Competition and Regulation in the Railroad Industry

Ioannis Kessides and Robert D. Willig
No. 1  Rhee, Katterbach, Belot, Bowring, Jun and Lee, *Inducing Foreign Industrial Catalysts into Sub-Saharan Africa*
No. 2  Mody and Wang, *Explaining Industrial Growth in Coastal China: Economic Reform...and What Else?*
No. 3  Biddle and Milor, *Institutional Influences on Economic Policy in Turkey: A Three-Industry Comparison*
No. 4  Lanjouw and Mody, *Stimulating Innovation and the International Diffusion of Environmentally Responsive Technology*
No. 5  Tan and Batra, *Technology and Industry Wage Differentials: Evidence from Three Developing Countries*
No. 6  Navarro, *Reversal of Fortune: The Ephemeral Success of Adjustment in Venezuela, 1989-93*
No. 7  Morales, *Bolivia and the Slowdown of the Reform Process*
No. 8  Ibrahim and Lofgren, *Successful Adjustment and Declining Governance? The Case of Egypt*
No. 9  Tan and Batra, *Enterprise Training in Developing Countries: Overview of Incidence, Determinants, and Productivity Outcomes*
No. 10 Sundaram, Teik and Tan, *Vision, Policy and Governance in Malaysia*
No. 11 Tzannatos, *Labor Policies and Regulatory Regimes*
Competition and Regulation in the Railroad Industry

Ioannis Kessides and Robert D. Willig

February 1996
# Contents

1. Introduction ......................................................................................................................... 1

2. Public Policy Issues in the Rail Industry ........................................................................... 3
   The Proper Criterion for Adequacy of Revenues ............................................................... 5
   The Regulatory Problem .................................................................................................... 6
   The Cost Allocation Problem ............................................................................................. 7
   The Problems in Fully Allocated Cost Pricing ................................................................. 8
   Long-Run Marginal Cost and Pricing Efficiency ............................................................... 10
   Economically Efficient Pricing ......................................................................................... 11
   Demand Based Differential Pricing ................................................................................. 12
   The Efficiency and Equity of Ramsey Pricing ................................................................. 13
   The Stand-Alone Cost Constraint ..................................................................................... 14
   Stand-Alone Costs: Protection Against Excessive Rates ................................................... 15
   Efficient Pricing and Regulatory Control ......................................................................... 17
   Contestability and the Scope and Structure of Regulation ................................................ 18

3. Options for Vertical Railway Restructuring .................................................................... 21
   The Need for Restructuring ............................................................................................... 21
   Analysis of Structural Separation .................................................................................... 23
   Analysis of Competitive Access ....................................................................................... 26
   Separation Versus Competitive Access ............................................................................ 27

4. Concluding Remarks .......................................................................................................... 31

Appendix A ............................................................................................................................... 33
   Technology and the Structure of Railroad Costs .............................................................. 33
      Fixed and Variable Costs ............................................................................................... 33
      Sunk Costs ..................................................................................................................... 33
      Economies of Scale and Scope ..................................................................................... 35
      Economies of Size Versus Economies of Scale and Density ......................................... 36
      Empirical Evidence on Scale Economies ...................................................................... 37

Appendix B ............................................................................................................................... 41
   Rail Costs, Profitability, and Structural Changes ............................................................... 41
      Railroad Costs and Infrastructure Variables ................................................................. 41
      Low-Density Lines and Profitability ............................................................................. 42

References ............................................................................................................................... 45
Introduction

1.1 The rail industry has been one of the most extensively regulated sectors in the economy (see Friedlaender 1969; Keeler 1983). Price, entry, exit, financial structure, accounting methods, vertical relations, and operating rules have all been subject to some form of governmental control. The public utility paradigm of governmental regulation has been applied to the rail industry on the assumption that the industry's economic characteristics preclude competitive organization and any need for market responsiveness.

1.2 Over the past three decades, however, economists and policymakers have become increasingly critical of the traditional public utility model (see Friedlaender 1971; Levin 1978, 1981a; and Boyer 1987). It has become common wisdom that there is often effective competition in the relevant economic markets in which rail carriers seek to meet demand. It is also generally agreed that governmental restrictions on the structure and conduct of firms in the rail industry impose considerable costs on society. The misallocation of freight traffic among competing transport modes, excess capacity, excessive operating costs, and poor investment decisions are often the result of misguided regulatory policies. Regulatory controls have, therefore, been held responsible in large part for the poor financial condition of the railroads, for the deterioration of the rail plant, for the suppression and delay of cost-reducing innovations, and for the mediocre quality of rail service.

1.3 The purpose of this paper is to suggest a set of principles for restructuring railroad regulation, and indeed for restructuring the orientation of railroad entities, for the sake of the public interest. We focus first on the economic characteristics of the rail industry and their implications for the design of efficient regulatory policy. Then we apply powerful sets of analytic tools to clarify the relevant principles for reform. Much can be learned about policies to promote the public interest, from an understanding of market demands for the industry's products and the nature of the productive techniques available to the industry's firms. Indeed, before the implications of policies aimed at rate regulation or infrastructure investments can be fully assessed, a full understanding of the nature of technology, costs, and demand facing the rail industry is required. The role of the government in relation to market behavior should therefore be based explicitly on the underlying economic characteristics of the industry and the technological conditions of its production.
1.4 We hope to impart the following message: Recent developments in industrial organization analysis and in regulatory practice call for a major reorientation of public policy toward railroads. We suggest in this paper a set of principles to guide such a reorientation.
Public Policy Issues in the Rail Industry

2.1 The economic characteristics of the rail industry make it a natural target for government intervention, yet also render it particularly difficult to regulate in the public interest. The old regulatory systems failed to handle the central regulatory problem arising in railroads and certain other major industries (for example, telecommunications, electric power, and postal services): the mixture of competition and monopoly elements in supply. Indeed, in these industries, just as in the railroad industry over the years, regulation has stifled competition in the provision of services, restricted the benefits of economies of scope, retarded innovation, and fostered inefficient service. Regulation thus has harmed the public interest while protecting it from the exploitation of monopoly power (see Willig and Baumol 1987). The first-best lesson of the perfect competition model, calling for prices to be set equal to marginal costs, has no doubt contributed to the common regulatory ethos that seeks to equate price to some measure of cost. This doctrine has frequently been used where it is completely inappropriate and without logical foundation—that is, in cases where prices should be based on demand as well as cost considerations.

2.2 This chapter focuses on the central pricing issues involved in the partial deregulation of railroad rates. It articulates principles to guide regulatory oversight of the rate setting of unsubsidized railroads—principles that are both consistent with economic analysis and essential for the protection of the public interest. Public interest regulatory oversight of railroad pricing involves two basic issues. The first of these is the adequacy of revenues, the criteria by which adequacy can be judged, and the means by which it can be achieved. The second issue is the choice of rates that are both consistent with adequate revenues and best for the public interest.

2.3 In a regime of deregulation, one of the key elements in protecting the public interest is the elimination of any residual regulation that effectively prevents the rail network from achieving financial viability. The public will hardly be well served by a set of regulatory rules that makes it impossible for the railroads to compete in the financial marketplace. The rail network that resulted would become increasingly deteriorated and obsolete, and cumulative abandonment of service would become the prevailing practice.

---

1 The authors acknowledge their debt to the thinking and writings of William J. Baumol on many of the subjects covered in this chapter. A summary of some of this material can be found in Baumol and Willig 1987.
2.4 In determining prices for the outputs of multiproduct railroad firms, regulators have long faced a number of difficult issues that flow inevitably from the basic economic characteristics of the industry discussed above. The endemic economies of scale and scope imply that straightforward measures of costs cannot be used to dictate pricing. Economies of scale imply that marginal cost pricing, absent subsidy to the firm or multipart tariffs, will not allow the firm to break even. Further, because the shared costs that are a concomitant of economies of scope cannot be unambiguously identified with individual products, any rule selected to associate shared costs with individual services will be arbitrary. Such arbitrary measures as fully distributed (or "fully allocated") costs, therefore, cannot substitute for marginal cost measures as decision rules for proper pricing. The misguided search for a purely cost-based substitute rule relies inappropriately on the model of perfect competition for guidance on regulation.

2.5 A system of rate regulation in which costs are apportioned on any basis other than demand is inappropriate because it is highly unlikely that prices set will permit railroads to achieve an adequate rate of return. Moreover, such a method leads to serious inefficiency by discouraging innovation and by generating prices that are too high to attract competitive traffic. The absence of competitive traffic in turn severely restricts the amount of services delivered by railroads and thus produces still higher rates for the remaining traffic (see Braeutigan 1977; Kahn 1988).

2.6 By contrast, there are sound pricing principles that promote economic efficiency while simultaneously removing impediments to adequate returns for carriers. These principles can be applied in a practical fashion to assess the reasonableness of those rates that are judged to require continued regulatory oversight (see Braeutigan 1979, 1984). The principles lead to demand-differentiated prices, sometimes referred to as Ramsey prices, which apportion all unattributable fixed and common costs of the railroad among its services on the basis of the values of those services to consumers, mathematically expressed as their elasticities of demand. Economically efficient differential pricing combines cost and demand factors in an optimal manner, by pricing each service at a markup over marginal costs that is inversely related to the elasticity of demand for that service. The resulting set of rates encourages the purchase of more rail transportation services by more shippers than would be the case with fully distributed cost-based pricing, thereby creating a larger traffic base over which unattributable costs can be apportioned and lowering prices for shippers generally. Ramsey pricing maximizes the opportunity for rail carriers to earn an adequate rate of return on capital and fosters innovation and efficiency in the provision of rail transportation services by rewarding carriers that achieve cost reductions.

2.7 Economically efficient differential pricing is entirely consistent with the hallmark of deregulation: that market forces, rather than regulation, should control the rates for transportation services. Thus, when a particular type of traffic is subject to competition, direct or indirect, regulatory intervention is unjustified. Furthermore, so long as a railroad's earnings fall short of its cost of capital, the need for regulatory constraints on any of that carrier's rates is minimal, and to the extent such a constraint prevents the carrier from earning an adequate return in the future, it is contrary to the public interest. By definition, there is no danger that such a carrier is
receiving excessive overall profits derived from market power or any other cause. In addition, if the rate for any service supplied by a railroad not yet earning adequate revenues overall is held down by regulation below the level that consumers of that service are prepared to pay rather than do without the service, then, in the long run, even those consumers will be harmed. The carrier will find it unprofitable to invest the necessary replacement and maintenance capital, causing a deterioration in, and ultimate withdrawal of, the service.

The Proper Criterion for Adequacy of Revenues

2.8 Since avoiding impairment of financial viability plays so crucial a role in any rational program of rate regulation, it is important to describe the criterion by which financial viability can be judged. What information is required to determine when a firm's revenues are adequate to cover its pertinent costs? While the answer would appear to be obvious, the history of regulation demonstrates rather forcefully that it is in fact widely misunderstood. The basic issue is that the cost of the firm's capital, including any capital it has generated internally, must always be included in these calculations.

2.9 The logic of this criterion is straightforward. Revenues are defined to be adequate when they are just sufficient to enable the firm to attract the capital needed for maintenance, replacement, modernization, and whatever expansion is justified by demand conditions. If revenues are below this level, the deterioration and eventual disappearance of the service are inevitable.

