OPTIMAL PUBLIC TRANSPORT PRICES, SERVICE FREQUENCY AND TRANSPORT UNIT SIZE

Kjell JANSSON
Ph. D. , Economics
Department of Economics
Stockholm University
Stockholm-Sweden

INTRODUCTION

A number of works on joint determination of price and output of public transport have been presented over the last decades. A common conclusion is that optimal price falls short of average variable operator cost. Mohring [1972], Turvey and Mohring [1975], J. O. Jansson [1984] deal with price and service frequency, using models which are most relevant for <u>frequent urban services</u> and assuming one passenger group. Nash [1978] optimizes price and output in terms of miles operated for <u>frequent urban bus services</u>, contrasting maximum profit and maximum welfare solutions and assuming demand in terms of passenger miles to be dependent on price and bus miles operated. Panzar [1979] analyzes <u>infrequent airline services</u>, assuming demand to be dependent on price and service frequency and allowing for a distribution of ideal departure times. These works consider demand from all passengers, or from one representative group travelling the average distance, with no concern for where passengers board and alight. This paper takes into account <u>a variety of passenger groups</u> - -where a group is defined as those travelling between one specific Origin-Destination (O-D) stop pair - - and in principle all public transport modes. It is shown (a) that optimal prices vary with distance at a higher rate for long-distance than for urban public transport, and (b) that optimal financial deficit per passenger is typically smaller for long-distance than for urban public transport.

Here we depart from the obvious fact that consumer demand for public transport typically relates to specific O-D stop pairs. Supply, on the other hand, relates to service volume, which typically meets demand for a number of O-D pairs on a route. Supply can be specified in terms of number of transport units (vehicles or trains) per hour as is done here. This approach differs from that usually taken in the literature. For example, the approach in Nash [1978] assumes that "the producer produces output in terms of passenger miles while the consumer consumes passenger miles."; Here we maximize consumers' plus producer's surplus with respect to the output variables of service frequency and unit size, and with respect to price for each of the passenger groups, i.e., price in each O-D pair.

The decision problems of the Public Transport Authority and its passengers are defined in section 1. Section 2 presents the objective function and optimum for the policy variables and includes interpretations of the results for the unconstrained and constrained optima. The main conclusions are stated in section 3.

The paper is a modified short version of chapter 2 in a work by Jansson K. [1991]. This work is also referred to for readers who are interested in the mathematical derivations behind the analysis and results presented here.

1. THE PUBLIC TRANSPORT AUTHORITY AND ITS PASSENGERS

1.1 The Public Transport Authority.

The PTA is assumed to deal separately with each route and each time period in a specific network, with given residential areas and activity centres and given distances between stops. Demand is thus taken to be specified for certain periods, such as the average weekday afternoon peak hour in wintertime, the average Saturday etc. Only one type of charge -- a per-trip price -- is considered.

The policy variables available to the PTA are assumed to be (a) prices, (b) transport unit size, where a unit is a bus, a plane, a train etc. and (c) service frequency, which is the number of units that depart per hour. The PTA reaches decisions about relevant inputs and prices well ahead of implementation because of a necessary planning lag. All factors of production that are variable between decision and implementation are therefore considered relevant for the joint decision on the magnitude of policy variables.

The route is a γ kms long round trip. The round-trip time of the service is $h=b\sum X_j/F+\gamma r^{\gamma}$ hours, where b is fixed boarding time per passenger, X_j is the number of passengers belonging to group j;(1,....,i,j,....,J), F is the number of departures per hour (frequency), $\sum X_j/F$ is the number of passengers per departure belonging to group j and r^{γ} is the remaining (passenger independent) run time per kilometre, assumed to be constant. A group is thus denoted with subindex i or j, but for the sake of simplicity indices are often omitted, e.g., in sum expressions. Arguments of functions are throughout delimited by [], while polynoms are delimited by ().The number of units (carriages for railway mode and vehicles for other modes) needed in each period is $F(b\sum X_j/F+\gamma r^{\gamma})$. If C[] denotes the cost per departure, total operating costs per hour are:

$$(1) \quad FC[\sum X_j, F, N,\sigma] \equiv F((c_l + c_c[N\sigma])(b\sum X_j/F+\gamma r^{\gamma}) + c_s + c_t + \gamma c^{\gamma}[N\sigma])$$

