THE ECONOMIC JOURNAL

MARCH 1975

A SIMPLE ANALYSIS OF OPTIMAL FARES ON SCHEDULED TRANSPORT SERVICES

In transport economics, much attention has recently been given to problems of competition between road and rail and between private and public transport. These problems, which involve income distribution, congestion and other important matters have, however, obscured our continuing failure to solve some older problems. The problem of railway rates is one which was with us before cars and lorries became important, and which involved no doubts about whether willingness to pay adequately represented social usefulness. This paper goes back to these older problems, and in order to concentrate upon them supposes that the income distribution is just, that there are no external diseconomies in transport operation and that prices in the rest of the economy generally reflect marginal costs.

In order to make the analysis as simple as possible, we initially assume that the transport service supplies only regular daily return journeys from A to B, that capacity is fixed and that there can be but one fare. In other words, we are considering a daily bus, train, boat or plane which will take no more than n passengers. There are two issues:

(1) The pricing question: What is the optimal return fare?
(2) The investment question: Should the daily service be provided?

The first question assumes that the service is to be provided and requires us to find that pricing rule which maximises the value of the service to users. (If there were any cost dependent upon the actual number of passengers, this would have to be deducted, but we assume that the daily avoidable cost of C is independent of their number.) The second question is then answered by examining whether the value of the service to the passengers with the optimal fare, *i.e.* daily fare revenue plus their consumer surpluses, exceeds or falls short of C.

An obvious answer to the pricing question is that the fare should be zero unless a higher fare is necessary to restrict daily demand to n passengers. When the fare thus determined is zero or low, it is apparent that it is possible for daily fare revenue to fall short of C even though the value of the service (fare revenue *plus* consumers' surpluses) exceeds it. Thus we have our old-fashioned problem: Optimal pricing and optimal investment may require a subsidy.

Since economists like classifying things in textbook boxes, it is tempting to try to describe this problem as one of a decreasing cost industry. But such an attempt might obscure the peculiarly significant feature of transport that the units of demand and of supply are different. Passengers demand passenger-journeys, while the enterprise supplies vehicle-journeys. To underline the distinction, let us speak of (passenger) "trips" and (vehicle) "runs". Demand is measured in trips; these are constrained by runs, and it is runs which determine costs.

The problem in terms of the simple transport system described above, is thus the following: If the fare which restricts daily trip demand to n is F (which may be zero), what should be done when $n.F < C$? One obvious answer is to change over to a smaller vehicle with a smaller capacity, n, and a lower daily cost, C. But sometimes there will be no vehicle size for which $n.F \geqslant C$ and even if there is, it may not be the vehicle size which maximises:

$$\text{daily fare revenue} + \text{consumers' surpluses} - C.$$

Thus the problem remains that there is a conflict between the aim of making the service pay on the one hand, and charging an optimal fare to avoid wasteful under-utilisation of the service on the other hand.

A recognised and sensible way of resolving, or at least reducing, this conflict is to have fare discrimination in order to "get at" some of the consumers' surpluses. Within the present framework this can take three broad forms. One consists of differentiating the fare according to the type of passenger—for example, by having lower fares for children or tourists. The second involves differentiating the service, as well as the fare—for example, by providing first-class passengers with more space than second-class passengers. The third, which is only sometimes possible, is to charge regular passengers a fixed periodical fee plus a low fare, while casual passengers are charged a higher fare. Season tickets are an example of this, the low fare being zero.

Traditionally, transport undertakings have attempted to cover their costs by a combination of:

(1) fare discrimination,
(2) fares high enough to cause some wasteful under-utilisation of services,
(3) cross-subsidies from services with revenue in excess of costs to services with a deficit.

