OPTIMAL PUBLIC TRANSPORT PRICE
AND SERVICE FREQUENCY

By K. Jansson*

1. INTRODUCTION

Because a Public Transport Authority in practice makes decisions about both price and
frequency of service well ahead of their implementation, this paper analyses the joint
optimisation of price and frequency.

In the literature two situations have been modelled separately. One concerns high-
frequency urban transport, where travellers go to the nearest boarding point and wait
there. For this situation Mohring (1972), Turvey and Mohring (1975) and J. O. Jansson
(1979, 1984), derive optimal price and frequency given that demand is not dependent on
price and frequency. For the same situation Larsen (1983) and Else (1985) argue that
demand should be a function of both price and frequency, but they define this dependence
very restrictively. The second situation concerns low-frequency, long-distance transport,
where travellers use timetables and spend the time before departure not at a bus stop or
airport but at a more convenient place like home, office, and so on. For this situation,
Panzar (1979), dealing with air transport, determines both optimal price and frequency,
specifying demand as a function of these two variables.

A basic result that holds for both situations in general terms, is that the optimal price
equals the operator’s marginal cost plus the marginal external effects on fellow passen-
gers. The optimal price (at least without a budget or a capacity constraint) also equals the
average variable operator cost minus a term that constitutes a financial deficit, reflecting
a positive external effect.

Here both situations are treated within the same framework, leading to some new and
more general results, due to the observation of two effects of frequency on passenger
behaviour which do not seem to have been fully taken into account in the literature. The
first effect reflects the following dual behaviour: for low frequencies, and given the
availability of a reliable timetable, people prefer to plan their trips according to the table.

* Department of Economics, Stockholm University. I am indebted to Ann Friedlaender, Massachusetts
Institute of Technology; to colleagues at the Department of Economics at Stockholm University, especially
Harald Lang, Kari Lantto, Jan-Eric Nilsson, Roger Pyddoke, Ingemar Svensson, Hans Wijkander; to Lars
Lundqvist, Transport and Planning Committee, Stockholm County Council, and to anonymous referees, for
useful comments on earlier versions. I am particularly grateful to my adviser, Professor Peter Bohn,
Department of Economics, Stockholm University. This work has been financed to a considerable extent by
the Swedish Transport Research Board (TFB) and also by Stockholm Transport.
For high frequencies they prefer to go to the stop or station spontaneously, rather than consult the timetable. The second effect relates to the fact that the disutility of waiting at stops is higher than that of waiting at home or at work, and that the passengers' waiting cost may vary with the duration of the wait; typically the cost increases with the duration for high frequencies and decreases for low frequencies.

It is shown that these effects imply the following: (a) two local optima may arise for most urban public transport: one low-frequency/low-demand optimum and another high-frequency/high-demand optimum, where one of them constitutes the global optimum; (b) perhaps contrary to intuition, the "optimal deficit per passenger" should generally be larger for typical frequent urban services than for infrequent urban and inter-regional services; (c) the optimal off-peak price may exceed the optimal peak price.

Finally it might be interesting to note the fact that the distinction between the two behaviours with respect to frequency not only has implications for price and frequency principles, but also has a considerable potential effect on network design and investment criteria.

The decision problems of the Public Transport Authority and its passengers are introduced in Section 2. Section 3 presents the maximisation problem and the solutions for each situation, and Section 4 contrasts the results for the two situations. Section 5 briefly discusses a few extensions, and Section 6 presents the main conclusions. The paper presented here is a modified version of part 1 in K. Jansson (1991).

2. THE PUBLIC TRANSPORT AUTHORITY AND ITS PASSENGERS

In this section, we specify the aspects which the (welfare-maximising) Public Transport Authority (PTA) is assumed to consider when deciding on price and frequency. We also define the passengers' decision problem with respect to price, travel time and service frequency.

2.1. The Public Transport Authority

The Authority is assumed to deal with optimal price and service frequency of one service (not a network of routes), with given distances between stops and given residential areas and activity centres. This allows us to assume that demand is completely defined by price and frequency. Only first-best pricing rules are considered, without any concern for environmental or modal split (auto/public transport) issues, and only one type of charge - a per-trip price - is considered.