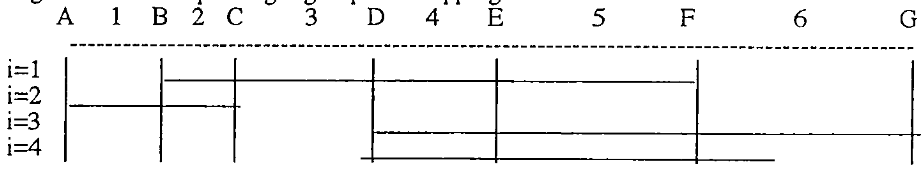
where N is the number of carriages used in a train (N=1 for all modes except train) and σ is the number of seats per vehicle or carriage.

In (1), c_l is the capital and personnel costs related to the transport unit (but not related to the unit size), i.e., costs for the locomotive, the driver, the captain etc. $c_c[N\sigma]$ are the costs directly related to the transport unit size, i.e., certain personnel costs (e.g. conductors in trains) and capital costs for vehicles and carriages. $c^{\gamma}[N\sigma]$ is the distance (kilometre) cost, which increases with the number of carriages due to track maintenance, a variable energy consumption (in order to maintain a constant speed, r^{γ}) etc. c_t is a fixed terminal cost per departure, related to the terminals (start and end points of the route), but not related to number of passengers, passenger groups, unit size, run time or distance. c_s is a stop cost related to terminals as well as intermediate stops, and thus to the number of passenger groups, but not to the number of passengers. For example, the terminal cost might be most relevant for rail transport, while the stop cost might be most relevant for water and air transport. For bus transport, terminal and stop costs may often be negligible.

1.2 The Passengers

A passenger group, supposed to meet a specific price, is defined as those travelling between one specific Origin-Destination (O-D) stop pair (see figure 1 below). Each passenger in a group i affects other passengers in the same group as well as passengers belonging to the other groups using common links, both in terms of crowding (in-vehicle congestion) and prolonged riding time.

Figure 1. Several passenger groups overlapping



In figure 1 the route starts at point A and terminates at point G, using the links 1 to 6, calling at the intermediate stops B, C, D, E and F.

Aggregate consumers' surplus is expressed as a function of "generalized cost", $G=p+\phi\varphi$, where φ is the vector of the travel time components and ϕ is the vector of monetary time values, i.e., the marginal rates of substitution between price and travel time components. Although ϕ is assumed to be the same for all individuals within a group, i.e., the same for all at each point $[p,\varphi]$, ϕ may be a function of φ, and vary among passenger groups. For the sake of simplicity, however, ϕ is here often written without index for group. The vector φ is here assumed to comprise riding time and "frequency delay", which is the time interval between ideal and actual departure time.

Riding time (including boarding) for group i is written

$r_i=\Sigma_{mi}h_m \equiv \Sigma_{mi}(b\frac{\Sigma X_j^m}{F}+\gamma_m r_m)$, where h_m is the total riding time on link m, r_m is the passenger-independent riding time on link m, γ_m is the link length, and Σ_{mi} is the sum over those links m where group i travels. X_j^m is the number of passengers in group j boarding on link m. The cost of riding time for group i is then:

$$(2) \quad T_i \equiv \Sigma_{mi}T_{mi} \equiv \Sigma_{mi}\phi[\frac{\Sigma X_{j/m}}{FN\sigma}](b\frac{\Sigma X_j^m}{F}+\gamma_m r_m)$$

T_{mi} is the riding-time cost of group i on link m, ϕ is the value of riding time, assumed to be dependent on the occupancy rate, $R_m \equiv \Sigma X_{j/m}/FN\sigma$, which is the number of passengers per seat in group j travelling on link m, and where we assume that $\partial\phi/\partial R_m>0$ and $\partial^2\phi/\partial R_m^2>0$. The number of passengers per departure on link m is here sometimes also written as $\Sigma q_{j/m}/N\sigma \equiv \Sigma X_{j/m}/FN\sigma$, and the number of passengers per departure as $\Sigma q_j/N\sigma \equiv \Sigma X_j/FN\sigma$.

The interval between departures is 1/F hours. Ideal departure times, t, are uniformly distributed within this interval, i.e., $0<t\leq 1/F$. Frequency delay is then $\tau \equiv 1/F - t$. The cost of frequency delay for a group i is $T^{\tau}_{i}[\phi^{\tau}_{i}, F, t] \equiv \phi^{\tau}_{i}[1/F - t](1/F - t)$, i.e., the delay multiplied by value of time, which can be a function of the delay. The cost of frequency delay has been shown to differ between the situations in which passengers use and do not use timetables respectively, and the impact of this will be discussed later on.