These devices often require a monopoly position, and it was the need for this monopoly revenue which motivated arguments against allowing competitors to "skim the cream". Where competition did develop, as it usually has, transport undertakings had to respond by closing down some services, by new kinds of fare discrimination, by increasing fare levels and by seeking external subsidies. At the same time they have complained that their competitors have been too cheap and that they should be restricted or made to bear a "fair" share of costs. This accounts for the restrictions on road transport which are so common in many countries and for the old rules about affinity groups for charter flights. In terms of optimal resource allocation there are only two justifications for such arguments and measures. One is a demonstration that competitors' private marginal costs are below marginal social costs. The main case of this is road congestion, but this is precisely the kind of second-best consideration which I am ignoring in this paper in order to concentrate on other, older problems. The other justification is a demonstration that optimal resource allocation does require certain services to be carried on which would involve a loss in the absence of fare discrimination or monopoly, coupled with the partly political judgment that it is better to allow fare discrimination and cross subsidisation than to have a government subsidy (*i.e.* higher taxes imposed by government). Unfortunately neither of these two kinds of justification is put forward as often as are silly arguments, such as much of the discussion as to whether heavy lorries bear their "fair share" of track costs.

Leaving aside for the moment these questions of fare discrimination, let us go back to the problem of a simple transport system where $n.F < C$. This problem was formulated deterministically. It was implicitly assumed that the same set of potential passengers existed on each and every day, each passenger having some maximum amount, M, which he would pay for the return trip rather than not make it. The demand curve was thus a simple ranking in diminishing size of the M's. The ith passenger took a trip if $M_i > F$ and achieved a consumer surplus equal to the difference.

Now this simple model obscures some very important matters. The set of potential passengers and their M's will change from day to day. Some of these fluctuations are regular, holiday traffic for example, and call for regular fluctuations in fares. But I want to concentrate on the irregular fluctuations, in other words to allow for the fact that demand is stochastic, and re-examine the pricing question. Assuming now that demand fluctuates from day to day but that one fare has to be fixed in advance to rule on all days, what is the optimal fare?

If the fare is fixed so high that the demand for trips *never* exceeds n, then on most days there will be a waste of resources in that some of the n seats will be empty while some potential passengers would have been willing to pay something for a trip. Alternatively, if the fare is fixed so low that the n seats are always filled, there will nearly always be some excess demand. This may, according to circumstances, take two forms. Where the number of seats is fixed and no standing is allowed, there will be some frustrated potential passengers who would have been willing to pay more for the trip than the passengers who did get a seat. As they have either wasted time in coming to the terminus or have to wait for the next service, besides not making their trips they actually suffer some inconvenience. Alternatively, when the seats can be put closer together (as sometimes happens between aircraft flights) or when passengers can stand, we have the congestion phenomenon where each extra passenger adds to the discomfort of all others. Thus a low enough fare to fill all n seats on all runs will, in one way or another, entail high inconvenience or discomfort on a large number of runs. High quantity, in other words, will make quality unacceptably low. Furthermore, a fare low enough to produce this result is more likely to bring a revenue which falls short of costs.

The right thing to do seems to lie somewhere between these extremes. Since the M's of different potential passengers and their valuations of discomfort or inconvenience are not knowable in practice, there is no point in any sophisticated algebraic construction. A crudely practical approach is simply to pick a target percentage load-factor, which is higher the less variable is demand and the smaller is the negative value put upon discomfort and inconvenience, and to try to set the fare at the level which achieves this load-factor *on average*. If this load-factor expressed as a fraction is L, there will be a subsidy problem if $L.n.F < C$.

In the earlier deterministic analysis it was pointed out that conflict between charging an optimal fare and avoiding a deficit could be resolved, or at least reduced, by using three forms of fare discrimination. Now that the stochastic nature of demand is recognised, we can recognise that there is a fourth and very important kind of discrimination. This is to divide the n seats into groups with different availabilities and different fares. Thus seat reservation charges can exceed the administrative costs of reservation, first-class fares are set for a lower load-factor, ABC fares are set so as to obtain 100 % load-factor on one group of seats and stand-by fares to achieve this on others. The four kinds of fare discrimination may be combined: first-class fares secure greater comfort as well as greater availability and cheap advance booking fares are set for return periods which, it is hoped, will exclude business travellers.

The word "discrimination" must not be understood pejoratively; it merely means that fare differences do not reflect differences in marginal costs. Such a lack of reflection could be said to be bad only if it were desired to have all fares equal to marginal costs. But this is not always the dominating factor; sometimes it is more important to avoid a deficit. In other words it may be preferred to have some customers subsidise other customers rather than to have the general taxpayer subsidise all of them.

