MARGINAL COST PRICING OF SCHEDULED TRANSPORT SERVICES

A Development and Generalisation of Turvey and Mohring's Theory of Optimal Bus Fares

By Jan Owen Jansson*

1. BACKGROUND, PURPOSE AND MAIN IDEA

A useful general price theory for the whole transport sector is difficult to imagine, in view of the many important differences within the transport sector in both demand and supply. The most common approach is to deal separately with each mode of transport. This is the most natural division from a transport technological point of view.

Another division of the transport sector can be more adequate from an economic point of view. Instead of considering the pricing problems of each mode of transport separately, it seems in many respects more appropriate to divide transport price theory into these three departments:

Price theory for track (road, rail, fairway) services

Price theory for services rendered at passenger and freight handling "terminals".

Price theory for services rendered by transport vehicles, or simply "transport services".

The third department should, in turn, be divided into the subdepartments of the "charter" market and the market for "scheduled transport services". The present discussion is about marginal cost pricing of scheduled transport services—sts for short.

The vehicles producing sts are charged either a price or an accounting cost for the use of tracks and terminals. Where the track and terminal costs appear as charges on the vehicles, but it can be expected that those charges do not equal the marginal costs of making use of the tracks and terminals concerned, a problem arises of limitation of an analysis of marginal cost pricing of sts. The present analysis does not go into this problem. It is in line with the above suggested division of transport price theory to "delegate the responsibility" for a total optimum between experts of each of the three departments, so to speak. For modes of transport like the railways where the tracks and terminals are owned by the producer of transport services, it could

*National Road and Traffic Research Institute, Linköping, Sweden.

be argued that a discussion of MC-pricing of the transport services should also go into the problems of costing or internal pricing of the constituent track and terminal services. This is not done here. One point of the suggested division of transport price theory is that the particularly complicated pricing problems facing railway companies and other vertically integrated transport enterprises will be separated in such a way that theoretical advances in other areas can be put to the maximum use. For example, the problem of rail track costing is more akin to the problem of road pricing, and the problem of railway terminal costing has more in common with the problem of port pricing, than with the problem of costing and pricing of rail transport services.

Apart from railway passenger transport, economists have not been much interested in the pricing problems of scheduled transport undertakings. Even for railway transport pricing, the vivid interest shown in the first half of this century by leading economists like Wicksell, Cassel, Pigou, Taussig, Hotelling and Dessus seems to have faded in modern times. The reason for the lack of interest of modern economists cannot be that the consensus of opinion has been attained on a method of optimal pricing of scheduled transport services. At least in the general debate, suggestions of reformation of public transport pricing policy range from the application of zero prices to self-financing full-cost pricing.

However, a path-breaking contribution, which should cause a renaissance of the interest of economists in this field, has been made by Ralph Turvey and Herbert Mohring (1975). I think they have located exactly where earlier contributions went wrong, and why it never seemed to be possible to arrive at a consistent theory of optimal pricing of scheduled transport services:

> The right approach is to escape the implicit notion that the only costs which are relevant to optimisation are those of the bus operator. The time-costs of the passengers must be included too, and fares must be equated with marginal *social* costs. (Turvey and Mohring, p. 280).

The main problem considered by Turvey and Mohring was, in their own words, "to find out what marginal cost pricing of bus services consists of", but also to give an indication of what the financial result of optimal bus fares would be. The present paper follows up and seeks to generalise some of the ideas put forward by them to the whole sector of sts. It is argued that the crucial (for optimal pricing) characteristics of urban bus services reappear in one form or another for all scheduled transport of passengers and freight alike.

A bird's eye view of every system of scheduled transport reveals a basic similarity: Regardless of the mode of transport, whether the vehicles are borne on a road, a rail track, water, or air, the vehicles are moving to and fro along given routes, making halts at more or less regular intervals for loading/unloading at predetermined "stops". However, for every mode of transport the size of the vehicles employed can vary in a very wide range. And, roughly speaking, there is a one-to-one trade-off between size and number of vehicles.

From an economic point of view, the salient feature is that there are significant *economies of number* of vehicles serving a given geographical area, and equally important *economies of vehicle size*. The number economies are manifest in the *user costs*, while the size economies exist in the *producer costs*. The number economies have to do

with the access to sts, and are normally reflected in both the frequency of service and the density of service.

Given the size and load factor of buses, it is clear that the frequency of service will be roughly proportional to the demand. A symmetrical source of variation in user cost is the positive correlation between the accessibility in space of sts and the density of demand. A study of maps of networks of railways, airlines and bus services as well as roads will give a vivid impression that the more sparsely populated a region is, the more coarse-meshed the network will be. Similarly, the density of the shipping lines connecting two continents is strongly correlated with the seaborne trade volume per mile of coastline.

Generally, the average feeder transport distance will become longer and longer as the network becomes more coarse-meshed or the density of sts lower. Therefore, the costs of users of scheduled transport services for transporting themselves or their shipments to and from the sts "stations" can be expected to be related to the density of demand, in much the same way as the user costs of infrequency of services are related to the density of demand.

The main transport policy angle of the matter is that the coexistence of significant economies of vehicle size and economies of vehicle number makes the sts sector a pronounced "decreasing-cost industry"; that is, the application of social MC-pricing will result in a financial deficit. This conclusion is particularly remarkable in view of the fact that the charter markets for sea, air and road transport alike, where basically the same sort of vehicles operate as those engaged in scheduled transport, are commonly regarded as some of the most perfectly competitive markets in existence.

The basis for the discussion is the following model of an "abstract mode" of sts, viewed in a bird's eye perspective in order to keep hold of the common characteristics of sts, and not get lost in specific details.

2. GENERAL MODEL OF AN "ABSTRACT MODE OF SCHEDULED TRANSPORT SERVICES"

Just as for investment problems of sts, the understanding of optimal pricing of sts needs a *system cost* approach. This implies:

1. equal treatment for the transport service producer costs and the transport service user costs; and
2. that a network of scheduled transport services should be considered, rather than an individual link in a network.

In cost/benefit analyses of transport investments the method of translating benefits to users into user cost savings has a long tradition, particularly in road investment appraisals. In the field of urban public transport, the CBA of the Victoria Line from 1962 was a pioneering contribution (Foster and Beesley (1963)).

In transport pricing, the importance of treating producer costs and user costs on a par has not been widely recognised outside the extensive literature of optimal road user charges. As has been mentioned, in the wide field of scheduled transport it has only very recently been realised that the principle of marginal cost pricing is prac-

tically impossible to apply correctly unless all user sacrifices and efforts are, at least conceptually, treated as costs on a par with the producer costs. By the following model it will also be shown that the level of optimal sts prices comes out differently when a sts system rather than an individual line is considered.

The model seeks to describe the main effects on producer and user costs of different designs of a system of sts. The system is defined by a given geographical area, which is traversed in different directions by a smaller or larger number of transport vehicles plying fixed routes according to pre-announced time tables. The aggregate output of sts per unit of time, Q, is given by the production function (1) below. A main simplification of the analysis is that cyclical fluctuations in demand are ignored. An alternative interpretation of the assumption of a homogeneous unit of output is that it represents a "composite unit" including α units of peak output and β units of off-peak output.

The production function is supplemented by a function for the total producer costs (2), and a function for the total user costs (3).

Then

$$Q = \phi NHV, \tag{1}$$

where

$$H = H(X), \tag{1a}$$

and

$$V = V(X, Y, \phi). \tag{1b}$$

$$TC^{prod} = NZ, \tag{2}$$

where

$$Z = g(X) + pY + c(X). \tag{2a}$$

$$TC^{user} = Qh(\phi, N, V). \tag{3}$$

Q = transport volume
ϕ = load factor or occupancy rate
N = number of vehicles
H = holding capacity of a vehicle
V = overall speed of vehicles
TC^{prod} = total producer cost
TC^{user} = total user cost
X = vehicle design vector
Y = current input vector (fuel, cargo handling labour, etc.)
Z = total producer costs per vehicle-hour
$g(X)$ = overheads, capital and crew costs per vehicle-hour
p = current input prices
$c(X)$ = track and terminal charges
h = total user cost per journey (passenger trip or freight transport).

Total output is the product of the load factor or occupancy rate, the number, the holding capacity, and the overall speed of the vehicles in the system. By "overall

speed" is to be understood the ratio of the distance travelled per operating year to the total hours of operations, including all the time spent in handling passengers or freight at terminals as well as the time spent in hauling. The holding capacity H and overall speed V are determined by a number of "primary" factors of production. These factors can be divided into two distinct classes: on the one hand there are a number of vehicle design variables such as the hull dimensions of a ship and the engine power, and on the other hand there are some current inputs such as fuel. The holding capacity cannot be varied by varying the current inputs. The overall speed is a function of the vehicle design, as well as of the actual input of fuel and labour for freight handling at terminals. A third determinant of V is the load factor, ϕ. This factor is a somewhat unconventional "factor of production". It appears both as an argument in the V-function and as a scalar of the production function. Its latter role is self-explanatory: the transport capacity is given by $\mathcal{N}HV$, and, in order to produce transport output, the capacity provided has to be utilised by passengers or freight. If the capacity were independent of the actual rate of utilisation, the output would be strictly proportional to ϕ. This is, however, not true in general. It is not that the load factor affects the hauling speed very appreciably. This effect is in fact almost negligible, at least so far as passenger transport is concerned. It is, in the first place, the handling time that depends on the load factor. Given the average trip length, the more passengers a bus carries, the longer the time at stops will be, *ceteris paribus*. In freight transport this effect can be quite strong. This is particularly true in break-bulk cargo shipping, where the ships often have to spend more time in port than at sea (unless the load factor is abnormally low).

