cost causation-based tariff cannot be designed. Consumers in region A clearly drive transmission costs over the long run, but under the available tariff options (the demand charge, energy charge and embedded tariff design), consumers in region B contribute to the recovery of transmission costs and are worse off than under autarky (no transmission investment).

Interestingly, the embedded tariff design, which is purported to "ensure that customers pay transmission charges that reflect cost causation and the long-term costs of using the transmission system going forward," fails to achieve either of its stated goals; consumers in region B pay for transmission cost that they do not cause, and the price paid by all consumers fails to reflect the long run cost of transmission which is \$140/MWh. The competitive benchmark is the only model that reflects cost causation (through congestion rents paid only by consumers in region A) and prices that reflect the long run cost of transmission (i.e., the price of \$140/MWh reflects the marginal cost of new transmission (\$140/MW) and the marginal cost of imported electricity from region B (\$0/MWh)).

4. Key implications for transmission rate design

The theoretical model and numeric parable presented in sections 2 and 3, respectively, offer several key lessons for transmission rate design.

First, the theoretical model shows in general that transmission policy objectives such as a no congestion transmission policy and postage stamp rate framework create residual network costs and move market outcomes away from the competitive benchmark. While the link between cost causation and the decisions of individual economic agents is complicated as a result of these objectives, a cost allocation approach that maximizes economic efficiency while considering the given transmission policy objectives remains possible, but its determination is an empirical question.

Second, while the numerical parable is but one example, it illustrates that there is no theoretical foundation to suggest that the embedded tariff design best provides efficient price signals and promotes the most efficient outcomes relative to other consumption-based, postage stamp tariffs, including a pure demand charge or pure energy charge. Furthermore, in a no congestion policy regime with postage stamp rates, the embedded tariff design does not establish rates the are truly reflective of cost causation. This represents a challenge to the view that transmission rate design should be primarily focused on the principle of cost causality and rely on cost causality studies to

[27] AESO (2021) at paragraph 87.

establish rates. Detailed efficiency analyses would provide regulators with a better understanding of the economic efficiency and equity implications of different proposed tariff designs.

Third, the economic efficiency attributes of a given set of consumption-based tariffs is an empirical question. In the numeric parable, a demand charge provided the highest net welfare gain followed by the embedded tariff design and the energy charge. However, this was a result of key assumptions, like the recovery of costs in only two periods, the relative size of the zero-price elastic residential demand in the peak period, the relative elasticity of industrial demand and supply. For example, if the numeric example was expanded from a two-period model to an annual model with 8,760 hours — 12 hours like the peak period and 8,748 hours like the off-peak period — it can be shown that the energy charge would provide the highest net welfare gain, followed by the embedded tariff design and the demand charge. This is because the energy charge could be allocated across a large amount of demand, thereby creating a smaller wedge between the price consumers pay and the marginal cost to produce electricity. The key point is that no one consumption-based design can be deemed more efficient than another *a priori*; the efficiency properties of any consumption-based tariff, which includes the embedded tariff design, depends on the general features of the electricity markets under review.

Fourth, alternative tariff designs that rely more heavily on the recovery of residual network costs through connection charges that do not vary with consumption — effectively fixed charges — do not reduce economic efficiency as energy and demand charges do. It is well established that fixed charges do not distort short-term consumption decisions and hence promote more efficient outcomes. For example, if it is assumed that there are 190 residential customers as in our numeric example, then a fixed connection charge of $182.29 applied to the 190 residential customers and two industrial customers would provide a higher net welfare gain than any of the consumption-based charges. The practical challenge that regulators face regarding fixed charges, however, is identifying a billing determinant that would allocate costs in a manner that is deemed fair and equitable. Several economists have recently advocated for a greater use of fixed connection charges for the recovery of residual network costs, including, Batlle et al. (2020) who consider a fixed charge based on a one-time measure of historic consumption and Borenstein et al. (2021) who consider the use of income-based fixed charges. Borenstein et al. (2021) further argue that the growing trend towards the decarbonization, decentralization, and digitalization of power grids is making the need to design rates that promote efficiency even more important.

5. Conclusion

This paper set out a simple theoretical model and numeric parable to consider the implications for economic efficiency of various transmission rate designs. The most important observation is that, in the presence of transmission policy objectives that move market outcomes away from the competitive benchmark, designing an economically efficient rate is an empirical question.

It is our view that the evolution of technology over time will make consideration of economic efficiency in the determination of regulated rates more important than in the past. In particular, the emergence of increasingly price-responsive consumers means that rates which were designed without consideration of this responsiveness will be economically inefficient by arbitrary and potentially enormous magnitudes.

