

Economic Regulation of Transport: Principles and Experience

Ian Savage
Northwestern University

Professor Ian Savage
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, Illinois 60208

Phone: (847) 491-8241
Fax: (847) 491-7001
ipsavage@northwestern.edu

Published in Michael Crew and David Parker (eds.) *International Handbook on Economic Regulation*. Cheltenham, UK: Edward Elgar, 2006

Introduction

Transport represents a huge proportion of the world's economic activity. The Eno Foundation's annual *Transportation in America* compendium of statistics estimates that passenger transport represents 10 per cent of the United States's Gross Domestic Product, and freight transport another 6 per cent. With the inclusion of some closely related industries, transport probably accounts for about a fifth of all economic activity. Not surprisingly, this has made transport a major focus of government regulatory activity. In the United States the railways were one of the first industries to be regulated by the federal government, in 1887. In most, if not all, countries the government is not only a regulator, but also an important producer of transport services. At a minimum, government provides most of the roads, and it is common for buses, subways, railways, airlines, airports and ports to be in public ownership.

The regulation takes (at least) three forms. This chapter will deal with only one of these - the economic regulation of prices, output, entry and exit. Transport is also subject to regulation of product quality, most notably safety. For a review of the economics of transport safety and its regulation see Savage (2001). Transport is also subject to environmental regulation concerning damage to surrounding property and watersheds during the construction of new infrastructure, and the moderating of the risks of externalities caused during the transport of dangerous or hazardous cargo (for a review of the latter see Savage, 2003).

Characteristics of Transport

There are a number of leading characteristics of transport that both engender the need for regulation, and present practical problems to the regulator. Probably the most important is that all forms of transport are dependent on extensive and expensive infrastructure, which is 'fixed' in both an economic and geographic sense. Roads, railway tracks, airports, waterway improvements, ports and pipelines require large amounts of capital to build. Moreover, the

infrastructure only comes in lumpy increments, and each increment (a single-track railway line, a highway lane) can accommodate many users. Therefore, average total costs will fall over wide ranges of levels of demand. With the exception of markets with very high demand, where it is feasible for multiple infrastructure providers to exist, there are natural monopolies in infrastructure provision. Consequently, there is the fear that the monopoly provider may charge prices that are high enough to produce excess profits and a decrease in social welfare, yet are not sufficient to encourage a rival firm to build a competitive infrastructure.

In transport economics, the cost conditions derived from large, lumpy, infrastructure are called economies of density. These economies are market specific, and do not arise simply because a firm serves many markets. Most analyses of transport costs have found limited economies of scale in firm size, that is to say economies due to the sheer scale of the organization (Braeutigam, 1999). Economies of density can also arise from the transport vehicles themselves. As soon as a bus company commits to running a 70-person capacity bus down a city street, an airline flies an Airbus A320 from New York to Chicago, or a container ship sails from Philadelphia to Rotterdam, the capital cost of the vehicle, staff costs and the majority of the fuel costs become a sunk fixed expense. The marginal cost of an additional passenger or container (up to the capacity of the vehicle) is close to zero. In relatively thin markets, these vehicle-level joint costs can constitute a natural monopoly even if infrastructure costs are not an issue (such as when there is vertical separation between infrastructure suppliers and vehicle operators).

As soon as a commercial firm has to recover vehicle and infrastructure costs, price has to exceed marginal cost. In setting prices, a complicating factor is that the common infrastructure and vehicles are shared by a wide variety of users with different price elasticities and willingness to pay. Business passengers sit next to leisure travellers on aeroplanes. Coal trains share the tracks with container traffic. Economists have long recommended that common costs can be recouped with the minimum of deadweight loss by price discrimination between users based on the marginal cost of serving them and their price elasticity (Ramsey, 1927). Price-inelastic consumers should pay a higher markup over marginal cost compared to price-elastic consumers. While allocatively efficient, such price discrimination infuriates consumers, especially if there are perceived distributional or equity concerns. While there may be few social objections to business airline passengers paying higher fares than leisure passengers, or shippers of bulk commodities paying higher rates than shippers of manufactured goods (albeit that those discriminated against will complain vociferously), the same is not true if commuters were asked to pay higher prices than those making discretionary trips at midday, or if transit riders in poorer neighbourhoods without access to a car had to pay more than the rider from an affluent neighbourhood with alternative options. In effect, Ramsey-style pricing discriminates in favour of optional journeys over necessary trips and in favour of those customers with a variety of transport options over those with few options.

Transport is also a network industry. People's residential location, workplace and leisure destinations are diverse. Industry, its suppliers and its customers form a complex web that expands across the world. Transport firms usually operate in multiple markets for reasons of economies of scope in production, and the desire of some customers to deal with one firm for their diverse transport needs. For these multi-market firms it is difficult to identify the costs or the revenues associated with a particular link in the network. Often there are implicit cross-subsidies between markets. Frequently these arise because prices are set by the network firm in a fairly uniform manner across markets, even though costs vary markedly. It is a standard result in economics that cross-subsidy results in a lower level of social welfare compared with when

prices are tailored specifically for each market. However, the interconnectedness of revenues and cost between markets makes it difficult to substantiate the magnitude of the cross-subsidy.

