

Understanding ICC Rate Structure Regulation: A Spatial Analysis

Kenneth D. Boyer

Published online: 16 July 2013
© Springer Science+Business Media New York 2013

Abstract A spatial equilibrium of a stylized railroad network is offered to illustrate that monopoly pricing requires prices on some links that are so high that traffic “does not move.” ICC price structure regulation, which was based on the idea that traffic should be priced to “move” in a privately owned and operated network, is modeled as maximizing market access subject to an AVC minimum and an aggregate break-even constraint in a network that is over-extended. Flows that result from such regulation are highly inefficient but provide more surplus to shippers and receivers than the unregulated price structure.

Keywords Railroads · Rate structure · Regulation · Spatial analysis

1 Introduction

Modern American railroads operate in an environment that was largely shaped by the Interstate Commerce Act of 1887. In order to better understand today’s freight industry, it is useful to review the history of railroad regulation—what problems regulation was intended to fix and how the Interstate Commerce Commission tried to correct those problems. This paper uses a simple stylized spatial model of a railroad network to show that monopoly pricing requires close attention to rate relationships.

It is not generally understood that the focus of railroad regulation in the late 1800s was not on the level but rather on the structure of railroad rates. To a modern economist, schooled in the idea that the problem of monopoly is that prices are set too high and that, as a result, quantities offered on the market by a monopoly are inefficiently low,

K. D. Boyer (✉)
Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA
e-mail: boyerkd@msu.edu

the preoccupation shown by the Interstate Commerce Act with the structure of railroad rates rather than their level seems incomprehensible.

In this paper, this focus on the structure of rates is traced to the fact that American railroads are organized as monopoly service providers over tracks that are owned by the railroad company. This vertical integration of infrastructure and operations gave nineteenth century railroads power that varied by commodity and location. The combination of ownership of the infrastructure and monopoly operations on it that gave the power to charge each origin–destination–commodity triad the profit maximizing rate—something that would have been impossible had the infrastructure owner not also been the monopoly operator. This paper demonstrates that profit maximization called for setting some rates at relatively high levels and others low (relative to marginal costs) rather than maintaining high prices to all customers, and more importantly, refusing to serve some origin–destination–commodity combinations. It was logical, then, that the control of market power was focused on the structure of rates rather than their level. This focus on the structure of rates is found whenever railroad track ownership and operations are within the same company. Had the nineteenth century technology permitted the practice of separating track ownership from carriage, with free entry of operators—as is the case in waterways and highways—railroad economics in North America would have developed quite differently from the way it did and it is unlikely that the 1887 Interstate Commerce Act would have put the type of limits on railroad behavior that it did.

When railroads first appeared in the 1830s, public policy toward the new technology was shaped by the common law on roads in which the fixed facilities were provided by a different party than the transporter. The first charters for railroads had assumed that the same common law would be applied to railroads as were applied to other modes of transportation. For example, the Boston and Providence railroad charter authorized the line “to erect toll houses, establish gates, appoint toll gatherers and demand toll upon the road” (Cleveland and Powell 1909, p. 160). But this organization of the railroad industry was soon seen as impractical—a disappointment since it was generally recognized that granting the monopoly right to operate over a track between two locations introduced market power issues that were not seen in other modes of transportation. Canal promoters in the first half of the nineteenth century pointed to the ability of having multiple operators on a canal as an advantage for public investment in that mode rather than railroads. But the technical advantage of railroads over road and canal transport was so large that public policy favored investment in railroads over other types of transportation as soon as the technology became economically viable.¹

All railroads built in the 80-year era of railroad building between the appearance of the Baltimore and Ohio railroad in 1830 and the beginning of World War I were constructed through the use of eminent domain. The benefit to the public welfare provided by railroads was thought to justify not only the taking of private property, but also the construction subsidies that were granted to the monopoly operators. The nineteenth

¹ In a controversial calculation that downplays the importance of user-costs of transportation, Fogel (1964) argue that the canal and rail costs were not substantially less than canal costs in the 1840s. Nonetheless, the rapid abandonment of canals in areas where railroads were built suggests that the two technologies were not close substitutes.

century technical impracticality of allowing more than one carrier to use the tracks that were built in the public interest then gave enormous power to that operator. With the current perspective of open access to highways, it is hard to appreciate that the new companies' ability to set rates to individual communities and vary them for particular shippers within those communities gave the operating companies market power that is rarely seen today. According to a contemporary source, "The power to make freight rates is the power to turn a wilderness into a city or a city into a wilderness" (Beale and Wyman 1915, p. 657). This power passed to the Interstate Commerce Commission under the Interstate Commerce Act. According to an observer in the early 20th century, "It is a common saying that it (the Interstate Commerce Commission) is the most powerful body in the world, administrative or otherwise, and when its authority is closely examined, there will probably be few dissenters from this view" (Fletcher 1923, p. 43).

Modern railroads have about \$60 billion of revenue in an economy with a GDP of \$15 Trillion (AAR 2012) and carry only one tenth as many tons as are moved on the highways (USDOT 2012). How could the regulators of such a small industry have such enormous power? A part of the answer, of course, is that the railroad industry, before the widespread use of trucking, constituted a much larger share of GDP. Combined passenger and freight revenues in 1914 were \$3 billion for US railroads against an estimated US GNP of \$38.6 Billion. But much more importantly, the economic power of railroads was exercised at a very local rather than national level—which was made possible by the monopoly that operators had over their own tracks.

A spatial model of transportation aids in comprehending how the public could have felt economically threatened by an industry that, even at the height of its influence, was a relatively small part of the economy. This paper uses an extremely simple stylized model of a freight network to illustrate how the control of the structure of freight rates to individual shippers and communities can be used to transfer social surplus to a monopoly operator. Section 2 of this paper shows that the control is exercised not simply by maintaining prices at a monopoly level, but by setting the structure of rates charged in the network to insure the profit-maximizing pairing of buyers and sellers.

A spatial model also helps to understand the destructive effect of excess railroad building in the nineteenth and early twentieth century as real resources were expended to respond to the manipulation in the cost of distance that was part of the monopoly rate structure. Section 3 shows that monopoly rate structures gave individual communities an incentive to build socially wasteful lines, which created excess capacity and exacerbated the need to find railroad revenues to pay for the uneconomically dense network that had far too little traffic to support profitable operations.

Sections 4 and 5 of this paper briefly describe the legal charge within the Interstate Commerce Act and how the Interstate Commerce Commission chose to interpret its mandate. Section 6 then shows the spatial equilibrium that results from implementing the regulated rate structure in the way that the ICC interpreted as its mandate—one that ignores the costs of particular traffic flows, but seeks to encourage market access for all sellers that can reach a buyer at a cost greater than or equal to average variable cost on a link, while covering all movement and track costs in the over-extended system through revenues from all links in the network. It is not surprising that, without efficiency as a criterion for setting freight rates, the resulting equilibrium contains considerable

inefficiencies. But at least in the stylized example used in this paper, all shippers and receivers are better off than in the monopoly equilibrium despite the substantially higher social costs of transportation. And in the regulated equilibrium, the railroad benefits from rates that are set at a level to allow it to cover its cost, despite excess capacity—a situation that would not have been possible in a competitive market for rail services.

2 Railroad Monopolies Exercised Power Through the Structure of Prices

Modern economists are bemused by the intense concern with the structure of freight rates rather than their level in the century after railroads appeared in North America. While issues of price discrimination have long been part of the economic analysis of market power, the issue has long been seen as something of a side issue—a problem of the distribution of the burden of monopoly rather than the fundamental problem of the restriction of output that causes the static inefficiency of monopoly outcomes.