2.10 Adequate revenues are those that provide a rate of return on net investment equal to the current cost of capital (that is, the level of return available on alternative investments). This is the revenue level necessary for a railroad to compete equally with other firms for the financing needed to maintain, replace, modernize, and, where appropriate, expand its facilities and services. If railroads cannot earn the fair market rate of return, their ability to both retain existing investments and obtain new capital will be impaired, because the existing and prospective funds could be invested elsewhere at a more attractive rate of return. Indeed, the market for funds is one of the most competitive in the economy. There is no escaping the following principles that determine the adequacy of revenues:

- The firm's overall rate of return must be equal to the returns currently earned by the typical firm with similar risks elsewhere in the economy. Otherwise, the required funds will be denied to it.

- The adequacy of a firm's revenues can be judged only by comparison with the earnings of firms outside of regulated industry. If the regulated industry's earnings are compared with the market value of the firm's equity, the market prices of those securities will automatically adjust themselves downward to match any act by the regulator that restricts the earnings of the firm below a compensatory rate of return. Such a comparison will thus appear to justify any earnings restrictions, no matter how inappropriate.
In determining the firm's revenue requirements for financial viability, the rate of return obtained by comparison with other industries must be applied to a rate base that covers the economic replacement cost (under regulation) of all facilities. (Suitably updated historic costs can be utilized instead of replacement costs if the allowed rate is expressed in nominal terms.)

With the rate base determined in this way and the rate of return on that rate base equal to the cost of capital, as given by earnings prevailing elsewhere in the economy, the result will be the total net earnings figure that can appropriately be considered to be adequate for the railroad to compete successfully in the capital market.

This earnings figure must not be applied as a rigid ceiling. Doing so would make it impossible for railroads to earn this figure over the long run, since they would be precluded from making up for any revenue shortfalls resulting from temporary downward fluctuations in demand for their services.

To make sense economically, prices must never be incompatible with this earnings level. Of course, no prices can guarantee that a railroad will earn adequate returns overall. If demands for its services are insufficient, operations are conducted wastefully, or services are poor, even appropriate prices cannot be expected to lead to profitable operation. But once the railroads are permitted to charge appropriate prices in a competitive environment, the regulatory impediments to financial viability will have been eliminated. It then will be up to the railroads to take advantage of the opportunity through economic operations, quality service, and effective marketing.

The Regulatory Problem

Indivisibilities, pervasive economies of scale and scope, high costs of entry, and small-numbers competition in the railroad industry are all consistent with the likely persistence of prices in excess of marginal cost. However, while scale economies go hand in hand with natural monopoly, a railroad may or may not have the price-setting discretion that characterizes the textbook monopolist. It all depends on whether the activities characterized by economies of scale and scope are shielded from other sources of competition in the relevant market and whether there are protective barriers to entry.

In the railroad industry, extensive capital sums must be sunk in way and structures and in a variety of ancillary facilities to create new rail lines. The sunk cost and longevity of railroad capital may suggest that the railroad industry cannot be conceived to be contestable. However, railroad services are far more contestable than these impediments to rail entry would suggest because other modes of transportation, such as trucking and water carriage, often exert
strong competitive pressures on the rates charged for shipment of a wide variety of commodities.²

2.14 The basic patterns of railroad regulation, established many decades ago in wholly different market conditions, are simply obsolete. Their premise was that railroads had a collective monopoly, or near-monopoly, in land transport. This condition disappeared long ago, if it ever existed. Nearly every sphere of rail freight service now faces intense competition. Rival products and rival sources of supply (including trucks, barges, and alternative rail routes) are likely to impose effective competitive constraints on many, if not most, rail activities. In those activities where there is no evidence that the railroad holds a position of market dominance, the industry should be offered freedom in pricing. Still, there remain instances in which the competitive checks of intramodal, intermodal, geographic, and product competition are weak or nonexistent. There is an understandable apprehension that in such cases market forces may not be relied on to prevent excessive pricing. The resulting monopoly power is the basic justification for the regulation of rail rates and earnings and defines the basic task with which regulation must grapple.

2.15 Before discussing the appropriate means to deal with this issue, however, it must be emphasized that, in practice, effective competition can assume a variety of subtle forms. Therefore, one must never proceed in haste to undermine the workings of the market through special intervention. Railroads do not face only the competition of trucks and barges. For example, oil and natural gas shipped by pipeline competes with coal shipped by rail; since coal shipment is profitable to the railroads, the competition of petroleum products limits the price railroads can charge for carrying coal. Also, the market served by one railroad may compete for the coal with a market served by another. This situation, too, can keep rates in line.

The Cost Allocation Problem

2.16 The presence of substantial economies of scale and scope in the railroad industry creates several problems for government regulation. Perhaps the most troubling is the fact that it is impossible to allocate, in any nonarbitrary way, a share of fixed and common costs to any one of a railroad's many activities. There is simply no way to subdivide those costs in a mechanical fashion that is unique and has any foundation in economic logic.

2.17 In practice, regulatory authorities historically have determined tariffs based on so-called fully distributed (or allocated) costs. Under this method regulators do (somehow) allocate shared production costs to individual services. Each service is then required to generate revenues that will cover all the costs associated with that service. Although it is often argued that there is

² It should be noted that in many instances the relevant competition is not just on the route invoked in the rail movement but also on alternative routes that offer economic substitute services for the shipper. For example, a manufacturer may find it equally desirable to ship output to two very different places for the purposes of sale, and will choose the option with the least expensive transportation.
no sound economic rationale for fully distributed cost pricing, this practice obviously has
economic consequences.

2.18 Traditionally, regulatory proceedings have focused on three types of fully distributed
cost rules. The first of these is the distribution of shared costs on the basis of a common measure
of utilization, such as gross ton-miles. Under this approach, termed the relative output method,
shared costs are allocated in proportion to the number of units of output of each service. A
second approach is the allocation of shared costs in proportion to the costs that can be directly
attributed to the various services. This attributable cost method has also been traditionally used
by many unregulated firms in their allocation of overhead costs. A third scheme requires
allocation of shared costs in proportion to the gross revenues generated by each service. This
gross revenue approach has been frequently used to allocate overhead costs between freight and
passenger services.

2.19 In addition to costs that are directly attributable, a service may also be assigned a
portion of those costs that cannot be clearly associated with any one service. Railroad track, for
example, is used in the transport of many kinds of freight. Shared costs may therefore constitute
a large portion of total costs. Thus, the method of allocating shared costs may significantly
influence the rate required for any particular service.

The Problems in Fully Allocated Cost Pricing

2.20 Fully distributed cost pricing rules suffer from several disabilities:

- Since fully distributed costs bear no direct relationship to marginal costs, there is no
  basis in economic efficiency for fully distributed cost pricing.

- On grounds of economic efficiency, it may sometimes be desirable to set a price for
  some service so that the revenues it generates do not cover its fully distributed costs.

- Because the determination of fully distributed costs is arbitrary, there is no economic
  basis for concluding that a service is being subsidized by other services if its revenues
  are less than its fully distributed costs.

- Fully distributed cost pricing is anticompetitive since it prevents a supplier from
  offering a service at a proposed tariff less than a fully distributed cost price,
  particularly if the proposed tariff exceeds the marginal cost of providing the service.
  In addition, there is circular reasoning behind the fully distributed cost practice.
  Tariffs that are determined to be “appropriate” at a given time may depend on the
  existing levels of output or revenues, which in turn depend on previous tariffs. Thus,
  fully distributed costs may depend on the acceptance of a prior tariff structure.

- The most serious defect of fully distributed costs as a basis for rate determination is
  that they do not necessarily measure marginal cost responsibility in a causal sense—
  they are costs that are averaged by an arbitrary method. They do not measure by what
  amount costs would be increased if additional quantities of any particular service
  were taken, nor do they measure by what amount costs would be reduced if the
service were correspondingly curtailed. Also, being apportionments of historical costs, even when they do accurately reflect historical responsibility for the incurrence of these costs among the respective users, they do not provide a reliable measure of what will happen to costs in the future if particular portions of the business are expanded or dropped.

- Finally, the fully allocated cost criterion completely neglects any demand data. Even if based on "relative use" as measured in tons or ton-miles, it cannot capture the role of demand, which economic analysis has shown to be vital in the choice of optimal prices. Even the best-intentioned of fully allocated cost standards must employ some rigid criterion to allocate the portion of a railroad's total costs that is not directly attributable to any one of its services in particular. But no such fixed allocation criterion can possibly reflect the subtleties, fine structure, and changes in patterns of demand for the railroad's services that are induced by external developments and that clearly call for adjustments in its prices. This, of course, is true not only of a standard fully allocated cost approach, but of any rigid formula that bases future prices on cost data of the past and thus cannot take account of changes in demand.

2.21 It may seem paradoxical that fully allocated cost criteria, which are apparently designed to ensure that all costs are covered by revenues, can in fact preclude rail carriers from achieving financial viability. The reason is that ceilings based on fully allocated costs are set so that unattributable costs are divided in an arbitrary manner among all types of traffic. Then, for these costs to be recovered, all types of traffic must actually move at the rates that include the arbitrary cost allocations. But traffic with transport value that is below average for its tons, ton-miles, or other allocator will not move by rail at those rates. That is, any service whose demand is insufficient to cover its allocated share of total cost at the fully allocated cost-determined price will have a revenue shortfall that fully allocated cost ceilings will prevent other services from making up. Consequently, if the unattributable costs are substantial, and if the values of rail services vary substantially, fully allocated cost rate ceilings will preclude attainment of adequate revenues.

2.22 The effects of fully allocated cost pricing on the utilization of transport resources are equally pernicious. In doing their best to earn adequate revenues despite the handicap imposed by fully allocated cost rate ceilings, rail carriers will be unable to preserve traffic whose value to the shipper exceeds its attributable cost but which falls sufficiently far below fully allocated cost. True, in the absence of fully allocated cost regulation, any such traffic could contribute revenues that exceed the costs that it causes and would provide social benefits greater than social costs. But with fully allocated cost rate ceilings, this traffic will reduce the net revenues of the rail carrier and will thus not be compensatory. The reason is that this traffic will be assigned its portion of unattributable costs on the basis of its tons, ton-miles, or some other arbitrary allocator, thereby reducing the share of those costs allocated to other traffic with higher value, and consequently reducing the ceiling and the rates on that traffic.
Fully allocated cost rate ceilings may also stifle the incentives of railroads to innovate and compete. A rail carrier cannot be expected to invest in new facilities, research and development, and marketing activities designed to elicit new traffic if the financial gains from the new traffic are counterbalanced by induced decreases in the ceilings on the rates charged to pre-existing traffic. Similarly, a rail carrier could not be expected to compete for freight by offering low rates if the necessary markups were much below the arbitrary allocations of unattributable costs; if it did so, it would never earn adequate revenues because its gain from the low-rated traffic would be outweighed by the induced decrease in the ceilings applied to more highly rated traffic.

**Long-Run Marginal Cost and Pricing Efficiency**

The indivisibilities, economies of joint production, and high fixed costs that make small-numbers competition in the railroad industry an inevitable consequence also render the traditional measure of static deadweight loss incomplete as a welfare indicator. A regime of marginal cost pricing would eliminate the deadweight loss. But marginal cost pricing is a questionable regulatory objective, since the railroads would incur substantial losses. If the regulator attempts to force rates to equal marginal costs, overall revenues will fall short of overall costs. Without subsidy, reduction of the short-run welfare loss to zero would cause long-run deterioration of the industry's capital stock. If revenues are to cover total costs, rates for rail systems that are characterized by scale economies must generally lie above the costs economically attributable to individual services.