Appendix: Mathematical Result

This Appendix derives the equilibria and social welfare for each of the scenarios discussed in the paper.

Scenario 1: Autarky

In autarky, the equilibria in region A and region B are determined independently by equating supply and demand in each region and for each period ($t = 1$ for peak and 2 for off-peak).

Region A

Peak residential demand is $450\ MW$ and peak industrial (inverse) demand is $p(q_A^1) = 400 - 2q_A^1$. Peak supply is $p(Q_A^1) = 50 + 0.5Q_A^1$. Setting residential plus industrial peak demand equal to supply provides the equilibrium quantities and price, $q_A^1 = 50\ MW, Q_A^1 = 500\ MW$, and $p_A^1 = \$300/MWh$.

Off-peak residential demand is 200 MW. Setting residential plus industrial off-peak demand equal to supply provides the equilibrium quantities and price, $q_A^2 = 100\ MW$, $Q_A^2 = 300\ MW$, and $p_A^2 = \$200/MWh$.

Region B

Peak (inverse) demand is $p(q_B^1) = 400 - 2q_B^1$ and supply is $p(Q_B^1) = 0, Q_B^1 \leq 50\ MW$. Given the available supply is limited to $50\ MW$, the equilibrium quantities and price on peak are $q_B^1 = Q_B^1 = 50\ MW$ and $p_B^1 = \$300/MWh$.

Off-peak demand is $p(q_B^2) = 400 - 2q_B^2$ and supply is $p(Q_B^2) = 0, Q_B^2 \leq 400\ MW$. Setting demand equal to supply, the equilibrium quantities and price off-peak are $q_B^2 = Q_B^2 = 200\ MW$ and $p_B^2 = \$0/MWh$.

Social welfare

Consumers' surplus in the peak period in region B is calculated as the area below the demand curve less the equilibrium price times the equilibrium quantity consumed. Dropping the unit measure notations, this is calculated as:

$$CS_B^1 = \int_0^{50}(400 - 2q)dq - 50x300 = 400(50) - 50^2 - 15{,}000 = \$2{,}500 \qquad (1)$$

Producers' surplus in the peak period in region B is equal to the equilibrium price times the equilibrium quantity sold less the area below the supply curve. This is calculated as:

$$PS_B^1 = 50x300 - \int_0^{50}(0)dq = 1{,}5000 - 0 = \$15{,}000 \qquad (2)$$

Consumers' and producers' surplus in the off-peak period in region B are calculated similarly as in equations (1) and (2) and equal:

$CS_B^2 = \$40{,}000$ and $PS_B^2 = \$0$

To calculate the consumers' surplus for the residential customers in region A, it is assumed that the willingness to pay or value of loss load for these costumers is equal to $1,000/MWh. Following equation (1) above, consumers' surplus for residential and industrial customers in region A in the peak period is calculated as:

$$CS_A^1 = 1{,}000x450 - 450x300 + \int_0^{50} (400 - 2q)dq - 50x300 = \$92{,}500$$

Producers' surplus in region A in the peak period is calculated as:

$$PS_S^1 = 500x300 - \int_0^{500} (50 + 0.5q)dq = \$62{,}500$$

Consumers' and producers' surplus in region A in the off-peak are calculated similarly, and equal:

$CS_A^2 = \$70{,}000$ and $PS_A^2 = \$22{,}500$

Social welfare, or total surplus, is calculated as the sum of consumers' and producers' surplus in regions A and B and for both peak and off-peak. That is:

Total Surplus = $CS_A^1 + CS_A^2 + CS_B^1 + CS_B^2 + PS_A^1 + PS_A^2 + PS_B^1 + PS_B^2 = \$305{,}000$

Scenario 2: Competitive benchmark

The SO's problem is to choose the level of transmission capacity, k, that maximizes social welfare subject to the constraints that demand equals supply in each region and for each period, and that transmission capacity is less than or equal to the available supply in region B (i.e., $\mathrm{k} \leq 400\ \mathrm{MW}$).

Social welfare includes the total welfare of consumers and producers (as in autarky) but also the total welfare of the transmission owner. The transmission owner's welfare is equal to the payments it receives from the SO (the congestion rents) less the cost it incurs to build and maintain the transmission line.