Demand is differentiated not only spatially, but also by time-of-day and day-of-week. Transport demand tends to be heavily peaked. Highways are built with an eye to the demand at 5 in the evening, and there is a chronic overcapacity at 3 am. The need for urban transit buses is typically 50 per cent to 100 per cent higher in the peak periods than at midday. Containerized freight demand peaks in the autumn months. It is much more expensive to serve peak consumers than those who use off-peak service. However, traditionally the price differential between the peak and off-peak service has not reflected the difference in costs. In part this is because many of the customers in the peak are also off-peak users and their demand is interrelated. A commuter on a crowded train to downtown in the morning values the fact that the railway provides a reasonable service in the middle of the day in case circumstances dictate that he or she has to leave work early.

In summary, transport can be characterized by huge fixed infrastructure costs, joint costs in the form of vehicles, multiple users with different valuations of the service, network firms serving many different markets both geographically and by time of day, and prices that have little connection with the cost of provision.

The Role of Technological Change

When viewed in a historical context, technology changes the relative advantages of different types, or 'modes', of transport and this in turn affects the need for regulation. Two centuries ago, legislators were worried about monopoly exploitation by canal companies. Then the advent of the railways led to the demise of the canals, and created concerns about railway monopolies. This concern then abated with the development of the internal combustion engine and the building of national networks of highways. Railways lost their monopoly power over passenger traffic and many (but not all) types of freight. Commercial aviation became regulated in an era when the technology was in its infancy, prices were high and the market was thin. The advent of the jet engine turned the aviation industry into a mass market where many firms can compete on certain city pairs. Widespread auto ownership in the 1950s meant that urban mass transit lost market share, and the market power that it may have enjoyed earlier in the century.

A problem emerges when the rigidities of the legal regulatory system lead to a continuance of regulation decades after technological change has blunted any monopolistic powers. Moreover, this continued regulation may financially harm the viability of the industry. For example, the United States Interstate Commerce Commission (ICC) refused to allow railways to abandon passenger service and freight service on secondary lines years after the ascendant highway and aviation industries had taken away most of the traffic. The financial burden of providing these loss-making services undermined the finances of railways and their ability to provide freight service on their main lines.

A Roadmap for the Regulator

How should a government decide on the appropriate regulatory regime (if any) for a particular transport mode? In Figure 1 is a roadmap showing a series of questions that a regulator should ask. This figure has been freely adapted and modified over the years since it was originally

published by my colleague Ronald Braeutigam in 1989 (his paper also contains numerous references to the underlying theoretical literature).¹ In the remainder of this chapter, we will walk our way through the roadmap.

Is There a Natural Monopoly?

The initial question is whether there are natural monopoly conditions that might lead to possible market power. All transport modes involve some form of an expensive infrastructure and relatively low marginal costs associated with each user. Therefore, natural monopoly is a likely outcome except in two situations. The first occurs in some large markets where total demand is many times the minimum efficient scale of an individual firm. For example, due to high demand two competitive railways traverse western Canada even though the mountainous terrain makes infrastructure provision very expensive. In contrast, light traffic has resulted in only one transcontinental railway in Australia. Of course, the high cost of the infrastructure will mean that the total number of firms in a market will be quite limited, frequently just a duopoly. Depending on the conduct of these firms, this may or may not lead to regulatory concerns.

The other situation occurs when the provision of the infrastructure is naturally, or forcibly, separated from the actual operation of the service. Historically the providers of the highway infrastructure are separate entities from firms who provide bus, truck or taxicab services. Similarly, in the maritime and aviation industries, the operation of ports and airports has historically been separated from the shipping and airline companies. The operating firms pay relatively low usage-dependent access fees in the form of licences, tolls and fuel taxes. Consequently, in the eyes of the operating companies, the infrastructure has been converted from a fixed to a variable cost. Provided that the operating companies do not have economies of scale in firm size, and that the markets are of sufficient size that the economies of density at the level of individual vehicles are unimportant, one might expect that there would be a highly competitive environment. This type of market would not need regulating. Examples are the trucking and maritime industries and heavily trafficked bus and airline routes. Taxicab service should also be competitive, albeit that regulation is often called for to control safety and to protect consumers from price gouging by cruising cabs. Of course, monopoly power remains with the infrastructure provider and it is therefore not surprising that roads, ports and airports are frequently provided by governmental or quasi-governmental organizations.

A recent development has been the forcible separation of previously vertically integrated state-owned railway companies into an infrastructure company (which remains in public hands or is regulated) and train operating companies. As described later in the chapter, rival firms can then bid for the rights to manage these train operating companies for a set period of time. While the European Union sees this as the future model throughout Europe, it has only been fully implemented in Great Britain and Sweden.

Is It a Public Good?

Assuming that it has been determined that a natural monopoly exists, the next step is to ask whether the mode has the characteristics of a public good. The definition of a classic public good is twofold. First, it must be difficult to exclude potential users and charge an access fee. Second, in non-congested situations, consumption must be non-rival, in the sense that one

¹ Ron and I have shared teaching duties for an undergraduate transport economics class for 15 years (at the time of writing), and we have both used variants of this roadmap in the class.

person's usage does not diminish the ability of others to consume. A prime example is the local road network where, unlike limited-access highways and bridges, it is difficult to charge a direct fee. Provision of navigational devices in the maritime industry and air traffic radio beacons and basic air traffic control services also have characteristics of public goods.