But this perspective that dismisses a concern with the structure of rates misunderstands how monopoly power can be exercised by railroads. It is not the case that railroads simply maintain high prices and thereby restrict output. Rather, if a railroad company is the sole seller of transportation services to a locality, the carrier will rationally quote separate prices for separate services to each individual based on a sizing-up an individual's ability to pay. Moreover, since the railroad also carries competitors' products to a destination market—perhaps from different geographic locations—it can manipulate the ability to pay by an individual shipper by setting the level of rates charged to competitors. In the nineteenth century, it was recognized that it is equally destructive to a business to charge it a high rate as it is to charge a low rate to its competitors. Thus market power is effectively exercised by the manipulation of the structure of rates rather than the setting of the overall level of rates. The regulation of minimum rates that was part of railroad regulation, while often assumed to facilitate cartel formation (Hilton 1966) is not necessarily evidence that ICC policies were designed to advance railroad interests; minimum rate regulation can be seen as a response to market power that was exercised by the manipulation of the spectrum of prices that were charged to different customers.

In order to understand why the structure of railroad rates was the focus of resentment against railroad companies and thus the focus of attention of the Interstate Commerce Commission when it was created at the end of the nineteenth century, it is useful to use a spatial model of transportation demand in the spirit of Samuelson (1952) and Takayama and Judge (1971).² This paper's calculations illustrating the exercise and control of railroad market power are based on the simple stylized network that is discussed below, but the basic ideas hold for any similar model of geographic competition in which different supply and demand locations are differentially affected by decisions of the carrier and in which arbitrage of atomistic agents in both demand as well as supply locations assure that prices between demand and supply areas differ by the full cost (rate charged plus user costs) of transportation. In the succeeding sections, the model

² For an introduction to the techniques of spatial equilibria, see Anderson and Wilson (2007).

is used to illustrate the market power problems that the Interstate Commerce Act was charged to deal with, some solutions that the Interstate Commerce Commission came up with, and the logical effect of those policies.

Railroads do not exist in a featureless plane, but rather connect localities that have their own geographic positions, sizes, strengths, and weaknesses. It is important for understanding the exercise of railroad market power that locations not be represented as symmetric and uniform, but instead be somewhat randomly distributed, with the railroad connecting places that trade with one another due to, for example, different endowments of non-appropriable resources, which is the traditional source of Ricardian rents. The amount of fertile land or the quality of ore deposits are easily placed in this framework, but in the short run an investment in plant and equipment made in one community rather than another generate differences that look like the equivalent of land where fertility varies with locations.

Figure 1 shows five supply origins and two demand destinations distributed at a few integer locations in a Cartesian coordinate system. Below the figure are the basic data about locations and the parameters of supply curves in the numbered cities and demand curves in the cities labeled with letters. There are a large number of demanders in cities A and B and a large number of producers in the numbered locations so that sellers and buyers act as price takers. City A sits at the origin. Supply location 1 is 1 unit above A, supply location 2 is 1 unit to the right of A. As is indicated in Fig. 1, production costs at location 1 start average and rise slowly. Those at 3 start high and rise quickly. To keep the analysis simple, the analysis ignores the fact that railroads were built to carry both passengers as well as freight. It also ignores the issue of backhauls and the fact that railroads carry multiple commodities. The data in Fig. 1 are used to compare the spatial equilibria generated under various policy assumptions.

Railroads are built to connect origins and destinations. (Again, I ignore the fact that origin cities can be destinations for other commodities or for passengers.) How many railroads will be built and their locations will depend on balancing the fixed cost of building and maintaining track against the costs of hauling goods that are incurred by the railroad operator and its customers. It is important to recognize that all transportation also involves costs that are borne by the purchasers of transportation services and not paid to the operator. The time costs of travel is the easiest of these costs to identify, but there are others—notably the degradation of cargo quality due to the time or distance of haulage. If we use the parameters in Fig. 1 and assume that it costs \$10 per unit of distance to build and maintain a railroad line and \$2 per unit of haulage-distance as movement cost, and we assume that un-charged user costs of transportation are half of the variable costs of movement (\$1 per ton-mile or equivalent), then it can be shown that the efficiently structured network is the one shown in Fig. 1 in which a single main line connects destinations A and B, with spurs connecting origins 1, 5, and 4 to the main line. Given the level and location of demands and supplies in Table 1, the movement cost savings of adding additional links within the network do not outweigh the additional fixed costs of maintaining a denser network of tracks.

The disadvantage of maintaining the slimmed-down network shown in Fig. 1 is the circuitry of some routes. For example, location 4 moves its freight first to the main line between A and B and then to the final destination rather than making a direct link. It is important to recognize that the data in Table 1 implicitly assumes the absence

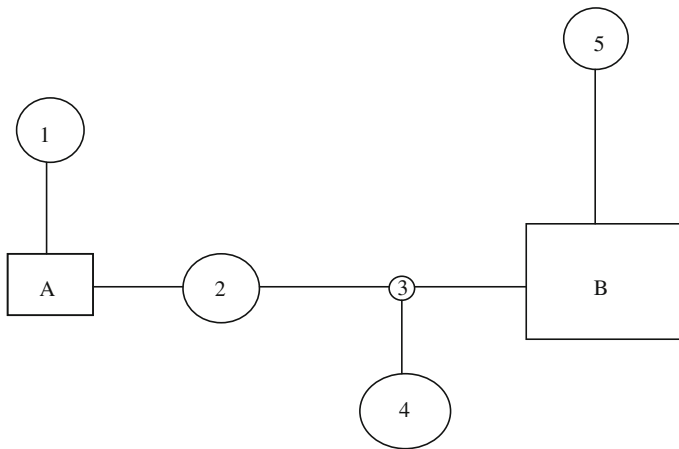


Fig. 1 An efficient rail network for connecting origins and destinations. *Supply locations and supply curves* Location of City 1: (0,1); Supply curve at City 1: $P_s = 42 + .5 * Q_s$. Location of City 2: (1,0); Supply curve at City 2: $P_s = 32 + 1 * Q_s$. Location of City 3: (2,0); Supply curve at City 3: $P_s = 45 + 4 * Q_s$. Location of City 4: (2,-1); Supply curve at City 4: $P_s = 30 + 2 * Q_s$. Location of City 5: (3,2); Supply curve at City 5: $P_s = 42 + .5 * Q_s$. *Demand locations and demand curves* Location of City A: (0,0); Demand curve at City A: $P_d = 70 - Q_d$. Location of City B: (3,0); Demand curve at City B: $P_d = 120 - 2 * Q_d$. *Transportation Cost Parameters* Average Variable Carrier Cost per unit of commodity and distance: 2. Costs of Right of Way per unit of distance: 10. User Costs of Transportation per unit of commodity and distance: 1

of binding capacity constraints that would cause movement and user costs to rise as traffic levels rose. If traffic is affected by capacity constraints, either user costs or movement costs (or both) per unit will rise with traffic levels. By assuming that user and movement costs are constant, Fig. 1 guarantees economies of density for the system.

With the network and parameters of Fig. 1, the efficient level of flows is given in Table 1. This is a spatial equilibrium that maximizes total benefit (the area under the demand curves for the amount of commodity delivered), minus total production cost (the area under the supply curve for each supply location up to the amount supplied), minus the variable costs of transportation borne by the railroad, minus the user costs of transport. The efficient spatial equilibrium is characterized by the supply price at each origin, plus rate costs plus user costs of transportation per unit, being equal to the demand price at each destination to which it ships. If this condition does not hold, it is possible to alter the flow on one link or another to increase social welfare.