Noting that transmission investment only affects the off-peak period, the SO's problem is written as:

$$\max_{\{q_A^2, q_B^2, Q_A^2, Q_B^2, k\}} TS = CS_A^2 + CS_B^2 + PS_A^2 + PS_B^2 + CR^2 - fk \qquad (3)$$

s.t. $\quad q_B^2 + k = Q_B^2 \qquad (4)$

$$200 + q_A^2 = k + Q_A^2 \qquad (5)$$

and $\quad Q_B^2 \leq 400$

where $f = \$140/MW$ is the marginal cost to construct a unit of transmission capacity, and $CR^2 = (p_A^2 - p_B^2)k$ are the congestion rents paid by the SO to the transmission owner.

The Karush-Kuhn-Tucker first order necessary conditions for optimization are:

$$400 - 2q_B^2 - u_1 = 0 \quad (6)$$

$$400 - 2q_A^2 - u_2 = 0 \quad (7)$$

$$u_1 - u_3 = 0 \quad (8)$$

$$u_2 - 50 - 0.5Q_A^2 = 0 \quad (9)$$

$$u_2 - u_1 - 140 = 0 \quad (10)$$

$$u_1(Q_B^2 - k - q_B^2) = 0 \quad (11)$$

$$u_2(Q_A^2 + k - 200 - q_A^2) = 0 \quad (12)$$

$$u_3(400 - Q_B^2) = 0 \quad (13)$$

and $$u_i \geq 0 \; \forall \; i = 1, 2, 3 \quad (14)$$

where u1, u2, and u3 are the Lagrange multipliers for the constraints given by equations (3), (4), and (5) respectively.

Equations (6) to (14) can be solved to determine the off-peak equilibrium quantities and prices in the regions A and B; $q_B^2 = 200\,MW, Q_B^2 = 350\,MW$ and $p_B^2 = \$0/MWh$, $q_A^2 = 130\,MW$, $Q_A^2 = 180\,MW$, and $p_A^2 = \$140/MWh$; and the optimal level of transmission capacity, $k = 150\,MW$. Furthermore, the Lagrange multipliers in equilibrium are $u_1 = 0, u_2 = 140$, and $u_3 = 0$, where u_1 represents the marginal cost of adding an additional unit of demand in the region B, u_2 represents the marginal cost of adding an addition unit of demand in region A, and together, $u_1 + u_2$ represents the marginal benefit of adding a unit of transmission capacity. Note, in equilibrium the marginal benefit of transmission capacity, $u_1 + u_2 = \$140/MWh$, is equal to the marginal cost of transmission capacity, $f = \$140/MW$.

Social welfare

Consumers' surplus in regions A and B in the off-peak can be determined as per equation (1) above to equal $CS_B^2 = \$40{,}00$ and $CS_A^2 = \$88{,}900$. Producers' surplus in regions A and B in the off-peak can be determined as per equation (2) above to equal $PS_B^2 = \$0$ and $PS_A^2 = \$8{,}100$. The transmission owner just recovers its cost with $CR^2 = \$140 * 150 = \$21{,}000$. Adding the peak and off-peak period surplus measures provides, $\text{Total Surplus} = \$309{,}500$.

Scenario 3: "No-congestion" transmission policy

Under a no-congestion transmission policy, the SO's problem is to facilitate investment in transmission capacity, k, to the point that there is no transmission congestion and prices (assuming there are no transmission losses) in the two regions are equal. The equilibrium conditions for the off-peak period are written mathematically as:

$$q_A^2 + 200 + q_B^2 = Q_B^2 + Q_A^2 \tag{15}$$

$$Q_B^2 = 400 \tag{16}$$

and $$k = Q_B^2 - q_B^2 \tag{17}$$

Equation (15) requires supply and demand to be equal across both regions in off-peak. Equation (16) ensures that the low-cost available supply in region B is fully utilized in a no-congestion policy regime. Equation (17) ensures that transmission capacity is built to allow the flow of excess supply from region B to region A.

Expressing the demand and supply functions as functions of price and dropping the regional subscript implies:

$$200 - \frac{p^2}{2} + 200 + 200 - \frac{p^2}{2} = 400 + 2p^2 - 100 \tag{18}$$

$$k = 200 + \frac{p^2}{2} \tag{19}$$

Solving equation (18) for p^2 provides $p^2 = \$100/MWh$, and with equation (19) provides k=250 MW. Inputting the equilibrium price into the demand and supply curves provides the equilibrium quantities, $q_B^2 = 150\ MW, Q_B^2 = 400\ MW$, $q_A^2 = 150\ MW$, and $Q_A^2 = 100\ MW$.