Not surprisingly, provision of all of these services usually resides in the public sector, occasionally with private entities providing management services. Typically they are indirectly financed through licence fees, fuel taxes, and sometimes from general tax revenues. However, that may change. Technological advances may eventually make it possible for direct pricing. For example, in Britain there has been a discussion that ultimately satellite-based vehicle location systems may permit collecting fees based on when and where a driver uses the system, rather than relying on licence fees and fuel taxes.

First-Best versus Second-Best Pricing?

For natural monopolies that are not public goods, prices can be charged directly to the user. The issue for a regulator then becomes, what price should be charged? Economic efficiency dictates that price should be equal to some concept of short-run or long-run marginal cost. This will obviously not generate the revenue necessary to cover the fixed infrastructure costs, except in cases where the price also includes the cost of a congestion externality, which is then used to finance capacity enhancements. At first best, the infrastructure would need to be funded either by some form of two-part or nonlinear tariff or from general tax revenues. The public funds for the latter incur a shadow value. The alternative is a second-best price equal to average total cost, which will be associated with a deadweight loss from users priced off the system. The deadweight loss is likely to be nontrivial as transport, with its large fixed costs and low marginal costs, is characterized by a large gap between marginal and average costs.

The crucial issue is the relative size of a potential deadweight loss from second-best pricing compared with the shadow value of the public funding necessary at first best. There are also subsidiary questions concerning the politics of pricing some users off the system, and whether a nonlinear tariff is practically or politically feasible. There are clearly some facilities where it is felt that it is politically inappropriate to charge a price that would deny mobility to some segments of the community. Examples might include bridges and tunnels to locations where there are no alternative routes, and provision of airport runways in remote locations where aviation is the only means of transport (such as airports in northern Canada and Alaska). In the case of urban mass transit, governments in many (most?) developed countries provide grants that cover the cost of new infrastructure construction and vehicle purchase. Users, often with the aid of additional subsidies, only have to pay for the cost of day-to-day operations. In this way, financially disadvantaged people are not priced out of access to jobs and denied basic mobility.

However, in other modes and markets where demand is more discretionary in nature, second-best pricing is more politically acceptable. Freight railways, pipelines, ports, urban airports, and airlines are supposed to price so as to recover both their operating and capital costs. Intercity limited-access toll highways are frequently allowed to price at second best provided that there are parallel non-tolled highways which price-sensitive drivers can use. The task for the regulator is to ensure that these firms make enough surplus over operating costs to fund the infrastructure, but are not earning supernormal profits.

Traditional Second-Best Regulation

Historically, at this point in the roadmap there would have been just two options: public ownership or regulation to constrain price to second best. In most parts of the world the period from 1945 until the mid-1980s was marked by nationalization of inherently commercial transport firms and implicit price setting by the government. In the United States provision remained largely in the hands of private firms, yet by 1940 most modes were regulated by government agencies.

Public ownership and economic regulation extended well beyond traditional natural monopolies (railways, pipelines, airports, ports) to included modes where competition was viable, such as maritime shipping, urban transit, taxis, aviation and trucking. Often these latter modes were regulated because other regulated modes argued that there should be a 'level playing field'. The railways argued in the 1930s that the newly emerged trucking industry should be regulated to prevent unfair competition. Similarly the urban transit industry argued that taxicabs should also be regulated to counter aggressive competition from jitneys (a hybrid of a taxi and a minibus) in the years after the First World War.

Irrespective of whether the government was directly dictating pricing and output by public ownership, or influencing it by regulatory commissions, there would appear to have been several common characteristics of the traditional regime.

The first feature is that regulators had a desire to directly influence prices and the structure of prices rather than indirectly affecting prices by controlling excessive profitability. While rate of return on capital regulation did exist for some transport modes (notably for oil pipelines in the United States), direct control over prices was more common. The basic motivation was to constrain perceived inequities from price discrimination. One of the famous lawsuits that prompted the regulation of the United States' railway industry dealt with agricultural shippers in Gilman, Illinois protesting that they had to pay a higher price to ship goods to New York than did similar shippers in Peoria, Illinois, which is located 100 miles further west on the same railway line. The cause of the differential was that the Peoria shippers were more price sensitive as they were located on a navigable river, which provided a competitive mode of transport.

Concerns about the fairness of the implicit allocation of common costs between users drew the regulator into the minutiae of the pricing schedule. The early regulators felt that it was their objective to reduce explicit price dispersion and to discourage a more subtle form of discrimination obtained by giving secret discounts or kickbacks to a subset of consumers. The form of price regulation adopted for the railways, and subsequently adopted in other modes, required firms to publish a list of their prices. To ensure that the prices applied to all consumers, firms were not allowed to offer discounts to select customers, and could not refuse to serve any customer willing to pay the price. The first of these effectively outlawed negotiations between transport firms and individual customers for bespoke contracts. The second, known as a 'common carrier obligation', was designed to stop a transport firm quoting a price but then deciding which customers it would allow to purchase at that price.