It should also be noted that the use of long-run marginal cost to measure pricing efficiency frequently leads to misguided rules that could force the railroad into a pattern of behavior in conflict with the dictates of the market. Indeed, the rigid requirement that each rate always cover the long-run marginal cost of service is tantamount to a prescription of pricing inefficiency for railroads. Moreover, such a misguided decision would be likely to impose a heavy penalty on the public by sometimes depriving it of a valuable service at a price it is willing to pay and that also best serves the interests of the company—namely, a price that lies between long-run and short-run marginal cost.

The role of a cost floor as a measure of efficiency is to determine whether the railroad would be better off without the traffic in question. There are two basic reasons why it will often be appropriate for a price to lie below the corresponding long-run marginal cost. First, many investment decisions that were entirely rational and appropriate when they were made will subsequently be affected by unexpected developments. Such eventualities may cast a shadow over the future of the service that utilizes the investment. A railroad is always better served by carrying any and all traffic that can cover its short-run avoidable costs and contribute to its fixed and common costs, than by abandoning the service. The test of efficient pricing above short-run avoidable costs is whether the railroad is pricing in accordance with market demand. So long as the revenue-inadequate railroad is charging profit-maximizing rates, it is necessarily pricing
efficiently; if the price maximizes the service's contribution to company profits, clearly no other price can conceivably bring that service closer to being compensatory in the long run.

2.27 The second reason why efficient prices will often fall short of long-run marginal cost affects even services whose financial viability is absolutely clear. Whether a railroad will be able in the long run to earn revenues that are sufficient to cover the replacement cost of a particular service or a group of services depends on the level of demand over time. The rail industry is strongly affected by business fluctuations in the economy, and demand for individual rail services and groups of services can and does vary widely over time. Even a service whose financial viability is absolutely clear will certainly encounter years in which business is good and other years in which business conditions are poor. In the less prosperous years, the firm's earnings will often fall short of long-run marginal cost because market conditions permit no alternative. Of course, the shortfall will then be made up during the prosperous periods. In this manner then, the firm will in the long run meet its revenue requirements. But to insist that prices always cover long-run marginal costs is effectively to undermine the market pricing process and, very likely, even the viability of the service. Doing so would clearly distort the intertemporal pattern of usage of the service and thereby reduce economic efficiency. In addition, innovation and improvements in operating efficiency over time could potentially reduce costs and enhance contribution. A rule that assumed assets would not be replaced simply because current revenues from a particular service were depressed would remove any incentive or ability to respond to upswings in demand or make improvements in efficiency that would otherwise permit the service to continue.

2.28 The long-run marginal cost should never be used mechanically as a rigid minimum cost floor in the pricing of an operating railroad. At the same time, it should be emphasized that the long-run marginal cost cannot serve legitimately to establish the level of efficient pricing above short-run costs at any point in time. Instead, efficient rates will always have to be consistent with demand. This is true regardless of whether a railroad has market dominance over a particular service or has achieved adequacy of revenues. The demand for each service always helps to determine the contribution that service should make to the railroad's overall costs if its behavior is to comport with the requirements of economic efficiency.

**Economically Efficient Pricing**

2.29 If there were no need for enterprises to be financially self-supporting, an ideally efficient allocation of society's resources would exist if the price of each good or service were equal to its marginal cost. At such prices, consumers elect to purchase all units of goods and services that yield them benefits larger than the costs of providing them. And, in response to such prices, consumers avoid purchasing units that yield them benefits smaller than the costs of providing them. As a result, the economy misses no opportunity to allocate resources to uses where they yield benefits greater than costs, and no resources are allocated to uses with benefits lower than costs.
2.30 In industries without substantial fixed costs, competition tends to result in prices that approximate marginal or incremental costs. In the railroad industry, however, the prevalence of large fixed and common costs makes it impossible for the supply of rail services to become financially self-supporting with marginal cost pricing. The financial infeasibility of marginal cost pricing rules out any sensible mechanical or formula-based procedure for regulatory determination of rates. In particular, compensatory rates cannot be determined by the regulator on the basis of cost data alone since the financial viability of any price depends also on the quantity of rail services customers are willing to buy at that price. This is true because there is no correlation between demand considerations and any cost accounting convention.

2.31 Allocation of fixed and common costs in accord with any non-demand-based apportionment rule will almost invariably produce inconsistencies with the patterns of shipper demands. Some rates will be too low, and consequently the railroad will receive less than the optimal contribution from those services. Other rates will be too high, so that the railroad will earn either less than the optimal contribution or no contribution at all. In short, in a multiproduct industry with uncongested fixed and common costs, the pricing of individual services on the basis of any cost allocation is contrary to the interests of both the operating entities and the shipping public. Rational determination of prices must be based on both cost and demand conditions to permit adequacy of revenues and achieve efficiency.

**Demand-Based Differential Pricing**

2.32 Non-demand-based cost apportionment methods do not necessarily reflect the railroad's ability (or inability) to impose the assigned allocations and cover its costs. Thus, they frequently overassign or underassign the carrier's unattributable costs to particular services. If a carrier sought to apply fully distributed cost pricing to all its traffic, it would lose that portion of the traffic for which demand could not support the price assigned. In that event, the remaining shippers would be saddled with a larger portion of the carrier's unattributable costs since they would no longer share those costs with the lost traffic.

2.33 Ramsey prices, in contrast, apportion all of the railroad's unattributable fixed and common costs among its services on the basis of their demand characteristics. Each service is priced at a markup over marginal cost that is inversely related to the elasticity of demand for that service. Services whose demands are highly elastic are assigned prices that are very close to their marginal costs, whereas services whose demands are very inelastic are priced well above those costs. The magnitude of these markups among all services must be sufficiently high to earn net revenues that cover fixed and common costs and, hence, achieve revenue adequacy.

2.34 The logic of this inverse elasticity rule and its implied allocation of unattributable costs is quite simple. The elasticity of demand provides a quantitative interpretation of the traditional concept of value of service, which has played an important role in public utility pricing. Consumers who place relatively high value on a service will have demands for it that are relatively inelastic, and vice versa. If a rise in the price of a service would lead to no significant reduction in quantity demanded (that is, if demand is inelastic) then the service must be worth at
least the higher price to its consumers, that is, the value of the service must be high. Conversely, if a rise in the price of a service would lead consumers to curtail their demand substantially (that is, if demand is quite elastic), then the service must be worth little more to its consumers than the original price, that is, the value of the service must be low.

2.35 In view of this correspondence between value of service and demand elasticity, the inverse elasticity rule of Ramsey pricing can be restated in terms of a familiar and long-used principle in railroad pricing. Services with relatively high values to their consumers should contribute relatively large net revenues to the coverage of unattributable, fixed, and common costs. Thus, the implicit allocation of unattributable costs should be based on value of service rather than any pro rata sharing or other arbitrary method. All factors that influence a rail carrier's elasticities of demand are relevant for the carrier's Ramsey prices. These factors may include the value of the commodity shipped, intermodal competition, intramodal competition, interport competition, and the substitutability of other commodities for the one shipped at its destination. Value of service is therefore properly construed as a market concept. It refers to the value of the rail carrier's service with all demand factors considered and generally cannot be evaluated by such measures as the ratio of a commodity's price to its weight alone.

The Efficiency and Equity of Ramsey Pricing

2.36 Under Ramsey pricing, the "nonmarginal" portion of total costs (that is, the total cost less the marginal cost of each service multiplied by the quantity of the service provided) is apportioned on the basis of demand. Equivalently, the nonmarginal portion of total costs is the shortfall between total costs and the revenues that would accrue from pricing each service at the level of its marginal cost. In the presence of economies of scale, this shortfall is positive. Ramsey prices, therefore, deviate from marginal costs only to the extent necessary to provide adequate revenues. They thus permit the railroad to achieve the goal of revenue adequacy with less sacrifice of economic welfare than does marginal cost pricing.

2.37 Increases above marginal cost in the price of an elastic service cause much traffic to be lost—traffic that would generate net benefits because it is valued above the cost it causes. However, less traffic is lost when the price of an inelastic service is raised, and the traffic that is curtailed is the least-valued portion. Consequently, when prices must be elevated above marginal costs to cover unattributable costs, it is economically efficient to increase the prices of inelastic services more than the prices of elastic ones. Such Ramsey prices are, on average, the lowest consistent with financial viability. As long as the price charged to the price-elastic service exceeds its incremental cost, the service is contributing to the carrier's overhead costs. Thus, Ramsey pricing principles benefit all shippers by establishing a set of rates that encourages the purchase of more transportation services by more shippers than would artificial prices based on fully distributed cost. By creating a larger traffic base over which unattributable costs can be apportioned, Ramsey pricing also benefits the so-called captive shippers; the expansion of rail traffic represents an increase in the flow of commodities to their markets at lower transportation
costs. As a result, social productivity is enhanced, and more consumers can obtain more of the goods they desire at lower costs of supply.

2.38 Since Ramsey prices are based on the relative values of the different services, they may seem to approximate the solution of the profit-maximizing monopolist, sometimes loosely described as "charging what the market will bear." However, only the firm's necessary costs, including the cost of capital, are covered by Ramsey prices. Monopoly prices, by contrast, are controlled by no such constraint. Ramsey prices therefore are very different both qualitatively and quantitatively from monopoly prices.

2.39 It should also be emphasized that Ramsey prices are equitable. First, they are nondiscriminatory in the sense that services with similar economic characteristics have similar prices, regardless of the commodities shipped, the route, or the identity of the shipper. That is, two different services with the same elasticities of demand will be priced at the same percentage markups above marginal costs. And two different services with the same marginal costs and demand elasticities will bear identical Ramsey prices. Second, while the Ramsey prices of different services are different proportions of the services' marginal costs, the burdens from these necessary markups that are borne by the consumers have roughly the same proportion to their respective values of service.

The Stand-Alone Cost Constraint

2.40 Ramsey pricing requires that both the marginal cost and the elasticity of demand be quantified for every movement in the carrier's system—which is all but impossible to do with any degree of accuracy. The amount of data and the analysis required are overwhelming. Thus, while the Ramsey formula is useful as a theoretical guideline for rate determination, a valid criticism is that its application would be administratively difficult and burdensome.

2.41 The Ramsey pricing rule has also been criticized because it does not constrain the railroad's pricing of traffic over which it possesses market dominance and thus fails to protect captive shippers. In addition, although Ramsey pricing minimizes the static welfare cost of the revenue adequacy constraint, output levels still are less than they would be if rates were set at marginal costs. This situation results in economic inefficiency because the value of the lost output to the shipper is greater than the value of the resources saved by reducing output. Under these conditions, it may be feasible for the parties to negotiate a contract with incentive clauses, volume-sensitive pricing, or two-part pricing, which would leave both parties better off than at the flat Ramsey price and consequently be yet more desirable for the public interest.

2.42 The critical issue from the standpoint of efficiency is the criterion used to set the ceiling on rates where there is market dominance. As noted above, rate ceilings derived from fully distributed costs are inimical to the public interest. Economically rational ceilings are obtainable from the stand-alone cost. The stand-alone cost of serving any captive shipper or group of shippers that benefit from sharing joint and common costs is the cost of serving that shipper or group of shippers alone, as if the shipper or its group were isolated from the railroad's other customers. A rate calculated by the stand-alone cost methodology represents the theoretical
maximum rate that a railroad could levy on shippers without substantial diversion of traffic to a hypothetical competing service. Thus, the stand-alone cost criterion serves as a surrogate for competition and leads to a simulated competitive price. The competing service could be either a shipper providing rail service for itself or a third party competing with the incumbent railroad for the traffic. In either case, the stand-alone cost represents the minimum cost of a possibly hypothetical alternative to the service provided by the incumbent railroad.

