Technical Appendix on Congestion

John E. Parsons*

this draft April 19, 2023

*MIT Center for Energy and Environmental Policy Research (CEEPR).

A Congestion Rent

This section presents a few fundamental points about congestion rent. Since FTR markets are funded by congestion rent, this provides a foundation for the later sections on FTR markets. Some readers may choose to skip over this section and refer back here as needed. Others may prefer to take this opportunity to refresh their understanding and thereby better orient themselves for what comes later.

A.1 Defining Congestion Rent

Electricity markets that employ locational marginal pricing charge customers more for the electricity consumed than is paid to generators for the electricity produced. The difference is variously called congestion rent, congestion surplus, congestion income, merchandising surplus and marketing surplus. Sometimes the terminology is shortened to congestion.

How congestion rent arises can be illustrated using a simple two bus network as shown in Figure 1. On the left-hand-side is Bus A where 200 MW of load is located as well as some

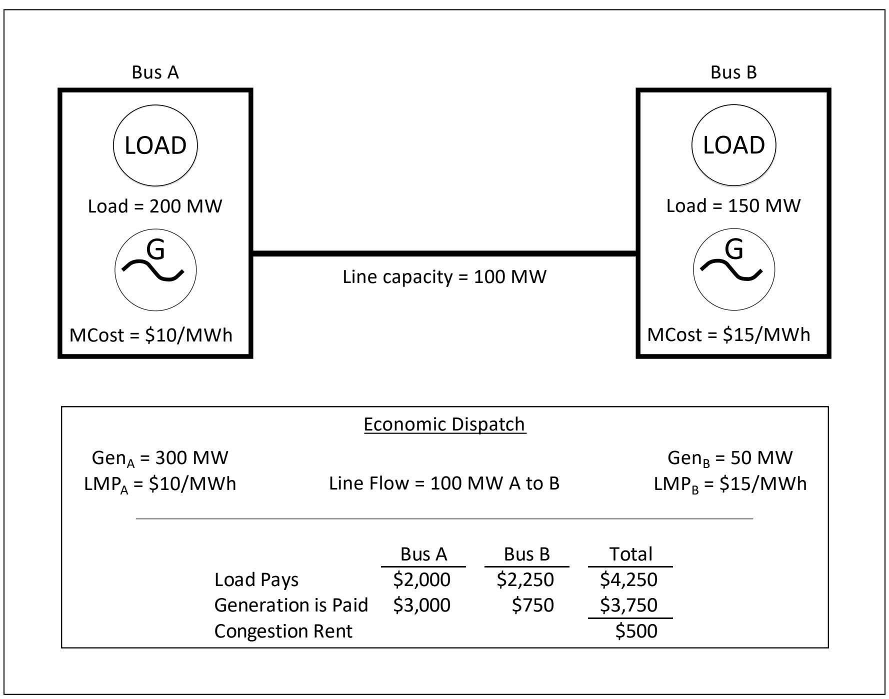

Figure 1: Two Bus Network Example of Congestion Rent.

inexpensive generation with a marginal cost of \$10/MWh. On the right-hand-side is Bus B where another 200 MW of load is located alongside some relatively expensive generation with a marginal cost of \$15/MWh. The two buses are connected by a transmission line with a maximum capacity of 100 MW in either direction. This limit is what gives rise to congestion rent. To keep the focus on the capacity of the transmission line, we assume the capacities of each generator are very large and we do not bother with them. The economic dispatch uses the cheap generator at Bus A for 300 MW, which supplies all of the load at Bus A and delivers the remaining 100 MW to Bus B. This utilizes the transmission line to its full capacity. The expensive generator at Bus B is dispatched at 100 MW to serve the remaining load at Bus B not supplied by the transmission line. The LMP at Bus A is \$10/MWh, and the LMP at Bus B is \$15/MWh, the marginal cost of generation at each Bus. The table at the bottom of the figure calculates the payments from load, which come to \$5,000, and the payments to generators, which come to \$4,500. The difference of \$500 is the congestion rent. The congestion rent arises because the customers at Bus B pay \$5/MWh more for the power delivered over the transmission line than is paid to the generator at Bus A for that power supplied to the transmission line.

A.1.1 A Formal Definition

We can usefully formalize the definition of congestion rent as an outcome of the economic dispatch problem. The following presentation is a condensed version of results from Wu et al. (1996), with some minor changes to notation choices. See also Kirschen and Strbac (2004), Chapter 6, Section 6.3.4.4. We model an electricity network with N nodes or buses, $i, j = 1, ..., N$, and a set of K transmission lines that connect some pairs of nodes, ij, $K \leq N(N-1)/2$.[1] Each node hosts generators and may host load. Denote the load at node i by $d_i \geq 0$, and the vector of nodal loads as d. Denote the dispatched generation at node i by $g_i \geq 0$ and the vector of nodal generation as g, which we call a dispatch. We take the perspective that the loads are fixed exogenously, and the problem is to choose a dispatch, in which case this economic dispatch problem is also the optimal power flow problem.

[1] The terms bus and node are interchangeable throughout this paper. The term bus comes from the electricity industry, where a bus (short for bus bar) is a piece of equipment used for making connections such as between high-voltage power lines, or a generator to a transmission line, or a low-voltage distribution network to a high-voltage network. The term node comes from the mathematical usage denoting the point of connection of paths in a network. In modeling electricity networks, where buses are nodes, the term bus is commonly used in some situations and node is commonly used in others, which can falsely suggest a meaningful distinction.