In practice, regulation went far beyond simply requiring a posting of prices. In a semi-judicial setting, aggrieved consumers could challenge a price increase, or protest about an existing price, on the basis that comparable consumers elsewhere were paying less. It became standard that the regulator had to approve any price changes. As one could imagine, the multiple combinations of commodity types and origins and destinations meant that a posted tariff could fill the equivalent of multiple telephone directories! The practical consideration of trying to

organize such a tariff coupled with the regulators' concern about excessive price discrimination led to a more uniform set of pricing. In the extreme, there was a set rate per mile for a carload of a given commodity irrespective of the parts of the network utilized, or a standard formula was adopted for setting passenger fares.

The second feature of regulation was control over entry and exit. The common carrier obligation resulted in the necessity for firms to obtain regulatory approval for abandoning service in poor markets. This restricted the ability of firms to exit some of their markets, as the remaining customers in that market would complain to the regulator that they were to be denied service for which they would be willing to pay the (uniform) price. The converse of this was restriction on entry, or the granting of legal monopolies to publicly owned firms. Of course, this is not really an issue for modes that are natural monopolies. Entry controls became binding for more theoretically competitive modes that had become subject to regulation such as taxicabs, buses, trucking and aviation.

There is evidence that regulators used entry and exit controls to give preference to firms that provided a network of services, to allow maximum connectivity both in a geographic sense and by time-of-day and day-of-week. Network firms argued that entry would only occur on the best parts of the network ('skimming the cream' or 'cherry picking'), leading to lower prices in these markets, a diminished ability to cross-subsidize thinner parts of the network, and ultimately higher prices or a discontinuation of service in these thinner markets. Inherently, cross-subsidies between markets were not just tolerated, but actually encouraged by the regulator. There is a school of thought that the implicit cross-subsidization of thin markets from thick markets was part of the regulatory deal between network firms and the government, given in return for protection from competition in markets where more than one firm could possibly operate.

The Questioning of Traditional Regulation

In the 1970s, a movement gained momentum that challenged the existing, all-pervasive, system of public ownership and regulation. The arguments of the challengers were basically fourfold.

The first was that standardized pricing and the resulting cross-subsidies were inherently bad and had led to poor resource allocation. Prices were 'too high' on parts of the network where high demand allowed extensive economies of density to be reaped, and 'too low' where lower demand allowed fewer economies of density. It was commonly claimed by critics in the 1970s that the uniform pricing system was supporting secondary airline and rail markets at the expense of the trunk routes. This can be seen as an ascendancy of those interests associated with the major or trunk routes, arguing that they had put up with poorer or more expensive service to cross-subsidize other markets that they did not use. One might also argue that the political power of the rural lobby had diminished.

This situation was made worse when exit controls forced network firms to continue to serve loss-making secondary services. In the extreme, the burden of these unremunerative services-can threaten to undermine the viability of the network firms and endanger continuation of provision in the strong markets. In the case of the railways, technological change led to a general decline in passenger traffic and the diversion of local freight traffic from branch lines to trucks. Yet regulators required that many routes were kept open for years (decades?) longer than economics would suggest. In countries where the railways were publicly owned, huge public subsidies became common, and there were protracted public debates on discontinuing rural

services. In the United States in 1970, the privately owned Penn Central Railroad, which represented a fifth of the industry and was a virtual monopolist in much of the populous Northeast of the country, declared bankruptcy, followed by other railways in the Midwest.

It is fair to say that the predictions of deregulation's proponents have proven to be true. Cross-subsidies have generally declined to the benefit of consumers in the busiest markets who have enjoyed lower prices or better service. However, fears that secondary services would disappear were misplaced. Alternative methods of service delivery have emerged that preserved service, such as the emergence of low cost, often non-union, firms using non-traditional technology. Examples include the development of commuter airlines flying small aircraft offering enhanced service from small airports to larger hub airports, and the emergence of 'shortline' railways, which took over operation of branch lines that were abandoned by the network carriers.

The second line of argument was that entry controls had led to the earning of rents that were then dissipated in the form of cost inefficiencies or excessive quality provision. In the airline and trucking industries, the number of firms had been fossilized at their equilibrium levels in the 1930s. However, both industries had witnessed enormous increases in demand over the following half a century, and had changed from infant industries to mature mainstream industries. Similarly, it is frequently argued that in many cities there are now 'too few' taxis to cater for an increased population. In markets where supply is artificially constrained, price can rise above an efficient level of costs. While it is possible in theory that owners of firms can thereby earn supernormal profits, it is generally the case that most transport firms made very modest returns during the period of regulation. (Although there is evidence of rents being earned by trucking firms and taxicab companies because they were able to sell their rights to be in the market, known as 'operating authorities' and 'medallions', respectively, to other firms for substantial prices.)

While regulation stopped the most blatant profit-seeking behaviour, it was unable to stop the appropriation of rents by organized labour. Transport was, and in many places still is, highly unionized. The ability of labour to strike and, in effect, halt commerce gave unions a strong position in negotiations. Firms also knew that the regulator would acquiesce to price increases to compensate for demonstrated increases in costs. Rose (1985) found evidence of rents earned by labour in the regulated trucking industry, which is quite remarkable as this is an industry that has few economies of scale or density, and has proven to be very competitive in the post-deregulation era. Not surprisingly, unionization of the trucking industry has dropped considerably in the United States since deregulation in 1980. In the airline industry, the regulation-era rents earned by labour have been eroded more slowly. Reports on the battles to remove these rents currently occupy the business pages (and the front pages) of the newspapers more than 25 years after airline deregulation in the United States, in 1978.