Stand-Alone Costs: Protection Against Excessive Rates

2.43 The stand-alone cost test rules out the possibility of abuse of monopoly power by enforcing a competitive standard on railroad rates. The hallmark of monopoly power is the elevation of the price of a service above the costs at which competitors could provide that service. The stand-alone cost test makes that impossible and imposes the same ceilings on rates for any traffic over which the railroad is dominant that the market would impose if it were subject to either active or potential competition. In the long run, in contestable markets, no group of shippers would agree to pay a carrier more for their transportation services than it would cost them to produce these services for themselves or more than it would cost a competitor to supply the services to them. In the short run, a rail carrier facing either active or potential effective competition could not obtain revenues from a group of shippers that exceeded their stand-alone costs, because those shippers could then be profitably served by a competitor charging lower rates. Thus, the stand-alone cost test affords shippers the same protection that effective competition would provide.

2.44 Clearly, the stand-alone cost is unnecessary and inappropriate where there is competition. In a competitive market, the price set by competitors (reflecting current costs of service) will set a market ceiling. If only potential competition exists, the regulatory test is still unnecessary; if the rates charged by the existing carrier exceeded stand-alone costs, that fact would constitute an invitation to entry by the potential competitors. However, for any shippers that are truly captive, which is to say that the rail carrier faces no effective direct, indirect, or potential competition for their freight, the stand-alone cost does provide an economically rational ceiling.

2.45 No regulatory ceiling is needed to act as a surrogate for active or potential competition from a mode that can operate through the market. Market pressures will enforce the stand-alone cost ceiling since no one will be able to sell at a higher price. Yet another consideration reduces further the likelihood that regulators will have to intervene, except on the rarest occasions, to enforce stand-alone cost ceilings on rates. This consideration stems logically from the very concept of stand-alone cost. If the rates for any service exceed those necessary to cover stand-alone cost, that fact by itself invites the sort of competition that automatically prevents the continuation of such excessive rates.

2.46 The stand-alone cost test does not apply, and cannot be made to apply without disastrous consequences, if railroads are denied the freedom to abandon unremunerative facilities or services. Without such freedom, a railroad cannot earn adequate revenues if it is constrained
by stand-alone cost ceilings on rates in the potentially remunerative portions of its activities. For this reason, any public policy that limits the freedom of railroads to curtail unremunerative services must also provide public funds to help defray the costs of those services.

2.47 The stand-alone cost ensures the equitable treatment of all of a railroad's shippers. By requiring each service or each group of services supplied by a rail carrier to contribute revenues less than stand-alone costs, the test ensures that each shares in the benefits derived from the economies of scope resulting from simultaneity of production. Thus, each shipper is guaranteed some benefit from the revenue that the carrier collects from others. The stand-alone cost offers assurance to each shipper that it will be better off with the existing rates than it would be if it had to fend for itself, as it would have to do in the long run if the rail carrier were denied adequate rates.

2.48 If the price paid by a shipper is no greater than the stand-alone cost of that service, then that price cannot possibly contribute to the cost of any facility from which the shipper derives no benefit. This must be true because the stand-alone cost of any facility used by a shipper includes only the (replacement) cost of those facilities after subtraction of any contributions made by any other railroad customers toward the cost of these services. Thus, together, all the customers that share the use of some facilities will provide revenue contributions that do not exceed the costs of the facilities they use. There will be no excess that the railroad can use to defray the cost of unused facilities. The stand-alone cost test therefore precludes cross-subsidies among the railroad's different customer groups.

2.49 The absence of cross-subsidies under the stand-alone cost test is an appropriate and accepted criterion of equity in the treatment of shippers. Cross-subsidies are properly of public policy concern because they generally lead to a misallocation of resources by encouraging inefficient investment. For the shippers, cross-subsidies may be of concern because they are perceived as unfair. If payments of one group of shippers help make up for shortfalls in payments by another, the first group might well believe it is being forced to cross-subsidize the second. Yet mere payment of a relatively higher rate is not evidence of a cross-subsidy where fixed and common costs must be covered. Rather, a cross-subsidy in an economic sense can occur only if a shipper (or a group of shippers) pays more than the total cost of serving it alone. If no shipper pays more than that amount, differences in their rates simply reflect differing contributions to the common costs of the system, not cross-subsidies.

2.50 Imposing stand-alone cost as a rate ceiling is a form of incentive regulation that avoids introducing distortionary incentives to the railroad with respect to its operations and costing decisions. Since the stand-alone cost is the cost of service by a hypothetical entrant that offers alternatives to the shippers at issue, it is not determined by any of the costs actually incurred by the regulated railroad. Consequently, under the system of stand-alone cost rate

---

3 This important property of stand-alone cost is not significantly undermined by the practice of determining stand-alone cost in a fashion that may provide guidance or even a model of the actual railroad. While these operations may provide guidance or even a model for the operations of the stand-alone railroad, the stand-alone cost need not reflect the same decisions as those made by the incumbent, especially if they lead to unnecessarily high costs.
ceilings, a railroad has no incentive to pad or otherwise increase its expenditures for the purpose of relaxing a regulatory constraint. Further, since the ceilings apply only to services over which the railroad has monopoly power, they do not interfere with the railroad's incentives to pursue aggressively additional traffic and other new business opportunities. Finally, while stand-alone costs may be calculated on the basis of detailed engineering studies and judgments, it is significant to note that they are consistent with the "price caps" that are becoming so popular today, inasmuch as they can be periodically updated on the basis of net measures of inflation and changes in productivity.

**Efficient Pricing and Regulatory Control**

2.51 For prices to be efficient, they must reflect implicitly all of the interdependencies that characterize a rail network. This could be taken to imply that to institute efficient prices for one segment of a railroad's activities (that requires regulatory oversight), it would also be necessary to regulate the prices for all of the railroad's other services. Convincing evidence that such a conclusion is unfounded is provided by the workings of the free market in unregulated industries. In such industries, although there exists no authority that coordinates pricing decisions, compatible and efficient prices nonetheless emerge, their consistency ensured by the forces of competition. This is precisely why free and unplanned markets perform so effectively in comparison with those operated by central planners, despite the latter's alleged ability to take interdependencies into account.

2.52 It is for this reason that no regulatory control need be exercised over the rates of competitive services. Here, efficient prices are automatically imposed by the market, and regulatory intervention can only impede the efficiency of the process of rate determination and resource allocation. Also, relatively little control need be exercised over rates set by a carrier whose revenues are still short of adequacy. If total revenue is not yet adequate, the best rates in terms of the long-run public interest are those that maximize the railroad's net revenues—Ramsey prices. Any railroad with inadequate revenues has powerful incentives to select such rates. In such a case, the railroad as a whole possesses no monopoly power that offers it excessive profits; for individual services for which competition is inadequate, the stand-alone cost test provides the requisite protection to shippers. Under these conditions, there is no possibility of unfair competition through cross-subsidy, with noncompetitive rates increased in order to permit noncompensatory prices in competitive markets. For where the railroad's overall revenues are inadequate, any internally subsidized service must be a drain on the railroad's already insufficient revenues and therefore self-destructive. Thus, where overall revenues are inadequate, only the stand-alone cost test need ever be used in the regulatory oversight of rate setting.

2.53 More than this minimal regulatory scrutiny may conceivably be required if (and only if) a railroad is in a position to earn revenues that are more than adequate. Here, there is at least the hypothetical possibility that high prices for one service will be traded off for price reductions in another. Consequently, it may be desirable to devote regulatory attention to prices for services sold on markets from which competition, direct or indirect, actual or potential, is absent. Yet
even here, the railroads have incentives to select the efficient Ramsey prices. That is, the
interests of the railroads are still likely to be served best by the prices that best serve the public
interest—although it must be admitted that the incentives to select Ramsey prices are apt to be
somewhat less powerful than those in the prevailing case of insufficient revenues.

2.54 There is one principal source of incentives for a carrier capable of earning adequate
revenues to adopt efficient pricing, even though its net revenues are constrained by regulation to
cover only its capital costs and no more. Such a rail carrier is motivated, perhaps more than
other firms in similar circumstances, to maintain its traffic base and to guard against substantial
diversion of its traffic to suppliers already in operation or to potential competitors. This is
because a large portion of a rail carrier’s capital stock is nonfungible, or sunk, so that significant
losses of traffic would cause losses of revenue far greater than the costs that would thereby be
saved. Consequently, a rail carrier with adequate revenues has a particularly compelling
incentive to set rates in a manner that will discourage defections of shippers and market erosion
to competing suppliers of transportation services, in both the short and the long run. It may be
clear intuitively that among the pricing policies that generate adequate revenues, Ramsey pricing
most effectively discourages such defections and market erosion. This is true simply because at
any one time the Ramsey prices yield shippers the greatest total net benefits possible from prices
that yield adequate revenues, and therefore offer shippers the smallest feasible inducement to
divert their traffic.

2.55 In sum, regulation need not take on the overwhelming task of controlling all of a
railroad’s rates, simply to ensure an appropriate choice of prices in those circumscribed arenas
requiring regulatory attention. Elsewhere, the forces of competition and the self-interest of the
railroads constitute powerful mechanisms that can do the job efficiently and automatically using
the crucial demand information possessed by the railroads—which is certain to be more complete
and more accurate than any demand data a regulatory agency could hope to assemble.

Contestability and the Scope and Structure of Regulation

2.56 Contestability is an apt benchmark for the railroad industry. By contrast, the familiar
benchmark of perfect competition is neither attainable nor desirable for the railroad industry, in
which economies of scale and scope are substantial. In this industry, attempts to approximate
perfect competition may in fact be highly inefficient and contrary to the public interest. In any
case, the theory of contestable markets demonstrates quite clearly that neither large size nor small
number of firms necessarily means that markets need function unsatisfactorily. Indeed, a variety
of market forms far removed from perfect competition may perform well for the public interest
so long as such markets are structurally contestable. If an industry is contestable, it is best left
alone without government interference, even if it is composed of a very small number of large
firms. Impediments to entry and exit, not concentration or scale of operations alone, are a
primary source of interference with the public interest workings of the invisible hand.

2.57 Contestability focuses increased attention on entry barriers and their defining
characteristics. High fixed costs and the consequent economies of scale, for example, have
traditionally been considered as impediments to entry; contestability analysis shows, however, that they need not permit excessive profits or prices or any of the other manifestations usually associated with market power. It is the presence of sunk costs rather than economies of scale that is of vital importance for market performance.

2.58 The theory of contestability offers an improved set of rules to be followed by regulators in those cases in which their intervention is appropriate. In addition, it provides economically sound criteria for distinguishing between cases in which intervention by the public sector is warranted and those in which it is not. The theory of contestability is the framework from which the following precepts for railroad regulation (discussed above) were derived:

- Permit a private sector railroad to have freedom of pricing and operations on services that face effective competition in the relevant market, whether from other railroads, other transportation modes, other origins, other destinations, or other commodities.