Power flow on the network is determined by the net injections at each of the nodes, $q_i \equiv g_i - d_i$, and by the physical properties of the transmission lines, such as the admittances, $Y_{ij} > 0$ and $Y_{ij} = Y_{ji}$. We discuss a simplified version of the physical model with linearized DC power flow which has no losses. Later in the paper we will mention some complications the arise in the more accurate and complicated AC flow model. Denote by q_{ij} the power flow from node i to node j, with $q_{ji} = -q_{ij}$ and by convention $q_{ij} > 0$ means the power flows from node i to node j. For node combinations, ij, not connected by a transmission line, $q_{ij} \equiv 0$. In this simplified model the power flows are given by $q_{ij} = Y_{ij}\left(\theta_i - \theta_j\right)$, where θ_i denotes the phase angle at node i. It is then convenient to restate the dispatch problem as a simultaneous choice of the net injections, $q = (q_1, ..., q_N)$, and a set of voltage angles, $\theta = (\theta_1, ..., \theta_N)$, which together satisfy the N real power flow equations

$$q_i = \sum_{j=1}^{N} q_{ij} = \sum_{j=1}^{N} Y_{ij}\left(\theta_i - \theta_j\right), \qquad i = 1, ..., N. \tag{1}$$

Only $N - 1$ of these equations are independent, which we will come back to later. Also, any choice that satisfies these also satisfies the power balance equation, $\sum q_i = 0$, so we do not need to separately define that constraint. Under our simplifying assumptions, the set of feasible dispatches is convex.

The transmission lines are also characterized by limits on the flow of power in each direction, $F_{ij} \geq 0$ and $F_{ji} \geq 0$:

$$q_{ij} = Y_{ij}\left(\theta_i - \theta_j\right) \leq F_{ij}, \qquad i, j = 1, ..., N. \tag{2}$$

These line flow constraints are only relevant for the K pairs ij connected by a transmission line. For each pair, there are 2 constraints, one for flow in the direction i to j, and one for flow in the opposite direction.

The cost of generation at node i is given by the increasing, convex function $C_i(g_i)$, with $C_i(0) = 0$ and $C_i(g_i) > 0$ for $g_i > 0$. We can also write this as the cost of the net injection at node i, $\hat{C}_i(q_i) \equiv C_i(q_i + d_i)$. Note $\partial\hat{C}_i(q_i)/\partial q_i = \partial C_i(g_i)/\partial g_i$.

The economic dispatch problem is:

$$\underset{(q,\theta)}{\text{minimize}} \sum_{i=1}^{N} \hat{C}_i(q_i) \tag{3}$$

subject to the real power flow equations, (1), and the line flow constraints, (2). This is a convex program with linear constraints. We can construct a Lagrangian, with multipliers p_i for the N power flow constraints and $\mu_{ij} \geq 0$ for the $2K$ line limit constraints:

$$\Phi = \sum_i \hat{C}_i(q_i) + \sum_i p_i \left[\sum_j Y_{ij}(\theta_i - \theta_j) - q_i \right] + \sum_i \sum_j \mu_{ij} \left[Y_{ij}(\theta_i - \theta_j) - F_{ij} \right] \qquad (4)$$

The optimal solution, (q^*, θ^*) known as economic dispatch, and the associated Lagrangian multipliers, p^*, μ^*, satisfy

$$p_i = \frac{\partial \hat{C}_i(q_i^*)}{\partial q_i}, \qquad i = 1, ..., N; \qquad (5)$$

$$\sum_{j=1}^{N} Y_{ij}[p_i^* - p_j^* + \mu_{ij}^* - \mu_{ji}^*] = 0, \qquad i = 1, ..., N; \qquad (6)$$

$$\mu_{ij}^* \left[Y_{ij}(\theta_i^* - \theta_j^*) - F_{ij} \right] = 0; \mu_{ij}^* \geq 0. \qquad i, j = 1, ..., N. \qquad (7)$$

The N Lagrange multipliers, p_i^* are the nodal prices. At each node, the price equals the marginal cost of generations, $p_i^* = \partial C_i(g_i^*)/\partial g_i$.

The $2K$ Lagrange multipliers, μ_{ij}^*, are the shadow prices for the line flow constraints F_{ij}. Denoting by C^* the value of the aggregate cost at the economic dispatch, $C^* = \sum_i^N C_i(g_i^*)$, and writing it as a function of the limit F_{ij}, $C^*(F_{ij})$, then

$$\mu_{ij}^* = -\frac{\partial C^*}{\partial F_{ij}}. \qquad (8)$$

Increasing the limit by one unit lowers the system cost of generation by μ_{ij}^*, which is why we say μ_{ij}^* is the marginal value of additional capacity on that line in that direction.

With this formalization in hand, we can now approach the definition of congestion from two directions. First, we calculate congestion rent from the shadow price of each constraint and the flow on that constraint. Second, we calculate the merchandising surplus from the nodal prices paid by customers and to generators. As it happens, these two different approaches come to the same result because of the duality property of the solution to the economic dispatch problem given by equations (5)-(7).

Define the congestion rent, R_{ij}, on constraint F_{ij} as the product of the shadow price on that constraint and the line flow at that limit:

$$R_{ij} \equiv \mu_{ij}^* F_{ij}. \tag{9}$$

If the line is unconstrained in the direction ij, then the congestion rent on that constraint is zero: if $q_{ij}^* < F_{ij}$, then $\mu_{ij}^* = 0$ and so $R_{ij} = 0$. If the line is constrained, then the congestion rent on that constraint is positive: $q_{ij}^* = F_{ij}$, then $\mu_{ij}^* > 0$ and so $R_{ij} > 0$. The aggregate congestion rent is,

$$R \equiv \sum_{i=1}^{N} \sum_{j=1}^{N} R_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{ij}^* F_{ij} \geq 0. \tag{10}$$

The summation includes both R_{ij} and R_{ji}, but only one of them can be positive: $\mu_{ij}^* > 0$ implies $\mu_{ji}^* = 0$ and $R_{ji} = 0$, while $\mu_{ji}^* > 0$ implies $\mu_{ij}^* = 0$ and $R_{ij} = 0$. Of course, the line may be unconstrained in both directions at the same time so that both are zero.

Define the merchandising surplus, S, as

$$S \equiv \sum_{i=1}^{N} p_i^* (d_i - g_i^*). \tag{11}$$

Given the assumptions of our simplified network model, using the power flow equations (1) and the set of equations (6) tying together the shadow prices on the line flow constraints and the nodal prices, one can show that:

$$\sum_{i=1}^{N} p_i^* (d_i - g_i^*) = \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{ij}^* F_{ij}, \tag{12}$$

i.e., that the merchandising surplus exactly equals the aggregate congestion rent

$$S = R. \tag{13}$$

Certain of our simplifying assumptions are key, so that without them there may be cases in which $S \leq R$, and indeed cases in which $S < 0$ as shown in Philpott and Pritchard (2004).