Fare discrimination is naturally a fine art. It is not sufficient to ascertain that one type of customer is ready to pay more than another. It is also necessary to ensure that the first type either may not pay a fare intended for the second group (discrimination by passenger type) or will choose not to pay it (discrimination by service or availability). When customers of the first type both may and do choose to pay the lower fare there is said to be "revenue dilution" in airline parlance. Since revenue dilution is a loss to the transport undertaking, it is a gain to someone else, and entrepreneurs attempt to seize part of this gain by passing on the rest of it to customers. Furthermore, when competing undertakings serve one route they may each be willing to turn a blind eye to selling their own cheap trips to customers who would otherwise buy an expensive trip from a competitor.

All that has been said so far relates to scheduled services. With tramp ships, and with genuine charters of trains, planes or buses, where a single aggregate charge is paid under a separate contract for a specially provided run, the problem under discussion does not arise; the pricing question and the investment question are then one and the same. They are separate for scheduled services because the trip decisions are taken individually by many passengers in the light of the fare structures they face, while the choice of the timetable (the investment decision about what services to offer) is made independently by the transport undertaking in the light of uncertainty about the passengers' trip decisions.

We have concluded that where social considerations, congestion and other problems which we are ignoring do not justify subsidy:

(1) fare discrimination in some mixture of the four forms distinguished is often desirable,

(2) at least one fare class should provide instant seat availability on most runs by a fare high enough to give an average load-factor of considerably less than 100 %.

We now go on to consider a network rather than simply a daily return run between *A* and *B*. This means talking about a fare structure and it would be too complicated to bring in both different trips and different fare classes. Solely for the sake of simple exposition, therefore, fare discrimination will now be ignored and we shall again suppose that there is but one fare for each kind of trip.

As a first step, consider the case where all passengers do not make daily return trips from A to B but where single trips from A to B are distinguished from single trips from B to A. If the two fares which now have to be fixed are to optimise resource allocation, they will have to be different if either the level or the variability of demand differs in the two directions. This may sound implausible for passengers, but it can often happen with freight. In fact Alan Walters' pioneer analysis of this case related to freight.[1] His contribution was couched in terms of cost allocation but was essentially to stress the importance of stochastic demands. The aim, in terms of the present approach, is to fix fares so that in both directions a target average load-factor is achieved. These load-factors are set at less than 100 % because the loss from sometimes having empty seats is outweighed by avoiding the loss either from often frustrating potential passengers from obtaining one or from frequent overcrowding.

Just as the optimal AB fare could thus differ from the optimal BA fare, at least in the case of freight, so the optimal BC and CB fares can differ from either of them, even if the distance between B and C is the same as that between A and B. We are now considering the simple network shown in Fig. 1, and it is clear that, assuming that all runs are from A to C and back and stop

[1] A. A. Walters "The Allocation of Joint Costs with Demands as Probability Distributions," *American Economic Review*, June 1960.

at B, a further question is whether the AC fare should equal the sum of the AB and BC fares.

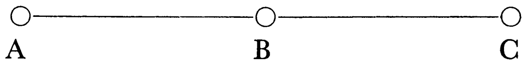

Fig. 1

The answer to this, which flows from the provisional ruling-out of any fare discrimination, is that the fares for trips over different segments should be additive. The probability of a passenger from A to C preventing someone else from obtaining a seat or causing crowding is the same as the probability of this happening when one passenger travels from A to B and another travels from B to C. It is this probability times the social cost of either preventing someone else from obtaining a seat or of increased crowding which is the expected marginal social cost of the AB trip, and this expected marginal social cost is the optimal fare.

Optimal fares on successive segments are thus additive. If the fare per mile is the same on all separate segments then that same fare per mile should rule for multi-segment trips. But if the optimal fare per mile differs between segments then we get the interesting result that no general formula is possible. In particular a structure where the rate per mile is tapered according to distance cannot be optimal! To see this, go back to the diagram and suppose that, per mile, the optimal AB and BC fares are not the same. Then additivity requires the optimal AC fare per mile to be a weighted average of these two fares per mile. It will therefore be higher than one of them and lower than the other; it can neither be higher than both nor lower than both. This means that the fare per mile can be neither an increasing nor a decreasing function of distance.