Table 1 computes the demand price of the commodity at each destination based on the total amount received and the supply price at each origin by the total amount produced there. In an efficient equilibrium, each city has a well-defined hinterland supplying it. In the case of homogeneous commodities, the supply areas around each demand location are especially starkly defined. Since cross-hauling is inefficient, each origin has a unique destination that is assigned to it in the efficient equilibrium. The exception is location 2, which divides its production between the destinations. Equilibrium requires that suppliers in location 2 be indifferent between sending production to city A or city B. It costs \$6 to ship to location B (\$4 of which is paid to the railroad

Table 1 Efficient spatial equilibrium for network described in Fig. 1

	To A Quantity	To B Quantity	Quantity sent	Supply price	Surplus	To A Rate	To B Rate
From 1	<i>12.4</i>	<i>0.0</i>	12.4	48.2	38.5	2.0	
From 2	<i>6.4</i>	<i>9.8</i>	16.2	48.2	131.3	2.0	4.0
From 3	<i>0.0</i>	<i>1.6</i>	1.6	51.2	4.8		2.0
From 4	<i>0.0</i>	<i>9.1</i>	9.1	48.2	82.9		4.0
From 5	<i>0.0</i>	<i>12.4</i>	12.4	48.2	38.5		4.0
Quantity received	18.8	32.9					
Demand price	51.2	54.2					
Surplus	176.6	1,082.2					
Social surplus	1,484.8						
Railroad profit	-70.0						
Variable transport costs	166.1						
Fixed transport costs	70						

Source Rail quantities in italics of the table are summed horizontally to find shipments from each supply area and are summed vertically to find quantities received in each demand area. Prices and surplus are computed from supply and demand curves in Fig. 1. Social surplus is calculated as the area under both demand curves minus the area under all five supply curves minus the variable cost borne by the railroad minus the user costs borne by the shipper. Quantities in italics, representing shipment quantities from the location on the left to the destination at the top of the column, are determined by numeric optimization to maximize social surplus subject to the condition that no negative quantities can be shipped. Rail rates are calculated as the difference between the demand price and the supply price for each pair of origin and destination for which there is positive movement less the user cost of transportation. Rail profit is calculated as the sum of rates multiplied by quantity shipped on each route less the sum of variable and fixed costs of rail transport

and \$2 of which is paid by the shipper but not to the railroad) so the \$54.20 price in B is \$6.00 higher than the \$48.20 price in origin 2. Similarly, it costs \$3 to ship from origin 2 to location A, so the demand price there is \$3 greater than in supply node 2. It then follows that commodity prices in city B are \$3 higher than in City A. The possibility of arbitrage of homogeneous goods requires that supply prices for each supply location must differ from the demand price in the city that it supplies by the full price (movement cost plus user cost) of transportation between them and the price differential must be less than the cost of transportation for any link on which there are zero flows.

The fixed costs of the railroad system, \$70, which is calculated as \$10 per unit of distance for each of the 7 units of track that are constructed, can easily be paid out of the surplus in the form of lump sum levies on suppliers, demanders, or a combination. If the fixed cost is to be paid by recomputing the spatial equilibrium to maximize social surplus so that the revenue from transportation is equal to the total cost of operating the system—what Baumol and Bradford (1970) named Ramsey Pricing—the result is shown in Table 2. Since fixed costs are relatively small in the efficiently configured network, the Ramsey prices in Table 2 are close to the efficient prices, with the difference between them identified as the Ramsey allocation of track costs to different flows. It should be noted that there seems to be nothing “fair” about the

Table 2 Second-best efficient spatial equilibrium for network described in Fig. 1

	To A Quantity	To B Quantity	Quantity sent	Supply price	Surplus	To A Rate	To B Rate
From 1	<i>12.1</i>	<i>0.0</i>	12.1	48.1	36.8	2.6	
From 2	<i>6.2</i>	<i>9.6</i>	15.8	47.8	125.3	2.8	5.9
From 3	<i>0.0</i>	<i>1.516</i>	1.5	51.1	4.6		3.7
From 4	<i>0.0</i>	<i>8.9</i>	8.9	47.8	79.1		5.9
From 5	<i>0.0</i>	<i>12.1</i>	12.1	48.1	36.8		5.7
Total received	18.4	32.1					
Demand price	51.6	55.7					
Surplus	168.5	1,032.9					
Social surplus	1,484.0						
Railroad profit	0						
Variable transport costs	162.2						
Fixed transport costs	70						

Source See description of calculations under Table 1. Shipment quantities in italics of the table are determined by numeric optimization to maximize social surplus subject to the condition that no negative quantities can be shipped and that the sum of rail revenue be equal to the sum of the fixed and variable costs of rail transportation. Rail rates are calculated as the difference between the demand price and the supply price for each pair of origin and destination for which there is positive movement less the user cost of transportation. Rail profit is calculated as the sum of rates multiplied by quantity for each route less the sum of variable and fixed costs of rail transport

allocation of the fixed costs back to the participants that derives from the exercise of finding second-best prices. Origin 5, which requires more miles of track to reach its destination than Origin 1 nonetheless sees a smaller reduction in surplus when fixed costs are paid for through a surcharge on freight prices despite the fact that it is in almost all ways similar to Origin 1. Destination B bears by far the largest share of the surplus reduction due to the fact that the demand curve in B is particularly high and steep.

While the efficient equilibrium maximizes the sum of consumers' and producers' surplus, it should be noted that there is a highly uneven distribution of surplus across the different locations. City B has much more surplus per unit consumed than A. Supply centers 2 and 4, with their low production costs, have much more surplus per unit shipped than the other cities. Part of this is the result of the random draw of supply and demand characteristics, but part is also due to geographic locations. Thus locations 2 and 3 are especially well positioned on the main line between the two demand locations. Prices are higher in location B than in A since A's hinterland is geographically smaller, requiring less transportation than B's hinterland. The marginal unit is shipped 1 mile to A, but 2 miles to B. The amount by which prices in B exceed A is the cost of moving a unit of production 1 mile.

Despite the fact that the benefits of the efficient system are unevenly distributed, the efficient spatial equilibrium is a useful benchmark for rate policies. This equilibrium, which is the result of charging prices equal to marginal costs for an efficiently-configured system, has no empirical example in the history of US railroad development.

Table 3 Monopoly priced spatial equilibrium for network described in Fig. 1

	To A Quantity	To B Quantity	Quantity sent	Supply price	Surplus	To A Rate	To B Rate
From 1	6.2	<i>0.0</i>	6.2	45.1	9.6	14.5	
From 2	3.2	<i>4.9</i>	8.1	40.1	32.8	19.5	45.0
From 3	<i>0.0</i>	<i>0.8</i>	0.8	48.1	1.2		38.0
From 4	<i>0.0</i>	<i>4.6</i>	4.6	39.1	20.7		46.0
From 5	<i>0.0</i>	<i>6.2</i>	6.2	45.1	9.6		40.0
Total received	9.4	16.4					
Demand price	60.6	87.1					
Surplus	44.1	270.5					
Social surplus	1,096.1						
Railroad profit	707.4						
Variable transport costs	83.0						
Fixed transport costs	70						

Source See description of calculations under Table 1. Shipment quantities in italics of the table are determined by numeric optimization to maximize railroad profit subject to the constraint that quantities moved from origin to destination be nonnegative. Railroad profit is calculated as the quantity shipped from an origin to a destination multiplied by the rate less the full cost (rate cost + user cost) of the movement minus fixed costs. The rail rate is calculated as the difference between the demand price at the destination and the supply price at the origin less the user cost of the transportation

Unlike highways, airlines, or waterways, there is no entry of new operators on existing infrastructure that would enforce cost-based pricing.

A profit maximizing monopoly operator of the core network shown in Fig. 1 would not charge rates equal to transportation costs. Table 3 shows the result of choosing flows to maximize railroad profit. The profit maximizing flows are found by numerical optimization subject to the condition that flows be non-negative and that the demand price and supply price differ by the full price of transportation (rate cost plus user cost) when there is a positive flow. The pattern of flows is similar, though not identical (since monopoly rates depend on demand elasticities in addition to costs) to those in Tables 1 and 2, though at considerably lower quantities. Rail rates for all cities are much higher than in the equilibria based on efficient allocations of traffic. As a result, supply prices are lower, demand prices are higher, and total surplus is diminished by deadweight loss. The social surplus is now shared between buyers, sellers, and the operator of the railroad with the railroad getting by far the largest portion. A decomposition relative to the second best efficient equilibrium shows that the railroad increases profit from 0 to 707.4; suppliers’ surplus is diminished from 282.6 to 74.0; consumers’ surplus is diminished from 1,201.4 to 314.7; and deadweight loss increases from 0.8 to 388.7.