- Permit a railroad to set prices that are responsive to differences in demands, as well as to differences in marginal costs, and further to enter into voluntary contracts with shippers that have individualized terms, conditions, commitments, and compensation mechanisms.

- Constrain the prices that a railroad sets to captive shippers over whom the railroad has monopoly power, by the stand-alone costs of the shipper's service (or by a comparison of the revenues and stand-alone costs associated with any larger group of shippers' services) and by the stipulation that the railroad's prices do not generate earnings that persistently exceed the railroad's replacement costs, including a competitive return on capital.

2.59 In addition, contestability is a fruitful framework for the analysis of issues pertaining to the vertical structure of an industry. For one thing, in a perfectly contestable market, survival against potential competition requires a firm to undertake efficient vertical relationships and to structure itself efficiently along vertical as well as horizontal and conglomerate dimensions. For another, contestability theory suggests consideration of the idea of separating firms vertically in order to segregate the portions that need regulation from those that do not because of their degrees of competition or contestability.

2.60 This idea emerges from the application of contestability theory to regulatory policy where sunk costs are not pervasive in an industry but centered in a particular sector of its operations, such as the track, way, and structures in railroading. By isolating the activities with which the heavy sunk costs are associated, their need for regulation can be quarantined. By placing relations with the remainder of the industry at arm's length, to the extent permitted by economies of scope, it may be possible to leave the operations of the bulk of the industry safely to the free market, permitting open entry and more flexible pricing, and to draw a regulatory net over only the segment of the activities that are inextricably associated with heavy sunk costs. Thus, contestability suggests a flexible, case-by-case regulatory approach.
Options for Vertical Railway Restructuring

3.1 The historical model of railway operations is the monolithic organization: A single entity controls all facilities and operating and administrative functions and determines what services to provide to significantly captive markets. This railway is an integrated entity that owns and operates its own facilities and vehicles. Typically, the monolithic entity lacks financial incentives and desegregated information on profitability, and is (at best) production-oriented, unresponsive to market demands for services, and hierarchical (if not bloated) in organizational architecture.

The Need for Restructuring

3.2 Although no one would deliberately choose the monolithic railway structure from the standpoint of public interest, it has nevertheless been chosen all too often, either for private interests in monopoly control or for the political benefits that could be collected and disbursed through a state-owned monolithic railway. It is predictable that a state-owned railway enterprise would fail to be responsive to the needs of shippers and would instead be politically responsive, at the expense of providing efficient operations and a stimulus to the economy (see Willig 1994). It is equally predictable that a privately owned railway that was exposed to excessively controlling and economically arbitrary regulation would also lack incentives for efficiency and market responsiveness. Financial deficits would be a natural consequence, as the railway entity failed to attract traffic from alternative modes and geography, as it expended inefficiently on costs, and as it allowed its facilities to suffer from deferred maintenance and replacement.

3.3 The conditions that generated the monolithic railway model no longer exist in most countries, and governments have had to consider fundamental restructuring of both the railway entity itself and the relationship between the railway and the state. The objectives for such restructuring have properly included injecting more innovative and efficient management, reducing railway deficits and burdens of public subsidies, increasing competition with other transport modes, and improving responsiveness to the needs of emergent capitalist enterprises.

3.4 Four generic options can be identified for the vertical restructuring of railways, addressing the set of relationships between the railway entity and other transportation entities (both rail and nonrail), the markets served, and the functions performed. These functions include ownership, improvement and maintenance of the fixed facilities, control of operations such as
dispatching and freight classification, train movement, equipment provision and maintenance, marketing, and financial control and accountability.

- **Option 1: Lines of business organization.** Railway entities can be reorganized and accorded financial responsibility for lines of business to foster comprehensive business planning, market-sensitive and cost-sensitive decisions, and greater responsiveness to demand for various services. British Rail, for example, has divided itself into five lines of business that are financially accountable to top management and that "purchase" service by contract from an operating department that is organized along a matrix of regional and functional lines. By so doing, British Rail hopes to give commercial sectors a profitability objective and to give noncommercial lines of business incentives to reduce their losses.

- **Option 2: Competitive access.** Under this option competing railway companies would have exclusive control over some trackage but also have (and give) the right of competitive access over the trackage of (to) other companies. Some forms of competitive access include joint terminal agreements and conferrals of trackage rights, whereby one railway obtains the right to use the freight-handling facilities or line haul tracks of another railway at a particular location or along a particular route. A further characteristic of this option is arrangements for interlining traffic that is handed off between distinct railroad entities, in their preference sometimes to utilization of trackage rights. In the United States, railroads do a great deal of interlining under terms that are largely unregulated, perform reciprocal switching under terms that are subject to regulation, and exercise trackage rights that are sometimes freely negotiated and that sometimes result from regulatory mandates (that were mostly put into place in the context of settlements of disputes over rail mergers).

- **Option 3: The "wholesaler."** Under this option the railway entity would own and operate the fixed facility and perform all operations on behalf of marketing entities that would be the "retailers." The railway itself would only haul trains, but it would do no marketing to shippers. In Australia, for example, freight forwarders function as retailers using the state railways' "wholesale" services. These forwarders provide multimodal transport and conduct a deregulated trucking business. They control their own rail terminal and yard operations and negotiate on the open market with the railways to charter unit trains with agreed-upon service specifications. This structure permits competition among efficient intermodal "retailers" to flourish, despite a state or private monopoly on railway ownership.

- **Option 4: The "toll rail" enterprise.** Under this option the entire fixed facility, except for exclusive facilities, would be the property and responsibility of one owner. There could be one or more authorized users, which would pay tolls for use of the facility. This approach differs from the competitive access approach (option 2) in the following respect: Under the toll rail approach, separate entities provide the fixed facility and conduct operations, whereas under the competitive access approach, more
than one entity operates in a given market over a particular fixed facility. Sweden has
implemented a separation of fixed facility from operating functions since 1988. The
United Kingdom recently moved in this direction by establishing a separate entity to
hold and manage the rail system's assets associated with the track and road bed. And
the European Union has articulated a policy principle that urges its members to move
in the direction of separating rail operations from fixed facilities.

3.5 It is clear today that a railroad organized and controlled according to the monolithic
model must be restructured in order to contribute best to the economy and to avoid being a
significant impediment to growth and prosperity, becoming responsive to shipper needs and
demands, as well as to marketplace opportunities for innovation. One key element of
restructuring is to develop internal organizations of rail entities that provide managerial
incentives, information, and decisionmaking decentralization that contribute to efficiency, market
responsiveness, and fiscal responsibility. Thus, option 1 is certainly crucial for restructuring,
whatever else is also entailed. It should be recognized that although an internally restructured
railroad enterprise may show lower technical operating efficiency by some traditional measures
(for example, coach-kilometers per locomotive-kilometer), it may succeed in making each
service more responsive to customers' needs and willingness to pay. Economic productivity and
customers' interests are best promoted by minimum total logistics costs, not the lowest railway
rates accompanied by minimum service quality.

3.6 Another key element of restructuring is to unleash the forces of competition to the
fullest extent. It is difficult to predict what are efficient and market-responsive vertical
relationships and combinations of logistical roles among various rail entities, truckers, barge
operators, port operators, warehouses, forwarders, and other players. The U.S. experience
confirms what theory predicts: Decentralized, market-oriented decisionmaking that is freed from
excessive regulatory control and energized by market incentives is the surest means of finding
and implementing efficient, innovative solutions to the problems posed by transportation needs
(see, for example, Baumol and Willig 1987).

3.7 Options 2, 3, and 4 are approaches to restructuring that have the potential for bringing
more competition and more market decisionmaking into the domain of railroading and its
vertical relations. Which of these options is the best choice is a complex policy decision with
many important dimensions that must be considered. The analyses in the two sections that
follow may clarify some of the important considerations.

Analysis of Structural Separation

3.8 Options 3 and 4, which separate ownership of facilities from other rail functions such
as train operations and marketing, have generated much attention of late and deserve serious
analysis. These options have considerable appeal because they seem to mitigate the difficult
problems blocking comprehensive rail deregulation that are associated with the roadbed costs,
which are largely sunk. Fixed costs are large because of the infrastructure (track and stations, for
example) that must be provided before any trains can run on a route. Because duplication of
infrastructure would generally be inefficient, natural monopoly cost conditions characterize physical network provision. These fixed infrastructure costs are largely sunk because the assets are of minimal value for other purposes. For example, embankments and cuttings, the rail formation, and the platforms are fixed in place and committed irreversibly to a specified market. The sunk nature of infrastructure costs creates significant entry barriers, especially where natural monopoly conditions also exist.

3.9 The cost conditions relating to the operation of services on the physical network, by contrast, may be more consistent with active and potential competition. To operate a service it is necessary (at least) to have trains, staff, support, and rights-of-way. Although there are inevitably some sunk costs in hiring staff and buying or leasing rolling stock, they are small in relation to the massive sunk costs of establishing network infrastructure. Locomotives and freight cars constitute capital on wheels, and most of their cost might be easily and quickly recovered by rolling them to other markets.

3.10 If ownership of track and trains were separated—with the track assets held by the government, by a consortium of the operators, or by a regulated private entity—there might be vigorous active and potential competition over railway services provided by operators with equal access to the utilization of the roadbed. There would be no need to regulate these operators, who would have all the powerful incentives that accompany competition to be efficient and responsive to the needs of shippers and a growing entrepreneurial economy.

3.11 Several links in this chain of policy reasoning, however, may be inapplicable or wrong in a given set of realistic circumstances:

- The provision of many innovative and market-responsive rail services may require specific investment in infrastructure, such as maintaining or upgrading way and structure facilities, constructing loading and transshipment facilities, and building spurs of track to reach a shipper's location. It may be difficult and inefficient for any operator (or retailer) to coordinate, as necessary, with the infrastructure monopoly (or wholesaler) entity, especially if their incentives with respect to investment behavior are not in harmony. The investment incentives of the infrastructure monopolist (or wholesaler) will, of course, depend critically on whether it is a state-owned entity or, if it is in the private sector, on the character of its regulation.

- Efficient, safe, and delay-minimizing utilization of track and yard facilities by trains, cars, and shipments requires close coordination in accordance with priorities that are driven by considerations of both operations and shipper sensitivities. Rival operators (or retailers) will compete vigorously and acrimoniously over scarce or congested infrastructure facilities (or wholesaler services). Constantly sorting out their claims will be important for the overall efficient and responsive operation of the rail system. This task would be difficult enough for an unintegrated system with a monopoly infrastructure entity; it seems virtually impossible to accomplish efficiently where there exist rules against discrimination and infrastructure (or wholesale service) pricing that is either tightly regulated or, for a state enterprise, politicized.
• The freight hauling operations on all or part of the rail system in question may well constitute a natural monopoly, even when disintegrated from the infrastructure. The economies of scale and scope that arise from running long trains, from blocking many different shippers' freight in classification yards, and from efficient utilization of yard facilities, crew, and rolling stock are all associated with operations rather than infrastructure. Consequently, a separated operations firm may be a monopoly, and it may have considerable market power unless potential competition is a powerful force.