In the V-function ϕ is consequently a factor which has a negative effect on output. The total effect of ϕ on Q is, of course, positive. The partial elasticity of Q with respect to ϕ equals 1 plus the partial elasticity of the overall speed with respect to ϕ.

$$E_{Q\phi} = 1 + E_{V\phi}. \tag{4}$$

$E_{V\phi}$ is negative, but its absolute value is normally much less than unity.

The total factor costs of the producer of sts are written as the product of $\mathcal{N}$ and the cost per vehicle-hour $\mathcal{Z}$, which consists of the costs of the current inputs, pY, and the costs of all those factors which are fixed in the short run, including "overheads", vehicle capital and crew costs as well as track and terminal charges. These charges are assumed, for a given type of vehicle, to be proportional to the time of use.

A common procedure in transport service costing is to make a sharp distinction between the "traffic operation costs" (including vehicle capital and crew costs) and those costs which are not directly associated with the traffic operations, that is, the overhead costs. It is argued here that this sharp distinction is overdone and, when it comes to marginal cost calculation, can be quite misleading. The overhead costs constitute, of course, a notorious problem in all costing, and deserve special discussion here. However, in order not to break the flow of the main argument, this discussion is postponed to the last section.

Given the location of activities generating demand for sts, the users' own contributions to the fulfilment of this demand can be summarised by these four efforts:

1. *Catching* the vehicles producing sts, involving "feeder transport" in the widest sense, as well as waiting by passengers and storage of freight;

2. *Queuing*[1] *for* these vehicles on occasions when there is excess demand;
3. *Going by* the vehicles.

So far as freight is concerned, a fourth cost can be very important, namely:

4. The direct effort of loading and unloading.

At this stage the only necessary specification of the user cost function is to mark that in the first place ϕ, N and V are important determinants of the corresponding costs. The load factor is the principal determinant of the queuing costs. In the case of passenger transport, a high occupancy rate can also cause "crowding costs", that is, raise the cost of travel time. Given the overall speed, the number of vehicles in the system N determines the accessibility of service in time and space. A given fleet of vehicles can be allocated to a large number of separate lines—to maximise the accessibility in space—or to a relatively small number of lines to give a high frequency of services. If the former course is chosen, the feeder transport costs will be low, but the actual or potential waiting time cost will be high, and vice versa if the latter course is chosen. A generic name for all user costs caused by the necessarily imperfect accessibility of sts is "catchment costs".

The appearance of the third argument in the user cost function—the overall speed V—is explained by the facts that the transport time cost is inversely proportional to V and that, given the number of vehicles employed on a particular route, the frequency of service is proportional to V.

This concludes the presentation of the general model. In what follows, the model will be put to use to discuss the contents of the optimal price of sts in terms of marginal costs. On the basis of this discussion, it is possible to get a rough idea of what the financial result of optimal pricing would be for different categories of sts.

3. THE OPTIMAL PRICE IN TERMS OF MARGINAL COSTS

MC-pricing of transport services, as well as of other services where the user cost is a non-negligible social cost component, implies that the optimal price P should consist of two items—the cost effect on the producer, and the cost effect on the original users of the facility concerned, of another unit of output. The following expression is the most general formulation of the optimal price, or "pricing-relevant cost" as it may be preferable to call it.

$$P = \frac{\mathrm{d}}{\mathrm{d}Q}(TC^{prod}) + Q\frac{\mathrm{d}}{\mathrm{d}Q}(AC^{user}). \tag{5}$$

This general formulation of the pricing-relevant cost does not commit us to any decision on the relevant "run" for pricing, that is, on which factors of production should be assumed to be variable.

Turvey and Mohring discussed optimal bus fares in terms of short-run marginal

[1]The queuing may be conspicuous or latent as in the case where seat reservations are made beforehand. The disadvantages for the passengers of (e.g.) an airline of sometimes finding that the desired flight is fully booked are viewed as queuing costs.

costs (SRMC). By the present model their discussion is generalised, and a certain vagueness of definition can be clarified. After that we present an alternative "medium run" formulation of the pricing-relevant cost, which—it will be argued—seems more useful for sts pricing in practice. It will be shown, finally, that under the normal efficiency conditions P, as stated above, comes out the same no matter which "run" it is calculated for.

3.1 The pricing-relevant cost in the short run

In Turvey and Mohring (1975), as well as in an earlier paper on bus fares by Mohring (1972), only the right-hand component of (5) is recognised as pricing relevant, or, in other words, only various costs imposed on fellow passengers enter the optimal price. In the present model this corresponds to assuming all producer inputs to be fixed, allowing only the occupancy rate ϕ to vary. Under this assumption the general expression for the optimal price (5) takes this shape:

$$P_{SR} = \frac{Q\left(\frac{\partial h}{\partial \phi} + \frac{\partial V}{\partial \phi}\frac{\partial h}{\partial V}\,\mathrm{d}\phi\right)}{\left(NHV + \phi NH\frac{\partial V}{\partial \phi}\right)\mathrm{d}\phi} = \phi\frac{\partial h}{\partial \phi} + \phi\frac{\partial V}{\partial \phi}\frac{\partial h}{\partial V}. \qquad (6)$$

Two separate "imposed costs" can be distinguished: one is the crowding and/or queuing cost caused by raising the occupancy rate, represented by the left-hand term, and the other is the imposed handling time cost. As seen from (6), the latter effect on the user cost of raising ϕ is working via the negative effect on V of an increase in ϕ.

The crowding or queuing cost appears in one form or another for all modes of sts. On buses or trains where a number of standing passengers are allowed, the average comfort will fall as soon as all passengers cannot be seated. As the number of standing passengers grows the discomfort will be aggravated, and when the passengers are packed like sardines conditions are hardly endurable. In the "pure crowding case" it can, *a priori*, be assumed that except in very exceptional cases buses or trains *should* not be so full that some passengers have to be left behind. (It is probably not optimal to operate at such a high rate of use of holding capacity that some passengers of their own will choose to wait for the next bus/train in the hope that it will be less crowded.)

There is also a "pure queuing case" where no crowding costs are present. In freight transport the pure queuing case is applicable, provided only that the risk of damage does not increase as the load factor goes up. In passenger transport it may be argued that the pure queuing case is applicable, too, in all cases where no standing is allowed. However, even though everybody has got a seat, most travellers probably consider comfort to be more or less impaired by a high occupancy rate.

The imposed crowding cost appears as an increase in the cost of travel time. A separate effect of raising the load factor/occupancy rate is that, *ceteris paribus*, the transit time will go up. Given the transport capacity provided, more passengers or freight means that more time has to be spent in cargo handling or in boarding/alighting of passengers. In Turvey and Mohring's discussion of optimal bus fares the boarding/alighting time cost imposed on fellow passengers plays a main role,

and in Mohring's earlier paper (1972) on urban bus transport it is assumed to be the *only* pricing-relevant cost.

From one point of view it may seem appropriate to reserve the epithet "short run" to the costs just dealt with, which arise when all producer inputs are fixed, and only ϕ is a variable. This is in the closest analogy with the short-run costs of (for example) road use or the utilisation of other fixed facilities. From another point of view this does not appear as the most appropriate convention. From the point of view of the actual decision-making in a sts undertaking, the *fixing of the schedule* can be considered to be the most important dividing line between questions of short-term operations and longer-term "policy" matters. Some of the most typical short-term questions concern the problem how the schedule is to be kept in the face of varying conditions, including random fluctuations in the demand from one day to another; that is, how to maintain a *reliable* service.

In this light Turvey and Mohring's short-run marginal cost concept appears as a half-measure. They assume the total bus-hours to be fixed, but not that a fixed schedule is adhered to. An increase in the demand will in their "short-run" cost model cause time losses for the original passengers through a lengthening of the time at stops. This must imply either that the bus line schedule is being adjusted in accordance with changes in the expected demand, or that the important desideratum of reliability of the service is ignored. For the former alternative (where the schedule is assumed to be a variable) the "short run" is not a very fitting designation. If the schedule can be adjusted, one might just as soon assume (e.g.) the vehicle input to be variable. To abandon reliability is to my mind an unrealistic representation of the behaviour of most sts undertakings (including urban bus services).

The very idea of sts is that a published time-table is to be maintained even under adverse circumstances. The main precondition for this is that the schedule is not made too tight, but that a time loss incurred at one stage of the round trip can be made up subsequently by speeding up operations. With reference to expression (6) for P, this means that the "imposed handling time cost" is zero in the short run. As long as the schedule is fixed, the potential handling time costs caused by additional traffic will be reflected in various "schedule-keeping costs". This is a generic name for a variety of minor cost items which make themselves known on the occasions when the actual demand is greater than usual.

For example, in liner shipping, when an unusually large quantity of cargo which is difficult to handle is presented for a particular sailing, the time lost in the handling operating will be regained by reducing both the dead port time and the cruising time on that particular sailing. By working overtime in port the ship's idle time can be substantially cut, and the time at sea can be reduced by steaming at maximum speed. This is obviously not done with impunity; the tighter the schedule is, the higher will the cost be of saving a day in port or at sea. In passenger transport the main possibility to regain time when the boarding/alighting time has been unusually long is to increase the running speed. In urban bus transport another frequently used method of time-saving when behind schedule is to cut the time at stops—to the detriment of the comfort of passengers. The time per stop can be shortened by abrupt stopping and starting, by rushing boarding and paying passengers, by ignoring the woman with the pram, by not waiting for the late-comer running to catch the bus, etc.