If **p** denotes the price vector for all groups, the generalized cost of travel for group i at time t is:

$$(3)\quad G_i[\mathbf{p},F,N,\sigma,t] \equiv p_i + \phi_i\varphi_i \equiv p_i + T_i + T^{\tau}_{i}$$

Demand at time t, x_i, and demand per hour, X_i, for each group i is a function of generalized cost:

$$(4)\quad X_i[\mathbf{p}, F, N,\sigma]] = F\int_0^{1/F}(\partial s_i/\partial G_i)dt = F\int_0^{1/F} x_i[G_i[\mathbf{p}, F,t]]dt$$

where s_i is the consumers' surplus of group i at time t. Note that demand in a certain group i is affected by the number of passengers in group i and other groups, because both riding time and crowding (in-vehicle congestion) are affected by passengers. We know that own-price elasticities, denoted ε_p, are negative, $\varepsilon_p<0$. We assume that demand elasticities with respect to frequency, vehicle size and train size, denoted ε_F, ε_N and ε_σ, are such that $0<\varepsilon_F<1$, $0<\varepsilon_N<1$, $0<\varepsilon_\sigma<1$. This rules out the possibility that an increase in frequency or unit size would generate so many passengers that occupancy rate is unchanged or increases, and these assumptions seem plausible for virtually all situations.

Aggregate consumers' surplus for group i is:

$$(5)\quad S_i[G_i] \equiv F\int_0^{1/F} s_i[G_i[t]]dt \equiv F\int_0^{1/F}\int_{G_i}^{G_{imax}} x_i[\rho]d\rho dt$$

where G_{imax} is the reservation price in generalized cost terms for the individual in group i with the maximum reservation price in generalized cost terms.

2. OPTIMUM FOR PRICES, FREQUENCY, VEHICLE SIZE, TRAIN SIZE AND DEFICIT

2.1 Welfare Maximization and Optimum

The maximization relates to one service during a period normalized to one (hour). The analysis may then be repeated for other periods and route.

We form a Lagrangian, L, composed of consumers' plus producer's surplus and the budget constraint $\mu(\Sigma p_i X_i[p_i, F, N,\sigma] - FC[\Sigma q_i[p_i, F, N,\sigma]] - \Pi)$, where μ is the multiplier and Π minimum profit:

$$(6) \quad L = \Sigma S_i[G_i[\mathbf{p}, F, N, \sigma]] + \Sigma p_i X_i[p_i, F, N,\sigma] - FC[\Sigma X_i[p_i, F, N, \sigma]] +$$

$$+ \mu(\Sigma p_i X_i[p_i, F, N, \sigma] - FC[\Sigma X_i[p_i, F, N, \sigma]] - \Pi)$$

First-order conditions with respect to prices, frequency, vehicle size and train size are assumed to represent an optimal solution. For the sake of simplicity we use an asterisk (*) for optimal values only when required to facilitate reading.

The following themes will be discussed under the assumption (made for the sake of simplicity and in order to focus on principal aspects) that the budget constraint is not binding: (i) optimum for financial deficit due to a positive external effect, (ii) optimal prices for various groups on a route and (iii) optimum for various modes and output variables. Finally we point to some implications of a binding budget constraint.

If the budget constraint is not binding (μ=0) the first-order conditions with respect to prices, frequency, vehicle size and train size are :

Prices:

$$(7) \quad p_i^* = bc[N\sigma] + \Sigma X_{j/\underline{mi}}\phi[R_m]\frac{b}{F} + \Sigma_{mi}\Sigma X_{j/mi}\frac{\partial\phi[R_m]}{\partial R_m}\frac{b\frac{\Sigma X_j^m}{F} + \gamma_m r_m}{FN\sigma} \equiv K+L+M \; ; \quad i=1..J$$

or, when summed over all passenger groups:

$$(7') \quad \frac{\Sigma p_i X_i}{\Sigma X_i} = bc[N\sigma] + \frac{\Sigma_i X_i \Sigma X_{j/\underline{mi}}\phi[R_m]\frac{b}{F}}{\Sigma X_i} + \frac{\Sigma_i X_i \Sigma_{mi}\Sigma X_{j/mi}\frac{\partial\phi[R_m]}{\partial R_m}\frac{h_m}{FN\sigma}}{\Sigma X_i};$$

The sum $\Sigma X_{j/\underline{mi}}$ denotes the number of passengers in groups j, who either board at (the start of) link m or pass link m, and who therefore get delayed by passengers in group i who board at link m. The underlining of $\underline{mi}$ denotes that group i passengers board at link m. The sum $\Sigma X_{j/mi}$ denotes the number of passengers in groups j, who travel on the link m where passengers in group i travel.