In the material above, we had identified the $2K$ transmission constraints with the indexes for the nodes the line connected, i and j. In later sections, it will be convenient to use a single variable k to index them, $k = 1, ..., 2K$. So, where we previously wrote the shadow prices for the line flow constraints as μ_{ij}^*, we now write them as μ_k^*, and where previously we wrote the line flow constraints as F_{ij}, we now write them as F_k.

A.1.2 Two-Settlement Markets

So far, we have informally and more formally defined congestion rent for a single-settlement dispatch. However, the standard market design employed in the U.S. is a two-settlement system. An initial, prospective dispatch is calculated for each hour in the day-ahead market. This dispatch satisfies the transmission line constraints as modeled in an approximation to the full AC-transmission model. Day-ahead financial settlements are made for this initial dispatch. Then, an actual dispatch is calculated for subsets of each hour in the real-time market. This dispatch incorporates updated and more granular load forecasts and generator availability. It also satisfies possibly different transmission line constraints as modeled in the true AC-transmission model. The differences between the actual dispatch and the initial dispatch are the balancing adjustments. Real-time financial settlements are made on these balancing adjustments. Total load charges are the sum of the day-ahead settlements and the balancing settlements, as are total generation credits. Total congestion rent is the difference between the total load charges and the total generation credits: it is the surplus of payments from load over the payments to generation. Total congestion is the sum of day-ahead congestion and balancing congestion.

Table 1 illustrates the calculation of congestion in a two-settlement market using the two-node system shown in Figure 1. Many changes can produce a divergence between the day-ahead and the real-time settlements, such as revised load or a generator trip. In this illustration the only change is to the transmission limits. The inputs for load and marginal costs of generation match those used in Figure 1, as does the real-time line limit and final dispatch. However, we assume that the day-ahead line limit is 1 MW more. Therefore, the day-ahead dispatch for the inexpensive generator at Bus A is 1 MW more as compared against what is shown in Figure 1, and the day-ahead dispatch for the expensive generator at Bus B is 1 MW less. The flow on the line is 1 MW more. Then, in the real-time market, the generation at Bus A must be reduced by 1 MW and the generation at Bus B increased by 1 MW. Total charges to customers remain unchanged. However, more of those charges must be paid out to generators than had been anticipated in the day-ahead market. That is, the marketing surplus is less than what had been anticipated in the day-ahead market. Where the day-ahead congestion had been \$505, the actual congestion rent is only \$500. Therefore, balancing congestion is -\$5.

Table 1: Illustrative Calculation of Congestion in a Two-Settlement Market.

Day-Ahead Settlement				
Line Limit = 101 MW				
		Bus A	Bus B	Total
Load	(MW)	200	150	350
Generation	(MW)	301	49	350
LMP	($/MWh)	10	15	
Load Charges	($)	2,000	2,250	4,250
Gen Credits	($)	3,010	735	3,745
Congestion Rent	($)			505
Real-Time Settlements				
Line Limit = 100 MW				
		Bus A	Bus B	Total
△ Load	(MW)	0	0	0
△ Generation	(MW)	-1	1	0
LMP	($/MWh)	10	15	
Load Charges	($)	0	0	0
Gen Credits	($)	-10	15	5
Congestion Rent	($)			-5
Total Congestion				500

The inputs for load and marginal costs of generation match those used in Figure 1, as does the real-time line limit and final dispatch. However, the day-ahead line limit is more.

This illustrates that balancing congestion is a true-up of the congestion rent account. While both day-ahead and total congestion are always positive, balancing congestion may be positive or negative. In practice, across most ISOs, balancing congestion has generally been negative, indicating that the day-ahead transmission model is not fully capturing all of the constraints that bind in the real-time dispatch.

A.1.3 Congestion Rent in Practice

Table 2 shows the size of the day-ahead congestion rents across the seven ISOs in the United States over the years 2008 to 2020, and places that in context with total billings in each ISO.[2]

The annual variance in the rents is large. For example, in PJM they range between $663

[2]All of these ISOs use the multi-settlement market structure that is standard in the U.S., with a day-ahead and a real-time market. The total congestion is a function of both markets, and may be more or less than the congestion in the day-ahead market. For example, in PJM total congestion has generally been less than day-ahead congestion by approximately 17%. I use day-ahead congestion in this table only because it is the quantity that is reported most consistently across all of the ISOs. Most of them report only day-ahead congestion.

Table 2: Day-Ahead Congestion Across Seven ISOs

	PJM		CAISO		ERCOT		ISONE		MISO		NYISO		SPP	
Year	($ millions)	share of wholesale cost	($ millions)	share of wholesale cost	($ millions)	share of wholesale cost	($ millions)	share of wholesale cost	($ millions)	share of wholesale cost	($ millions)	share of wholesale cost	($ millions)	share of wholesale cost
2008	2,597	7.6%					125	1.0%	500	1.7%	953	5.7%		
2009	901	3.4%					27	0.4%	305	1.8%	376	4.4%		
2010	1,713	4.9%					37	0.4%	498	2.4%	419	3.9%		
2011	1,245	3.5%			473	2.5%	18	0.2%	503	2.6%	407	4.3%		
2012	780	2.7%	486	5.8%	516	5.4%	29	0.5%	778	5.3%	301	3.6%		
2013	1,011	3.0%	445	4.2%	466	4.0%	46	0.5%	842	5.1%	664	5.7%		
2014	2,231	4.5%	458	3.8%	528	3.7%	34	0.3%	1,444	5.2%	578	4.5%	292	3.8%
2015	1,632	3.8%	231	2.8%	300	3.0%	31	0.3%	751	3.9%	540	6.0%	161	3.1%
2016	1,100	2.8%	250	3.4%	408	4.4%	39	0.5%	737	3.8%	438	6.1%	298	5.3%
2017	733	1.8%	358	3.9%	787	7.3%	41	0.5%	743	3.8%	415	6.1%	506	8.6%
2018	1,379	2.8%	628	5.8%	1,097	7.6%	65	0.5%	701	3.1%	501	5.9%	453	6.2%
2019	714	1.8%	355	4.0%	1,081	5.5%	33	0.3%	528	3.0%	433	7.0%	457	7.4%
2020	663	2.0%	488		1,343	12.4%			660	4.2%	297	5.2%		
Average	1,285	3.4%	411	4.3%	700	5.4%	44	0.5%	691	3.5%	486	5.2%	167	5.7%

Sources and Notes:

PJM: Monitoring Analytics, "2020 State of the Market Report for PJM, Volume 2: Detailed Analysis", Tables 11-11 and 11-12.