Social welfare

Consumers' surplus in regions A and B in the off-peak can be determined as per equation (1) above to equal $CS_B^2 = \$22{,}500$ and $CS_A^2 = \$102{,}500$. Producers' surplus in regions A and B in the off-peak can be determined as per equation (2) above to equal $PS_B^2 = \$40{,}000$ and $PS_A^2 = \$2{,}500$. Under a non-congestion policy, the SO does not recover congestion rents which means, absent a regulated tariff, the transmission owner incurs a loss of $\$140x250MW = \$35{,}000$. Adding the peak and off-peak period surplus measures provides Total Surplus = $\$305{,}000$.

Scenario 4: No-congestion transmission policy and cost recover through a regulated tariff

As in the previous scenario, the SO's problem is to facilitate investment in transmission capacity, k, to the point that there is no transmission congestion and prices (assuming there are no transmission losses) in the two regions are equal. Additionally, the SO must establish a regulated tariff that recovers sufficient funds from consumers to cover the transmission owner's cost.

The SO's problem can be written generically to include a demand charge, t_d levied on consumption in the peak period, and an energy charge, t_e levied on consumption in both periods. Let α be the share of transmission costs, fk recovered through a demand charge. The equilibrium conditions are:

$$q_A^1 + 450 + q_B^1 = Q_A^1 + Q_B^1 \tag{20}$$

$$q_A^2 + 200 + q_B^2 = Q_A^2 + Q_B^2 \tag{21}$$

$$Q_B^1 = 50 \; and \; Q_B^2 = 400$$
(22)

$$k = Q_B^2 - q_B^2 \tag{23}$$

$$t_d(450 + q_A^1 + q_B^1) = \alpha f k \tag{24}$$

and $$t_e(200 + 450 + q_A^1 + q_B^1 + q_A^2 + q_B^2) = (1-\alpha) f k \tag{25}$$

Following the approach taken to derive equations (18) and (19), the equilibrium conditions can be written as:

$$200 - \frac{p^1}{2} - \frac{t_e}{2} - \frac{t_d}{2} + 450 + 200 - \frac{p^1}{2} - \frac{t_e}{2} - \frac{t_d}{2} = 50 + 2p^1 - 100 \tag{26}$$

$$200 - \frac{p^2}{2} - \frac{t_e}{2} + 200 + 200 - \frac{p^2}{2} - \frac{t_e}{2} = 400 + 2p^2 - 100 \tag{27}$$

$$k = 200 + \frac{p^2}{2} + \frac{t_e}{2} \tag{28}$$

$$t_d(850 - p^1 - t_d - t_e) = \alpha(2{,}800 + 70p^2 + 70t_e) \tag{29}$$

and $$t_e(1{,}450 - p^1 - p^2 - t_d - 2t_e) = (1-\alpha)(2{,}800 + 70p^2 + 70t_e) \tag{30}$$

Solving equations (26) and (27) for price provides setting $p^1 = 300 - \frac{(t_d + t_e)}{3}$ and $p^2 = 100 - \frac{t_e}{3}$. Inserting prices into equations (29) and (30) further implies:

$$k = 250 + \frac{t_e}{3} \tag{31}$$

$$t_d\left(550 - \frac{2t_d}{3} - \frac{2t_e}{3}\right) = \alpha(35{,}000 + 46.\dot{6}t_e) \tag{32}$$

and $$t_e\left(1{,}050 - \frac{2t_d}{3} - \frac{4t_e}{3}\right) = (1-\alpha)(35{,}000 + 46.\dot{6}t_e) \tag{33}$$

Note, setting $t_d = t_e = 0$ in equations (26) to (28) provides the no-congestion equilibria and transmission capacity level derived in Scenario 3.

Note further, that by equation (31) a positive energy charge will lead to an increase in the amount of transmission capacity required to permit no congestion in equilibrium.

Demand charge

To determine the equilibria with a demand charge and full transmission cost recovery, set $t_e = 0$ and $\alpha = 1$. Equations (26) to (28) imply $p^1 = \$147.58/MWh$, $p^2 = \$58/MWh,$ and k=258 MW. The demand charge is determined by solving the quadratic equation (29), with $t_d = \$14.39/MW$.

Furthermore, as per (1) and (2), consumers', producers', and total surplus can be solved as:

$CS_B^1 = \$723$, $CS_A^1 = \$19{,}838{,}751$, $CS_B^2 = \$10{,}082$, $CS_A^2 = \$3{,}998{,}482$

$PS_B^1 = \$6{,}879$, $PS_A^1 = \$49{,}992$, $PS_B^2 = \$19{,}200$, $PS_A^2 = \$336$

$\text{Total Surplus} = \$23{,}924{,}450$

Energy charge

To determine the equilibria with an energy charge and full transmission cost recovery, set $t_d = 0$ and $\alpha = 0$. Equations (26) to (28) imply $p^1 = \$148.35/MWh$, $p^2 = \$56.35/MWh,$ and k=267 MW. The energy charge is determined by solving the quadratic equation (29), with $t_e = \$10.33/MWh$.