In order to prevent traffic from moving when it is not in the profit-maximizing equilibrium, the monopoly will charge rates that makes the full prices for unused links exceed the difference between the demand price in A or B and the supply prices in any of the supply locations. Since these prices are not paid, they are not defined and thus not shown in Table 3. Relative to Table 2, the railroad rates in Table 3 are several times

higher, and the quoted prices on unused links—those designed to prevent movements that are not in the profit maximizing rate structure—must be an even higher multiple of costs.

In the efficient spatial equilibrium, it is uneconomic for location 1 to do business with city B. The distance is so great that, given the cost of transportation, a shipper in location 1 could not profitably pay the cost of transportation to city B and sell there at the going price. Geographically, city B is not in location 1's hinterland since B has other cheaper and closer suppliers that can supply its needs.

Once the transportation network is governed by principles of monopoly pricing, it is no longer the case that market forces prevent sellers in city 1 from doing business with customers at B. Rather it is the rates charged by the railroad for shipping from 1 to B that is the problem. As part of the profit maximizing rate structure, the railroad keeps the rates from 1 to B high enough to prevent the flow of commodities between the two locations; this enhances the profitability of its own movements from cities 3, 4, and 5 to B. It makes no money on shipments between 1 and B because it chooses not to make the rate low enough for traffic to move between the two locations. The reward to the railroad is its ability to charge higher prices for the traffic that it does choose to serve.

3 Railroad Building in Response to Pricing that Does Not Let Traffic Move

It is worth belaboring the point that standard market analysis is not helpful to understanding the issue of monopoly pricing on link 1-B. The flow from 1 to B seems to be a market. It is clearly different from shipments from 1 to A, and a shipper in 1 will not be indifferent if the product placed on the rail car suddenly shows up in a different place. But rather than charging the apparent monopoly price, determined by the familiar Lerner markup formula $(P - MC)/P = 1/\epsilon$ where ϵ is the elasticity of demand for transportation services, the monopoly operator will refuse to provide service—perhaps by charging a prohibitively high rate—thereby forgoing apparently profitable sales. The key is that, since the railroad is both a monopoly infrastructure owner and monopoly operator, it can set different prices on each link. Prices set on one link affect the yield on other links, so the profit maximizing strategy is to refuse service that would be apparently be profitable to provide in order to increase the profitability on other links in the network. A monopoly infrastructure operator without having a monopoly on operations, would not be able to use this strategy. When infrastructure is owned by the monopoly operator, monopoly power is not exercised by setting an optimal price-cost markup for each link but rather by adjusting the structure of prices on all links. Railroad prices set so high that no shipments are made under them is an essential part of monopoly pricing in railroad networks since it is through these prohibitively high prices that a monopoly carrier can maintain a spatial equilibrium that restricts traffic on each link to the profit maximizing level. This practice of pricing so that traffic did not move was seen in the nineteenth century as the fundamental problem of railroad monopolies.

Before the introduction of regulatory oversight of the rate structure, the first response of communities affected by the exercise of railroad market power was to

Table 4 Monopoly priced spatial equilibrium for network described in Fig. 1 with a new diagonal link between Origin 1 and Destination B

	To A Quantity	To B Quantity	Quantity sent	Supply price	Surplus	To A Rate	To B Rate
From 1	<i>10.4</i>	<i>7.4</i>	17.9	50.9	79.8	7.6	
From 2	<i>0.0</i>	8.8	8.8	40.8	38.9		17.6
From 3	<i>0.0</i>	<i>1.0</i>	1.0	48.8	1.8		10.6
From 4	<i>0.0</i>	4.9	4.9	39.8	24.1		18.6
From 5	<i>0.0</i>	7.6	7.6	45.8	14.6		12.6
Total received	10.4	29.8					
Demand price	59.6	60.4					
Surplus	54.3	887.5	941.7				
Social surplus	1,355.6						
Railroad profit	254.5						
Variable transport costs	108.3						
Core railroad fixed costs	70						

Source See description of calculations under Table 1. Shipment quantities in italics are determined by numeric optimization to maximize railroad profit subject to non-negative quantity flows and the constraint that the difference between the supply price at 1 and the demand price at B can be no greater than the full movement cost between them using the diagonal route with distance 3.16. Railroad profit is calculated as the sum of quantity shipped from an origin to a destination multiplied by the rail rate minus fixed and variable costs. Rail rates are calculated as the difference between the demand price and the supply price for each pair of origin and destination for which there is positive movement less the user cost of transportation

build independent railroads. The monopoly pricing scheme in Table 3 distorts the price of distance and in the process creates opportunities for railroad building that would not exist in the efficient spatial equilibrium. For example, supply center 1 can see that it would be profitable to build a diagonal line to connect it directly to destination B, rather than using the efficiently structured system that is monopoly priced. A diagonal line is shorter, with lower movement cost and lower user cost than using the spur and the core system pictured in Fig. 1. While the diagonal line is not part of the system that maximizes the surplus of the entire community, it is in the private interest of city 1 to build it once the system’s flows are determined by maximizing the railroad profit rather than maximizing social surplus.

After the diagonal line is built by a competing railroad, the entire structure of rates charged by the railroad operating the core system of main line and spurs will need to be recalibrated. For example, the core railroad can no longer charge the old monopoly price between location 1 and A since that price was based on 1’s having no preferable alternative. Table 4 shows the effects on the monopoly pricing structure of the introduction of the additional constraint that the rate that is charged between 1 and B must equal the full movement cost between these two cities using the diagonal link. It is worthwhile tracing through the effects of this change in some detail to see the gains and losses from building the new line.

Building one single link outside of the monopoly railroad changes the entire structure of the profit maximizing rates. With the rate between 1 and B now equal to the

movement cost, location B now has a much lower-cost source of supply in location 1. Similarly supply center 1 can now deliver to B more profitably than to the demand center (City A) to which it is closest. In response, all rates must be lowered throughout the system in order for the incumbent railroad to retain its business. Prices fall at both destinations, and supply prices rise at all production centers as output, consumption, and shipments increase throughout the system.

Economically, supply center 1, with its cheap link to B, is now geographically better positioned than location 2, so it provides the marginal units to each of the two destinations, taking over that role from location 2. Both destination cities' surplus increases, as does surplus at each of the supply locations, despite the fact that the new link is used only to ship from origin 1 to destination B. Rail profits, not surprisingly, shrink when the diagonal link is completed between 1 and B.

But the big winner from opening the new link between 1 and B is location 1. Its surplus in the second best equilibrium had been 36.8, but was shrunk in monopoly pricing to 9.6. But with the opening of the new link, its surplus swells to 79.8. It is now better off than in the efficient equilibrium. Economic interests in city 1 can easily pay for the building of the new link out of the increase in surplus. The new link is inefficient in the sense that it would not be part of a system designed to maximize social surplus for the entire group of supply and demand locations. That is, the savings in variable costs due to the shorter diagonal route is not sufficient to offset the increase in fixed costs due to having another track. But railroad building in this case is not about efficient transportation but rather a way of taking advantage of the distortion in the cost of distance that monopoly pricing has created. Through railroad building, location 1 has managed to alter geography in its favor, making it an apparently low cost supply center.