Frequency:

$$(8) \quad \frac{\Sigma p_i X_i}{\Sigma X_i} = \frac{FC}{\Sigma X_i} - \frac{F\Sigma y_i}{\Sigma X_i} \equiv bc[N\sigma] + \frac{F(\gamma r^{\gamma} c[N\sigma] + c_s + c_t + \gamma c^{\gamma}[N\sigma])}{\Sigma X_i} - \frac{F\Sigma y_i}{\Sigma X_i}$$

$$(8') \quad F^* = \frac{\Sigma_i X_i X_{j/\underline{mi}} \phi[R_m] \frac{b}{F} + \Sigma_i X_i \Sigma_{mi} \Sigma X_{j/mi} \frac{\partial \phi[R_m]}{\partial R_m} \frac{h_m}{FN\sigma}}{\gamma r^{\gamma} c[N\sigma] + c_s + c_t + \gamma c^{\gamma}[N\sigma] - \Sigma y_i}$$

Vehicle size:

$$(9) \quad \frac{\Sigma p_i X_i}{\Sigma X_i} = bc[N\sigma] + \frac{\Sigma_i X_i X_{j/\underline{mi}} \phi[R_m] \frac{b}{F}}{\Sigma_i X_i} + \frac{\sigma F(\frac{\partial c}{\partial \sigma}(b \Sigma X_j / F + \gamma r^{\gamma}) + \gamma \frac{\partial c^{\gamma}}{\partial \sigma})}{\Sigma X_i}$$

$$(9') \quad \sigma^* = \frac{\Sigma_i X_i \Sigma_{mi} \Sigma_j X_{j/mi} \frac{\partial \phi[R_m]}{\partial R_m} \frac{h_m}{FN\sigma}}{F(\frac{\partial c}{\partial \sigma} h + \gamma \frac{\partial c^{\gamma}}{\partial \sigma})}$$

Train size (noting that marginal boarding time, b, is assumed to be negligible and set to zero for this mode):

$$(10) \quad \frac{\Sigma p_i X_i}{\Sigma X_i} = \frac{NF(\frac{\partial c}{\partial N} \gamma r^{\gamma} + \frac{\partial c^{\gamma}}{\partial N} \gamma)}{\Sigma X_i}$$

$$(10') \quad N^* = \frac{\Sigma_i X_i \Sigma_{mi} \Sigma X_{j/mi} \frac{\partial \phi[R_m]}{\partial R_m} \frac{h_m}{FN\sigma}}{F(\frac{\partial c}{\partial N} \gamma r^{\gamma} + \frac{\partial c^{\gamma}}{\partial N} \gamma)}$$

According to (7), the optimal price equals the increase in social cost due to an additional passenger, i.e. the marginal cost of production plus the negative external cost borne by fellow passengers. The first term on the right hand side, *K*, is the producer's marginal cost, proportional to the boarding time per passenger. The second term, *L*, reflects the marginal boarding-time cost borne by passengers, i.e. the cost of prolonged riding time due to boarding. The third term, *M*, reflects the marginal crowding (in-vehicle congestion) cost, i.e., the cost of increased occupancy rate caused by an additional passenger. (8), (9) and (10) say that the average of optimal prices equals: the increase in operators' cost minus the passenger benefit, per passenger, due to a one unit increase in frequency according to (8) ; according to (9) it equals the increase in operating cost and riding time cost of an additional passenger plus the cost per passenger of one more seat; according to (10) it is the operating cost per passenger of one more carriage.

2.2 Optimum for financial deficit due to a positive external effect

The purpose of this section is to combine the role of the distinction made in Jansson [1991], between situations where passengers use time tables (case I) and situations where they do not (case II), with empirical findings concerning frequency delay, in order to indicate the magnitudes of the deficits for urban, regional and inter-regional transport.