Proponents of deregulation were able to use comparisons of regulated and unregulated markets to support their contention that regulation had induced cost inefficiencies. For example, in the United States, regulation was applied at a federal level and the constitution limits regulation to services that cross state lines. In the larger states, such as Texas and California, there was an active intrastate air transport industry and these markets were more liberally regulated. Proponents of deregulation observed that the prices charged in these markets were considerably below those charged in interstate markets. Similarly, trucking rates for historically unregulated agricultural products were less than those for other commodities. Moreover, many

shippers were finding it cheaper to operate their own trucking fleets to move their goods (known as 'private carriage' or 'own-account haulage') so as to avoid regulated prices.

Potential profits were also eroded by competition in product quality. Regulation was mainly concerned with price. Firms could only obtain a larger market share by providing a higher quality of service than their competitors. In the aviation industry, airlines competed by providing the latest models of aircraft, fancier in-flight amenities, and a larger number of daily flights. Increased service frequency is valued by consumers as it allows for more convenient departure times and less waiting around. However, demand is usually inelastic with respect to service frequency. Therefore, higher frequency depresses the average number of people per flight ('load factor') and increases the average cost per passenger. Longtime air travellers will have personal experience that quality has declined since deregulation as load factors have risen, the seats have been moved closer together, and full meals have been replaced by pretzels.

The third line of argument questioned the longstanding belief that regulation protected the nation from the disruption to mobility and the free flow of goods if a large network firm failed. The proponents of deregulation took the view that the bankruptcy and exit of inefficient or poorly run firms would not have disastrous long-run consequences. They argued that newly-formed firms or existing competitors would take over the assets and continue to provide service. To a great extent this has proven to be the case. In the United States, Eastern Air Lines, Pan-American World Airways and Consolidated Freightways (the fifth-ranked general trucking firm) all ceased operation and exited the industry without merging with another firm, and were quickly forgotten. The national airlines of Switzerland and Belgium have ceased operation. However, there is still resistance in some quarters to such a Darwinian outcome. The United States government nationalized the bankrupt Penn Central system from 1973 to 1987; the New Zealand government renationalized the ailing Air New Zealand that it had privatized a few years previously; and in effect the British government took the bankrupt railway infrastructure company, Railtrack, back into the public sector. There is continued discussion in the United States about the desirability of providing a government guarantee for loans made to stave off the liquidation of major airlines.

Finally, proponents of deregulation argued that even a natural monopoly could be subject to market forces. Some of this competition may come from alternatives modes or neighbouring firms who provide consumers with imperfect substitutes. Alternatively, in some markets, discipline may be imposed by the threat of entry if monopoly rents are earned. In addition, an alternative regulatory structure emerged in the 1970s whereby rival firms compete for the rights to provide the monopoly service for a set period of time. The common characteristic is that market power may be limited even though it is unlikely that multiple firms would be competing in delivering the service to the consumer. The next part of this chapter considers four of these alternative market structures, all of which provide an affirmative answer to the question 'can there be competition *for* the market rather than *in* the market?'

Alternatives to Traditional Regulation

Intermodal Competition

The first form of competition for the market arises from competition between modes. Even though a railway or a pipeline may be a natural monopolist in its own mode in a specific market, other modes may provide competition. Some of this competition only emerged after the Second

World War as a result of technological change. Trucking now provides active competition to the railways for short-distance movements of all types of freight. It is also a competitor for non-bulk and high-value commodities over longer distances. When rivers and canals parallel railways or pipelines, there has always been strong barge competition for bulk commodities. It would be hard to argue that railway intercity passenger services have much market power nowadays given that the private automobile, commercial aviation and express bus services have captured a high market share. While these other modes are imperfect substitutes, in that they are either more expensive or slower, they do have the advantage that they are highly competitive. The traditional economic model of the dominant firm with a competitive fringe of firms is particularly suitable for analysing this type of competition. The model demonstrates that the existence of a competitive fringe greatly reduces the rents that can be extracted by a dominant firm.

Spatial Competition

In some circumstances, a natural monopolist may face competition from firms in the same mode who operate in neighbouring markets. For example, if a port or airport started to charge outrageously high fees then traffic would move to neighbouring facilities. While users may find the alternative facility less convenient, some would be willing to switch if the differential in price became large enough. This rationale was used in the privatization of the Mexican state railways in the late 1990s. The state monopoly was split into seven regional vertically integrated companies. It was argued that any individual regional monopoly would not abuse its position as it would cause industry to relocate to other parts of the country, and importers and exporters would change their port of entry. (This is not to say that Mexican railways are completely unregulated. The government retains powers to intervene to resolve disputes over rates.) In Canada, rival railways have mandated access over each others' tracks for 30 kilometres around the junctions where they meet. Prices are kept in check because a shipper in this 30-kilometre zone, whose facility is only served by one firm, can request service from a second railway.