But there is nothing about location 1 that suggests that it alone could benefit from building a new link. Clearly locations 4 and 5 could also gain from having more direct routes to A and B, taking advantage of the relatively high delivered prices that railroad monopoly pricing has created to justify building infrastructure that is not part of the efficient capital structure. Simple self interest will cause all of the diagonal routes to be built, thereby generating excess capacity throughout the system.

Throughout the period of railroad building, it was clear that localities needed to have multiple railroads in order to protect themselves from monopoly pricing.³ From the perspective of the twenty first century, railroad price discrimination is often considered as having primarily distributional effects, the evaluation of which require the use of controversial concepts of fairness rather than the well-defined concepts of efficiency. But in fact uncontrolled monopoly pricing had profound real costs in the form of the placement of excessive quantities of rail infrastructure to combat monopoly behavior.

Throughout the nineteenth and early twentieth centuries, construction companies, which made money through building rather than operating railroads, were happy to provide services to communities that were willing to provide capital to build new rail links. They also were happy to receive contracts for railroad building from existing railroad systems that wished to build lines to gain strategic advantages over their

³ In the United States, charters were granted freely, unlike most nations overseas where a more systematic attempt to control excess building was undertaken. See [Hilton \(1966\)](#).

rivals—investments that were designed to diminish the profitability of another railroad rather than make money from operating the new lines. There is no evidence of limit pricing to prevent railroad building, but many examples of rate wars to weaken new railroads in the hope of making them fail. Railroad building was an essential element of the nineteenth century rate wars that resulted from unstable railroad competition and attempts to control it.

It should be recognized as well that, while monopoly power was exercised primarily by setting freight rates, passenger train miles exceeded freight train miles well into the twentieth century. Passenger traffic generally has a much higher user cost of transportation than does freight, thus placing a much higher value on reducing miles of movement. Thus an inefficiently dense network of railroads that was the result of the excess railroad building through World War I, provided some benefits for passengers on the diagonal routes of the model in Fig. 1. Of course, when passengers began to disappear from railroad stations in the 1920s, the problems of overcapacity were even more apparent.

4 The Charge to Regulators

There is no stable competitive equilibrium when multiple sellers offer a homogeneous product and an industry has excess capacity. As noted in the previous section, the excess capacity in the freight market is implicit in the assumption of constant movement costs, but is also traceable to communities' attempts to escape from railroad monopoly power; to communities' attempts to gain a geographic advantage over their neighbors; to motives for building railroads that were beyond the expected profits of operating the extended system; and to the fact that capacity was put in place to serve passengers as well as freight.

The rate wars in the nineteenth century that were the logical result of the excess capacity in the rail freight system have been extensively studied; see, for example, Porter (1983). In the era before antitrust laws, railroads used revenue pooling to maintain a collusive equilibrium at high rates (Hilton 1966). This pooling of course defeated the advantage in Table 3 of having an alternative route to a destination and allowed the industry as a whole to reestablish the monopoly rate structure; pooling was thus widely condemned by shippers. It is then no surprise that railroad building was related to the success of pooling (Harley 1982). But the fundamental problem remained excess capacity combined with the ability of railroads to tailor prices to individual communities if there was no independent link serving it.

The Interstate Commerce Act promised benefits both to railroads, through inhibiting rate wars that were the natural result of over-capacity, and to the communities that were vulnerable to the exercise of market power.⁴ Excess capacity increased the urgency of finding revenues to cover the cost of the infrastructure that was owned by the

⁴ Authors on the left, for example, Kolko (1970) and those like MacAvoy (1965) on the right have interpreted the fact that railroads recognized the commercial advantages provided to them by regulation as evidence of regulatory capture. This paper, by contrast, emphasizes that more literal readings of regulatory decisions are capable of explaining how regulation could have benefitted the railroad industry in ways other than facilitating cartel coordination.

railroad companies. The Interstate Commerce Act gave the promise to communities of protection against this search for revenue as well as offering railroad companies a solution to the lack of profitability that resulted from overcapacity.

The Interstate Commerce Act can be thought of as a continuing series of legislative rules and clarifications that began with the named Act in 1887 and that were placed in final form by the Transportation Act of 1920. There were seven key sections of the law:

1. All rates shall be just and reasonable.
2. Personal discrimination is unlawful: this outlawed charging different prices for “like and contemporaneous service in the transportation of a like kind of traffic under substantially similar circumstances and condition.”
3. It is unlawful for any common carrier to give any undue or unreasonable preference or advantage to any particular person, company, or firm; it is also unlawful to grant any undue or unreasonable prejudice or disadvantage in any respect whatsoever.
4. Long-haul short-haul discrimination is outlawed.
5. Pooling is prohibited.
6. Rates must be published.
7. Interstate Commerce Commission was established to carry out the previous six sections of the law.

5 How the ICC Interpreted its Charge

The Interstate Commerce Commission was the first of the independent regulatory commissions and is recognized as a developer of the American institution of administrative law. Through the regulatory decisions, the ICC provided substance to what the law required.

It is well-known to economists that a rate that is set equal to marginal cost will lead to an efficient allocation of resources (as was done in Table 1) but (if marginal costs are constant) will not cover the fixed cost of an overextended system and that a rate set by the formula $(P-MC)/P = k/\epsilon$ (the so-called Ramsey Price structure) will cover system cost with minimum deadweight loss in the case of independent demands and costs. (Note however that this formula, based on independent demands for each link, will not generate the prices in Table 2 with the extensive cross elasticities of demand among all flows).

But there is no evidence that the ICC interpreted the criteria that it was to use to pass on rates to be simple efficiency or second-best efficiency. In fact, there was nothing in the list of the provisions of the Act that suggested that efficiency should be among the factors considered when setting rates. For example, while efficiency requires knowledge of the cost of individual shipments between specific endpoints, the ICC did not attempt to make this measurement. The cost to the railroad of serving a customer was simply not part of the criteria that the ICC saw itself as charged to use in setting rates.

It should also be noted that cost accounting in the nineteenth century did not permit the isolation of the marginal cost of a particular movement. At best, a calculation could

be made of the average variable cost of all movements system-wide, which might be quite far from the costs on any particular line.

Sections 2, 3, and 4 of the Interstate Commerce Act all ask the ICC to compare the rates charged to different shippers. And even Sect. 1, requiring that rates be just and reasonable, was based on common law, which in turn judged the justness and reasonableness of rates by making comparison with rates that were charged for similar shipping situations. Section 3 seems to be the one that is clearest: A shipper's competitors' rates should not give it an unreasonable advantage, nor should the shipper's own rates give it an unreasonable advantage. A glance at the monopoly prices in Table 3 shows many instances in which one community was charged higher rates than another: Location 2 is charged more than Location 1 to ship to City A, for example, despite the fact that the shipping distance is the same. All of the non-used links in the monopoly equilibrium must have arbitrarily high rates to prevent them from undercutting sellers that are part of the monopoly price structure. This will appear to the excluded shipper to be undue advantages that are granted to those shippers that are allowed to compete in the destination market. In the monopoly equilibrium, the railroad chooses which sellers to be allowed access to which markets and blocks the rest with rates so high that the profit maximizing structure of flows is maintained.

The ICC, without the ability to use the costs of particular movements to set rates, but charged to insure that rates were just, reasonable, and non-discriminatory and that gave no undue advantage or disadvantage to any particular users of the system, saw its job as trying to make the railroad system like the system of highways of today, in which shippers compete with one another without the intermediary of a separate company that determined which supplier was paired with which customer. Under monopoly railroad pricing, traffic could not be profitably delivered to destinations outside of the monopoly spatial equilibrium because the rates on those movements were too high relative to the rates charged to other suppliers. The U.S. Supreme Court, in its approval of a decision of the ICC, declared that rates should be set so that "the conditions which at any given time prevail in our several industries should be considered in so far as it is legally possible to do so, to the end that commodities may freely move."⁵ In another case approving a decision of the ICC, the Court held that a rate that was set to meet competition with a water carriers must be reasonably compensatory, meaning that it "(1) cover and more than cover the extra or additional expenses incurred in handling the traffic to which it applies; (2) be no lower than necessary to meet the existing competition; (3) not be so low as to threaten the extinction of legitimate competition by water carriers; and (4) not impose undue burden on other traffic or jeopardize the appropriate return on the value of carrier property."⁶ All of these criteria can be seen as approving of rates as low as necessary to allow traffic to access as many markets as is economically feasible, subject to a variable cost threshold, and that will allow existing carriers to survive.