Conversely, when an unusually small quantity of cargo is lifted, the ship does not

normally steam ahead of schedule, but keeps the schedule by going slowly and avoiding all overtime work. In passenger transport it is particularly important not to frustrate travellers by early departures.

In the present model this means that an additional constraint should be imposed, namely, that the overall speed V should be assumed to be given in the short run. In that case the existence of some current inputs Y has to be recognised. Otherwise it would not be possible to maintain a fixed schedule in the face of random fluctuations in the demand.

It is true that delays occur in practice in all modes of sts, since schedules are not made so loose that they would be possible to keep under any circumstances. It seems, however, wrong in a general sts model to take no account at all of the quality represented by the reliability of service.

The assumption that the current producer inputs Y are variable but that a fixed schedule is maintained, implying that not only N and X but also the overall speed V is to be regarded as given in the model, yields this variant of the short-run pricing-relevant cost:

$$P'_{SR} = \frac{Np\,\mathrm{d}Y + Q\dfrac{\partial h}{\partial \phi}\mathrm{d}\phi}{NH\bar{V}\,\mathrm{d}\phi}. \tag{7}$$

The condition that $V(\bar{X}, Y, \phi) = \bar{V}$ means that

$$\frac{\partial V}{\partial Y}\mathrm{d}Y + \frac{\partial V}{\partial \phi}\mathrm{d}\phi = 0. \tag{8}$$

Inserting the resulting value for $\mathrm{d}Y/\mathrm{d}\phi$ in (7) gives the following result as to P'_{SR}:

$$P'_{SR} = \frac{E_{V\phi}}{E_{VY}}\frac{pY}{\dfrac{Q}{N}} + \phi\frac{\partial h}{\partial \phi}, \tag{9}$$

where $E_{V\phi}$ is the elasticity of V with respect to ϕ, and E_{VY} is the elasticity of V with respect to Y. The ratio of pY to Q/N stands for the current input cost per unit of transport output. It corresponds to the average variable cost (AVC) of the sts producer. The whole left-hand term of (9) represents the "schedule-keeping cost" component of P'_{SR}. It is consequently found that AVC times the ratio of $E_{V\phi}$ to E_{VY} plus the imposed crowding and/or queuing cost $\phi(\partial h/\partial \phi)$ makes up the optimal price in the short run.

3.2 The pricing-relevant cost in the medium and long run

In actual practice, decisions about pricing of sts are *not* taken in a short-run (="fixed schedule run") perspective. The tariff of freight rates or passenger fares of most sts undertakings is regarded as possessing a degree of "fixity" at least comparable to that of the vehicle input. Price-making of sts is a "medium-run" concern. Let us see which appearance the medium-run pricing-relevant cost takes in the present model. Forming a wholly general expression for P—without any precondition as to which factors of production are fixed or variable—gives the following result:

$$P = \left[\left(Z + Q\frac{\partial h}{\partial N}\right)\mathrm{d}N + \left(N\frac{\partial g}{\partial X} + \frac{\partial c}{\partial X} + Q\frac{\partial h}{\partial V}\frac{\partial V}{\partial X}\right)\mathrm{d}X + \left(NP + Q\frac{\partial h}{\partial V}\frac{\partial V}{\partial \Upsilon}\right)\mathrm{d}\Upsilon + Q\left(\frac{\partial h}{\partial \phi} + \frac{\partial h}{\partial V}\frac{\partial V}{\partial \phi}\right)\mathrm{d}\phi\right] \div \left[HV\,\mathrm{d}N + \phi N\left(V\frac{\partial H}{\mathrm{d}X} + H\frac{\partial V}{\mathrm{d}X}\right)\mathrm{d}X + \phi NH\frac{\partial V}{\mathrm{d}\Upsilon}\mathrm{d}\Upsilon + \left(NHV + \phi NH\frac{\partial V}{\partial \phi}\right)\mathrm{d}\phi\right]. \quad (10)$$

In many sts undertakings such as urban bus companies, the number of vehicles N is definitely regarded as variable when it comes to fixing the prices. In railway companies the holding capacity H, as represented by the number of carriages per train, is even more "variable" in the pricing-relevant run. Some sts operators may not recognise any fixity whatever in vehicle and system design when decisions about pricing policy are to be taken.

A question arising is now: which marginal cost is the most appropriate one for pricing from a social welfare point of view? Before going into this well-trodden area of controversy, it is pertinent to ask if it really matters whether prices of sts are based on short-run, medium-run or long-run marginal costs?

3.3 Under the efficiency conditions the degree of factor fixity makes no difference to the pricing-relevant cost

The efficiency conditions for the sts system design imply that the total system costs should be the lowest possible at each level of output. Forming a lagrangian equation, the efficiency conditions are obtained as follows:

$$\Pi = TC^{prod} + TC^{user} + \lambda(Q - \phi NHV). \quad (11)$$

Taking the partial derivatives of Π with respect to N, X, Υ and ϕ gives:

$$\frac{\partial \Pi}{\partial N} = Z + Q\frac{\partial h}{\partial N} - \lambda\phi HV = 0 \quad (12)$$

$$\frac{\partial \Pi}{\partial X} = N\left(\frac{\partial g}{\partial X} + \frac{\partial c}{\partial X}\right) + Q\frac{\partial h}{\partial V}\frac{\partial V}{\partial X} - \lambda\phi N\left(V\frac{\partial H}{\partial X} + H\frac{\partial V}{\partial X}\right) = 0 \quad (13)$$

$$\frac{\partial \Pi}{\partial \Upsilon} = Np + Q\frac{\partial h}{\partial V}\frac{\partial V}{\partial \Upsilon} - \lambda\phi NH\frac{\partial V}{\partial \Upsilon} = 0 \quad (14)$$

$$\frac{\partial \Pi}{\partial \phi} = Q\left(\frac{\partial h}{\partial \phi} + \frac{\partial h}{\partial V}\frac{\partial V}{\partial \phi}\right) - \lambda\left(NHV + \phi NH\frac{\partial V}{\partial \phi}\right) = 0. \quad (15)$$

Using the efficiency conditions (12)–(15) to specify the general expression for the optimal price P according to (10), it is found that the denominator of (10) equals the numerator times $1/\lambda$. Under the efficiency conditions the optimal price is consequently equal to the lagrangian multiplier λ, *no matter which factors are fixed or variable*. Setting any number of the factor increments $\mathrm{d}N$, $\mathrm{d}X$, $\mathrm{d}\Upsilon$ or $\mathrm{d}\phi$ equal to

zero has the same effect on the numerator and on the denominator of (10). Even when all producer inputs are fixed, and only the load factor ϕ can be varied to produce different outputs, the end result is one and the same:

$$P^{opt} = \lambda. \tag{16}$$

It thus appears that the question whether the short-run, medium-run, or long-run marginal cost should be the basis for pricing is not an issue *as long as the efficiency conditions are fulfilled.* It becomes an issue only in a situation where the facility design is inefficient for the production of the actual output. Received welfare economic theory states that in those circumstances the best is made of a non-optimal situation by basing the price on the short-run marginal cost. This secures an optimal utilisation of the short-run fixed resources. In the discussion of road pricing this argument has been driven home forcefully by leading economists (see, for example, Walters (1968)). It should be observed that the "short run" has no exact self-explanatory definition. However, so far as most production taking place at fixed plants is concerned, this does not cause very much confusion. The "putty/clay" metaphor pinpoints the key dichotomy: when the long-run costs are considered, it implies that the factors of production are like putty, which is true only at the planning stage, that is, before the production plant is built. The short-run costs apply in the clay state. Certain problems of definition remain (for example, how the costs of different categories of salaried staff and labour are to be treated), but the main idea is clear enough.

In the case of transport vehicle services, on the other hand, things look quite different. For example, where a bus transport system is inefficient—too many buses produce too few passenger-miles, or vice versa—a sensible economist would hesitate to suggest that the bus fares should be set on the basis of marginal costs, which would assume that the input of vehicles was given. To reform a structure of bus fares is quite an involved business. A change in fares is hardly justified unless the new structure is likely to be a fairly permanent one, or the change is at least a step towards rather than away from a more permanent solution. The more sensible thing to do is to adjust the size of the bus fleet to be consistent with the level of output which corresponds to the point of intersection between a marginal cost curve, assuming the vehicle input to be variable, and the demand curve.

Calling the latter marginal cost the "medium-run marginal cost" (MRMC), the conclusion is in other words that positions along any SRMC-curve *off* the MRMC-curve are irrelevant for pricing policy. Positions off MRMC are likely to be so temporary that there is no time to adjust prices to them, and it would not be worth the extra costs of pricing.