In a similar way, competition can occur between national networks. For traffic between North America and secondary cities in Europe and beyond, Lufthansa, the German national airline, provides competition through its Frankfurt hub to services provided by British Airways through London. Such competition relies on the ability of the airlines of one nation to have access to destinations in other countries. The liberalization of air travel in Europe initially focused on permitting 'open skies' access on international routes between member countries. The liberalization culminated in 1997 with the granting of rights for any European airline to operate domestic routes in any member country (known as 'cabotage'). It is frequently argued that more competition could be encouraged in the United States airline industry if foreigners were allowed to purchase domestic firms (there are currently limits on the proportion of foreign ownership), or if foreign airlines could start domestic services.

Contestability

An intellectual development in the 1970s called *contestability theory* (Baumol et al., 1982) had a considerable impact on the deregulation debate in transport, especially with regard to airlines. There are many secondary airline markets where there are only one or two firms providing service. The source of the natural monopoly is not economies of density due to the infrastructure, because airports are vertically separated - literally and figuratively - from the operation of the aircraft. The economies of density occur at the level of the aircraft itself. There

are also economies in ‘station costs’ because an airline wishing to start a service at an airport has to incur expenses such as providing ground handling staff and facilities, customer service staff, and local marketing. Often these expenses are the same whether an airline is offering two flights or ten flights a day.

Baumol et al. argued that under certain conditions the incumbent firms cannot extract monopoly, or duopoly, rents because there is always the threat that this may cause other firms to jump into the market. For example, United Airlines and Northwest Airlines would appear to be duopolists in the market for nonstop service between Chicago and Memphis. Baumol et al. would argue that regulation is not necessary because both firms know that raising the price above average cost would encourage, say, AirTran Airways to enter the market and initiate a price war. Much of the debate subsequent to deregulation is how applicable the theory has been, especially in thin markets, such as Chicago to Des Moines, Iowa.

The basic requirements for a contestable market are that potential entrants can replicate the costs of the incumbent firm(s), can turn up almost instantaneously when incumbents try to extract rents, and suffer no sunk entry or exit costs (because price is competed down to the operating cost of the incumbent firm(s) the entrant can never recoup any entry costs). There is no doubt that potential entrants have similar costs to incumbent firms as the technology is ubiquitous across firms. Also, relative to many industries, entry can occur quite quickly with a lag of perhaps three months. Consequently, proponents of contestability described aeroplanes as ‘marginal cost with wings’, implying that airlines could move their equipment quickly and costlessly from one market to another. While this is probably true to some extent, contestability has not proved the panacea that some, if not most, economists had argued in 1978.

In many airports in the United States, one or two airlines serve 85 per cent or more of the passengers. These are typically airports that an airline has selected as its hub airport. There is a debate among economists as to whether the observed higher prices at these airports are due to rent seeking or are merely a result of the higher quality service provided at major hubs. At these hub airports there is often competition on the major routes to other hubs, but monopolies on the feeder routes. Part of the explanation is that there really are sunk costs of entry. There are investments in station costs that cannot be recouped, marketing costs to advertise the new service and, what is more important, financial losses in the initial period of operation due to low initial load factors until people become aware of the new service. This is a deterrent to ‘hit and run’ entry as hypothesized by Baumol et al.

Not surprisingly, this has allowed incumbent airlines to protect their position and earn rents in certain markets. These are often known as ‘fortress hubs’. In the United States this has been facilitated by the practice that airlines own or have long-term exclusive leases on gates. Other barriers to entry have included alleged predatory pricing against small new entrants to force them out and deter others, employing loyalty schemes such as frequent flier programmes to deter consumers from defecting to rival carriers, and offering discounts and kickbacks to travel agents and corporate travel arrangers to ensure that they book exclusively with the large incumbent firm.

Consequently, there is occasional lobbying by passengers at certain fortress hubs and users in secondary markets for a return of regulatory intervention. However, these concerns do not appear to attract widespread political support. Undoubtedly the increased inroads made by new entrant low-cost carriers into the market share of the incumbent ‘legacy’ carriers in the last decade, coupled with the poor financial performance of these latter firms, has blunted the protests. The North American experience has been repeated in Europe and increasingly in Asia.

Indeed, there is an active debate worldwide about the continued existence of the large network airlines. Their future would appear to hinge on whether a sizeable number of consumers place a value on large airlines offering a network of services to numerous destinations, as compared with a fragmented industry with different firms serving different markets.

Demsetz Competition

The final form of competition for the market has gained a lot of attention in recent years, especially in urban transport and rail passenger markets. In academic circles it is known as Demsetz competition, after its originator Harold Demsetz (1968). In the transport industries, it is more commonly referred to as competitive contracting, tendering or franchising. This model argues that even if there is a natural monopoly in provision, it is possible to hold an auction for the monopoly rights to serve the market. The rights are then re-auctioned sufficiently frequently so as to provide an incentive for the winning firm to provide high-quality service and to excise control over cost increases.

Demsetz competition has become common in the provision of urban bus services in developed countries. While one might imagine that actual competition, or contestability, should characterize urban bus service, this has not proved to be the case. Since 1986 the British bus industry has been completely deregulated excepting services in London. In practice, actual competition has occurred infrequently and typically has not been sustained. The thinness of most public transit markets nowadays in developed countries reduces the attractiveness to potential competitors due to economies of density at the level of the individual vehicle. In contrast, thick urban markets in many developing countries have witnessed active competition, especially when high unemployment encourages many owner-operated jitneys.