• For potential competition to be powerful, an entering operator must perceive that significant sunk investments in rolling stock and in specialized facilities can be avoided. Locomotives and freight cars may indeed be an example of capital on wheels so long as they can be transported to alternative points of gainful utilization without substantial costs. While this is likely to be the case for services provided in the middle of a landmass with a rich rail network ready to accommodate the cars, it may not be the case for more specialized cars or for a more isolated market. Also, the entering operator may not have yard, loading, car maintenance, or spur facilities available unless it makes new and significant sunk investments. For these to be available on equal terms with the incumbent operator, it must be the case that the infrastructure entity made the needed investment as part of its role in the system. But the greater the entrepreneurship and risk-taking investment that the infrastructure entity (or the wholesaler, under that option) must undertake, the less is gained by the separation, since the infrastructure (or wholesaler) entity is either a state-owned or a tightly regulated private sector monopoly.

• Efficient pricing to cover replacement costs is made more difficult by separation. Where economies of scale are important, efficient pricing to cover replacement costs requires that shipments of different commodities on different origin-destination routes bear prices with different relationships to marginal costs. If it is the case that the operator (or retailer) firms can readily evade price discrimination by the infrastructure entity (or wholesaler)—so that different prices cannot be collected by the infrastructure entity (or wholesaler) for facility utilization (or for wholesale service utilization) by different shippers of different commodities—then it will be difficult if not impossible for the costs of the infrastructure to be defrayed by Ramsey prices. At the extreme, a regulated infrastructure (or wholesaler) entity charging competitive operators (or retailers) an equal price for each ton or each ton-mile of freight that utilizes each of its facilities is, in essence, recreating a system in which prices are set according to fully allocated costs. As discussed above, such pricing can be a prescription for inefficiency and financial disaster.

3.12 Thus, it is clear that separation of operations from infrastructure in a railroad system is no panacea for regulatory problems. Instead, as a policy direction, it must be compared with the leading alternative.
Analysis of Competitive Access

3.13 Option 2 is most clearly distinguished from the separation options just discussed by the fact that the competitive access option allows integrated operations by the rail entity. It is superficially easy, albeit mistaken, to identify an integrated carrier with the case of the monolithic carrier, because it is tempting to jump to the conclusion that an integrated carrier would make it difficult for other entities to participate in its business. This option implies a requirement that the integrated carrier make its facilities available to other entities on a "fair and equal basis." However, if the integrated carrier has strong incentives to keep other entities out, it is unclear how effective such equal access mandates are likely to be. In the United States the rail industry, like other regulated industries (for example, gas pipelines, telecommunications, and electric power), has seen many disputes with claims of "unfair" and "unreasonable" exclusion from a carrier's facilities, despite rules of equal access.

3.14 Thus, an assessment of this option must include an analysis of the incentives of the integrated carrier to accommodate others wishing to participate, and able to participate efficiently, in the provision of service. If the integrated carrier is regulated in a fashion that permits the carrier to charge higher prices to captive shippers if it does more of the business, then the carrier clearly would have incentives to exclude other participants.\(^4\) Likewise, if the integrated carrier is constrained by regulation in the amount it can earn from the portion of service it provides when it does cooperate with another entity, then the carrier has incentives to undermine or avoid efficient cooperation in order to enlarge its portion of service.\(^5\) In addition, the integrated carrier would be motivated to exclude an efficient participant if by so doing the carrier would weaken, in a predatory manner, the competitive impact of that entity in another market. Under classic rate-of-return regulation or under a system of regulated "divisions" specifying what an integrated carrier can earn from a cooperative movement—both features of U.S. rail regulation at one time—an integrated carrier does have incentives to undermine efficient cooperation.

3.15 In sharp contrast, under the regulatory system that has been described above as serving the public interest well, an integrated carrier would generally have a real profit motive to cooperate with an efficient participant in its business. Here, it is not "divisions" that are specified by regulation, even on the service provided to a captive shipper. Instead, the described stand-alone cost rate ceiling applies to the price charged to the shipper, and cooperation with an efficient entity enlarges the pot of returns available from the service, enabling more money rather than less to be earned by the integrated carrier. Consequently, except for the rare possibility of predation, an integrated carrier would have ordinary business incentives to find and cooperate with efficient participants in its business and to negotiate mutually beneficial terms. This is just a railroad version of business "make-or-buy" decisions in other industries.

\(^4\) One example of this effect arises under rate-base rate of return regulation, as was understood by Averch and Johnson (1962) in their seminal paper.

\(^5\) For a more complete discussion of these cases, see Ordover, Sykes, and Willig 1985.
3.16 Despite the prevalence of efficient incentives on the part of integrated carriers under the form of regulation described here, it is useful and wise to augment the system of regulation with a fallback set of standards to apply should disputes about predation through competitive access arise. In short, an integrated carrier that possesses a "bottleneck"—a facility without which the complainant cannot reasonably offer its services to the shipper—should not exclude the complainant by refusing an agreement that would be fully compensatory of all its costs, including opportunity costs. For example, if another carrier, or an operator, sought to participate in a freight movement that represented new business for the integrated carrier, then it is to be expected that the latter would negotiate in good faith and not exclude the other entity if an agreement could be found that would at least cover the incremental costs of the integrated carrier. If another carrier sought to handle some freight part of the way that the integrated carrier would otherwise handle itself, then it is to be expected that the integrated carrier would accept an agreement that earned it a larger net contribution of revenues above incremental costs than it would earn if it handled the freight without the other participant. Here, the contribution that the integrated carrier would earn on its own is part of the opportunity costs it faces from cooperating with the other participant. These same principles apply to interlining, trackage rights, car hire, or any other form of cooperation or participation through the use of a bottleneck.

3.17 "Efficient component pricing," or "parity pricing," are alternate names that have been given to the principle that an integrated carrier should offer the services of its bottleneck at a price that yields it the same contribution that it would earn from performing the end user's service itself. Behavior consistent with this pricing of bottleneck services, or more generally with the antipredation rule just articulated, leads to efficient vertical relations and is thereby consistent with nonpredatory incentives under the regulatory system we have described. Such pricing of bottleneck facilities does not place additional competitive pressure on pricing to shippers, since it is based on the contribution that could be earned from the shipper's service at the extant shipper's price. However, it does generate incentives for efficient combinations of transport services to make it to the market; it does provide quality and cost competition among potential and actual participants for the role of being part of the efficient combination; and it does help to ensure that those with efficient innovations in logistics or in marketing of transport services will be able to work with carriers to implement their ideas.

Separation Versus Competitive Access

3.18 The primary virtue of separation as a policy option is that it may permit active or potential competition to reign among rail operators or retailers—with corresponding assurance of efficient selection among them for provision of their services at efficient prices. At best, separation will accomplish this end, but it will leave unresolved the difficulties with regulating the provision of the services of the infrastructure, or bottleneck, assets of the railroad network. Prices charged to shippers will be at least the sum of the competitive prices for the services of the

---

6 This standard was first developed in Ordover and Willig 1981.
operators (or retailers) and the regulated prices for the services of the infrastructure entity (or wholesaler). They are unlikely to be fully Ramsey-efficient prices for the coverage of replacement costs because of the difficulties of reflecting shippers' differences in demands in the prices charged for infrastructure services. At the same time, separation may create serious coordination problems, loss of economies of scope, and otherwise unnecessary transactions costs. In addition, rail operators may not face effective active and potential competition, undermining the potential for realizing the primary benefit of the option.

3.19 The competitive access option could also be fraught with problems when the incentives of bottleneck holders are adverse to efficiency and competition. A variety of solutions to competitive access problems have arisen in industries seeking to replace regulation with competition. Typical examples include mandatory interconnections with competitors and line-of-business restrictions in the telecommunications industry, " unbundling" of the transportation and energy components of price in natural gas markets, and equal access to marketing channels (for example, computer reservations systems) in the airline industry. In designing rules that govern vertical relationships among competitors formerly subject to economic controls, regulators must address a common basic problem: How to implement pricing and terms of access by "nonintegrated competitors" to the restricted portions of the network so that competition on the merits will work to ensure that the efficient alternatives successfully participate in the provision of end users' services. The compensation for and terms of access should not distort the process by which prices are adapted to consumer preferences and demands for transportation service. Prices should be sufficiently high to be compensatory to the "landlord" railroad yet not so high as to preclude efficient operations by the "tenant" railroad. Where incentives are significantly adverse to these goals, experience teaches that rules are too easily evaded and disputes seemingly never-ending.

3.20 It is thus fortunate that under rail regulation that focuses on the levels of rates charged to shippers—rather than on other prices, such as those charged for access to bottleneck services—incentives are generally for the promotion of efficient vertical relationships. As a result, if integration is permitted under this system of price regulation, then the outcomes are predictably consistent with efficient participation by the integrated carrier and by other, nonintegrated carriers as well, on terms that permit compensatory support for the efficient participants. Further, prices to shippers can be selected in accordance with Ramsey efficiency, even as they are constrained by regulation where the carrier has monopoly power. Moreover, unlike the virtues of separation, the efficiency of the outcomes of competitive access does not depend on the absence of economies of scope, on the absence of coordination problems without integration, and on the competitiveness or contestability of rail operations.

3.21 Separation of track assets from operations is likely to be a particularly attractive option where a dense and extensive rail network permits many operators to function and to provide both active and potential competition to each other. Another favorable factor is a mature and well-developed set of fixed facilities, so that there is relatively little extent to the domain of new infrastructure investments, where incentive problems are more likely to arise. Where this factor does not apply, it will be important for regulation of the infrastructure entity to permit it to
enter into medium- or long-term contracts with shippers or with operators that themselves have contracts with shippers, so that the risks and rewards from investments can be efficiently shared by shippers, operators, and the infrastructure entity. The impediments to Ramsey pricing that separation might cause would be rendered insignificant to the extent that the infrastructure entity does not attempt to recover its sunk capital costs from "tolls" levied on traffic. If the infrastructure entity is expected to seek recovery of its replacement costs, then it should be permitted and even encouraged to implement forms of price discrimination that help to bring shippers' prices in line with principles of Ramsey efficiency. Finally, there may well be circumstances where a monolithic railway system cannot be converted to one with functioning competitive access because of embedded business culture and entrenched management. Here, the act of separation is so revolutionary that it may unsettle the business culture in a productive fashion and force reassignments of management that permit implementation of the necessary internal reorganizations of responsibilities, roles, incentives, and information flows.
Concluding Remarks

4.1 This paper has outlined a set of principles that together add up to a program for restructuring the relationships between government and railroad entities. These principles point toward a great deal of reliance on market forces to shape prices and logistics of services. At the same time, the principles include economically appropriate protections for any captive shippers and for any carriers that may be excluded or foreclosed from participation for anticompetitive reasons.

4.2 On the subject of restructuring, we have pointed out that internal managerial reforms are necessary, as are policies that address railway vertical relationships. The two leading candidates—separation of track from operations in different business entities, and incentives and fallback rules for competitive access—were compared on several dimensions, and their relative levels of appeal were found to depend on a variety of characteristics of the business environment.

4.3 Restructuring along the lines suggested here, providing a greater emphasis on marketing effectiveness, can be expected to result in a more profitable railway that is better able to cover the costs of its commercial services. Any noncommercial services that are needed should be carried out on the basis of an explicit agreement between the railway and government that views public service obligations as a business relationship between a customer (government) and the contract supplier (railway). This agreement would help to ensure that noncommercial services are more effective in fulfilling public policy objectives, remove an insuperable drain on revenues that would condemn the railroad to insufficient investment, and eliminate cross-subsidies that hamper the railroad from maintaining its competitive position against other modes.
Appendix A

Technology and the Structure of Railroad Costs

The output of the rail industry is multidimensional. Railroad firms produce different types of transportation services for different users at different origins and destinations at different times and at different levels of quality. The mix of output and shipment characteristics can have a major impact on the costs of any given firm. For example, railroads specializing in coal traffic have very different cost characteristics than those specializing in movements of general manufactured commodities.