In the days when railways were regarded as potentially avaricious monopolists rather than as deficit-producing claimants for government subsidy, simple rules relating fare to (rail) distance were obviously convenient, at least for the regulatory authority. Such rules survive in some cases, very often with a tapering element whose original defence lay in some notion that terminal costs were unaffected by trip length. We can now see that cost arguments of this kind are of little relevance and that universal tapering, which conflicts with additivity, cannot be optimal. This is not to deny, however, that a simple formula may still be simple to administer and that optimisation is exceedingly difficult.

What we have shown, then, is that any more fundamental justification of tapering than mere administrative convenience must be related to discrimination. There may be some reason why a departure from the optimality rule in the direction of making the fares for multiple-segment trips less than the sum of the fares for the several segments enables a railway to extract more revenue from its passengers. So we have to enquire what this reason may be.

In terms of standard economic theory we have to consider whether the elasticity of demand for long trips is higher than the elasticity of demand for short trips. This will be likely if there are more substitute transport modes for long trips; if the value of long distance travellers' time is less than that of short-distance travellers so that their ratio of money to money plus time cost is higher, or if, in some sense, long distance trips are less necessary than short-distance trips. However, I can think of no general reason why any of these circumstances should rule. In the case of freight on the other hand, there *is* an *a priori* reason for supposing the elasticity of demand to be greater for long-haul than for short-haul traffic. This is that the demand for freight is a derived demand whose elasticity will be greater the larger the share of the freight costs in total costs c.i.f. in the market where the goods are sold. This share must be larger for those goods which come from farthest away.

I conclude then, that tapering rates are non-optimal and that, as a generalisation, they are not a useful discriminatory device for raising more revenue than the optimal charges would provide from passenger traffic but that they are useful in this way in the case of freight traffic.

In the absence of any extra subsidy for an extra train, the fare revenue from each train will generally have to cover its costs; for a majority of trains it will have to do more than this, since there are other accounting costs to be covered in addition to avoidable train costs. When the time-table decisions are constrained by financial considerations in this way it is tempting to approach the fares decisions in terms of costs. If fares are fixed by dividing the avoidable costs of a train by the average number of passengers it carries and if trains are only run when the fares thus calculated seem reasonably in line with fares on other trains, a railway has a rough and ready way of making both sets of decisions together. Demand considerations can then be brought in by introducing discrimination in various ways in order to bring in a greater total revenue. Thus known costs are used and the much less knowable aspect of demand is brought into decision-making only when there is reason to believe that a departure from the cost-calculated fares will produce good results. The whole approach is not intellectually elegant, scarcely qualifying as constrained optimisation, but it is practical. Information about demand is used when it is available, but does not have to be systematically sought for all journeys.

In practice, it appears that the approach followed by some railways is even simpler than this on the cost side. Costs are not estimated for each and every train, but some sort of weighted cost is estimated for each of a number of different types of train. While it may be a function rather than a single figure, the point is that one estimate is made and used for all trains of a particular type, *e.g.* long-distance express pullmans. This can be used to calculate a marginal cost per passenger on a probabilistic basis by accepting that the initial level of traffic is uncertain.[1] The mathematical

[1] This is M. Hutter's solution to the "passenger to Calais" problem given in his paper "Qu'est-ce que le coût marginal," *Revue Général des Chemins de Fer*, Février 1950, p. 57.

expectation of the increase in expenses due to an extra passenger is of the form:

$$\begin{pmatrix}\text{probability that marginal passenger}\\ \text{will necessitate an extra carriage}\end{pmatrix} \times (\text{cost of extra carriage})$$
$$+ \begin{pmatrix}\text{probability that marginal passenger}\\ \text{will necessitate an extra train}\end{pmatrix} \times (\text{cost of extra train})$$

A similar formulation obviously applies to an air shuttle service even in the very short run, since such services guarantee seats to all passengers who turn up by the scheduled departure time.