The PTA reaches decisions about price and output in terms of frequency (the two policy variables) well ahead of their implementation because of a necessary planning lag. The reason is either administrative/political, or that the Authority is required to inform people well in advance. Thus, in practice, there is an opportunity to decide on prices and frequency simultaneously. Since both prices and frequency affect demand, there are efficiency reasons for taking the interaction between prices and frequency into account.
All factors of production that vary with demand and frequency and are variable during the months between decision and implementation are relevant for the PTA’s joint decision on price and frequency.¹ Such factors give rise to what we call variable costs, including all costs that may be related to frequency. Fixed costs are thus costs for factors that cannot be changed within this time period. Changes in demand due to our price and frequency variations are assumed not to exhaust what is considered to be the overhead (fixed factor) capacity. To simplify the exposition we may thus, without loss of generality, assume fixed costs to be zero.

The PTA is assumed to know demand for the next decision period (let us say a year) as a function of price and frequency. Demand is taken to be specified for certain periods, such as the average weekday afternoon peak hour in wintertime, the average Saturday, and so on. Price and frequency are assumed to be determined separately for each period, since the use of factors of production and/or effects on passengers are specific for each period.

In order to focus on the principal aspects, we confine ourselves in this paper to bus transport only, and as in previous studies, to the case of one homogeneous passenger group. This would in practice correspond either to the case where one passenger group travels from terminal A to terminal B and another equally large group from B to A (A-B-A constitutes the route), or the case where one group is subsequently replaced by another equally large group travelling for an equally long time along the whole route in both directions.

Operating costs
The round-trip time of the service is $bX/F + \gamma r^T$, where $b$ is fixed boarding time per passenger (assumed to dominate over alighting time); $F$ is the number of departures per arbitrarily chosen period of time (frequency), which we may — in order to focus our ideas — think of as an hour; $X$ is the number of passengers per hour; $\gamma$ is the round-trip distance in kms; and $r^T$ is the remaining (passenger independent) run time (including layover), assumed constant, per kilometre.² The number of vehicles needed is $F(bX/F + \gamma r^T)$. In the following, arguments of functions are delimited by [ ], while polynomials are delimited by ( ). If $c$ denotes the hourly capital and personnel cost per vehicle, $c^R$ denotes the distance (kilometre) cost and $C[X, F]$ denotes the cost for one departure, the variable cost for $F$ departures is written:

---

¹ Frequency is thus treated analytically as a continuous variable. In practice, there may be several reasons for the actual frequency to deviate from the theoretically optimal one. One may be that the optimal frequency implies a non-integer number of vehicles, another that the operator wants an integer number of departures per hour to facilitate memorising of schedules. We assume that such reasons would not force the actual frequency to deviate too much from the ideal. Furthermore, such deviation is often reduced by scheduling and rostering schemes which reallocate vehicles and staff among routes in order to obtain efficient usage.

² We assume that the number of stops at which the number of boarding passengers may vary between zero and a positive figure is negligible. This means that $r^T$ is exogenous and that riding time is determined solely by the variation in the number of boardings $\geq 1$ at each stop.
2.2. The passengers

We specify our use of the consumers' surplus as a function of generalised cost and define our concept of time values, riding time and frequency delay, and make the crucial distinction between two different behavioural cases.

We assume there is only one homogeneous group of passengers, each consuming one journey. Thus, all individuals are sufficiently alike for the ratio between price and travel time components to be valued in an identical manner. Passengers differ however with respect to preferences for non-transport consumption, implying a continuous variation in reservation prices and a continuously variable demand. Assuming also that the utilities which the individual derives from the journey and other consumption is additive, the income effect of price and frequency variations is zero, and consumer's and consumers' surplus can be expressed as functions of the "inclusive price" or "generalised cost", \( G = p + \phi_p \), where \( \phi \) is the vector of travel time components and \( p \) is time values, that is, the vector of marginal rates of substitution between price and travel time components (see for example Domencich and McFadden 1975 for further details on the link between individual preferences and generalised cost). Aggregate demand, \( x \), at a point in time, \( t \), equals \( x[p + \phi_p(t)] \). Although \( \phi \) is assumed to be the same for all individuals, that is, the same for all at each point \( [p, \varphi] \), \( \phi \) may be a function of \( \varphi \). The vector \( \varphi \) is here assumed to comprise riding time and frequency delay, allowing for the fact that both components may be valued differently depending on the circumstances under which they are spent.