CAISO: Data on Day-Ahead congestion for 2012-2019 from CAISO Department of Market Monitoring "Report on results of 2019 congestion revenue rights auction", January 27, 2020.
Data for 2020 from CAISO Department of Market Monitoring, "Q4 2020 Report on Market Issues and Performance", April 28, 2021.
Data on wholesale cost is from Department of Market Monitoring, "Annual Report on Market Issues and Performance", various years.

ERCOT: Data is from Potomac Economics, Independent Market Monitor for ERCOT, "State of The Market Report for the ERCOT Electricity Markets", various years.
Congestion shown is the "Congestion Rent". Although the accompanying text often labels this the Day-Ahead Congestion Rent, it is clear from the 2013 text that it also reflects balancing congestion.
Data on wholesale cost is calculated as the product of the "All-in Price" times Annual Load. The "All-in Price" is the Real-Time price plus small adders, as well as the average unit cost of ancillary services and uplift. Because the average Real-Time price is generally slightly less than the Day-Ahead price paid by the majority of load, this will be slightly less than the actual wholesale cost paid by customers.

ISONE: Data is from ISO New England, Internal Market Monitor, "Annual Markets Report", various years.

MISO: Data is from Potomac Economics, Independent Market Monitor for the Midcontinent ISO, "State of The Market Report for the MISO Electricity Markets", various years.
Data on wholesale cost is calculated as the product of the "All-in Price" times Annual Load. The "All-in Price" is the Real-Time price plus small adders, as well as the average unit cost of ancillary services and uplift. Because the average Real-Time price is generally slightly less than the Day-Ahead price paid by the majority of load, this will be slightly less than the actual wholesale cost paid by customers.

SPP: Data is from SPP Market Monitoring Unit, "State of the Market" reports, various years.
Wholesale cost is calculated as the product of annual load and the "All-in Wholesale Cost" per MWh.

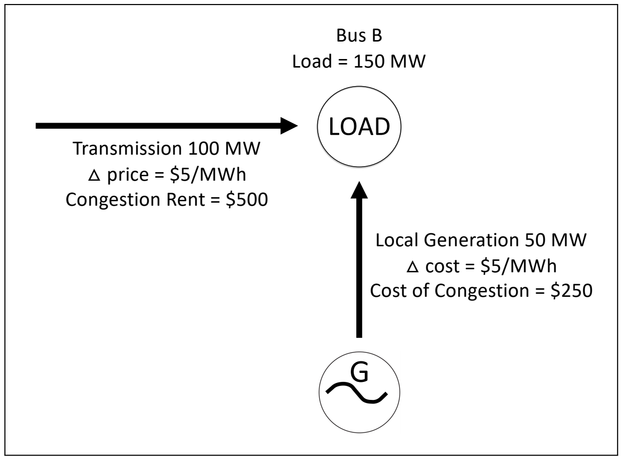

Figure 2: Distinguishing the Cost of Congestion from the Congestion Rent.

million and $2.597 billion, and from 1.8% to 7.6% of billings. In ERCOT, the share reached 12.4% last year. For individual customers, the congestion rent's share of their bill will vary around the average, so for some customers congestion paid will be much more than these percentages.

The variance across ISOs is notable as well. ISONE has the lowest day-ahead congestion as a share of the total wholesale cost, by far, at 0.5%. PJM is next. SPP has the highest average day-ahead congestion as a share of the total wholesale cost at 5.7%.

A.2 Congestion Rent versus the Cost of Congestion

Congestion is costly because the system is unable to fully utilize the lowest cost generation capacity. Higher cost capacity must be utilized while lower cost capacity lies idle. This cost of congestion is entirely distinct from the congestion rent discussed earlier. Figure 2 helps to make this distinction, using information from the example already displayed in Figure 1. Figure 2 focuses on the power delivered to load at Bus B. The transmission line delivers 100 MW of that power from Bus A. Although the cost of generating that power at Bus A is $10/MWh, the price paid at Bus B is $15/MWh. Figure 2 shows the difference, the Δ price, of $5/MWh, which gives rise to the congestion rent of $500. Also shown in Figure 2 is the local generation of 50 MW from the expensive generator at Bus B. The unit cost of this generation is more than the unit cost of the generation available at Bus A. Figure 2 shows the difference, the Δ cost, of $5/MWh. Utilizing this higher cost generation for 50MW of load is the cost of congestion, which comes to $250. Altogether, load at Bus B pays $750 more for power as compared against a counterfactual in which there were no transmission constraints and all load was served by the lowest cost generation.[3] The congestion rent and the cost of congestion are two distinct elements of the $750 cost premium paid by load at Bus B.

The cost of congestion is a system cost. Higher cost generation must be used despite its cost. Congestion rent, in contrast, is not a system cost at all. It is a transfer of money from load at Bus B to someone else. The money may be paid to the transmission owner or to an independent system operator. The money could be returned back to load at Bus B, in which case the load will only have paid the full cost of generation and nothing more. Of course, if the congestion rent is not returned to the load that paid it, then it is a cost to that load and a gain to whoever captures it. For economists, the nomenclature 'rent' is a term-of-art that connotes the fact that it is not a social cost.