4. THE AVERAGE COST OF A MARGINAL UNIT OF CAPACITY

When it can be assumed that the efficiency conditions are fulfilled as a general rule, one property of the pricing-relevant cost is very useful for a theoretical analysis of the contents of the optimal price: namely, that the equality $P^{opt} = \lambda$ is independent of which or how many factors of production are assumed to be fixed. As seen from (10) above, a rather complicated expression for P would result if values different

from zero were assumed for all of $d\phi$, dN, dX, and dY. A particularly useful formulation of the pricing-relevant cost is obtained by considering the input of another unit of capacity, either another carriage per train or an additional complete vehicle of the same design as the existing ones, while keeping the load factor/occupancy rate and (in the latter case) current inputs per vehicle constant. A problem of factor indivisibility can arise under certain circumstances, or more exactly where a marginal unit of capacity constitutes a *relatively* large addition to capacity. The crucial criterion for deciding whether or not it is a reasonable approximation to treat the number of least-capacity units as a continuous variable is not whether 5, 25 or 75 passengers can be carried by an additional bus, or whether 5,000, 25,000 or 75,000 tons of cargo can be carried by an additional ship, but the number of vehicles engaged in sts production in the system concerned. Treating this number as a continuous variable when calculating the pricing-relevant cost is the same thing as approximating the marginal cost by the "average cost of the marginal plant". If there is only one plant serving a particular market, the average cost of another plant is certainly a poor marginal cost proxy in wide output intervals. If there are 100 plants originally, it can be a nearly perfect MC proxy. There is obviously no "magic number" the limit below which it can be deemed illegitimate to ignore factor indivisibility. Let it only be said that, by comparison with other industries for which the "average cost of the marginal plant" is a widely accepted approximation to the marginal cost, sts production where each vehicle corresponds to a separate plant looks with a few exceptions like quite a suitable area of application.

The average cost of a marginal unit of capacity takes entirely different shapes in the case where each vehicle consists of a single unit, and in the case where trains of carriages are formed. In the latter case it is really an increase in the holding capacity per vehicle rather than in the number of vehicles that is the relevant factor increment. These two cases are considered below. We start with the train case.

4.1 The train case

The holding capacity H has been written as a function of an unspecified vehicle design vector X. Consider now a case where the holding capacity can be written as the product of the holding capacity of each individual carriage S (="size") and the number of carriages per train L (="length" of the train).

$$H = SL \tag{17}$$

The running speed will be decreased by a lengthening of the train unless an offsetting increase is made in the current inputs Y (fuel and freight-handling labour). It is both realistic and convenient to assume that this is in fact done, because it is then possible to assume that the overall speed V is kept constant. The boarding/alighting time per train is by and large independent of the train length, since the boarding/alighting capacity (=number of doors) can be assumed to be proportional to length. The same goes for the handling time of a freight train. It can be mentioned that in the latter case the loading/unloading of the railcars often takes place while the individual railcars are being uncoupled from the engine, so that the hauling capacity of the engine is fully utilised.

Note that the chosen value of the factor increment dY is inconsequential for the value of P (as long as the increment is small), provided that the efficiency conditions

are fulfilled. Therefore, no further limitation is imposed by choosing a value of $d\Upsilon$ that just offsets the negative effect on V of adding a carriage. The costs of capital, crew and charges per vehicle-hour can in the train case be assumed to equal $a + bL + c(L)$. That is to say, the capital and crew costs are equal to a constant plus a length-proportional component. In addition there are the "track user charges" on the train, which are a function of the train length. On these assumptions the production function (18) and the total social costs (19) are written:

$$Q = \phi NSLV(L, \Upsilon) \tag{18}$$

$$TC = N[a + bL + c(L) + p\Upsilon] + Qh(\phi, N, V). \tag{19}$$

On the further condition that the overall speed should stay constant, the pricing-relevant cost comes to:

$$P = \frac{N\left(b + \frac{\partial c}{\partial L}\right)dL + Np\, d\Upsilon}{\phi NS\bar{V}\, dL}. \tag{20}$$

The side-condition $V(L, \Upsilon) = \bar{V}$ gives:

$$\frac{\partial V}{\partial L} dL + \frac{\partial V}{\partial \Upsilon} d\Upsilon = 0. \tag{21}$$

Using (21) to eliminate $d\Upsilon$ from (20) yields:

$$P = \frac{b + \frac{\partial c}{\partial L} + p\left(-\frac{\partial V}{\partial L}\right)\frac{\partial V}{\partial \Upsilon}}{Q/NL}. \tag{22}$$

The total revenue from optimal pricing amounts to:

$$PQ = N\left(bL + cE_{cL} + p\Upsilon\frac{-E_{VL}}{E_{V\Upsilon}}\right), \tag{23}$$

where E_{cL} stands for elasticity of the track user charge with respect to L, E_{VL} for the (partial) elasticity of the overall speed with respect to L, and $E_{V\Upsilon}$ for the (partial) elasticity of the overall speed with respect to Υ.

A comparison of the total revenue with the total producer costs $N[a + bL + c(L) + p\Upsilon]$ shows that a contribution to the recovery of the train-length-independent capital and crew costs aN will be obtained if E_{cL} and/or the ratio $-E_{VL}/E_{V\Upsilon}$ exceeds unity. If the track user charges were proportional to the train length, E_{CL} would be equal to unity. Nothing is known about this, since—to my knowledge—no empirical work on rail track congestion costs has been done. Given the engine, the value of $-E_{VL}/E_{V\Upsilon}$ is initially well below unity. However, as the train is made longer and longer, this ratio will most likely exceed unity sooner or later. It is quite conceivable that the continued addition of carriages will at some point result in a value of P which equals AC^{prod}. The point is, however, that such a long train is unlikely ever to be consistent with the efficiency conditions. Well before the break-even point is reached, the input of another complete train is probably indicated.

The result of considering carriage additions in the pricing-relevant cost formula is consequently somewhat inconclusive. The following single vehicle case is in fact relevant also for the train case, since the optimal train length cannot be determined without regard to the costs and benefits (of increasing the frequency of service) resulting from an increase in the number of trains.

In passing, it can, finally, be mentioned that a common misunderstanding is that the pricing-relevant marginal cost of railway transport services will rise with increases in the size of capacity additions. Given the number of carriages in a train, the marginal cost of another passenger is practically zero; if it is taken into account that another carriage may have to be added to the train, the marginal cost will appear to be substantially greater; if a complete new train has to be put in, the marginal cost can be very high indeed; and if the central station has to be rebuilt in order to accommodate more and longer trains . . . and so on. The fallacy of this view of an escalating marginal cost is that it ignores the user cost savings obtained by each successively larger addition to capacity. Adding another carriage will reduce user queuing and crowding costs, and the input of another train will in addition save waiting time for the users. And the rebuilding/modernisation of railway stations can yield substantial user benefits. The last point belongs, however, to "another department": it is a question for the theory of MC-pricing of transport terminal services.

4.2 The single vehicle case

Setting dX, dY, and $d\phi$ all equal to zero in the general expression for P according to (10), and thus letting only dN take a value different from zero, yields the same value for the pricing-relevant cost as that obtained for λ from the efficiency condition (12). If we consider a change only in N, it is inevitable that user cost will be affected. By contrast with the train case, the pricing-relevant cost will in the present case not be restricted to producer cost items only. In the train case this was possible by also assuming $dY \neq 0$ so as to offset the negative effect on V of a lengthening of trains. It would, of course, be theoretically possible to do the same thing here, that is, to offset the effect on user cost on an increment dN by simultaneously considering the factor *decrements*, dX and dY, and/or an increment $d\phi$, choosing such values of these differentials that the user cost stays the same. This is, however, both unrealistic and inconvenient. The following attractively simple formulation of the pricing-relevant cost is much to be preferred:

$$P^{opt} = \lambda = \frac{Z + Q\dfrac{\partial h}{\partial N}}{\phi HV} = \frac{ZN}{Q} + N\frac{\partial h}{\partial N}. \tag{24}$$

This result may appear too "simple". I believe that it is of fundamental importance for the issue at stake. Provided that the efficiency conditions are fulfilled, the optimal price of sts is equal to the transport producer costs per unit of output (passenger trip or freight transport) ZN/Q plus the product of the number of vehicles N and the derivative of the average user cost with respect to N. *A priori* it is clear that the derivative $\partial h/\partial N$ is negative, which means that the optimal price will fall short of the average producer cost by an amount equal to the absolute value of $N\partial h/\partial N$.

5. THE FINANCIAL RESULT OF MC-PRICING

The standard way of predicting the financial result of MC-pricing is to consider whether there are economies of scale. For sts production several studies have been made of the relationship between the total *producer* costs and the "size" or turnover of individual sts firms. A typical point of departure has been that sts firms are to be regarded as markedly "multi-plant firms", on the assumption that each individual transport vehicle represents a production plant.

Significant economies of firm size are mostly to be found in industries consisting of single-plant firms. Economies of plant size and economies of firm size are then synonymous. Multi-plant firms are presumably operating with plants of optimal size (at the time of construction of the plants); in this case the economies of scale in production can be expected to be largely exhausted. Mergers of multi-plant firms can rarely be justified on production technological grounds; the rationale is marketing considerations, or simply a wish to restrict competition.