For the situation in which the value of frequency delay is constant, i.e., independent of the delay, it has been shown (in Jansson [1991]) for the optimal deficit per passenger, Fy/X (index omitted), that $\max\{Fy^I/X\} = \phi^{\tau I}/2F^I$ (case I), while $Fy^{II}/X = \phi^{\tau II}/2F^{II}$ (case II). The question of whether the deficit per passenger is higher or lower when passengers use or do not use timetables is thus dependent on the value of frequency delay and optimal frequency in each case. Let us now, assuming that the value of frequency delay is constant, calculate the deficit term using empirical findings.

For urban transport, a revealed-preference study in Sweden (Algers, Colliander and Widlert [1985]) shows that the value of frequency delay (as an average over persons and delays) is around four to five times higher for journeys where people do not use timetables than for journeys where they do. Let us assume for example that $\phi^{\tau II}/\phi^{\tau I} = 4$. Let us further assume that passengers use timetables if the interval is over 10 minutes ($F < 6$). The set of frequencies, Ω, for which the deficit per passenger is higher where passengers do not use timetables than where they do is then $\Omega = \{F^I, F^{II} : 0.25F^{II} < F^I < 6\}$; e.g., $\Omega = \{F^I = 3, F^{II} < 12\}$. The conclusion that no use of timetables corresponds to a higher deficit than use of timetables thus seems to apply to a great proportion of urban public transport. In monetary terms the same study indicates that the value of frequency delay for frequent service (waiting time) is about £10 (SEK100) per hour. If we assume a typical frequency in urban bus service to be 8 per hour, the optimal deficit would thus be $£\phi^{\tau II}/2F^{II} = £10/2 \cdot 8 = £0.67$. For operating costs in Stockholm this value corresponds to approximately 50 percent of variable costs for an inner-city bus journey taking 8 minutes and to 35 percent of variable costs for a suburban bus journey taking 15 minutes. In a stated preference (SP) study in England on commuter-rail journeys with 30 minutes headway (Fowkes and Preston [1989]), for which most passengers can be assumed to use timetables, the value of riding time was found to be £1.0/hour and the value of frequency delay £0.65/hour. The average distance was 15 kms and the average price £1.0. If price is taken to be a proxy for average variable cost, the deficit is $£\phi^{\tau I}/2F = £0.65/2 \cdot 2 = £0.16$, i.e., 16 percent of the average variable cost.

For inter-regional transport, a SP study in England (Fowkes and Wardman [1987]) of inter-city rail journeys indicates that the average value of frequency delay is £3.60/hour. The average journey time is 2.5 hours and the headway is 1 hour. If the estimated average price equal to £20 is taken to be a proxy for average variable cost, the optimal deficit would be $\phi^{\tau I}/2F = £3.60/2 \cdot 1 = £1.80$, corresponding to 9 percent of the average variable cost. In a SP study in Sweden of inter-city rail journeys (Widlert and Lindh [1989]), the value of riding time was found to be £5 and the value of frequency delay 30 percent of this, i.e., £1.50. The average distance travelled was about 300 kms and the average fare £15. If this fare is a proxy for average variable cost, the optimal deficit is $\phi^{\tau I}/2F = £1.50/2 \cdot 1 = £0.75$, i.e., 5 percent of average variable cost.

Although studies of this kind are few, they seem to be consistent enough to indicate the preliminary conclusion that optimal deficit in relation to average variable cost is in the range 5-10 percent for long-distance rail services, somewhat higher for urban commuter-train services and substantially higher, 35-50 percent, for urban bus services.

2.3 Optimal prices for various groups on a route

In analyzing the variation in optimal prices along a route, we are helped by the first-order condition with respect to prices (7), since first-order conditions with respect to the other variables express only the average of optimal prices.

Observing that the optimal price is comprised of three parts, one of which is related to riding time and distance, this leads to the following conclusions: (i) the first part, the producer's marginal cost, *K*, is independent of distance travelled but is higher during the peak period, due to the fact that vehicle-related costs enter the cost parameter, c; (ii) the second part, *L*, reflects boarding cost and increases with the load at the boarding stop, but is unaffected by riding time; (iii) only the third part, *M*, which reflects crowding cost, is dependent on riding time, and is in fact proportional to riding time and number of passengers.