Riding Time

Riding time (including boarding time) is \( h = bX/F + \gamma r \). The monetary value of riding time per hour is assumed to be dependent on the occupancy rate, \( R = X/F \sigma \), where \( \sigma \) is the number of seats. It is assumed that \( \phi/\partial R > 0 \). The cost of riding time is then:

\[
T = \phi[X/F\sigma](bX/F + \gamma r)
\]  

Frequency delay

First we note the importance of the concept of an ideal departure time. For example, one may want to go home as soon as the theatre ends. People may, on leaving the airport, want the city bus to depart right away. People may prefer to have access to a public transport facility to a certain destination on Sundays, not just on weekdays. Frequency delay is defined as the difference between ideal and actual departure time. The smaller the frequency delay, the more inclined the passenger is to choose this service instead of other modes of transport or other activities. We assume that the periods for which price and frequency are determined are defined in such a way that individuals' ideal departure times are uniformly distributed over this period. Since we have normalised the period to 1 hour, the interval between departures (all intervals assumed identical) is \( 1/F \) hours.
OPTIMAL PUBLIC TRANSPORT PRICE AND SERVICE FREQUENCY

K. Jansson

Use of timetables (Case I) or no use of timetables (Case II)
We will now distinguish between two cases relating to passenger behaviour, I and II, specify the generalised cost and analyse the endogenous choice of case. In case I there is a reliable timetable and passengers plan their trips according to this table. In case II frequency is so high that the passengers prefer to go to the stop randomly rather than consult the timetable. It may be useful, when memorising what I and II stand for, to remember that the low number stands for low frequency and the higher number for high frequency.

Use of timetable (case I) is valid in two different situations. The first is when the passenger must choose the first departure after ideal departure time. This situation arises when the passenger's ideal departure time is based on the fact that he wants to go immediately after breakfast, when the meeting is finished, "right now", and so on. The frequency delay is then defined as $\tau$ equals $1/F - t \geq 0$, where $t$ is measured from the preceding departure. The second situation arises when the passenger must choose the last departure that precedes the ideal one. This occurs when the passenger has to be at his destination at a certain time, when work or the theatre, or whatever, begins. Assume a person wants to be at work at 9.00 a.m. and the bus ride takes 30 minutes. His ideal departure time is then 8.30, but he has to take the departure before 8.30. In this situation, the frequency delay is again $\tau = 1/F - t$, but $t$ is measured from the following departure. There is thus no analytical difference between the two situations. Since we assume that ideal departure times are uniformly distributed over the interval $1/F$, one may arbitrarily measure $t$ from the preceding or the following departure.3

Where timetables are not used (case II), the passenger has to choose the first departure after ideal departure time, which is when he arrives at the stop. This means that, irrespective of whether passengers use timetables or not, frequency delay can be expressed as $\tau = 1/F - t$.

The value of time (per time unit) for the frequency delay, $\phi_\tau$, is assumed to vary with the delay itself, that is, $\phi_\tau(\tau)$. Aggregate demand $x$ at $t$ is a function $x[1/F - t]$. We assume here that gains and losses from changes in exact departure times are valued the same irrespective of to whom they may accrue. The cost of the frequency delay is $T_\tau[F, t] = \phi_\tau[F, t]x[F, t] = \phi_\tau(1/F - t)(1/F - t)$. The generalised cost of travel for passengers at time $t$ is:

$$G[p, F, t] = p + \phi[X/F\sigma](bX/F + \gamma\nu) + \phi_\tau(1/F - t)(1/F - t)$$  (3)