The pertinent question is, therefore, what will happen to the vehicles used by a particular sts firm following an increase in its market share? Studies of (e.g.) mergers of shipping lines are rather inconclusive on this. A merger can be a means of facilitating the containerisation of a particular trade route, but in other cases nothing in particular seems to happen as far as production is concerned. The same ships in the same number will be plying the same route in much the same way as before. There is also the possibility that savings in "overhead costs" will be achieved by increasing the size of a firm. In sts undertakings the costs which are not directly associated with traffic operations—the costs of general management and administration, etc.—make up an appreciable portion of the total costs. A common experience is, however, that hopes of savings in overheads as a result of agglomeration are illusory. There is ample empirical evidence showing that overhead costs develop basically in proportion to the size or turnover of sts companies. As a representative example, a chart of the relationship between overhead costs and fleet size for a number of bus companies in Britain is presented in Figure 1. The hypothesis of proportionality is borne out also in liner shipping. A cross-section analysis of a large number of shipping lines undertaken in 1960 (Ferguson *et al.*, 1961) showed that the administrative costs amounted to about 10% of gross revenue, regardless of the fleet size. A similar result is reported to have been reached in an investigation of American shipping lines ten years later (J. W. Devanney III *et al.*, 1972). Cross-section studies of the *total* costs of (American) railroad companies and bus transport companies do not generally point to any other conclusion. The evidence by and large supports the hypothesis of proportionality.[2]

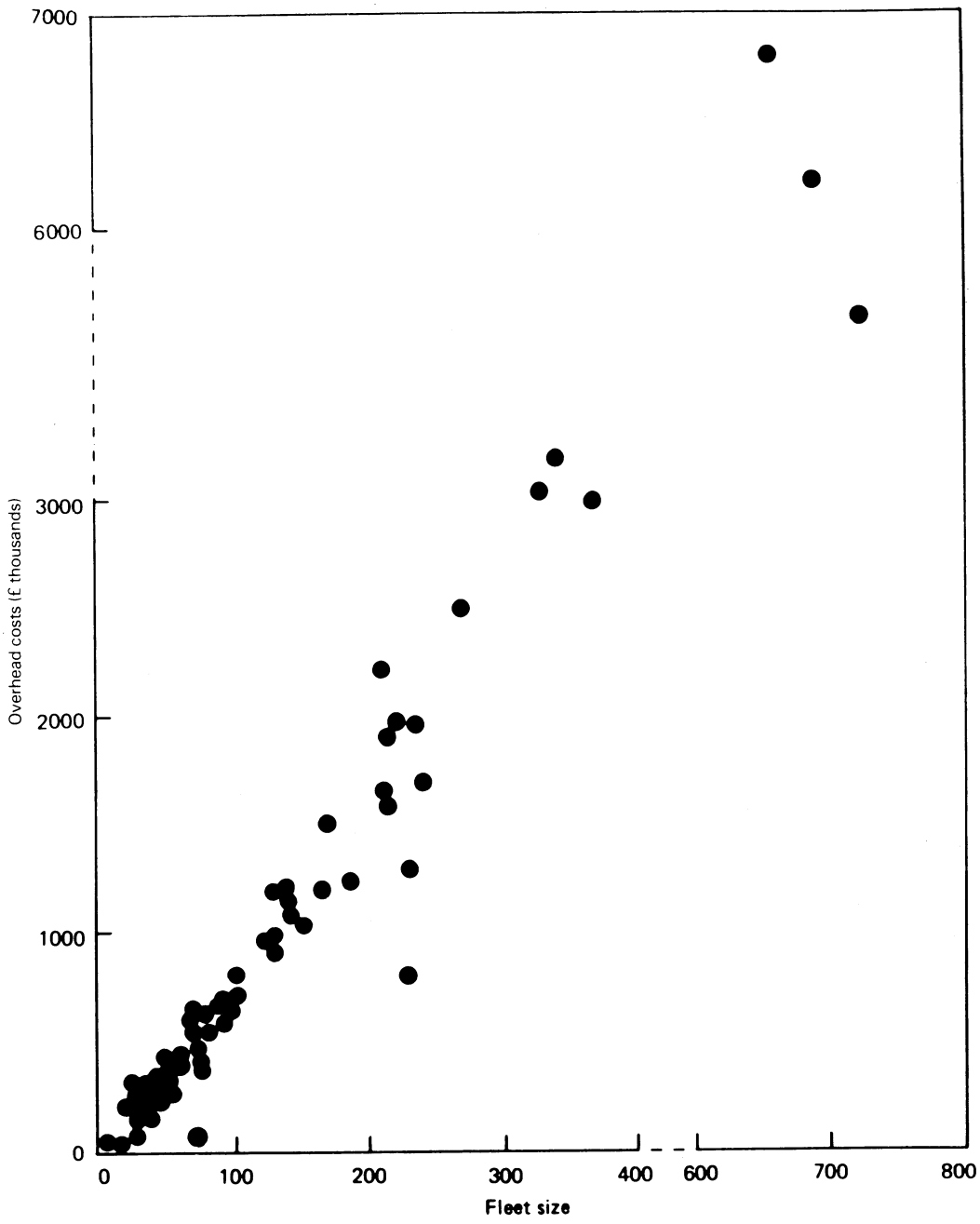
A main point made here is, however, that findings on the elasticity of transport producer costs with respect to output cannot be used for judging what would be the likely financial result of marginal cost pricing. Disregard of the user costs makes the cost picture wholly inconclusive.

In the present model it has been shown that a financial deficit will be the result of marginal cost pricing, irrespective of whether or not the total producer costs develop proportionally to the volume of transport. The question is only how large

[2]See Zvi Griliches (1972), G. Borts (1960), J. Meyer *et al.* (1959), N. Lee and N. Steedman (1970) and R. K. Koshal (1970) and (1972).

FIGURE 1

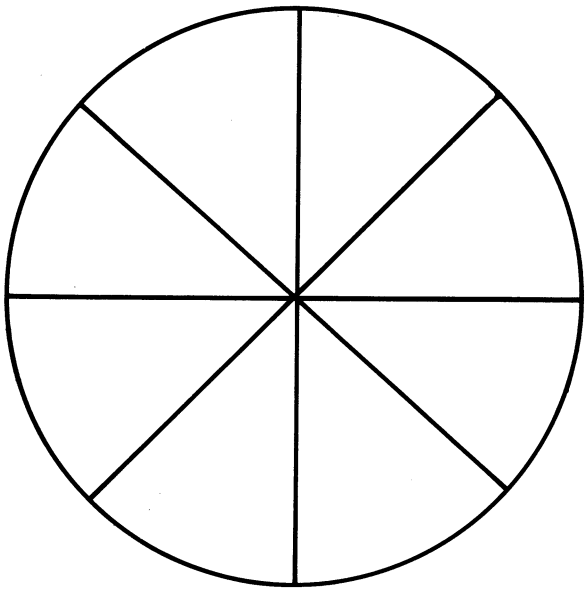
Relationship between overhead costs and fleet size for 63 *municipal bus undertakings.*



Source: APPTO Annual Summary of Accounts and Statistical Information, year ended 30 March 1973.

FIGURE 2

Schematic picture of the radial bus services of Circletown



the deficit will be. As is clear from (24) above, the relationship between the user cost and the number of vehicles in the system is of crucial importance for the financial result of optimal pricing of sts. Let us look a bit closer at this relationship.

5.1 Vehicle number economies in the user costs

As was mentioned in the presentation of the general model, the total user cost per unit of transport can be represented by the sum of (1) the catchment costs, (2) the queuing cost, (3) the transit time cost, and so far as freight is concerned (4) the direct handling cost.[3] It can be assumed straight away that item (4) is independent of N. The relationship between items (1) and (2) and N is better understood by introducing the sts qualities of "density of service" (D), and "frequency of service" (F), and by dividing the catchment cost into (1.1) the "feeder transport" cost, and (1.2) the "waiting cost". The concept of service density can be illustrated by the radial bus services of a centralised city (Figure 2), where D equals the number of the "spokes of a wheel" divided by the area of the city.

The feeder transport[4] costs of the users of sts are a function primarily of D, while the costs caused by the *in*frequency of service, including waiting time of passengers and storage time of freight, is obviously a function of F. Note the distinction made between waiting and queuing. Waiting occurs on account of infrequency of service, while queuing is caused by occasional excess demand. It is clear that the queuing

[3]That is, the effort of getting the objects of transport on and off vehicles; this is normally minute for passengers, but can be quite appreciable for freight, which may incur high handling charges.

[4]Feeder transport includes all modes of catching the sts concerned, such as walking or driving one's own car to/from bus stations, and hauling seaborne freight by road or rail to/from seaports.

cost, besides being a function of the load factor/occupancy rate, depends on F in much the same way as the waiting/storage cost does. The more frequent a particular service is, the less inconvenient it will be that excess demand occurs from time to time.

Provided that the routeing of the services in the system adheres throughout to the shortest-distance principle,[5] item (3)—the transit time cost—can be assumed to be independent of N.

In summary, the total user costs per unit of transport can be written in this way:

$$AC^{user} = f(D) + w(F) + q(\phi, F) + t(\phi, V) + c_h, \tag{25}$$

where

$f(D)$ = feeder transport cost as a function of the density of service D
$w(F)$ = waiting cost as a function of the frequency of service F
$q(\phi, F)$ = queuing cost as a function of the load factor/occupancy rate ϕ, and F
$t(\phi, V)$ = transit time cost as a function of ϕ, and the overall speed V
c_h = handling charges.

The point is, finally, that, given the vehicle design and the load factor/occupancy rate, the number of vehicles in the system N is by and large proportional to the product of the density of service D and frequency of service F.

$$N = \alpha DF. \tag{26}$$

The user cost term, $N(\partial h/\partial N)$ of the pricing-relevant cost according to (24) above, can now be developed in the following way:

$$N\frac{\partial h}{\partial N} = D\frac{\partial f}{\partial D} + F\left(\frac{\partial w}{\partial F} + \frac{\partial q}{\partial F}\right). \tag{27}$$

Inserting this in (24), and multiplying by Q to get the total revenue from optimal pricing, gives:

$$\lambda Q = \mathcal{Z}N + fQE_{fD} + qQE_{qF}, \tag{28}$$

where

$$E_{fD} = \frac{\partial f}{\partial D}\frac{D}{f}$$

$$E_{wF} = \frac{\partial w}{\partial F}\frac{F}{w}$$

$$E_{qF} = \frac{\partial q}{\partial F}\frac{F}{q}.$$

[5]This is not necessarily true for all categories of sts. In freight transport in particular, it is not unusual for the vehicles carrying out the scheduled transport services to do also a good part of the feeder transport work, implying that more or less substantial deviations are made from the shortest-distance route pattern. An additional cost-reducing effect of an increase in the density of demand can consequently be that individual services will be more and more in line with the shortest-distance principle.