The most striking feature of the cost structure of railroads is the high incidence of costs that cannot properly be attributed to any particular service at any particular point in time. That is, a significant portion of costs are incurred on behalf of several activities and do not vary with the amount of the service provided. These unattributable expenditures reflect both joint and common costs. Common costs are costs shared by two or more services in variable proportion. For example, a terminal represents a common cost; it is used by different services in varying proportions. More generally, the outlay on track and way and structures between points A and B is a common cost for all movements of whatever commodities are shipped between A and B over that route. Joint costs are costs shared by two or more services in fixed proportions. A backhaul movement is the classic railroad example.

The structure of railroad costs has important implications for the competitive structure of rail markets. It is sometimes mistakenly inferred from statistical evidence of constant returns to firm size that a competitive equilibrium with marginal cost prices covering total costs would be sustainable in the rail industry. Such reasoning neglects the critical fact that indivisibilities in rail technology make increasing returns to scale in total costs endemic and competition among a limited number of players inevitable. A rail link between two points requires lumpy investment in way and structures with associated highly significant economies of traffic density. Unit costs fall with output, letting all factors of production vary on a given route or route structure.

**Fixed and Variable Costs**

A fixed cost is one that is necessary to provide a service or group of services, and whose magnitude does not vary with changes in the quantity of a service provided. For example, if a railroad is to run between points A and B, a minimum outlay on track and roadbed must be incurred, even if the trains run virtually empty. Even in the longest run, the roadbed cost cannot be reduced to a negligible level if service is provided. Similarly, a loading facility may be necessary to transport coal efficiently between points A and B, but its cost may be unchanged if the amount of coal transported is doubled or halved. Common costs are often fixed. For example, the basic portion of the outlays on track and way and structures between A and B may be both fixed and common costs.
Fixed and common costs are quite different from variable costs. Economists employ two fundamental cost concepts in defining variable costs—marginal cost and incremental cost. The marginal cost of a service is the additional cost that would be incurred to supply an additional unit, or the saving in total cost by supplying one less unit. As such, the marginal cost of a rail service is the per-unit opportunity cost to the rail carrier of the level of a service's volume. The term "opportunity cost" refers to the value a resource could contribute if it were used in some alternative occupation rather than the one to which it is currently assigned by the railroad. Thus, marginal cost is similar in meaning to unit incremental cost and to the true economic variable cost. However, its definition makes clear that marginal cost should include the traffic-sensitive costs of capital facilities that are fungible and economically attributable to the service, as well as the more obvious cost components such as fuel, labor, and traffic-sensitive maintenance and replacement costs.

For example, locomotives and other rolling stock used to provide a particular rail service have a significant opportunity cost for a rail carrier. If not utilized to supply the service, they could instead be gainfully utilized elsewhere in the rail network, by either the rail carrier at issue or another carrier. Assuming that at least some carriers do not have excess supplies of the equipment in question, or their functional equivalents, a decrease in the quantity supplied of the service would release equipment that could decrease or delay the need to lease or purchase stock for replacement or expansion. Consequently, it follows that the opportunity cost of the rolling stock is its replacement cost, at the current cost of capital. Thus, the marginal cost of a given service includes the costs of fungible capital goods that are utilized, such as locomotives and other rolling stock, at the current cost of capital for the period of time during which they were so employed.

Of course, the marginal cost of a service also includes the wear and tear on capital assets and the required maintenance expenses that the supply of the service causes. (For example, it is clear that the passage of rail traffic causes wear and tear on track, ties, and ballast, which in turn shortens the lives of the assets. Consequently, one element of the marginal cost of rail traffic arises from the hastening of the time in which the assets it utilizes must be replaced—the present discounted value of the capital cost of the assets' value—over the time period that their needed replacements are advanced.) However, the costs of facilities that are fixed or common are not included in the service's marginal costs.

The incremental cost of a service is the cost per unit of service necessary to provide the entire service, or the cost avoided by not providing the service, given all the other services supplied. The term "avoidable cost" is also used to describe the cost per unit of service that could be avoided by not providing a particular service.

The important conceptual point here is that a railroad's total costs are composed of some costs that vary with the amount of a particular service provided and of others that do not. This is obvious enough, but considerable confusion is often engendered when the additional point is made that in the long run virtually all fixed and common costs can be varied. The reason is simply that in the long run virtually all assets must be renewed or replaced. At the date when
the decision regarding renewal or replacement of the fixed factors of production required to supply a service or group of services is under consideration, the costs involved are incremental to that service or group of services. If the decision were made to discontinue those services, the costs would not be incurred.

This obviously does not mean that there is no economic distinction between variable costs and fixed and common costs. What it does mean is that the perspective of the decisionmaker is very important. When a railroad is making decisions regarding the incremental costs of adding a particular service (or the avoidable costs of eliminating a service) given existing capacity, the short-run variable costs of service will include only the additional costs of production imposed by that service. Rarely will the short-run variable costs of service include the full measure of long-run fixed costs. In contrast, when a railroad is making the long-term decision whether it is economic to replace a portion of its rail network (or to make an entirely new addition to its network), the long-run variable costs of the service or services the railroad plans to offer will include all the fixed costs that will become sunk (that is, irreversible for a significant period of time) once they are incurred.

_Sunk Costs_

Long-run fixed costs are those costs that are not reduced, even in the long run, by decreases in output (see Baumol and Willig 1981; and Baumol, Panzar, and Willig 1988). But such costs can be eliminated in the long run by total cessation of production. Sunk costs, in contrast, are costs that (in some short or intermediate run) cannot be eliminated, even by total cessation of production. As such, once committed, sunk costs are no longer a portion of the opportunity cost of production.

Sunk costs need not be fixed, and even more important, fixed costs need not be sunk. To operate with current production techniques, a railroad requires at least a locomotive and one car, the costs of which must be included among its fixed costs. Yet because they constitute capital on wheels, most of their cost can easily and quickly be recovered by rolling them to another market, should the railroad's management decide (and be permitted) to close down the line in question. Thus, little or none of this portion of fixed cost is sunk, in contrast to the roadbed cost, which typically is sunk. While bridges, ballast, rails, and ties can also be moved from one route to another, they can be moved only at considerable expense.

The distinction between sunk and fixed cost is not a mere technological quibble. It makes a substantial difference for the design of appropriate public policy if the costs of the firms in an industry include the one rather than the other. Sunk costs contribute to entry barriers which, as is well known, can give rise to monopoly profits, resource misallocation, and inefficiencies. By contrast, fixed costs neither constitute barriers to entry nor entail the misallocation problems to which entry barriers lead. Fixed costs are not, and do not raise, entry barriers unless they also happen to be sunk.
Economies of Scale and Scope

The issue of whether a firm's total costs will be recovered from prices that are equal to the firm's marginal costs of supply is logically equivalent to the question of whether the firm's operations are characterized by economies of scale, or, in alternative terminology, increasing returns to scale.

For multiproduct railroad firms, economies of production could exist due to either the level of supply of all the firm's outputs (economies of scale) or the breadth of the set of services supplied (economies of scope). Economies of scale are exhibited where equiproportionate changes in the levels of all services provided would require a less-than-proportionate change in the level of efficient costs. In addition to economies deriving from the size or scale of a firm's operations, cost savings may also result from simultaneous production of several different outputs in a single enterprise, as contrasted with their production in isolation, each by its own specialized firm. That is, there may exist economies resulting from the scope of the firm's operations.

Substantial economies of scale in the provision of some rail services, whether focused on particular routes or types of freight, result from the heavy fixed costs associated with rail operations. To transport even small amounts of freight, a railroad must generally incur the costs of track, right-of-way, locomotive power, crew, and certain facilities. These costs do not rise proportionately with traffic volume. As more traffic uses a section of a roadway, very few additional fixed costs are incurred, and the extant costs are spread over more traffic. A single track line can handle large amounts of traffic before a full second track must be added or advanced signaling systems installed. Scheduled trains can be made longer to accommodate more cars on the same origin-to-destination route without proportional increases in the costs of locomotive power and crew. The more freight that is scheduled to traverse the same route, the larger can be the preblocked movements, with correspondingly less reclassification yard activities and time needed, and with more opportunities to run efficient through-train service. In short, additions to the levels of rail services supplied do not entail proportionate additions to the levels of expenditure required for fixed plant, for equipment investment, and for operating expenses. This is precisely the hallmark of economies of scale. Fixed costs, of both the sunk and fungible varieties, per ton of freight fall as traffic volume increases. Cost efficiencies therefore may be associated with provision of rail services by a single carrier.

Another advantage of firm integration in the rail industry arises from potential economies of length of haul. With fixed terminal expenses, longer hauls normally imply lower costs per mile. In the presence of such economies, a railroad with an integrated nationwide system will sometimes have a cost advantage over competitors that make and accept interline shipments to and from other railroads.

Increased firm size may convey cost advantages because of specialization and massed-reserves economies. A large firm may employ a more richly specialized array of accounting, finance, marketing, engineering, research, and legal talent than a smaller competitor. This specialized talent may be reflected in lower administrative costs, higher productivity, or
both. The large firm can amass its cash balance reserves and spread production, market, and financial risks over a larger volume of activity. The diversification of the portfolio of transportation services offered by a large railroad creates an overall system risk factor that could be substantially less than the risk associated with investment in just one of those services.

A large railroad firm with an integrated network may also realize economies in equipment investment. In general, railroads attempt to minimize the need for new equipment purchases by using equipment interchangeably throughout the system. When cars and locomotives are needed at some shipping point, the railroad can immediately send them out of the most convenient distribution center. Operations with assigned equipment require more switching than those that draw their equipment from common pools. In addition, the ability to use locomotives interchangeably reduces the number of reserve locomotives needed to protect against equipment failures, repairs, and inspection. A larger railroad firm can, therefore, obtain the same degree of protection at lower cost relative to total capacity carrying costs.

Another pertinent feature of the railroad industry is that substantial economies of scope result from the common costs of rail operations. Outlays on rails, ties, rights-of-way, yard facilities, locomotion, and train crews are among the many common costs of rail operations incurred in carrying a variety of types of freight between a variety of origins and destinations. These shared costs confer economies of scope on carriers offering a multiplicity of transportation services: A carrier that provides an array of services can do so at a lower total cost than a set of carriers producing each service separately.

**Economies of Size Versus Economies of Scale and Density**

The overall size of a railroad is likely to be quite independent of the amount of traffic that travels on any of its routes. That is, a large firm may have short or long hauls and high or low traffic densities between different points. There has been serious confusion about economies of size and economies of scale and density, and a concomitant failure to specify clearly which is being measured (see Harris 1977). Economies of scale are carefully defined to refer to a long-run average cost curve that declines as the quantity of the firm's output of a given collection of services increases.

Comparing the average costs of railroads that have different sizes of route networks, as many have done, does not provide information directly relevant to economies of scale, because such railroads do not supply different amounts of a given collection of services. Instead, they likely offer quite different collections of services as a result of their different route mileage and architecture. The correct and relevant measure in railroading is the extent of scale economies that relates to traffic volume on each route, rather than to traffic volume over an entire and possibly growing system. To emphasize this point, these economies have come to be termed economies of density. Thus, the critical determinant in pricing and (dis)investment policies is whether there are economies of density. It is therefore important to assess the degree to which unit costs decline as output increases while holding the route system, or miles of rail line,
constant. A small firm with high traffic density could potentially have lower average costs than a large firm with low density.