The distinction between the cost of congestion and the congestion rent is important to keep in mind, but often overlooked. Congestion rent is a well defined number for any dispatch, and ISOs regularly report the cumulative amount over a period, whether it be quarterly or annually. In contrast, to calculate the cost of congestion requires a counterfactual, and probably for this reason it is not regularly reported. Overbye (2003) provides an estimate for TVA. Forward-looking models of capacity expansion which focus on how to minimize system costs implicitly reflect the cost of congestion, although it is often not identified separately. An example where both the cost of congestion and congestion rents are reported is Duenas-Martinez et al. (2021), which models the U.S. midcontinent power systems and estimates that currently the congestion rent is much larger than the cost of congestion, while in a decarbonized system the cost of congestion is larger than the congestion rents. Discussions about congestion volatility often overlook the distinction between the two, focusing almost exclusively on hedging the congestion rent while overlooking entirely the problem of hedging volatility in the cost of congestion.

A.3 Who Pays Congestion Rent?

Loads at different nodes pay different amounts of congestion rent. In the simple two node network example shown in Figure 1 it is easy to see the attribution across loads: the load at

[3] Of course, realizing this counterfactual would require investing to expand transmission capacity, which would itself be costly.

Bus B pays all of the congestion rent, while the load at Bus A pays none. Loads downstream of a congested transmission line pay the congestion rent associated with that line. The rent paid equals the marginal cost of the constraint at that line, multiplied times the power flow across that line, which is at its limit.

In the two node network shown in the figure, there is only one point of congestion and it is unambiguous which load is downstream of the congested line. In larger, more complex networks, there can be multiple points of congestion, each with its own marginal cost. Also, identifying which nodes are downstream of each point of congestion is less obvious due to the properties of network power flow. Some nodes will be downstream of one subset of congestion, while others are downstream of other, perhaps overlapping subsets of congestion. Therefore attribution of which loads paid how much of the congestion rent requires a more extensive set of calculations to which we now turn.

A.3.1 Congestion Payments versus Congestion Rents

The standard accounting and settlement practice in U.S. ISOs employ a decomposition of the nodal LMPs into three components: an energy component which is the same across all nodes, and congestion and losses components which may vary by node. PJM calls the common energy component the System Marginal Price, SMP, and uses the abbreviations $CLMP$ and $MLMP$ for the congestion and losses components, respectively. So, where $LMP_i(t)$ is the nodal price at which energy is bought or sold at node i in hour t, the decomposition formula is:

$$LMP_i(t) = SMP(t) + CLMP_i(t) + MLMP_i(t). \tag{14}$$

The name/acronym for congestion prices vary across ISOs. PJM uses the acronym $CLMP$, MISO and SPP calls it the Marginal Congestion Component (MCC), NYISO CAISO the Marginal Cost of Congestion (MCC). This three-part decomposition is a variation on the original formulation of locational marginal pricing in, for example, Bohn, Caramanis, and Schweppe (1984) and Schweppe et al. (1988) in which one node was chosen as a reference point and the prices at other nodes were expressed relative to the price at the reference node. For example,

looking back to the solution of our optimal power flow problem given in equations (5)-(7), we can choose as our benchmark cost of energy the LMP at node j, p_j^*, and then write the

LMPs at all other nodes, $i = 1, ..., N, i \neq j$, as,

$$p_i^* = p_j^* + \sum_{k=1}^{2K} S_{i,k}^*(j)\mu_k^*, \tag{15}$$

where $S_{i,k}^*(j)$, known as the shift factor, is a function of the chosen benchmark node. This decomposition writes the price at node i in terms of the price at node j. Because shift factors are a function of the chosen benchmark node j, the decomposition changes with the choice of benchmark. Originally, when locational marginal pricing was first introduced, it was common to utilize the pre-existing terminology and practice of identifying a particular 'slack-bus' and denoting the LMP at that bus as the benchmark price. Often, a particular bus was chosen because it was a central focal point of the system–for example, a bus located at the city of Santiago for the Chilean system, or a bus located at the city of Buenos Aires for the Argentinian system, as discussed in Rivier and Pérez-Arriaga (1993) and Littlechild and Skerk (2008). Later, decompositions were developed using a benchmark calculated as the weighted average price across many buses–see, for example, Meisel (1993) and Rivier and Pérez-Arriaga (1993). The term 'reference bus' is now commonly used for this choice. Several U.S. ISOs, such as PJM, use a load-weighted average LMP as the reference bus. The NYISO uses a substation near Utica, New York as it is near the "electrical center" of the system, among other reasons.

ISO billing systems record credits and charges for electricity decomposed into these same three parts. A generator receives (i) an hourly 'energy' credit, (ii) an hourly congestion credit, and (iii) an hourly losses credit.[4] A load customer pays an (i) hourly 'energy' charge, (ii) an hourly congestion charge, and (iii) an hourly losses charge. The hourly congestion charge may be positive or negative. Abstracting from losses, as we are in this paper, the hourly congestion credit or charge will be positive for generation and load located at nodes with an LMP greater than the benchmark energy price used for the decomposition. It will be negative for generation and load located at nodes with an LMP smaller than the benchmark price.

Table 3 illustrates this billing convention for the example displayed in Figure 1. Column [A] shows how LMPs and bills are decomposed when Bus A is chosen as the reference bus. As shown in row [2], the congestion component of the price (CLMP) at Bus A is zero. As can

[4]We put the term 'energy' in quotation marks here to distinguish this 'energy' component from the total credit, which in the nomenclature of U.S. electricity markets is also a payment for energy.

Table 3: Illustration of Bill Decomposition for Alternative Reference Buses.