These conclusions may be illustrated with the help of a simple example: assume a route with consecutive stops a, b, c, d, links ab, bc, cd, with passengers groups travelling in O-D pairs ab, ac, ad, bc, bd, cd.

a ------------b----------------------------c--d

Assume that ab is the shortest link (measured in riding time), followed by bc and cd. If the load is constantly high all along the route from a to d, then all prices increase with riding time, that is, the cd price is highest, followed by bc and ab. The prices are, however, less than proportional to riding time, since the operator's cost and the boarding cost are independent of riding time. If the load is much higher for ab than for the other links, the congestion cost may dominate, i.e., the optimal fare for the short distance ab may be higher than for the longer distances bc and cd. On the other hand, if the load is high along bc and cd and low along ab, it may be that the optimal price for bc and cd is higher than for ab, in a manner that is more than proportional to riding time. Passengers travelling all along ad are, however, charged less per riding-time minute than ab and cd passengers.

Assume next that passenger group 1 travels between a and c and that passenger group 2 travels between b and c. It can then be shown (see K. Jansson [1991]) (i) that the optimal price for group 1 is likely to be higher than the optimal price for group 2 the more passengers there are in group 1 compared to group 2; (ii) that the price for group 2 may be higher than the price for group 1 even though both groups actually travel on the same section and group 1 in addition travels on another section.

2.4 Optimum for various modes and output variables

The optimum level of the output variables, frequency, vehicle size and train size, F^*, σ^* and N^*, expressed by (8'), (9') and (10'), yield the following relationships between parameter values and optimum for frequency, unit size and crowding: (a) the higher the unit value of frequency delay is, the higher optimal frequency is in relation to the optimal size and vice versa; (b) the smaller the marginal costs with respect to unit size are, the lower optimal frequency is in relation to the optimal size and vice versa; (c) the higher the unit value of frequency delay is and the lower the marginal costs with respect to unit size is, the lower the optimal level of crowding is. Typically ship, train and airline services are characterized by relatively low marginal costs with respect to unit size for a wide range of demand, thus implying relatively low optimal frequency and large size. Buses, on the other hand, will reach the legal limit for unit size at a fairly low level of demand, thus implying relatively small optimal size and high frequency.

We know from (8), (9) and (10) that the smaller the boarding cost is in relation to the crowding cost, the less likely it is that optimal prices per kilometre decrease with distance. Consequently, for modes that stand still at stops for a certain amount of time more or less independent of the number of passengers, e.g., underground, train, aeroplane, long-distance bus and ship, the boarding time is negligible, so that optimal price for each passenger group is the sum of the crowding cost on each link used, a cost that is directly proportional to riding time. It follows that where links on these modes are unevenly loaded, longer distances may have lower optimal prices than shorter distances, whilst where loads are the same, optimal prices are in fact directly proportional to time travelled. On the other hand, for urban buses, where boarding time is not negligible, optimal prices are less than proportional to distance travelled even if the passenger load is constant. Where urban bus services in off-peak are characterized by a low occupancy rate, the crowding cost part of optimal prices may be relatively low, implying that optimal prices per kilometre decrease with distance. Taking into account the fact that a differentiated fares system is more costly both for the passengers and the operator, the consequence is that a flat bus fare, independent of distance, has a much greater potential to be optimal in off-peak than in peak.

In the preceding section we concluded, on the basis of empirical data, that optimal fares for Swedish long-distance rail services cover nearly all variable costs. If the empirical findings relating to the value of frequency delay, our conclusion is that optimal train length should be greater and optimal frequency and price lower than they currently are in inter-city services in Sweden. The reason is that since Swedish Rail utilize much shorter trains (6-8 carriages) than the platforms have capacity for (14-15 carriages) on virtually all day-time departures, marginal costs with respect to train size are in fact fairly small.

2.5 Optimum with a binding budget constraint

The first-order condition with respect to price, (7), is now modified to:

$$(7') \quad p_i^* = bc[N\sigma] + \Sigma X_{j/\underline{mi}} \frac{\phi_j[R_m]\frac{b}{F}}{(1+\mu)} + \Sigma_{mi} \Sigma X_{j/mi} \frac{\frac{\partial \phi_j[R_m]}{\partial R_m} \frac{h_m}{\sigma F N}}{1+\mu} - \frac{\mu X_i}{(\partial X_i/\partial p_i)} \frac{1}{1+\mu}$$

The binding constraint means, according to (7'), that the optimal price is now the social marginal cost, composed of the operator's marginal cost and the marginal boarding and crowding costs borne by the passengers, plus a fourth term (in which $\partial X_i/\partial p_i < 0$). The marginal costs borne by the passengers are, however, discounted by $1/(1+\mu)$, which means that the PTA discounts the valuation of passenger costs by the rate μ, i.e., by the social marginal cost (shadow price) of deficit reductions.