Economies of density are normally attributed to declining average capital costs. However, the provision of rail service entails more than simply installed capacity; it includes minimal (and often indivisible) amounts of crew, engines, maintenance, and other variables. Indeed, recent empirical studies indicate that the maintenance of way and structure and transportation expense (mainly fuel and crew wages) account for a significant portion of the estimated economies of density. Approximately two-thirds of these economies are due to variations in unit operating costs per route-mile.

Under significant economies of density, the cost-minimizing market structure for a given route might call for a single firm—that is, the route would be a natural monopoly. In the absence of any other scale economies, the national railway system could be made up of a large number of small firms, each with a local monopoly. Alternatively, if there were substantial economies of firm size without economies of traffic density, it would be economic to have a number of integrated nationwide railroads that competed on all their routes. However, with economies of density, and with economies of scope, and with some economies of end-to-end long hauls, the cost-effective structure of the rail industry is likely to be characterized by very few firms.

**Empirical Evidence on Scale Economies**

There are at least two approaches to measuring cost-scale relationships in the rail industry. The first way is to use the expertise of those with intimate knowledge of railroad operations in ascertaining whether the costly inputs required to supply rail services must be expanded proportionately to accommodate increases in the quantities of services provided. This is known as the engineering approach. The second approach, statistical cost analysis, is to estimate econometrically the relationship between railroad costs and the levels of rail services provided. There is no conflict between the conclusions reached by using these two different approaches in the railroad industry. Both indicate quite clearly that railroad operations are characterized by increasing returns to scale and that the recovery of railroad costs consequently requires that prices exceed marginal costs.

The first approach has been followed by a long succession of industry observers, who have provided a knowledgeable overview of how economies of scale arise in rail operations. First, economies are created for the system as a whole by operations that are directly common to all traffic, such as network planning and management. If network management and control (for example, billing, payroll, systemwide insurance, and other housekeeping functions) involves a fixed cost regardless of network size (above a certain threshold), these costs will be spread over a larger user base in a larger integrated rail system. Similar integration economies arise in communications and dispatching activities and from increases in workforce specialization within the repair facilities of larger systems. Finally, large railroads benefit from capital-raising and other pecuniary economies (for example, price concessions from suppliers). Indeed, this appears
to be one of the most persistent advantages of firm size, with small incremental capital cost savings enjoyed out to very large scales. However, the capital-raising economies of scale are also associated with real resource savings. Negotiating a loan or a new stock issue or obtaining necessary regulatory clearances entails transaction costs, some portions of which are nearly fixed. Clearly, the larger the issue, the lower those costs per unit of capital raised.

Second, the integration of the railroad system permits economies that directly benefit some traffic and indirectly benefit other system activities. Most ancillary plant (for example, storage and marshaling yards, sidings, switches, and fueling and repair stations) can be utilized by more and more shippers without causing a corresponding increase in the amount of investment required. A coal shipper might need a storage and marshaling yard to hold its cars until a trainload volume is accumulated. If a mine produces only 20 carloads a day and holds them until 100 cars are available, a yard that could store and switch 100 cars would be required. However, on an independent operation basis, only 20 percent of the yard would be utilized in the first day, 40 percent in the second, 60 percent on the third, and so on. Yet a railroad that connected with more mines might receive 20 cars a day from each of five mines and send a trainload every day. The railroad would still need only a 100-car yard, but it would have five times as many cars to share in the coverage of the investment and operating costs of the yard.

Similarly, a full siding is necessary if each day only one train will meet one other train coming in the opposite direction. The same size siding would be necessary if four trains were meeting four other trains at the same place. Crossing protection must be built and maintained in a densely populated area whether the railroad sends one train a day or three trains a day over the track at the crossing. The same is true for switches, fueling stations, and all other fixed plant investment. Once the plant is installed, a railroad can utilize it far more heavily with very little additional fixed investment cost. Also, a train of 40 cars needs a crew of the same size as a train of 60 cars. The ability to marshal cars of different shippers into a larger train also cuts other operating expenses. The engine power necessary for a longer, heavier train is not commensurate with the additional cars that have been added.

The statistical or econometric approach to analyzing railroad economies of scale also has a long history. This history is rife with academic controversy and with steadily improving research methods. For example, some econometric studies, because they were founded on arbitrary allocations of costs between freight and passenger services, found no evidence of rail economies of scale. Other studies failed to distinguish economies of scale from possible economies stemming from the geographic extent of a railroad's operations. Such studies, finding that railroads covering more territory do not necessarily enjoy lower costs per ton-mile of freight, incorrectly concluded that increasing returns to scale are absent.
Recent econometric studies conducted in the United States have avoided these pitfalls, and their important conclusions warrant discussion here. First, most of the rail system is subject to increasing returns to scale and has elements of natural monopoly, whether considered in a single-product or a multiproduct setting. Second, as Figure A.1 indicates, although unit costs decline sharply with density, at some point between 35 million and 40 million annual gross ton-miles per route-mile, depending on the commodity mix, the cost curve flattens out and a large part of the traffic in the system flows over this range of flat (constant) costs. This range (flat part) represents the level of minimum efficient density, which one can think of as the capacity of a single track between two points, the fundamental indivisibility in the rail cost structure. Higher traffic density can be served at approximately constant cost by adding segments of parallel track and signaling devices. Third, for very short-haul, terminal-oriented railroads, the long-run cost curve seems to flatten out much sooner (at less than 2 million net ton-miles per route-mile). Fourth, there are considerable economies for longer hauls.

Overall, these studies establish the presence of substantial economies of scale in the freight operations of railroads. They indicate that pricing at short- and long-run marginal costs would recover less than 80 percent of total long-run costs. Also, high-density traffic seems to exhaust the economies of scale experienced at lower densities, but significant diseconomies of scale do not occur as densities grow larger. Consequently, since all railroads have relatively low-density traffic on many segments, and since most traffic flows on low-density track while it is gathered and distributed, rail services exhibit substantial economics of scale overall. As a result, prices set at marginal costs would leave uncovered a substantial portion of total efficient railroad costs.
Appendix B

Rail Costs, Profitability, and Structural Changes

Most of the statistical and econometric studies estimating rail costs and production functions suffer from two fundamental weaknesses. First, they generally fail to differentiate between way-and-structures capital, which is a measure of the quantity and quality of the capital utilized in the roadbed, and track, which in addition to being a proxy for the roadbed capital is also a measure of common carrier obligations to haul commodities. Second, they generally fail to take into account the effect on costs of the route network and to differentiate between high-density, fully utilized track and light-density, underutilized track (see Friedlaender and Spady 1980).

Way-and-structures capital is a measure of the capital utilized in the roadbed and as such should be treated as a conventional factor of production. An increase in the fixed factor, way-and-structures capital, should lead to a reduction in other factors and hence a reduction in variable costs. In contrast, general track and low-density track should be treated as technological variables that affect the costs of the railroad firm in a way that is not necessarily associated with conventional production theory. An increase in low-density route-miles or total track represents an increase in common carrier obligations and should therefore be associated with increases in expenditures on other factors of production.

A ceteris paribus reduction in way-and-structures capital will reduce the quality of the existing track and hence lead to cost increases by requiring increased amounts of variable factors. This is to say that more money must be spent on equipment maintenance and train crews as the quality of the roadbed deteriorates and speeds are reduced. Similarly, a ceteris paribus reduction in track will be correlated not only with a reduction in common carrier obligations and improvements in the quality of the existing track, but also with increases in its utilization. The first two considerations will tend to reduce costs whereas the latter will tend to increase them, making the impact of reduced track somewhat ambiguous. Reduction in low-density track, in contrast, will reduce common carrier obligations and their associated costs and will therefore tend to generate cost savings.

Railroad Costs and Infrastructure Variables

To assess the possible savings that would accrue from policies aimed at changing the railroad infrastructure, it is important to quantify the impact on rail costs of changing the three main infrastructure variables—the amount of way-and-structures capital, general track, and low-density track.

Ceteris paribus increases in way-and-structures capital will raise the amount of capital embodied in each mile of track and thus lead to reductions in variable costs. Indeed, econometric estimates by Friedlaender and Spady (1980) reveal that a 10 percent increase in way-and-structures capital leads to more than a 4 percent decrease in variable costs, consisting of
decreases of 11 percent in equipment usage, 3 percent in general labor, 3 percent in yard and switching labor, 2 percent in on-train labor, and 0.6 percent in fuel and materials. These estimates seem to indicate that the main effect of an increase in way-and-structures capital is to decrease equipment requirements, with somewhat lesser savings in the labor categories. This confirms the intuition that the source of the savings in variable costs that result from an increase in way-and-structures capital is train speeds.

Ceteris paribus reductions in light-density track are correlated with increases in the amount of capital embodied per mile of track and reductions in the proportion of low-density mileage; both of these factors should be associated with cost reductions. Econometric estimates (Friedlaender and Spady 1980) indicate that a 10 percent reduction of low-density route-mileage would reduce total variable costs by approximately 3 percent. This reduction comes about through reductions in yard and switching labor costs of somewhat more than 4 percent, in general labor and equipment expenditures of somewhat more than 3 percent, and in fuel and materials expenditures of less than 1 percent. Thus, the primary savings arising from the abandonment of low-density line are concentrated in transportation and switching categories associated with moving trains over lightly utilized track.

Finally, ceteris paribus reductions in general track are correlated not only with increases in capital embodied per mile of track, but also with increases in the proportion of low-density track. While the first factor should tend to reduce costs, the second should increase them. Econometric estimates (Friedlaender and Spady 1980) reveal that a 10 percent reduction in general track or route-miles leads only to a reduction of total costs of less than 1 percent. In terms of factor utilization, reductions in general route-miles lead to sizable reductions in equipment and materials expenditures but increases in labor expenditures. Thus, as the same volume of traffic is moved over a smaller network, increased expenditures on labor and switching are required, whereas savings on fuel and equipment are achieved.

**Low-Density Lines and Profitability**

Rail costs are quite sensitive to changes in way-and-structures capital and in light-density route-miles but not to changes in general route-miles. A change in general track or route-miles without a concomitant change in low-density route-miles has a small impact on variable costs but a significant effect on factor intensities. What distinguishes the provision of low-density service from that of general network expansion is the greater labor intensity of the former. Thus, efforts to adjust amounts of way-and-structures capital through roadbed maintenance or to abandon light-density lines are likely to have a rather large impact on costs, whereas the abandonment of general track per se will lead to relatively few economies.

Econometric estimates (Friedlaender and Spady 1980) reveal quite clearly that low-density lines are a significant drain on railroad profitability and seriously impede the attainment of static and dynamic efficiency in the industry. The avoidable losses recoverable by abandonment appear to be quite significant. In addition, the burden of excess capacity seems to have a dynamic impact on efficiency. The abandonment of low-density lines stimulates the
formation of new capital on the high-density portions of the rail network. First, since abandonment reduces the need for cross-subsidization, rates on the high-density lines are permitted to fall toward marginal cost. The lower rates attract additional traffic and thus raise the level of desired capital. Second, the abandonment of low-density lines lowers the cost of capital to rail firms by improving their long-run profitability and reducing the risk of bankruptcy.
References