			Reference Bus			
			Bus A	Gen Weighted	Load Weighted	Bus B
			[A]	[B]	[C}	[D]
[1]	Benchmark Energy Price	($/MWh)	10.00	10.71	12.14	15.00
[2]	$CLMP_A$	($/MWh)	0.00	-0.71	-2.14	-5.00
[3]	$CLMP_B$	($/MWh)	5.00	4.29	2.86	0.00
	Generation Credits at Bus A					
[4]	Energy	($)	3,000	3,214	3,643	4,500
[5]	Congestion	($)	0	-214	-643	-1,500
[6]	Total	($)	3,000	3,000	3,000	3,000
	Load Charges at Bus A					
[7]	Energy	($)	2,000	2,143	2,429	3,000
[8]	Congestion	($)	0	-143	-429	-1,000
[9]	Total	($)	2,000	2,000	2,000	2,000
	Net Charges at Bus A					
[10]	Energy	($)	-1,000	-1,071	-1,214	-1,500
[11]	Congestion	($)	0	71	214	500
[12]	Total	($)	-1,000	-1,000	-1,000	-1,000
	Generation Credits at Bus B					
[13]	Energy	($)	500	536	607	750
[14]	Congestion	($)	250	214	143	0
[15]	Total	($)	750	750	750	750
	Load Charges at Bus B					
[16]	Energy	($)	1,500	1,607	1,821	2,250
[17]	Congestion	($)	750	643	429	0
[18]	Total	($)	2,250	2,250	2,250	2,250
	Net Charges at Bus B					
[19]	Energy	($)	1,000	1,071	1,214	1,500
[20]	Congestion	($)	500	429	286	0
[21]	Total	($)	1,500	1,500	1,500	1,500
	Generation Credits Systemwide					
[22]	Energy	($)	3,500	3,750	4,250	5,250
[23]	Congestion	($)	250	0	-500	-1,500
[24]	Total	($)	3,750	3,750	3,750	3,750
	Load Charges Systemwide					
[25]	Energy	($)	3,500	3,750	4,250	5,250
[26]	Congestion	($)	750	500	0	-1,000
[27]	Total	($)	4,250	4,250	4,250	4,250
	Net Charges Systemwide					
[28]	Energy	($)	0	0	0	0
[29]	Congestion	($)	500	500	500	500
[30]	Total	($)	500	500	500	500

Corresponds to Example Shown in Figure 1 be seen in rows [4]-[12], generation credits and load charges at Bus A are entirely recorded as 'energy' charges. As shown in row [3], the congestion component at Bus B is $5/MWh, which is just the premium of the LMP at Bus B over the LMP at Bus A. As can be seen in rows [13]-[21], generation credits and load charges at Bus B are a combination of 'energy' credits and congestion credits. As shown in row [28], systemwide net 'energy' charges systemwide are zero. This is just reflects the fact that total generation equals total load, and when the same 'energy' price is used to book the 'energy' credits and charges for generation and load, the total net payments zero out. As shown in row [29], systemwide net congestion charges equal $500. As shown in row [30], systemwide net total charges equal $500, which is the merchandising surplus.

It is important to take note of the fact that the choice of a reference bus is, for most purposes, an arbitrary one. Any bus will do. The LMPs are fixed. Changing the reference bus does not change the LMPs, and so does not change any customer's total bill. However, changing the reference bus does change how each customer's total bill is decomposed into 'energy' credits or charges and congestion credits or charges.

This arbitrary element in this decomposition is shown in Table 3 by recalculating the credits and charges for different choices of the reference bus as we move across the columns. Column [B] uses a generation-weighted reference bus, column [C] a load-weighted reference bus, and column [D] uses Bus B as the reference bus. As shown in rows [1]-[3], the benchmark 'energy' price is gradually increasing as the weighting shifts towards Bus B, while at the same time, the congestion components are decreasing. Consequently, the decomposition of the bills to customers shifts. The total billings do not change–which can be seen in rows [6], [9], [15] and [18]–just the decomposition. As shown in rows [4] and [13], generators see their 'energy' credits increasing. In order for their total credits to remain fixed, their bills must also show decreasing congestion credits–which can be seen in rows [5] and [14]. As shown in rows [7] and [16], loads see their 'energy' charges increasing. In order for their total charges to remain fixed, their bills must also show decreasing congestion charges–which can be seen in rows [8] and [17].

Using a reference bus with a higher 'energy' price causes the 'energy' component of every individual bill to be larger and the congestion component to be smaller. However, as can be seen in rows [28] and [29], it has no impact whatsoever on the decomposition of the net charges systemwide. The net 'energy' charges always equal zero, regardless of the choice of reference bus. This is just a definitional fact: the benchmark energy price, which is used to calculate all 'energy' credits and charges in this decomposition, is the same at all buses, and total generation exactly equals total load. The net congestion charges always equal the congestion rent. This is because any increase in the 'energy' charge to load somewhere is offset in the net charge by an increase in the 'energy' credit to generation somewhere, and any decrease in the congestion charge to load somewhere is offset in the net charge by an increase in the congestion credit to generation somewhere.

While changing the reference bus does not change congestion rents at all, it does changes how congestion appears on individual bills and aggregations by subsets. Consequently, it is incorrect to take the congestion payment for any subset of the system as an attribution of the congestion rent paid by that subset. For example, consider a breakdown of congestion payments by node A and node B. This shows up in rows [11] and [20], and moving from column [A] to column [D], the net congestion charges appear to shift from Bus B to Bus A. For another example, consider the breakdown of congestion payments between generation and load. This shows up in rows [23] and [26]. Moving from column [A] to column [D] shows generation goes from initially receiving a congestion credit to paying congestion charges and shows load goes from initially paying congestion charges to receiving congestion credits.

A.3.2 Constraint-Based Attribution

A proper attribution of congestion rent paid by different loads is possible, however. As detailed in Monitoring Analytics (2020b), this requires (i) calculating congestion charges on a constraint-by-constraint basis, and (ii) using the upstream node of the constraint as the reference node when calculating congestion charges for that constraint. It will be useful here to detail the calculation using the notation from our earlier solution to the optimal power flow problem with which we defined congestion. Using the index k for constraints, equation (10), with which we defined aggregate congestion rent as the sum of the constraint-specific rent is rewritten now as:

$$R \equiv \sum_{k=1}^{2K} R_k = \sum_{k=1}^{2K} \mu_k^* F_k \geq 0. \tag{16}$$

For this constraint-based calculation, we use the shift factor (power transfer distribution factors, DFAX) for flow on line k from an injection of power at node i given the optimal dispatch g^*, denoted here by $\xi^*_{k,i}$. For any binding constraint, k, denote the upstream node, $m(k) \in \{1, ..., N\}$. We will use this upstream node as our benchmark price as we calculate the congestion charge at node i due to constraint k,