By introducing the following notations:

$$m_i \equiv bc[N\sigma]+\Sigma X_{j/\underline{mi}}\frac{\phi_j[R_m]\frac{b}{F}}{(1+\mu)}+\Sigma_{mi}\Sigma X_{j/mi}\frac{\frac{\partial\phi_j[R_m]}{\partial R_m}\frac{h_m}{F\sigma}}{1+\mu}$$ for marginal social cost,

ε_i for own price elasticity, ε_{iF} for frequency elasticity, $\varepsilon_{i\sigma}$ for vehicle size elasticity,

ε_{iN} for train size elasticity, we can the rewrite the optimal conditions as follows:

$$(14)\quad \frac{p_i^* - m_i}{p_i^*} = \frac{-\mu}{\varepsilon_i(1+\mu)}$$

$$(15)\quad \frac{\Sigma p_i X_i}{\Sigma X_i} = bc[N\sigma]+\frac{F(\gamma r^\gamma c[N\sigma]+c_s+c_t+\gamma c^\gamma[N\sigma])}{\Sigma X_i}-\frac{F\Sigma y_i\frac{1}{1+\mu}}{\Sigma X_i}+\frac{\Sigma X_i(m_i-p_i)(\varepsilon_{iF}-1)}{\Sigma X_i}$$

$$(16)\quad \frac{\Sigma p_i X_i}{\Sigma X_i} = bc[N\sigma]+\frac{\Sigma_i X_i \Sigma X_{j/\underline{mi}}\phi[R_m]\frac{b}{F}\frac{1}{1+\mu}}{\Sigma_i X_i}+\frac{\Sigma X_i(m_i-p_i)(\varepsilon_i\sigma-1)}{\Sigma X_i}+\frac{\sigma F(\frac{\partial c}{\partial\sigma}h+\gamma\frac{\partial c^\gamma}{\partial\sigma})}{\Sigma X_i}$$

$$(17)\quad \frac{\Sigma p_i X_i}{\Sigma X_i} = \frac{\Sigma X_i(m_i-p_i)(\varepsilon_{iN}-1)}{\Sigma X_i}+\frac{NF(\frac{\partial c}{\partial N}\gamma r^\gamma+\frac{\partial c^\gamma}{\partial N}\gamma)}{\Sigma X_i}$$

Expression (14) reflects the Ramsey pricing rule (see e.g. Baumol and Bradford [1970]), whereby the percentage deviation of prices from marginal costs should be inversely proportional to the elasticities of demand for the groups in question. The marginal cost is then understood as the marginal social cost, in which the passenger-cost element is discounted by the marginal value of deficit reductions. There is empirical evidence that peak journeys are less price elastic than off-peak journeys and, for certain distance intervals for local and regional transport, also that long journeys are less price elastic than short journeys.[1] Given that the budget constraint is binding, we conclude: (i) that there is an additional argument (besides capital costs) for peak prices to exceed off-peak prices; (ii) that there is an additional argument for prices to increase with distance for local and regional transport, which did not follow from the unconstrained solution.

Let us interpret the constrained optimum, bearing in mind that the budget constraint, $\Pi=\Sigma p_i X_i$ - FC, can be met either by price increases, in order to raise revenues, or by reductions in frequency and/or unit size, in order to reduce costs. The terms including the elasticities with respect to frequency, vehicle size and train size respectively, show that the smaller these elasticities are, the smaller the difference between prices and marginal costs can be while still fulfilling the constraint, implying that the constraint is met by reduced frequency or unit size or train size to a greater extent than by increased prices.

---

[1]See for example Transport and Road Research Laboratory [1980]: i) the price elasticity declines with the length of the journey, at least from short (0-10 kms) to medium long (15-20 kms) journeys, ii) the average peak elasticity is -0.11, while the average off-peak elasticity is -0.26.