With a given timetable, *i.e.* a given number and timing of trains and a given number of carriages per train, these probabilities will be increasing functions of the average level of passenger flow. Thus the mathematical expectation of marginal cost per passenger is an increasing function of the number of passengers, while that number is a decreasing function of the fare. There must therefore be a fare which causes the average number of passengers to be such that the mathematical expectation of marginal cost equals the fare.

So long as we hold to the assumption of a given timetable, in the sense both of a given number and timing of trains and a given number of carriages per train, the argument of this paper is that such a fare is not necessarily optimal. The optimal fare is the marginal social cost of an extra passenger and this is the probability that his taking a seat will frustrate or inconvenience one or more other passengers times the loss to each such passenger. This loss is the maximum that a frustrated passenger would have paid rather than not travel or the amount which an inconvenienced passenger would pay to avoid the delay or discomfort involved.

But once we allow the timetable decision to be optimised too, it is obvious that the combination of the optimal fare and the optimal timetable would involve equality between the marginal social cost of an extra passenger and the mathematical expectation of marginal cost. Thus we can say that, in the absence of any financial constraint which may necessitate both a sub-optimal number of trains and price discrimination, the optimal fare will equal the mathematical expectation of marginal cost. This is a manifestation of the general principle that the combination of an optimal short-run pricing rule and an optimal investment rule will yield a price equal to long-run marginal cost.[1]

While this theoretical result is intellectually pleasing, it is not very helpful in practice for two important reasons. One is that its application would demand far more information than is available and require the pricing and investment (fares and timetable) decisions to be taken simultaneously, an impossible intellectual effort. The other is that, even in principle, the result is irrelevant when a financial constraint necessitates a sub-optimal number of trains and price discrimination.

[1] *Cf.* my *Economic Analysis and Public Enterprises*, pp. 91–3.

The fact that fares and the timetable are both so complicated that they have to be considered iteratively rather than simultaneously means that fares do have to be fixed upon the assumption that the timetable is given. So far, I have suggested two ways of approaching this pricing question. The first was to use some or all of the four kinds of price discrimination and to aim at an appropriate load factor or load factors. This requires much information about demand but none about costs. The second approach was to start by basing fares upon cost figures and then to introduce variation and discrimination when known demand considerations suggest it.

The first of these may sound the more difficult, yet airlines do it. British Airways, for example, has twelve round-trip fares for individual adult travel from London to Bermuda (three of them with weekend surcharges) and eleven for group travel. They range from £407·10 first class to £68·40 for off-peak contract bulk inclusive tour passengers. Fares from Bermuda to London are different. Similarly, British Rail fares are far from simple. Here, for example, are the current second-class fares from London to Manchester and York as an example:

	Distance (miles)	2nd Class standard return	Day return	Weekend return	17-day return	Economy return	Students weekend
Manchester	183·5	£10·23	£5·73	£7·78	£8·49	£5·37	£2·65
York	188·0	£9·41	£5·30	£6·04	£7·52	£5·12	Not available

Weekend and Day return tickets do not allow travel on all trains and Economy return tickets have to be booked three weeks in advance and involve mid-week travel. In addition, there are Bargain returns such as £5·00 for *two* ladies travelling together from Manchester on a day trip to London on Thursdays, a classical example of fare discrimination!

Conclusions

Discrimination raises problems of equity, not touched upon here, and when there are competing carriers on a route, it causes competition to take curious forms. This paper has concentrated on the resource-allocation aspect, however, analysing the characteristics of optimal pricing from this point of view and pointing out that such pricing may require a subsidy. If this is ruled out, discrimination may be desirable.

It has been shown that the marginal cost relevant to the optimal fare is the short-run marginal social cost of frustrating or crowding the journeys of other passengers, that this is a probabilistic concept, that it rules out the traditionally acceptable tapering of fares with distance and that, given optimal investment decisions, it equals long-run marginal cost (which is also probabilistic).

Ralph Turvey

Scicon and London School of Economics.

Date of receipt of final typescript: July 1974.