The ICC wished to remove from the railroads the ability to use their rate-setting power to create winners and losers in competition, trying to set the railroad network

⁵ *Ann Arbor Railroad Co. v. United States*, U.S. 658, 667 (1930).

⁶ *Transcontinental Case*, 74 ICC 48 (1922).

up as a neutral structure over which commerce would flow. The ICC believed that the railroad system should support competition for markets by different sellers. The commission suspected that by favoring strong sellers over weak ones, railroads were encouraging the concentration of production in the nation into fewer hands. By allowing traffic to move over the railroad system, the ICC saw itself as promoting competition throughout American industry. According to the ICC, “Rates should be a medium for effecting the movement of commerce from one point to another and no carrier has the right, by erecting barriers of prohibitive rates, to restrict the sources from which a consuming point may supply its needs.”⁷

One device for doing this was the oversight of classification of commodities, insuring that analogous articles were treated analogously.⁸ This led the ICC to ensure a proper rate relationship between intermediate goods and finished products—for example, between building stone that had been cut near the quarry and stone that was shipped uncut, thereby allowing both to compete with each other in destination markets.⁹ If two goods did not compete with each other, then the rate comparison between them need not be made since lower rates on one did not inhibit the movement of a product with higher rates. This allowed the rates for manufactured goods to be set independently of the rates on grain and bulk goods, for example—except that the aggregate revenues from all transportation services needed to be large enough to cover the costs of the system. Since bulk goods sales were more price sensitive than manufactured goods, the ICC was happy to allow higher rates on the latter since they would discourage the free movement of commerce less than would higher rates on agricultural commodities.¹⁰ The ICC, however, was highly critical of the Hoch Smith resolution passed by Congress in 1925 that asked that agricultural rates be set as low as possible. The commission saw itself as creating a level playing field for all American industry and was uncomfortable at putting a finger on the scales to benefit one industry over another, even if it was an industry that was economically depressed.

Long-haul/short-haul rules were used as a device to encourage the use of a neutral element of the shipping conditions—distance moved—to insure that weak shippers would not be disadvantaged relative to strong shippers. Distance-based rates were simple and understandable and did not seem to introduce the possibility of favoritism. The long-haul/short-haul rules, however, did not prohibit charging the same rate for longer and shorter moves over the same line. In fact, such blanket rates were encouraged since they allowed distant suppliers to compete with local producers. The commission did note that such blanket rates would encourage uneconomic supply decisions, depriving local producers of their natural advantage relative to more distant ones and encouraging excess amounts of transportation; but the commissioners considered this to be less of a concern than the fact that strict distance-based rates would prevent competition

⁷ *Coal to South Dakota*, 47 ICC 750, 754 (1917).

⁸ This section only briefly describes some of the most important factors that the ICC took into account in setting the rate structure. For a more complete list see [Locklin \(1972\)](#).

⁹ *O’Meara v. Baltimore and Ohio RR Co.*, 183 ICC 3, 15 (1932).

¹⁰ Other examples of ICC actions that led to inefficiencies can be found in [Boyer \(1981\)](#).

between distant and suppliers and those closer to the destination.¹¹ It is important to recognize, however, that the attraction of distance-based rates to the ICC was not based on the desirability of linking prices to costs in a way that would encourage efficient choices by shippers. To the ICC, a railroad, while privately owned, is really a public thoroughfare that should be accessible by all under the same conditions; tying rates to distance but also creating blanket rate areas both seemed to decrease the possibility that advantages were being given to one shipper over another through the use of discretion exercised by a railroad.

Just as the railroads' local monopoly power was exercised primarily through properly structuring rates, so too the ICC's regulation was almost exclusively about rate relationships rather than the absolute level of rates. Average variable cost, measured quite crudely, provided a floor below which a particular rate could not fall, but otherwise the cost to the railroad of providing transportation services was not of concern to the regulator. There was no encouragement by the ICC to look for ways to reduce costs, especially if these were tied to the introduction of new technologies since these would disrupt the rate structure that the ICC had carefully built up to ensure that traffic could "move." For example, the development of unit trains, with their inherently lower costs, were discouraged because some shippers would not have enough volume to use them.¹² Unit trains would allow large shippers to deliver commodities more cheaply to a destination, depressing prices at the destination, thus depriving small shippers of the ability profitably to move their goods to market.

Railroad costs were indirectly involved in setting rates because, under the Transportation Act of 1920, the regulators were directed to insure that "in the exercise of its power to prescribe just and reasonable rates, the Commission shall initiate, modify, establish, or adjust rates so that the carriers as a whole.... will earn an aggregate annual net railway operating income equal.. to a fair return upon the aggregate value of the railway property of such carriers held for and used in the service of transportation."¹³ Thus while railroad costs were not used to set the rate structure (except to put a floor under low rates that were justified by inter- or intra-modal competition,) the rate level had to be high enough to allow a normal rate of return on invested capital for the industry as a whole.

According to Sharfman (1931, p. 284) in his four volume history, the ICC "adjusts the relativity of charges as a means of establishing an equitable distribution of transportation burdens, and it prescribes their general level, as a means of controlling the flow of revenue that results therefrom." In short, the basic charge to the ICC is to fairly allocate the burden of financing a system in which freight can freely flow. The shareholders of railroad companies were brought into this fairness balance for the first time in 1920, just as the construction was beginning on the highway system. The timing was fortuitous for railroads. Throughout the inter-war period, the ICC continued to act to make the railroad industry the neutral agent of economic development while guaranteeing the profitability of a system that was massively over-built.

¹¹ *Paper and Paper Articles to New Orleans*, 88 ICC 345 (1924).

¹² *Limestone to Baton Rouge*. 270 ICC 584 (1948).

¹³ Section 15a of the Interstate Commerce Act, as amended by the Transportation Act of 1920.

During this period, the railroads first lost their passenger traffic to the highways and then much of their freight traffic. The Motor Carrier Act of 1935, which subjected truck and bus transportation to the same logic of ratemaking as was applied to railroads, was designed to support the ICC's control of railroad rates. A system with vast over-capacity—the result of incentives that localities had to protect themselves from monopoly price setting by the railroad industry—quite rapidly lost much of the traffic that supported that over-extended system. Since the ICC saw its role as maintaining a system within which traffic could move, it discouraged widespread abandonment of tracks, reasoning that this would inhibit the flow of commerce that the railroad system was supposed to support.

6 The Consequences of Regulating Rates so that Commodities May Move Freely

The ICC saw its role as supporting a railroad system that encouraged structural competition among suppliers by maintaining a railroad rate structure that allowed traffic to “move.” Much of this was an attempt to maintain traditional rate relationships that developed before the regulated era. Such concepts of equity based on history and tradition are impossible to introduce in the model developed previously. But the regulatory urge to allow traffic to “move” can be modeled by analyzing a rate structure that allows all locations to trade with each other subject to the condition that each price is greater than the average variable cost of movement and that the railroad system as a whole has revenues that are equal to costs.¹⁴

Table 5 shows an equilibrium in which the railroad system is overbuilt for demand, with all diagonal links in place, and in which rates are set to guarantee that each supply location can sell to each demand location. In other words, the rates are set to allow the traffic to “move” so long as each rate is above average variable cost for that link. The concern of the regulator is now that rates be set so that sellers in each location can deliver to both demand locations, and similarly that each demand location sees equal delivered costs from each supply location, thus guaranteeing geographic competition. To do this, the regulator can set the ten rates between the two demand locations and five supply locations.