It is immediately seen that a financial deficit would result, which would fall short of the total producer costs $\mathcal{Z}N$ by the sum of three products, each made up of a total user cost item and the corresponding user cost elasticity with respect to D and F, respectively.

It should be pointed out that the first term of (27) may be inoperative under certain circumstances. It can happen that, for some geographical reason, the density of services in the system is fixed even in the long run. For instance, the density of the shipping lines between two continents can in most realistic cases be regarded as having an upper limit determined by the existence of suitable ports on either coastline. In such a case the financial deficit of MC-pricing will be restricted to $Q(wE_{wF} + qE_{qF})$.

What can be said about the relative order of magnitude (relative to the producer cost) of the financial deficit? It is well known that the feeder transport cost constitutes a substantial proportion of the total system costs for practically all scheduled transport systems. This proportion certainly varies a great deal; it tends, for example, to be higher for local transport systems than for systems of very long-distance transport. To determine the value of E_{fD}, it is helpful to develop this elasticity in terms of two other elasticities—the elasticity of the average feeder transport *distance*, d with respect to D, and the elasticity of the feeder transport *cost* with respect to d. The average feeder transport cost per passenger or freight ton is directly determined by d, which in turn is determined by D.

$$E_{fD} = E_{dD}E_{fd}. \tag{30}$$

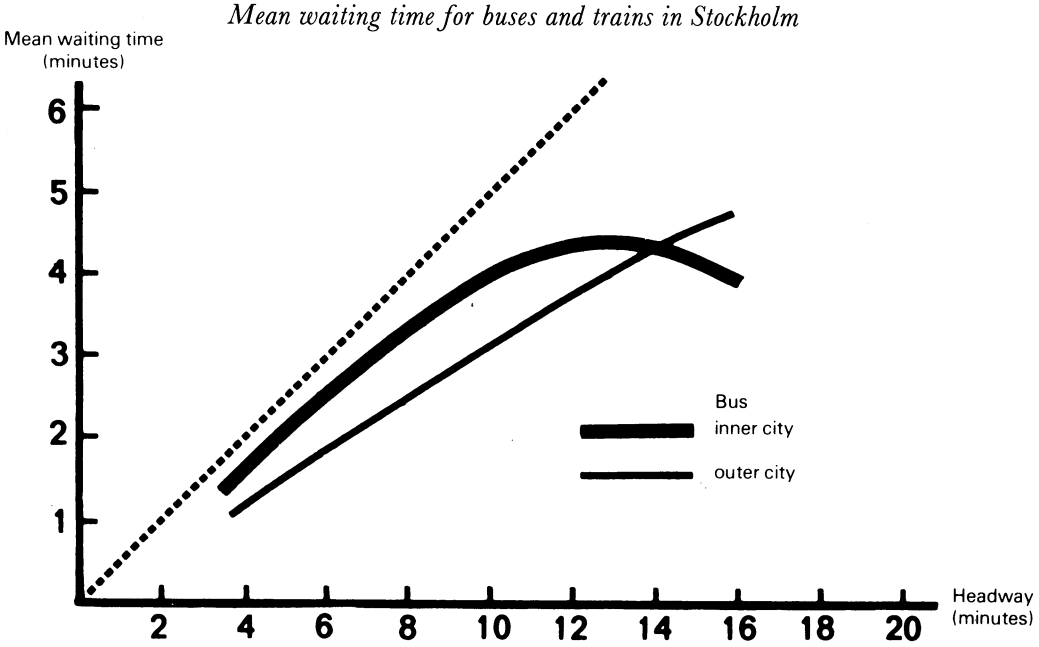
If the points of origin and destination are uniformly spread in an area covered by a system of scheduled transport services, E_{dD} tends to equal minus one. The value of E_{fd} for passengers is unity as long as the feeder transport is carried out by walking or by private car. If public transport is used (i.e. the total trip includes both scheduled feeder transport and scheduled trunk line transport), E_{fd} will be less than unity only if the feeder transport fares taper off with distance. For freight, feeder transport costs tapering off with distance are the rule rather than the exception. Thus, in summary, E_{fD} is typically equal to minus unity for passenger feeder transport, and somewhat greater than minus unity (i.e. between -1 and 0) for freight feeder transport.

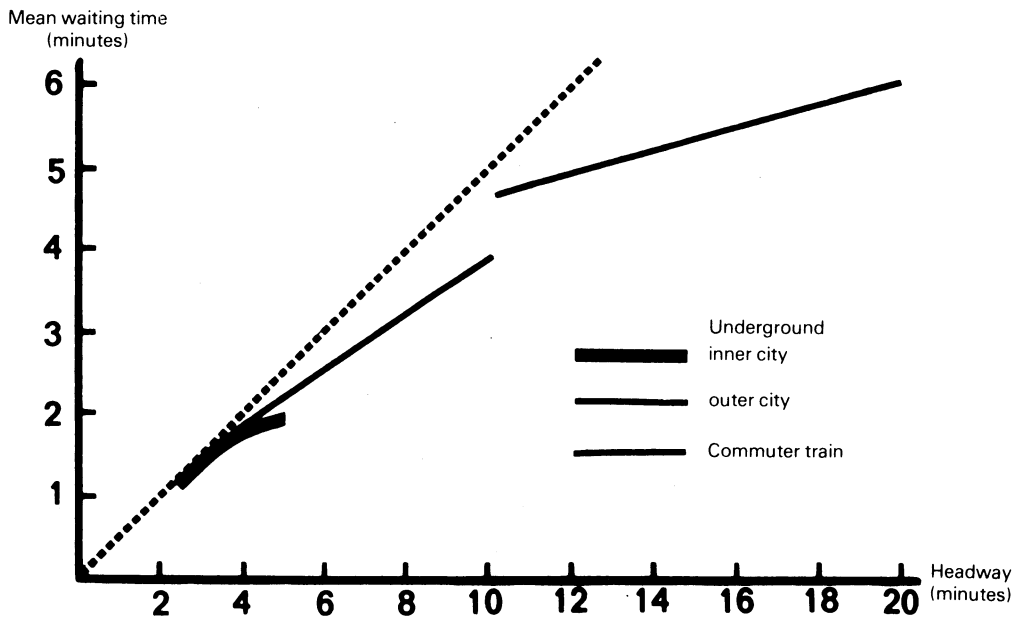
The waiting cost $w(F)$ and queuing cost are rather difficult to calculate unless they take the form of actual waiting and queuing time at stops/stations. In urban public transport this is what happens; it is easily established that the total waiting cost of passengers makes up an appreciable proportion of the total system costs in urban public transport, while the total queuing cost (rightly) constitutes a marginal item. In an urban transport context a useful benchmark value for the elasticity of the waiting cost with respect to the frequency of services, E_{wF}, is obtained by the observation that, if passengers arrive at random to stops/stations, the mean waiting time will be inversely proportional to the frequency. (See Figure 3.) The same goes obviously for E_{qF}.

But users of local bus services in rural areas, and of inter-regional scheduled transport, take the trouble to learn the time-tables, and the waiting costs caused by infrequent connections do not take the form of actual waiting time at stops/stations. It may be that the "waiting cost" is still inversely proportional to the frequency of

FIGURE 3

Mean waiting time for buses and trains in Stockholm





Source: TU 71: Trafikundersökningar i Stockholmsregionen hösten 1971, resultatrapport nr 1. Stockholms Läns Landsting, Trafiknämnden.

services; i.e. E_{wF} may still take the value minus one (although the waiting cost is certainly less than it would be if passengers were arriving at random to stops/stations). It is, however, difficult to give a simple rationale for this particular value of E_{wF} outside urban public transport.

So far as freight transport is concerned, some idea of the values of $w(F)$ and E_{wF} can be obtained by an inventory theoretical approach. It can, for example, be shown that under fairly general conditions the required level of safety stock tends to be inversely proportional to the square root of the frequency of service (Baumol and Vinod (1970)).

In summary, thus, in the interval of values of E_{wF} between $-\frac{1}{2}$ and -1, a range of values towards the (absolutely) higher limit is likely to be applicable to passenger transport, and a range of values towards the (absolutely) lower limit is likely to be applicable to freight transport. Moreover, since the value of time of passengers is many times higher than the value of time of freight, the absolute values of the total user costs fQ, wQ, and qQ are much greater in passenger transport than in freight transport. In proportion to the producer costs, this difference between passenger transport and freight transport user costs is presumably somewhat less large. Still the main conclusion is that the coexistence of vehicle size economies in producer cost, and vehicle number economies in user costs, can, regardless of the mode of transport, be predicted to make scheduled passenger transport a pronounced decreasing-cost industry in the sense that optimal pricing will result in a relatively large financial deficit. Concerning scheduled freight transport, the picture appears less clear. The total revenue from MC-pricing can always be predicted to fall short of the total producer costs. The relative size of the financial deficit is likely to vary substantially between sts for low-value and very high-value goods. On average, however, the character of a decreasing-cost industry is much less pronounced than in passenger transport.