$$\Delta p_i^*(k) = \mu_k^* \left(\xi_{k,m(k)} - \xi_{k,i}\right). \tag{17}$$

The increment to demand charges at node i due to constraint k is $\Delta p_i^*(k) d_i$, and the aggregate increment to demand charges due to constraint k is $\sum_{j=1}^{N} \Delta p_j^*(k)\, d_j$. Only a portion of this increment is congestion rent, however. That fraction is $\mu_k^* F_k / \left(\sum_{j=1}^{N} \Delta p_j^*(k)\, d_j\right)$. So, the congestion rent paid at node i is,

$$R_{k,i} = \Delta p_i^*(k)\, d_i \, \frac{\mu_k^* F_k}{\sum_{j=1}^{N} \Delta p_j^*(k)\, d_j}. \tag{18}$$

We can reorganize this equation to express the congestion rent paid by load at node i as a weighted share of the total constraint-specific rent,

$$R_{k,i} = \frac{\Delta p_i^*(k)\, d_i}{\sum_{j=1}^{N} \Delta p_j^*(k)\, d_j} \, \mu_k^* F_k = w_{k,i} \, \mu_k^* F_k. \tag{19}$$

Note that $\forall k$, $0 \le w_{k,i} \le 1$, and $\sum_i^N w_{k,i} = 1$, which is why we call them weights. Total congestion rent paid by load at node i is the sum of these constraint-specific rent payments,

$$R_i = \sum_{k=1}^{2K} R_{k,i} = \sum_{k=1}^{2K} w_{k,i} R_k = \sum_{k=1}^{2K} w_{k,i} \, \mu_k^* F_k \ge 0. \tag{20}$$

We can illustrate this attribution methodology using an example originally detailed in Monitoring Analytics (2020b). Figure 3 shows the network for a twelve node, nineteen line example. For each line connecting two nodes, the figure gives the corresponding transmission limit which we assume to be the same in both directions. Table 4 details the market clearing results for a single hour. Table 5 details the attribution calculation.

This presentation of the attribution algorithm overlooks a couple of complexities arising from how locational marginal pricing is implemented in PJM and other organized wholesale markets. Most importantly, it assumes a single settlement market, where a complete attribution must also incorporate balancing congestion. For the full details on the attribution of balancing congestion, see Monitoring Analytics (2020b). Other complexities arise from special cases in the market clearing program. These special cases are discussed in the Independent Market Monitor's *2020 State of the Market Report for PJM*–see Monitoring

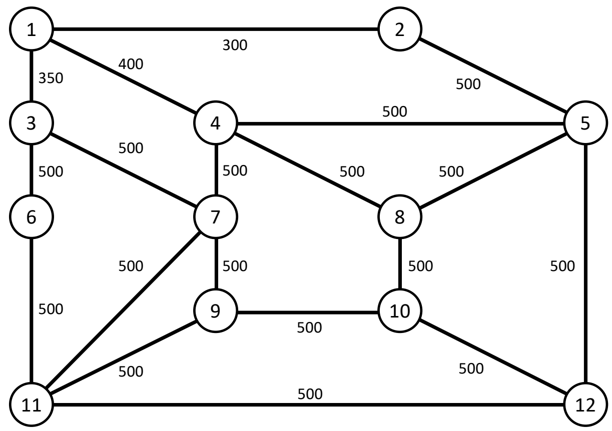

Figure 3: Example Network with Twelve Nodes and Nineteen Lines.

Source: Monitoring Analytics (2020).

Table 4: Market Clearing Solution for the Twelve Nodes and Nineteen Lines Example.

Node	LMP p_i^* ($/MWh)	Load d_i (MWh)	Gen g_i^* (MWh)	Payments: Load ($)	Payments: Gen ($)
1	16.24		450.0		7,308.74
2	14.12		500.0		7,060.41
3	17.60		0.0		0.00
4	17.00		165.4		2,811.33
5	12.00	100.0	531.7	1,200.00	6,380.58
6	18.00		582.9		10,492.43
7	18.57	200.0		3,714.13	
8	16.19	290.0		4,694.43	
9	19.45	180.0		3,501.55	
10	19.56	610.0		11,933.50	
11	20.23	350.0		7,078.91	
12	23.05	500.0		11,524.27	
Total		2,230.0	2,230.0	43,646.80	34,053.48
Surplus, S					9,593.32

Line k	Nodes I,j	Limit F_k (MW)	Flow q_{ij}^* (MWh)	Shadow Price μ_{ij}^* ($/MWh)	Constraint-Specific Congestion R_k ($)
1	1,2	300	-107.83		
2	1,3	350	229.55		
3	1,4	400	328.28		
4	2,5	500	392.17		
5	3,6	500	-82.91		
6	3,7	500	312.46		
7	4,5	500	0.00		
8	4,7	500	213.72		
9	4,8	500	279.94		
10	5,8	500	323.88		
11	5,12	500	500.00	(17.36)	8,678.54
12	6,11	500	500.00	(1.83)	914.78
13	7,9	500	221.55		
14	7,11	500	104.63		
15	8,10	500	313.82		
16	9,10	500	158.48		
17	9,11	500	-116.93		
18	10,12	500	-137.70		
19	11,12	500	137.70		
Total					9,593.32

Source: Monitoring Analytics (2020).

Table 5: Attribution Calculation for the Twelve Nodes and Nineteen Lines Example.