Several complications must be determined before the calculation can be made. First, some origin–destination pairs have more than one route connecting them. Thus, for example, to get from 1 to B, one can either take the diagonal route or take the spur-and-main-line route of the core railroad. Since the non-diagonal route is more circuitous, the user cost of taking this route is higher than that for the diagonal route. ICC policy was to allow a rate discount to the more circuitous route in recognition of the lower quality of service inherent in a less direct route. This is followed in Table 5 in which the core railroad is assumed to give a rate break relative to the diagonal route

¹⁴ It should be recognized that, the ICC did not actually set the rate structure; it adjudicated rate disputes between carriers, acting collectively, and shippers using the criteria described in the body of this paper. The idea that the Commission actions could be summarized into a single maximand would have been rejected by the ICC as well. The criterion used here is intended to show the effects of biasing the rate structure in the direction of expanding the number of viable trade corridors.

Table 5 Price and quantity data for rate regulated network described in Fig. 1

	To A Quantity	To B Quantity	Quantity sent	Supply price	Surplus	To A Rate	To B Rate
From 1	<i>2.5</i>	<i>6.1</i>	8.6	46.3	18.5	9.9	8.7
From 2	<i>5.0</i>	<i>12.0</i>	17.0	49.0	144.0	7.3	7.6
From 3	<i>0.4</i>	<i>1.0</i>	1.5	50.8	4.2	4.4	6.8
From 4	<i>2.7</i>	<i>6.4</i>	9.1	48.1	81.9	6.5	8.8
From 5	<i>2.2</i>	<i>5.2</i>	7.4	45.7	13.7	7.2	10.9
Total received	12.8	30.7					
Demand price	57.2	58.6					
Surplus	81.6	943.4					
Social surplus	1,287.4						
Railroad profit	0						
Variable transport costs	185.7						
Fixed way costs	174.2						

Source See description of calculations under Table 1. Quantities in italics are determined by numeric optimization to maximize social surplus subject to the constraints that (1) the delivered price (supply price plus movement cost plus user cost) to each destination be the same for each origin so long as the rate charged is greater than the variable movement cost, (2) that each origin delivers an equal proportion of its output to each destination, and (3) the sum of railroad revenue be equal to the fixed plus variable cost of operating the revenue. Rates are calculated as the difference between the supply price less demand price less the user cost of transportation. For links on which there is more than one route, the calculation assumes that traffic is shared equally between the two routes. Rail profit is calculated as the sum of rates multiplied by quantity for each route less the sum of variable and fixed costs of rail transport

equal to the increased user cost of that route so that the full cost rate by each route are the same.

A second complication derives from the fact that, in order that traffic can “move,” each supply location must see the same delivered price in each demand location and each demand location must see the same delivered price regardless of the origin of the goods. Since the model deals with a homogeneous commodity where buyers in demand locations and sellers in supply area are price takers, there is no principle for allocating quantities to each route. This apparent mimicking of the competitive outcome was the goal of the stated ICC policy. Without any principle for pairing buyers and sellers, each of which makes an identical offer, Table 5 arbitrarily allocates deliveries from each supply location to each demand location according to the relative sizes of the two demands. In the example shown, City A has 30% of total consumption, so each supply location is assumed to deliver 30% of its output to city A and 70% to city B.

A final complication is that the condition that the delivered price to each demand area be the same from each supply location does not uniquely define the rate structure. While there is no evidence that the regulator had efficiency in mind when setting rates, to maintain comparability with Table 2. Table 5 shows a rate structure that maximizes social surplus subject to the break-even constraint for the railroad, as well as the conditions that the delivered price from each supply location must be the same for each destination and the destination price less the full price of transportation must be the same for each supply location.

Table 5 shows that a rate system that encourages geographic competition among supply locations bears little resemblance to those motivated either by efficiency or profit maximization. This is not surprising since the actual movement costs of each link are not used in setting rates. Railroad costs enter the rate making process only at the end when adjustments are made to the entire rate structure to guarantee that the total revenue of the railroad is equal to the total movement and way costs.

Not surprisingly, there is considerable excess transportation in Table 5 compared with Table 2. This is due to the fact that there is little in the criteria for regulated price making to save on transportation costs. In fact, the ICC saw the well-defined hinterlands supply area around each city as evidence that cost-based pricing encouraged the development of market power among commodity sellers. For example, city A in Table 2 is supplied by only two locations while in Table 5, all suppliers are able to compete for its business. However, in the process, there is considerable cross-hauling, with some of A's demand satisfied by city 5, despite the fact that there is no economic logic in receiving from such a distant source of supply. Since the regulator froze the over-expanded system and then did not economize on transportation within the system, both the variable and the fixed costs are higher under the regulatory scheme of Table 5 than in the second-best spatial equilibrium of Table 2.

In an efficient spatial equilibrium, there is a tradeoff between production costs and movement costs to guarantee that each destination is served by the least (delivered) cost origin. In Table 5, there is neither economizing in production costs nor in the costs of transportation. As a result, there is a reduction in social surplus of 196.6 between the second best equilibrium and the regulated equilibrium. Of this, 104.2 is accounted for by excess trackage and 23.5 by excess distance that is moved by a much smaller quantity over the many more miles of track. The remaining deadweight loss is due to the fact that demand locations are not matched with the lowest cost supply locations, causing overproduction in some areas and underproduction in others, together with the excess burden that results from diminished levels of production and consumption that result from higher overall transportation costs.

All of the regulated rates in Table 5 are higher than those listed in Table 2, which is the result of considerably higher costs of transportation that the system as a whole is required to cover. The rates are also manipulated to keep prices differentially high in the demand locations so that all suppliers can reach them so long as the rate that they pay is greater than average variable cost. The surplus in both demand cities is diminished by this process, especially in city A. Part of the process by which the equilibrium in Table 5 allows locationally disadvantaged supply areas to serve both cities is to keep supply prices low, for example in cities 1 and 5, making them losers relative to the second-best spatial equilibrium in which they are locked into a single economic destination through cost-based rates. However some more centrally located cities actually gain from regulated ratemaking relative to the second-best efficient equilibrium despite the fact that their rates rise to pay for excessive transportation used by others. The reason is that delivered prices are higher under regulation, allowing the centrally located cities to expand production, even as they share destination markets with others.

One element of ICC regulation that cannot be illustrated in Table 5 is the prohibition in long-haul/short-haul pricing, under which it should cost more to ship through a city

than to it. This is violated in Table 5 since it costs 7.3 to ship from city 2 to City A but it costs 4.4 to ship from city 3, through city 2, to A. Mathematically, a violation of the principle must occur in one direction or the other since the delivered full cost from 2 to A must be the same as from 3 to A and similarly the delivered full price from 2 to B must be the same as from 3 to B, but the user cost of going from 3 to A is higher than the user cost of going from 2 to B. But the ICC gave many exceptions to the prohibition on long-haul/short-haul discrimination, and presumably one would have been granted here

The evaluation of the success of the ICC's attempt to regulate railroad rates so that traffic can move depends on the benchmark against which it is evaluated. The results in Table 5, which shows the results of regulated ratemaking, are far inferior to those in Table 2, which shows the second-best traffic flows; the result is a lower surplus for all, excessive transportation, and excessive costs of production. All locations are worse off under regulation than under the ideal system. But the ideal system based on efficiency calculations was never seen in the United States since railroads developed as monopolies over their own tracks rather than a system that used cost-based pricing. And compared to the monopoly equilibrium, all locations are better off, despite the considerable waste in the system.