APPENDIX

The Problem of the Overhead Costs

The conclusion that MC-pricing of sts would result in a substantial financial deficit probably does not come as a surprise to a cost accountant brought up in the modern "direct" or "marginal" costing school of thought. If it is accepted that the vehicles are neither a "common cost" nor a fixed cost in the pricing-relevant run, the result of basing the MC-pricing on direct costing principles will typically be a financial deficit equal to the total overhead costs.

It is true that the preceding analysis has led to the conclusion that optimal pricing of sts will as a general rule result in a financial deficit, which may be of the same order of magnitude as the total overhead costs, but which can, of course, be either smaller or substantially larger. The deficit is, however, due to the number economies in the user costs. It seems appropriate to attempt to refute the alternative "explanation", and show that the existence of overhead costs plays no part in the financial deficit of MC-pricing.

The problem of the overhead costs has been avoided so far. Now, at last, it is time to face it.

In the present model the total producer cost per hour is assumed to equal the product $N\mathcal{Z}$, where $\mathcal{Z}$ represents the total cost per vehicle-hour, including vehicle capital and crew costs, fuel costs, and charges, but also the overhead costs per vehicle-hour. On this assumption, the *incremental* producer cost per additional vehicle-hour equals the total average producer cost per vehicle-hour, that is, $\mathcal{Z}$.

$$IC^{prod} = \mathcal{Z} \tag{A.1}$$

This is in line with the evidence cited above of proportionality between overhead costs and fleet size. However, one problem which so far has been avoided is that $TC^{prod} = \mathcal{Z}N$ is really a long-run relationship. It can be argued that in the pricing-relevant "medium run" most overhead factors are fixed, and a different relationship between TC^{prod} and N should be used when discussing optimal pricing. The justification for the procedure adopted in the previous analysis is that, although the incremental cost per additional vehicle $\mathcal{Z}$ is obtained from a relationship applicable in the long run, $\mathcal{Z}$ seems still to be the best available proxy for the pricing-relevant medium-run value of the incremental cost per additional vehicle-hour. The following discussion seeks to substantiate the claim that it is a sounder proxy for $MRIC^{prod}$ than the average traffic operation cost per vehicle-hour, which would be used under the application of modern costing principles.

A common procedure in transport service costing is to distinguish the "cost of traffic operations" from those costs which are not directly associated with traffic operations, which are usually grouped under the catch-all name of "overhead costs". The costs of traffic operations are as a matter of course assumed to be proportional to the capacity provided, e.g. to the total vehicle-hours of operation, while the overheads are considered to be "unallocable", that is, to represent factors of production which cannot be allocated by any measure of capacity or output.

This may seem reasonable *per se*. The blunder is made when it comes to marginal or incremental cost calculations.

It can be helpful to give a name to the costing fallacy, which I shall attempt to put my finger on. In my view it boils down to a rather *disparaging treatment of overhead costs*. For the sake of brevity, let the total overhead costs be represented just by the total costs of the salaried staff, which can be designated sM (s for "salary", and M for "man-hours"). Denoting in addition the traffic operation cost per vehicle-hour by z, an expression for total producer costs typical of costing methods takes this appearance:

$$TC^{prod} = sM + zN. \tag{A.2}$$

On the crucial assumption that z is independent of N, the incremental cost per additional vehicle-hour is apparently equal to z. It is claimed here that this underestimates the relevant incremental cost, and that $\mathcal{Z} = z + sM/N$ is a closer approximation to the true value under normal conditions.

The basic problem of the costing approach is that all factors, the cost of which are considered to be "unallocable", are treated as if they were useless. As opposed to other factors of production, the overheads appear as pure "deadweight": they swell the cost side and do not seem to yield any benefits. This is illogical. The input M is bound to appear not only on the debit side but also in the production function, and/or as a cost-reducing factor in the function for traffic operation cost per vehicle-hour,

and/or in the user cost function. On the production or operation side M would be a factor of production just like the others given in (1) on page 6. The achievements of at least some categories of the administrative staff should result in a higher output from any given fleet of vehicles. The contributions made by other categories of administrative staff are likely to show up as traffic operation cost *savings* or user cost *savings*. For example, an increase in the staff in garages and repair workshops could lead to a lengthening of the economic life of the vehicles; and a strengthening of the OR-department could lead to improvements in the routing and scheduling of the vehicles, with the result that the feeder transport and waiting time costs for users are reduced. The most typical relationship is, in my view, that the traffic operation cost z is a function of M in the same way as of other factors of production and design variables.

A principal "staff function" is certainly to serve those in the operating line in various ways. In the absence of these services, line operators would perform their duties less efficiently.

Incorporating the empirical finding that the overhead costs tend to be proportional to the fleet size, the simplest way of writing the operation costs is the following:

$$\mathcal{N} = z\left(\frac{\mathcal{N}}{M}\right)\mathcal{N}. \tag{A.3}$$

The higher the ratio of fleet size to salaried staff, the higher the traffic operation cost per vehicle-hour will be, and vice versa. It is easily shown that when neither $\mathcal{N}$ nor M appears as a separate argument in the function for z, but only the ratio of $\mathcal{N}$ to M, it is optimal to expand the overheads proportionately to the number of vehicles. The total producer costs are written:

$$TC^{prod} = sM + z\left(\frac{\mathcal{N}}{M}\right)\mathcal{N}. \tag{A.4}$$

It is further assumed that M does not appear as an argument either in the production function or in the user cost function, but only in the z-function. Designating the ratio $\mathcal{N}/M = k$, the efficiency condition for the input of M then takes this simple shape:

$$s - k^2 \frac{\partial z}{\partial k} = 0. \tag{A.5}$$

This implies that M should be proportional to $\mathcal{N}$, or that the efficient ratio of the salaried staff to the fleet size is constant and equal to k^*. Total differentiations of TC^{prod} gives:

$$dTC^{prod} = \left(s - k^2 \frac{\partial z}{\partial k}\right)\mathrm{d}M + \left(z + k \frac{\partial z}{\partial k}\right)\mathrm{d}\mathcal{N}. \tag{A.6}$$

The incremental producer cost per additional vehicle-hour equals the ratio of dTC^{prod} to $d\mathcal{N}$.

$$IC^{prod} = \frac{dTC^{prod}}{d\mathcal{N}} = \frac{\mathrm{d}M}{\mathrm{d}\mathcal{N}} \quad s - k^2 \frac{\partial z}{\partial k} \quad + z + k \frac{\partial z}{\partial k}. \tag{A.7}$$

In the medium run $dM = 0$ by assumption, and only the last two terms are operative.

$$MRIC^{prod} = z(k) + k\frac{\partial z}{\partial k}. \tag{A.8}$$

On the assumption that $\partial z/\partial k$ is positive throughout, it follows that the average traffic operation cost per vehicle-hour z is always below the medium-run incremental cost. In the long run dM is different from zero, but under the efficiency condition the first term of (A.7) comes to zero all the same. $MRIC^{prod} = LRIC^{prod}$ for $k = k^*$.

It remains to be demonstrated that these incremental costs also equal the total producer cost per vehicle-hour, $\bar{z}$. Along the expansion path $dM/d\mathcal{N} = 1/k^*$. Multiplying both terms within the bracket of (A.7) by $1/k^*$, it is found that the second and fourth terms cancel, and an alternative formulation of the long-run incremental cost is:

$$LRIC^{prod} = \frac{s}{k^*} + z(k^*). \tag{A.9}$$

That this is equal to $\bar{z}$ is clear, since $\bar{z} = TC^{prod}/\mathcal{N}$. Dividing TC^{prod} according to (A.4) by $\mathcal{N}$ yields the same expression as that given for $LRIC^{prod}$ in (A.9).

The key point of the argument can also be illustrated diagrammatically in the following way:

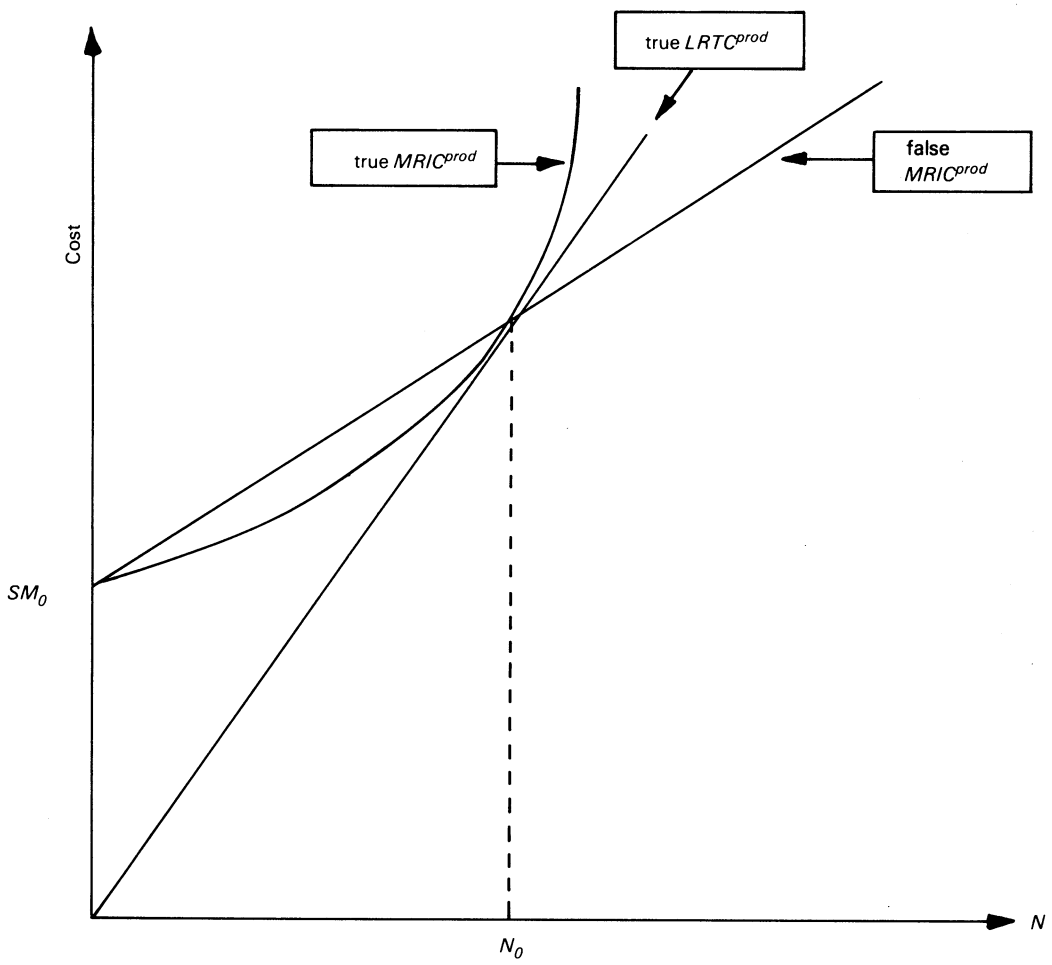
Suppose that the long-run total producer cost, including overheads, is proportional to the fleet size, as illustrated in Figure A:1. The medium-run total producer cost which is drawn on the assumption that a particular level of the overhead costs, sM_0 in the diagram, is fixed, takes the shape of the "true $MRTC^{prod}$" in Figure A:1. The linearisation of all cost and output relationships in cost accounting will lead to a differently shaped curve, or that given as the "false $MRTC^{prod}$" in Figure A:1. The starting-point at the vertical axis and the level at $\mathcal{N}_0$ are the same for the false and the true $MRTC^{prod}$. The main difference will become more and more pronounced the further away one looks to the right of $\mathcal{N}_0$. It cannot be true, as a general rule, that $MRTC^{prod}$ is below $LRTC^{prod}$ beyond $\mathcal{N}_0$. $MRTC^{prod}$ has to be above $LRTC^{prod}$ in its whole range except for a fleet size equal to $\mathcal{N}_0$, for which the medium-run fixed overhead costs sM_0 are optimal. It must be a disadvantage that the overheads cannot be varied as the output varies, and this disadvantage will manifest itself in increases in the costs which are variable in the medium run. To assume a linear shape of $MRTC^{prod}$ even beyond $\mathcal{N}_0$ is an (unconsciously) rather disparaging treatment of the resources representing the overhead costs. Surely, an undersized administration, too small repair facilities, etc., will cause increases in traffic operation cost which could be avoided by adequately matching the overheads to the traffic volume.

This point about the curvature of $MRTC^{prod}$ may seem rather subtle in a total cost context. It will, however, show up as quite important in a marginal cost context. In the diagram of Figure A:2 the systematic downward bias of marginal cost calculations, which is inherent in direct costing, is illustrated by portraying the slopes of the three total cost curves of Figure A:1. The conclusion of the discussion of this appendix can be summarised with reference to the chart in Figure A:2.

Of the three curves given there, $LRIC^{prod}$ is the only one which is empirically founded. An estimation of $MRIC^{prod}$ by statistical methods may prove to be an over-

FIGURE A1

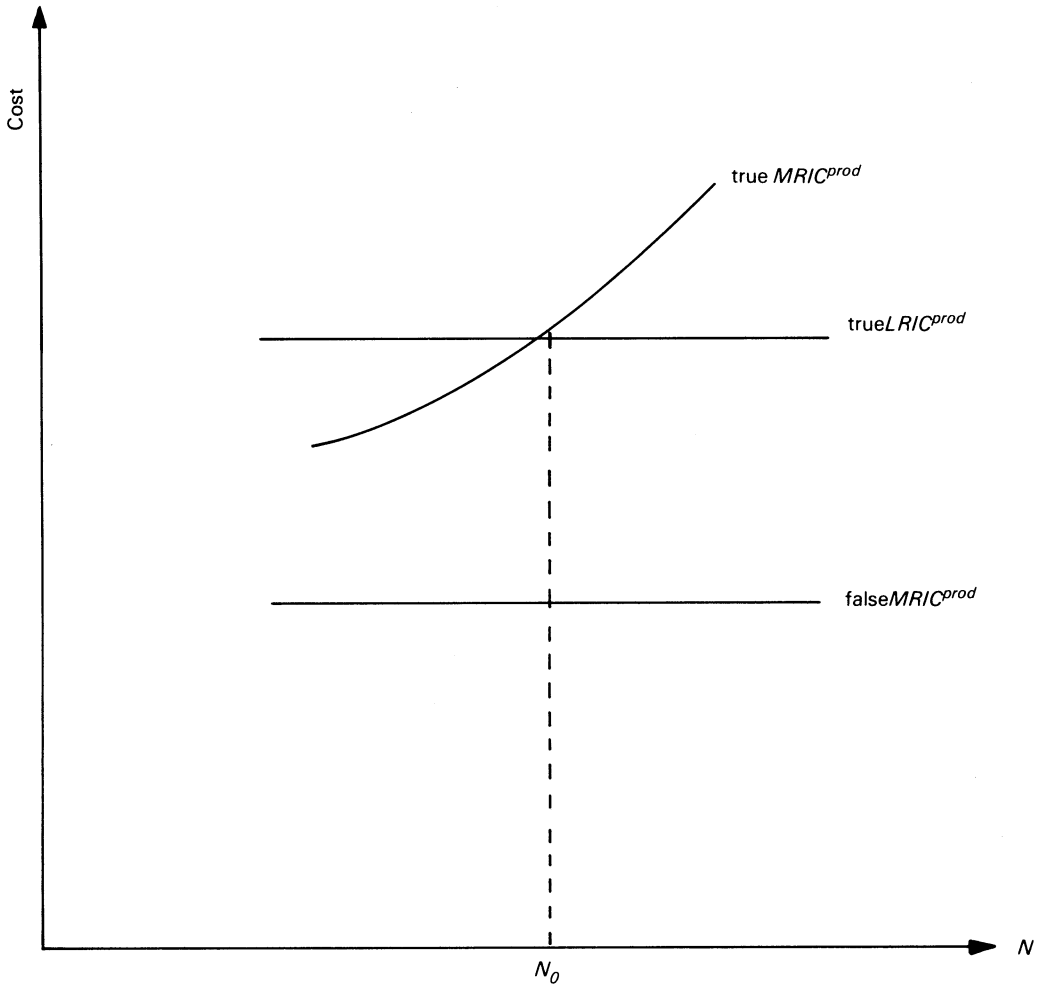
A diagrammatic representation of the "disparaging treatment" of the overheads inherent in direct costing



whelming task. In the absence of solid empirical evidence, theoretical reasoning has to be resorted to. The point made here is that the disparaging treatment of overheads inherent in direct costing appears to lead to a systematic underrating of $MRIC^{prod}$. It seems preferable to assume that even the input of overhead factors can be chosen according to normal economic criteria, and thus that $MRIC^{prod} = LRIC^{prod}$ under the standard efficiency conditions.

FIGURE A2

The result for the marginal cost calculation of the "disparaging treatment" of the overheads



REFERENCES

Baumol, W. J., and Vinod, H. D.: "An Inventory Theoretical Model of Freight Transport Demand." *Management Science*, March 1970.

Borts, G.: "The Estimation of Rail Cost Functions." *Econometrica*, 1960.

Devanney, J. W., III, *et al.*: *Conference Rate-making and the West Coast of South America.* Commodity Transportation and Economic Development Laboratory, Massachusetts Institute of Technology, 1972.

Ferguson, A. R., *et al.*: *The Economic Value of the United States Merchant Marine.* The Transportation Center, Northwestern University, 1961.

Foster, C. D., and Beesley, M. E.: "Estimating the Social Benefits of Constructing an Underground Railway in London." *Journal of the Royal Statistical Society*, Series A, 126, 1963.

Griliches, Zvi: "Cost Allocation in Railway Regulation." *The Bell Journal of Economics and Management Science*, Spring 1972.

Koshal, R. K.: "Economies of Scale in Bus Transport: Some Indian Experience." *Journal of Transport Economics and Policy*, January 1970.

Koshal, R. K.: "Economies of scale in Bus Transport: Some United States Experience." *Journal of Transport Economics and Policy*, January 1972.

Lee, N., and I. Steedman: "Economies of Scale in Bus Transport: Some British Municipal Results." *Journal of Transport Economics and Policy*, January 1970.

Meyer, J., *et al.*: *The Economics of Competition in the Transportation Industries*. Harvard University Press, 1959.

Mohring, H.: "Optimization and Scale Economics in Urban Bus Transportation." *American Economic Review*, September 1972.

Turvey, R., and H. Mohring: "Optimal Bus Fares." *Journal of Transport Economics and Policy*, September 1975.

Walters, A. A.: *The Economics of Road User Charges*. World Bank Staff Occasional Papers, No. 5, 1968.