In London, events took a different course. The existing government operator, London Transport Buses, was split into a planning and marketing division and multiple operating divisions associated with the various garages. The operating divisions were then sold off to the private sector. They then competed among themselves and against other firms for the monopoly rights to operate individual routes for periods of three to five years. In general, this has been successful in that costs have fallen considerably, demand has increased, and there are multiple bidders for each contract despite consolidation in the industry. The London model has been adopted worldwide in the past 15 years, particularly in Australia, New Zealand, Argentina, and Sweden. However, apart from San Diego and Denver there has been little movement in this direction in the United States.

A more recent application has been the privatization of passenger rail services. Typically, the former unified state railway is vertically separated into infrastructure and train-operating components. The infrastructure company typically remains publicly owned or is subject to regulation, while competitive contracts are issued for the provision of service over the infrastructure. Perhaps the most discussed example has been in Great Britain. The infrastructure remained with one company, Railtrack, which as described above had a rather checkered history, but the operation and marketing of passenger trains were split into an initial 25 franchises based on geography and type of service (local versus intercity). The franchises were then bid out for periods of seven to 15 years. Longer franchises were let in cases where more up-front investment was necessary. Some of these franchises were profitable and bidders had to indicate how much they would be willing to pay in the form of a concession fee. Others were loss-making and the bidders competed for who would need the least subsidy. The European Union already requires all its member countries separate their publicly owned national railways into

infrastructure companies and operating companies, and there are proposals that competition will be permitted in international and domestic markets and that new entrants must be charged the same infrastructure access fees as the incumbent operator. Currently, only Sweden has gone as far as Britain in terms of privatization. Countries such as France have established accounting separation between the infrastructure and operating companies, but there has not been any competition for service provision.

In theory, Demsetz competition can result in the elimination of monopoly rents by driving price down to average cost. The requirements for this to happen are that inputs are available to all bidders at prices determined in an open market, and that there are sufficient bidders so that collusion is unlikely. The first of these is usually not a problem. In Britain a series of leasing companies was established to own and lease railway rolling stock to operating companies. In urban transit there is an active second-hand market in vehicles, and the contracting agency always has the option of owning the vehicles itself and leasing them to successful bidders. The second requirement has become more of a problem. As Demsetz competition in passenger transport has spread around the world, market concentration has increased by the formation of a small number of multinational conglomerates who comprise the major bidders.

Unlike the other three alternatives which argue in favour of liberalized, or zero, regulation, Demsetz competition usually amounts to a greater amount of regulatory control compared with previous regulatory regimes. The contracting agency, which is typically governmental, is setting the terms of the contract on which the competitors bid. Usually this amounts to specification of the frequency of service, types of vehicles to be used, and the fares that will be charged. Especially in urban bus operations a common fare scale is used for all routes, and very little of the revenue is collected on the vehicle because most users purchase passes or multi-ride tickets from the central marketing agency. Demsetz competition is therefore all to do with costs, and the avoidance of inefficiencies. It is not surprising that Demsetz competition has been deployed in circumstances where a previously publicly owned operation was believed to have become inefficient. Demsetz competition allows the combination of the benefits of privatization with the continuance of public involvement either to moderate monopoly power or to meet other social goals, such as mitigating highway congestion or providing mobility to residents of disadvantaged neighbourhoods.

The Continuation of Second-Best Regulation

Ultimately the proponents of deregulation or alternative forms of regulation were largely successful. The United States and the United Kingdom were the pioneers in regulatory reform. In the United States there was complete deregulation of the airlines and trucking, in 1978 and 1980 respectively, and considerable liberalization of the regulation of the railways, primarily in 1980. Starting in the mid-1980s, Britain privatized the nationalized airports, airline, bus companies, ferries and railway. Other countries followed suit, notably New Zealand, Australia, Canada, Sweden, Chile and Argentina.

However, it would be wrong to assume that there has been a complete abandonment of economic regulation in transport. Regulation (or direct control when the companies are publicly owned) still persists where there is undeniable natural monopoly and market power, such as with pipelines, rail infrastructure, major airports which do not have close spatial competitors, toll

highways, and for specific consumers who are 'captive' to a particular transport mode and company.

In the United States, the federal Surface Transportation Board can still take action where a railway is market dominant for specific flows, and price is greater than 80 per cent above average variable cost. Action is initiated on receipt of complaints from shippers. If regulation is called for, price is set on the principle of a 'constrained market price', which is defined as the theoretical price that an efficient new entrant would have to charge. This is calculated as the stand-alone cost of a competitor constructing and operating a rival route. This is really only applicable for certain markets for bulk commodities such as coal, grain, ores and some chemicals.