Furthermore, as per (1) and (2), consumers', producers', and total surplus can be solved as:

$CS_B^1 = \$854$, $CS_A^1 = \$19{,}842{,}179$, $CS_B^2 = \$8{,}888$, $CS_A^2 = \$3{,}995{,}553$

$PS_B^1 = \$6{,}917$, $PS_A^1 = \$50{,}779$, $PS_B^2 = \$18{,}539$, $PS_A^2 = \$212$

$\text{Total Surplus} = \$23{,}923{,}920$

Embedded cost tariff

To determine the equilibria under an embedded cost tariff and full transmission cost recovery, as per Table 2 in section 3, set $\alpha = 0.827$. Equations (29) and (30) can be used to solve t_d and t_e. The authors used the solver function in Excel to derive $t_d = \$11.83/MW$ and $t_e = \$1.76/MWh$. Substituting these values into equations (26) to (28) implies $p^1 = \$147.83/MWh$, $p^2 = \$57.72/MWh,$ and k=259 MW.

Furthermore, following equation (1) and (2), consumers', producers', and total surplus can be solved as:

$CS_B^1 = \$744$, $CS_A^1 = \$19{,}839{,}328$, $CS_B^2 = \$9{,}873$, $CS_A^2 = \$3{,}997{,}977$

$PS_B^1 = \$6{,}891$, $PS_A^1 = \$50{,}242$, $PS_B^2 = \$19{,}087$, $PS_A^2 = \$313$

$\text{Total Surplus} = \$23{,}924{,}374$

References

Alberta Electric System Operator. 2021. *AESO Bulk and Regional Rate Design and Modernized DOS Rate Design Application*. October 15. https://www.aeso.ca/assets/Tariff-2021-BR-Application/26911_X0001.01_AESO-Bulk-and-Regional-Rate-Design-Application-Clean_000602.pdf

Batlle, C., Mastropietro, P., and Rodilla, P. 2020. "Redesigning residual cost allocation in electricity tariffs: A proposal to balance efficiency, equity and cost recovery," *Renewable Energy* 155 (August), 257-266. https://doi.org/10.1016/j.renene.2020.03.152

Biggar, D. and Hesamzedah, M. R. 2014. *The economics of electricity markets*. Wiley-IEEE Press.

Borenstein, S., Fowlie, M., and Sallee, J. 2021. "Designing electricity rates for an equitable energy transition," *Energy Institute at Haas WP 314*. February. https://haas.berkeley.edu/wp-content/uploads/WP314.pdf

Church, J. R., Rosehart, W., and MacCormack, J. 2009. "Transmission policy in Alberta and Bill 50," School of Public Policy, University of Calgary.

Government of Alberta. 2003. "Transmission development: The right path for Alberta: A policy paper." https://open.alberta.ca/dataset/0db52c69-eed1-4f4a-997c-a47d57cc9788/resource/7238f12e-2a43-41dc-856e-c623a9fc57a3/download/3103222-2003-transmission-development-policy.pdf

Hogan, W., Rosellón, J., and Vogelsang, I. 2010. "Toward a combined merchant-regulatory mechanism for electricity transmission expansion," *Journal of Regulatory Economics* 38 (June), 113-143. https://doi.org/10.1007/s11149-010-9123-2

Joskow, P. and Tirole, J. 2005. "Merchant transmission investment," *The Journal of Industrial Economics* 53(2), 233-264. http://www.jstor.org/stable/3569878.

Lazar, J., Chernick, P., and Marcus, W. 2020. "Electric Cost Allocation for a New Era." *Regulatory Assistance Program.* January 2020. https://www.raponline.org/wp-content/uploads/2020/01/rap-lazar-chernick-marcus-lebel-electric-cost-allocation-new-era-2020-january.pdf

NERA Economic Consulting. 2021. *AESO Bulk and Regional Tariff Design: Export Report. September 24.* https://www.aeso.ca/assets/Tariff-2021-BR-Application/X0038.01-Appendix-D-NERA-Expert-Report-AESO-BandR-Tariff-Design-Clean.pdf

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., and Bohn, R. E. 1988. *Spot pricing of electricity*. Springer.

Stoft, S. 2002. *Power system economics: Designing markets for electricity*. Wiley-IEEE Press.