The intuitive reason for this is that when passengers are fairly insensitive to frequency and unit size, the value of the loss of consumers' surplus is fairly small compared to the cost reductions obtained by fewer runs or smaller units.

3. CONCLUSIONS

This paper has analyzed the joint optimization of prices, frequency, vehicle size and train size, taking into account a variety of passenger groups in the presence of a budget constraint and a flat fare constraint.

We have seen that the unconstrained optimal prices do not fully cover variable operator costs, due to a positive external effect, and that coverage is typically less for peak than off-peak periods. However, some groups may be charged a price over total average cost and other groups a price below this cost.

It was found that passengers travelling a shorter distance may be charged more than passengers travelling a longer distance, and that this was so even if the short-distance passengers' journey constituted a part of the long-distance passengers' journey. The reason is that only the crowding cost depends on distance travelled.

The dependence of optimal prices on distance varies considerably between modes. At the one extreme, optimal fares per kilometre for local bus services decrease with distance travelled, while at the other extreme, optimal prices for non-stop trains, long-distance buses, boats or flights are directly proportional to distance.

In terms of optimal financial deficit per passenger, it was found that this varies considerably between situations and modes.The results indicate a deficit of around 30-50 percent for urban bus service, around 15-20 percent for local trains and around 5-10 percent for inter-city trains. Contrasting the results with the policy of the state-owned Swedish Rail indicates that optimal train length should be greater and optimal frequency and fares lower than they currently are in inter-city services.

The consideration of a binding budget constraint implies Ramsey-pricing. This also implies in this context that where frequency elasticity is fairly low compared to price elasticity, the constraint should be met by reductions in frequency or transport unit size to a greater extent than by increases in prices. For urban transport a binding budget constraint when a variable valuation of frequency delay is taken into account implies two additional arguments are introduced for peak prices to exceed off-peak prices.

To summarize, current policies such that fares are more or less proportional to distance travelled irrespective of mode are in no context efficient, but are less damaging for inter-city transport than for urban public transport. The efficient fare structure that follows from the analysis in this paper seems, however, simple in principle, since the only thing it requires is the setting of a specific price for each specific Origin-Destination (O-D) stop pair for each departure on each route. If optimal prices imply a large variation among O-D pairs and routes, it may be optimal to use a limited number of differentiations, but still maintain the basic optimal fares structure. For off-peak urban and regional transport optimal price differences between journeys may be so small that a flat fare may actually be optimal when the cost of differentiation is taken into account.

REFERENCES

Algers, S., and S. Widlert. Applicability and Stability of Logit Models in Sweden - some recent findings with policy implications.The 1985 International Conference on Travel Behaviour. Haag. 1985.

Algers, S., J. Colliander and S.Widlert. Logitmodellens användbarhet och generaliserbarhet. Byggforskningsrådet. Stockholm. 1985.

Baumol, W.J., and D.F. Bradford. Optimal Departures from Marginal Cost Pricing. American Economic Review vol. 60 (June). 1985. pp. 265-283.

Fowkes, T. and J. Preston. Novel Approaches to the Design of Surveys with Hypothetical Choice Data. Institute for Transport Studies, University of Leeds. Presented at the Third International Conference on Survey Methods in Transportation. Washington. 1990.

Fowkes, T. and M. Wardman. The Values of Overcrowding and Departure Time Variation for Inter City Rail Travellers, Technical Note 229, Institute for Transport Studies, The University of Leeds. December 1987.

Jansson, J. O.: Transport System Optimization and Pricing. John Wiley&Sons, New York. 1984.

Jansson, K. Efficient Prices and Quality in Public Transport. Department of Economics, Stockholm University. 1991.

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

Nash, C. A. Management Objectives, Fares and Service Levels in Bus Transport. Journal of Transport Economics and Policy, vol XII. 1978.

Panzar, J. C. Equilibrium and Welfare in Unregulated Airline Markets. The American Economic Review, May. 1979.

Transport and Road Research Laboratory. The demand for public transport - report of the international collaborative study of the factors affecting public transport patronage. Crowthorne, U.K. 1980.

Turvey, R. and H. Mohring. Optimal Bus Fares. Journal of Transport Economics and Policy, vol IX. 1975.

Widlert, S. and Lindh C. SJ-resenärernas kvalitetsvärdering (Quality Valuation of Swedish Rail Passengers), Department of Transport Planning, Stockholm Technical Institute. 1989.