Regulatory reform of pipelines, perhaps the most archetypal natural monopoly, took a slightly different course in the United States. The explanation is that regulatory authority resides with a division of the federal Department of Energy rather than with a traditional transport regulatory body. Historically, liquid pipelines (i.e. pipelines for oil and oil products) were regulated by specifying a maximum rate of return on the replacement cost of assets. Starting in the early 1990s, this was replaced with price caps which stated that prices could only increase by an amount equal to the Producer Price Index on Finished Goods less one percentage point (PPI-1%). The low inflation rate in the late 1990s meant that in some years prices had to be reduced! Not surprisingly, industry lobbying led to a revision of the formula to PPI-0%, in 2001. But even in the pipeline industry there is evidence of regulatory reform. Studies have suggested that at least 70 per cent of traffic moves in markets where there are multiple pipelines or where there is competition from another mode such as barges. Consequently, pipeline companies can apply for an exemption from regulation in markets where there is evidence of actual or potential competition, or where pipeline companies and shippers have freely agreed on a price.

Paradoxically, in many countries there has been a need to establish new regulatory bodies to oversee former publicly owned companies that have been privatized. Explicit regulation has been necessary to supersede the former implicit regulation in public ownership. In many ways the cutting edge of innovative regulatory mechanisms has moved from the United States to the United Kingdom and other countries which have taken the most extreme amounts of privatization. For example, a new regulatory structure was necessary in Britain when three major airports near London were privatized under a single owner, BAA plc, in 1987. Landing fees and other charges were only allowed to increase by the retail (consumer) price index (RPI) less X percentage points, where X is set by the regulator and revised every five years. The RPI-X price cap form of regulation has become very popular in Britain as many areas of the public sector have been privatized (see Rees and Vickers, 1995).

When Britain's railways were vertically separated and privatized in the mid-1990s, the government established an Office of the Rail Regulator whose mission was to oversee the prices paid by operating companies for access to the track, rather than regulating the prices paid by the ultimate consumer. The regulator was also charged with setting cost reduction goals for the infrastructure company and ensuring that these savings were passed on to the train operators. The final task was to ensure that the various train operators could obtain access to the infrastructure without discrimination. Certain post-privatization rail fares were regulated by another regulatory body using the RPI-X model, presumably because rail fares are politically sensitive, and in some commuting markets rail does have a sizeable market share.

A transport mode that has only recently come to the regulators' attention is the private toll highway. Traditionally most tolled facilities were operated by governmental or quasi-

governmental bodies. More recently, private operation has been encouraged, sometimes as part of a public-private partnership. Private firms employ their own capital to build new facilities, and are then granted rights to collect tolls for a period of years. In some cases the road ultimately passes back to the public authority under a 'build, operate, transfer' (BOT) agreement.

The regulatory structure for toll roads varies. In France and Indonesia there is direct regulation. The operator has to apply to the government to change the tolls. In Spain, a price cap is employed where toll increases are tied to changes in the national price indices for steel, petroleum and labour. In the United States, toll highways in Virginia and California are or were regulated using rate of return regulation. In Illinois an existing tolled facility was recently leased to a private operator with a set schedule of the maximum allowable price for different years. In Britain, a new tolled facility that was built north of Birmingham is not subject to price regulation because it parallels a non-tolled, but congested, existing highway.

Conclusion

During the middle third of the twentieth century, competition had been eradicated from most transport markets. In many parts of the world, transport companies were taken into public ownership. Even the United States, with its strong tradition of private enterprise, heavily regulated most forms of transport. There has been a dramatic change in the past 25 years. Public ownership is receding. In some modes the possibility of direct competition or the potential for competition for the market has rolled back regulation. However, do not be fooled. Regulation is alive and well. The potential for market power remains due to economies of density in infrastructure provision and the operation of the vehicles. Moreover, as privatization has gained ground there has been a need to establish a whole new set of regulatory structures to replace the previous implicit regulation associated with public ownership.

References

Baumol, W., Panzar, J. and Willig, R. (eds) (1982) *Contestable Markets and the Theory of Industrial Structure*, New York: Harcourt Brace Jovanich.

Braeutigam, R.R. (1989) 'Optimal Policies for Natural Monopolies', in R. Schmalensee and R.D. Willig (eds) *Handbook of Industrial Organization*, vol. 2, Amsterdam: North Holland.

Braeutigam, R.R. (1999) 'Learning about Transport Costs', in J.A. Gómez-Ibáñez, W.B. Tye and C. Winston (eds) *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*, Washington DC: Brookings Institution.

Demsetz, H. (1968) 'Why Regulate Utilities?', *Journal of Law and Economics*, 11 (1), 55-65.

Ramsey, F.P. (1927) 'A Contribution to the Theory of Taxation', *Economic Journal*, 37 (145), 47-61.

Rees, R. and Vickers, J. (1995), 'RPI-X Price-Cap Regulation', in M. Bishop, J. Kay and C. Mayer (eds), *The Regulatory Challenge*, Oxford: Oxford University Press.

Rose, N.L. (1985) 'The Incidence of Regulatory Rents in the Motor Carrier Industry', *RAND Journal of Economics*, 16 (3), 299-318.

Savage, I. (2001) 'Transport Safety', in D.A. Hensher and K.J. Button (eds) *Handbook of Transport Systems and Traffic Control*, Amsterdam: Elsevier Science.

Savage, I. (2003) 'Safety', in D.A. Hensher and K.J. Button (eds) *Handbook of Transport and the Environment*, Amsterdam: Elsevier Science.

Figure 1: Road Map Showing a Series of Questions a Regulator Should Ask