Obviously, timetables are used for scheduled infrequent services such as rural bus transport, but are not used for frequent peak-demand city services. For most urban public transport, however, we cannot tell a priori whether it is worthwhile to use timetables or

3 Panzar assumes a third situation, where the passenger can choose the nearest departure, before or after ideal time. This situation requires the passenger to have an ideal departure time but no ideal arrival time and that he has time enough to plan whether to choose the preceding or the following departure. Frequency delay is then $\tau = 1/2F - t$, where $t$ is measured from the midpoint of the interval. In note 11 we comment on the implications of also incorporating the third situation.

37
not. Here the individual passenger is assumed to behave in accordance with case I or case II, on the basis of whichever minimises his expected time cost. The value of frequency delay spent as waiting time at the stop, \( \phi^{\text{II}} \), is typically higher than the value of frequency delay spent at home, at work, or wherever, \( \phi^{\text{I}} \). When the passenger considers case I, that is, spending the time at some more pleasant place than the stop, he is assumed to meet an information cost for looking at the timetable (\( \eta \)). In addition, he is assumed to arrive at the stop a couple of minutes before the announced departure time. This security margin (\( \kappa \)) is valued as waiting time at the stop. We assume that the passenger knows the frequency of the service. His decision whether to use the timetable or not must be taken before he can actually use it. For this decision, the cost of frequency delay is thus in both case I and case II an expected delay, independent on \( \tau \), that is, \( T^{\text{II}}[F] = E[\phi^{\text{II}}[\tau] \kappa] \) and \( T^{\text{I}}[F] = E[\phi^{\text{I}}[\tau] \kappa] \) respectively. The passenger may then be assumed to choose case I or case II according to:

\[
\min \{ I, II \} = \min \{ \eta + E[\phi^{\text{II}}[\tau] \kappa] + \phi^{\text{I}}[\kappa] \kappa, E[\phi^{\text{II}}[\tau] \kappa] \} \tag{4}
\]

Assuming that all individuals place the same value on all public transport attributes, there is a common limiting \( \bar{F} \), below which the analysis should follow case I and above

---

4 Empirical studies indicate that frequency delay spent at a stop is valued several times higher than if it is spent at home, work and so on. See Algers, Collander and Widlert (1985).
OPTIMAL PUBLIC TRANSPORT PRICE AND SERVICE FREQUENCY

K. Jansson

which the analysis should follow case II. There may thus be one local optimum \( F^* \leq F \) and another \( F^u* > F \), where typically only one of them is a global optimum (see Figure 1).

Note that once the decision between case I and II is made, the cost of frequency delay in both cases is dependent on \( t \), that is, \( T^*[F, t] = \phi^*[1/F - t](1/F - t) \).

Note that the situation illustrated in Figure 1 may not be the only one. We may also find that an analysis assuming case I yields \( F^* > \bar{F} \), while at the same time an analysis assuming case II yields \( F^* < \bar{F} \). In this situation we have a corner solution at \( F^* = \bar{F} \), where optimum is determined by the one of the two analyses that yields the maximum welfare. The two constrained maxima are typically not equal, since demand, and subsequently riding-time cost and operating cost, may shift when moving between case I and II.

Demand

Where passengers use timetables, demand \( x \) at \( t \) varies within the interval \( 0 < t \leq 1/F \), implying that demand per hour, \( X \), is defined by the integral over the interval \( 1/F \). The three concepts of demand, per hour \( (X) \), per departure \( (q) \) and at time \( t \) \( (x) \) are thus related according to:

\[
X[p, F] = F q = F \int_0^{1/F} x[G[p, F, t]] dt
\]  

(5)

Where passengers do not use timetables they have no information about departure times, implying that demand \( x \) is constant within the interval \( 0 < t \leq 1/F \). This implies that demand per hour, \( X \), is independent of \( t \), that is:

\[
X[p, F] = F q = (1/F) x[G[p, F]]
\]  

(6)