The system of regulation is widely seen as having led to the decline and eventual bankruptcy of much of the American railroad system.¹⁵ But again, whether this was true or not depends on the benchmark against which the development of the system is to be compared. The railway network that the Interstate Commerce Commission inherited at the end of the nineteenth century was far denser than was economically justified. This was especially true beginning in the 1920s when automobile and then truck traffic on the new highway system shrunk railroad traffic on the overextended system of tracks, despite the intent of the Motor Carrier Act of 1935. Private carriage by truck, without the institution of monopoly operators over specific roadways, gave to localities the protection against railroad rate-making that the ICC had been trying to provide. A competitive market in transportation services would not have permitted the railroad industry to earn a return on its investment at the end of the nineteenth century due to overcapacity. After the development of the system of paved roads, the railroad system was even less capable of being funded by competitive pricing.

The Interstate Commerce Commission, by allowing regulated rates to rise as traffic fell, provided profits to railroad shareholders over a far longer period than there was an economic justification for, postponing the wave of railroad bankruptcies from the 1920s and 30s to the 1960s and 70s. A policy of efficient railroad pricing would have shrunk the system dramatically throughout the twentieth century and would have resulted in a core system of an appropriate size so that traffic pressed on capacity constraints, allowing marginal cost pricing to be economically viable. Had efficiency been an important element of price regulation, railroad prices would have been much lower, and the shift of highway freight from the railroads to the highways would have been slower and less extensive. But the lack of regulation and the presence of

¹⁵ See, for example, the discussion in Boyer (1981).

efficient pricing should not be conflated. Efficient spatial equilibria are at odds with the organization of railroads as a single firm operating on a track that it owns.

7 Conclusion

The form of regulation that shaped modern American railroads was primarily concerned with the structure of freight rates rather than their level. This paper identifies the organization of the railroad industry in which railroads were granted operating monopolies over geographically defined trackage systems as the fundamental reason that the Interstate Commerce Act and the Interstate Commerce Commission focused on the structure rather than the level of freight rates. The profit maximizing strategy in this situation calls for carefully setting different prices for different customers and for setting some prices so high that buyers and sellers find it attractive to do business with firms that are more profitable for the railroad to serve rather than those that would be part of an efficient equilibrium. In short, when railroads have a monopoly on serving routes, market power is primarily exercised by adjusting the structure rather than the level of freight rates.

This narrowly targeted pricing caused damage in the nineteenth century by creating incentives to build excess trackage as communities sought refuge from railroad market power; in the twentieth century, the Interstate Commerce Commission did more damage as it struggled to define a rate structure that did not contain the railroads' targeted rate advantages and disadvantages, but that funded the over-extended legacy system.

Using a simple stylized transportation network, this paper has analyzed the effects of the form of price structure regulation used by the Interstate Commerce Commission. The Commission did not try to set prices at efficient levels: the cost information was not available to set rates in this way and also the ICC did not see itself authorized to set efficient prices. Rather, it saw its charge as guaranteeing that commodities could "freely move" on the railroad network, mimicking what would have been observed had railroads been organized like highways or waterways, without a monopoly operator on each line. This paper models ICC behavior as maximizing market access for different communities subject to an average variable cost minimum and subject to a break-even constraint for the railroad system as a whole. In a spatial equilibrium, maximizing market access requires setting rates for different commodities and communities that have little relation to the costs of providing the service.

The system of maximizing market access clearly leads to more transportation than is necessary and leads to inefficient location decisions for production and consumption. Combined with a far more extensive track system than is necessary for providing efficient transportation services, it is not surprising that the cost of rail transportation is higher under the regulated system modeled in this paper than it could have been in an efficient equilibrium. Excessive trackage built in response to the locally-exercised market power was an important source of the inefficiencies of ICC regulation. The ICC, in guaranteeing the US railway system a normal rate of return on a capital stock that by the 1950s was much greater than necessary, shielded the US railway sector from the financial losses that a free market imposes on industries with excess capacity; in this sense, regulation was to the benefit of the railroad industry and its

shareholder-owners, despite its meager profitability, for most of the twentieth century.

The conclusions of this paper are based on a simple stylized example of a railroad network, making one-way connections between a fixed number of heterogeneous supply and demand areas that are defined geographically, with railroads carrying a single homogeneous commodity. The numeric conclusions depend on the particular assumptions of the model. There is, however, no reason to think that a more realistic multi-commodity simulation with two way movements, backhaul pricing, and the other complications that are familiar to students of railroad economics, would lead to fundamentally different conclusions.

Empirical verification may eventually be found in an interesting experiment that is taking place in other parts of the world: the separation of track ownership and railroad operations.¹⁶ To date there has been little successful entry by operators that are not connected with the former national railways, and thus it is too early to tell whether the introduction of contract and private operations on existing tracks will fundamentally change the way that railroad pricing is seen by customers. The experience of foreign railways, with much shorter distances and extensive passenger service in almost all cases, is also not a perfect predictor of what would be seen with true liberalization of rail service in the US. But when developing future railway policy, it will be useful to remember the legacy of ICC regulation that ignored efficiency and instead tried to develop pricing rules that were designed to allow traffic to “move” in a vertically integrated system of geographic railroad monopolies.

References

- Anderson, S. P., & Wilson, W. W. (2007). Spatial modeling in transportation. In J. F. Plant (Ed.), *Handbook of transportation policy and administration* (pp. 255–280). Middletown, PA: CRC Press.
- Association of American Railroads. (2012). *AAR Stats 2012-5-10*, PDF document (Washington, DC).
- Baumol, W., & Bradford, D. F. (1970). Optimal departures from marginal cost pricing. *American Economic Review*, 60(3), 265–283.
- Beale, J. H., & Wyman, B. (1915). *Railroad rate regulation* (2nd ed.). New York, NY: Baker, Voorhis and Co.
- Boyer, K. D. (1981). Equalizing discrimination and Cartel Pricing in transport rate regulation. *Journal of Political Economy*, 89(2), 270–287.
- Cleveland, F. A., & Powell, F. W. (1909). *Railroad promotion and capitalization in the United States*. New York, NY: Longmans.
- Fletcher, R. V. (1923). *The growth of American administrative law*. Berkeley, CA: Thomas Law Book Company.
- Fogel, R. (1964). *Railroads and American economic growth: Essays in econometric history*. Baltimore, MD: Johns Hopkins Press.
- Harley, K. (1982). Oligopoly agreement and the timing of American railroad construction. *Journal of Economic History*, 42(4), 797–823.
- Hilton, G. W. (1966). The consistency of the interstate commerce act. *Journal of Law and Economics*, 9(Oct), 87–113.
- Kolko, G. (1970). *Railroads and regulation, 1877–1916*. Princeton, NJ: Princeton University Press.
- Locklin, D. P. (1972). *Economics of transportation* (7th ed.). Homewood, IL: Richard D. Irwin.

¹⁶ For a good discussion of current European attempts to control market power through the separation of infrastructure and operations, see Nash (2005).

- MacAvoy, P. (1965). *The economic effects of regulation: The Trunk-Line Railroad Cartels and the interstate commerce commission before 1900*. Cambridge, MA: MIT Press.
- Nash, C. (2005). Rail infrastructure charges in Europe. *Journal of Transport Economics and Policy*, 39(3), 259–278.
- Porter, R. (1983). A study of Cartel stability: The joint executive committee, 1880–1886. *Bell Journal of Economics*, 15(2), 301–314.
- Samuelson, P. A. (1952). Spatial price equilibrium and linear programming. *The American Economic Review*, 42(3), 283–303.
- Sharfman, I. L. (1931). *The interstate commerce commission: A study in administrative law and procedure* (Vol. 1). New York: The Commonwealth Fund.
- Takayama, T., & Judge, G. G. (1971). *Spatial and temporal price and allocation models*. Amsterdam: North-Holland Publishing.
- US Department of Transportation, Bureau of Transportation Statistics. (2012). *National Transportation Statistics, 2012*. Washington, DC.