	Constraint k=11				Constraint k=12				
	$DFAX_{k,i}$	$\Delta p_{k,i}$	$w_{k,i}$	$R_{k,i}$	$DFAX_{k,i}$	$\Delta p_{k,i}$	$w_{k,i}$	$R_{k,i}$	R_i
Node		($/MWh)		($)		($/MWh)		($)	($)
[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]
1	0.0908	4.44	0.0%	0.00	(0.1170)	0.80	0.0%	0.00	0.00
2	0.2187	2.22	0.0%	0.00	(0.1705)	0.90	0.0%	0.00	0.00
3	0.0000	6.01	0.0%	0.00	0.0000	0.59	0.0%	0.00	0.00
4	0.0538	5.08	0.0%	0.00	(0.1804)	0.92	0.0%	0.00	0.00
5	0.3465	0.00	0.0%	0.00	(0.2240)	1.00	4.1%	37.88	37.88
6	(0.0566)	7.00	0.0%	0.00	0.3203	0.00	0.0%	0.00	0.00
7	(0.0343)	6.61	7.9%	686.73	(0.2034)	0.96	8.0%	72.89	759.62
8	0.1049	4.19	7.3%	631.85	(0.2209)	0.99	11.9%	109.24	741.09
9	(0.0777)	7.36	7.9%	688.55	(0.2737)	1.09	8.1%	74.41	762.96
10	(0.0856)	7.50	27.4%	2,377.16	(0.2584)	1.06	26.9%	245.69	2,622.85
11	(0.1131)	7.98	16.7%	1,450.82	(0.3593)	1.24	18.1%	165.55	1,616.37
12	(0.2841)	10.95	32.8%	2,843.44	(0.2806)	1.10	22.9%	209.10	3,052.54
Total				8,678.54				914.78	9,593.32

Source: Monitoring Analytics (2020).

Analytics (2021). Altogether, the congestion calculated for these special cases represents 1.3% of the total congestion.

PJM's Independent Market Monitor reports the attribution of congestion rent at the level of the Control Zone in its regular *State of the Market* reports. Table 6 shows this attribution for the years 2017-2020.

Table 6: PJM's Congestion Attribution by Control Zone, 2017-2020.

Control Zone	Total Congestion ($ million) 2017	2018	2019	2020	Coefficient of Variation
AECO	8.7	16.0	7.4	4.9	0.52
AEP	108.9	223.8	100.4	92.7	0.47
APS	27.8	76.5	41.8	36.2	0.47
ATSI	39.5	97.4	43.3	42.4	0.50
BGE	24.7	45.9	20.2	20.0	0.44
ComEd	156.9	163.2	80.3	66.7	0.43
DAY	11.3	26.5	11.4	10.1	0.53
DEOK	20.7	48.7	18.1	14.6	0.61
DLCO	6.9	16.2	6.4	6.1	0.55
Dominion	67.3	152.1	68.5	67.7	0.47
DPL	35.2	85.5	27.9	27.7	0.63
EKPC	9.9	23.4	8.9	7.7	0.59
EXT	-4.4	-3.8	-2.9	12.4	
JCPL	19.7	37.5	15.9	11.4	0.54
Met-Ed	15.0	29.9	13.0	13.0	0.46
OVEC			0.4	1.1	
PECO	33.7	59.6	23.1	18.1	0.55
PENELEC	14.6	29.9	15.6	13.5	0.42
Pepco	22.5	42.6	18.6	16.1	0.48
PPL	37.5	67.3	32.8	24.0	0.46
PSEG	39.8	69.4	31.4	21.7	0.51
RECO	1.4	2.1	1.0	0.8	0.43
Total	697.6	1,309.9	583.3	528.6	0.46

Source: Monitoring Analytics (2021), Table 11-17, Monitoring Analytics (2020a), Table 11-19, Monitoring Analytics (2019), Table 11-13.

References

Bohn, Roger E., Michael C. Caramanis, and Fred C. Schweppe (1984). "Optimal Pricing in Electrical Networks over Space and Time". In: *The RAND Journal of Economics* 15.3, p. 360. ISSN: 07416261. DOI: 10.2307/2555444.

Duenas-Martinez, Pedro, Karen Tapia-Ahumada, Joshua Hodge, Raanan Miller, and John E. Parsons (2021). "Challenges and Opportunities for Decarbonizing Power Systems in the US Midcontinent".

Kirschen, Daniel and Goran Strbac (2004). *Fundamentals of Power System Economics*. Wiley. DOI: 10.1002/0470020598.

Littlechild, Stephen C. and Carlos J. Skerk (2008). "Transmission expansion in Argentina 1: The origins of policy". In: *Energy Economics*. ISSN: 01409883. DOI: 10.1016/j.eneco.2007.12.003.

Meisel, Jerome (1993). "System Incremental Cost Calculations Using The Participation Factor Load-Flow Formulation". In: *IEEE Transactions on Power Systems*. ISSN: 15580679. DOI: 10.1109/59.221220.

Monitoring Analytics (2019). *2018 State of the Market Report for PJM, Volume 2: Detailed Analysis*. Tech. rep.

— (2020a). *2019 State of the Market Report for PJM, Volume 2: Detailed Analysis*. Tech. rep.

— (2020b). *Constraint Based Congestion Calculations: Measuring Congestion Paid by Zone*. Tech. rep. URL: https://www.pjm.com/-/media/committees-groups/task-forces/afmtf/2020/20200722/20200722-item-03a-constraint-based-congestion-calculations.ashx.

— (2021). *2020 State of the Market Report for PJM, Volume 2: Detailed Analysis*. Tech. rep.

Overbye, T. J. (2003). "Estimating the actual cost of transmission system congestion". In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003*. ISBN: 0769518745. DOI: 10.1109/HICSS.2003.1173857.

Philpott, Andy and Geoffrey Pritchard (2004). "Financial transmission rights in convex pool markets". In: *Operations Research Letters*. ISSN: 01676377. DOI: 10.1016/j.orl.2003.06.002.

Rivier, Michel and Ignacio J Pérez-Arriaga (1993). "Computation and decomposition of spot prices for transmission pricing". In: *11th PSC Conference*.

Schweppe, F.C., M. Caramanis, R.D. Tabors, and R.E. Bohn (1988). *Spot Pricing of Electricity*. Kluwer.

Wu, Felix, Pravin Varaiya, Pablo Spiller, and Shmuel Oren (1996). "Folk theorems on transmission access: Proofs and counterexamples". In: *Journal of Regulatory Economics* 10, pp. 5–23. ISSN: 0922680X. DOI: 10.1007/bf00133356.