Consumers' surplus

The reservation price in generalised cost terms for the individual with maximum reservation price is called \( G^{\text{max}} \), The consumers' surplus for passengers having the ideal departure time \( t \) is denoted \( s[G] \). The total consumers' surplus \( S[G] \) at actual \( G = p + \phi \phi \) is then:

\[
S[G] = F \int_0^{1/F} s[G(t)] dt = F \int_0^{1/F} G^{\text{max}} (x(\rho) d\rho) dt
\]  

(7)

3. DETERMINATION OF OPTIMAL PRICE, FREQUENCY AND DEFICIT

In this section welfare is maximised with respect to price and frequency for one period including \( F \) equally spaced departures. Only objective functions and results are presented. Mathematical derivations are found in the Appendix.
The objective function is the welfare, \( w \), composed of consumers’ plus producer’s surplus, : 

\[
    w = w[S[G(p, F)] + \int_0^{T_F} s[G(p, F, t)] dt + pX[p, F] - FC[X[p, F]] \quad (8)
\]

Meeting first-order conditions with respect to price and frequency is assumed to yield an optimal solution. For simplicity we use an asterisk (*) for optimal values only when it is required to facilitate reading. The first-order conditions with respect to \( p \) using (1) and (2) yields:

\[
    p^* = X \frac{\partial T}{\partial X} + F \frac{\partial C}{\partial X} = X\phi[R] \frac{b}{F} + X \frac{\partial \phi[R]}{\partial R} \frac{b}{\sigma F} + cb \quad (9)
\]

The section of (9) to the right of the equivalent sign demonstrates that the optimal price comprises three parts, one of which is related to riding time. The first part is the boarding cost, which grows with the passenger load at the boarding stop, but is unaffected by riding time. The second part is the crowding cost, which grows with the load through the value of time, that is, through \( \phi/\partial R \), and which is proportional to riding time.\(^5\) The third part is the producer’s marginal cost, which is proportional to boarding time and time cost, meaning that it is higher in peak than off-peak periods due to the fact that vehicle-related costs enter the cost function in peak periods.

The first-order condition with respect to frequency yields:

\[
    p^* = F \frac{C}{X} - F \frac{\gamma}{X} \quad (10)
\]

We thus arrive at a second relation between optimal price and frequency, where the optimal price is composed of two parts. In the first part, \( C \) is the operator’s cost of an additional departure, and in the second part \( \gamma \) is the passengers’ benefit of an additional departure. The optimal price is thus the marginal social cost with respect to frequency, per passenger. The first part also corresponds to the average variable cost, since our cost formulation assumes that there are no economies of scale related to the number of departures. The second part, \( F\gamma/X \), also represents the optimal (financial) deficit per passenger, where \( \gamma \) is the optimal total deficit.\(^6\) The total deficit corresponds to the direct effect (not via demand variations) on consumers’ surplus of a marginal increase in frequency (see Appendix). It may also be interpreted as a positive external effect, in the sense that more passengers induce a higher optimal frequency, thus benefiting intra-

\(^5\) If we had assumed a variety of passenger groups and a variable load along the route, optimal prices would vary in a more complicated manner.

\(^6\) We prefer to call the difference between price and average cost a (financial) deficit instead of a subsidy (which is sometimes used in the literature for the same thing), because the financial loss here is reconcilable with the fact that optimal price equals marginal social cost.
marginal passengers. Note that while the first-order condition with respect to price yields that the optimal price equals the producer's marginal cost with respect to passengers plus a negative external effect, the first-order condition with respect to frequency says that the optimal price equals the operator's marginal cost with respect to frequency minus a positive external effect.\(^7\) Note that this motive for a financial deficit exists independently of a second-best motive for a subsidy, that is, insufficient pricing of car travel, a matter analysed for example by Glaister (1974, 1978).

The important observation is that the deficit term differs between the two situations where passengers use and do not use timetables respectively, a fact which constitutes the basis for the main theme in this paper, that is, the contrasting of the two situations dealt with in the next section. For the situation where passengers use timetables (case I) and do not use them (case II), the deficit term \( y \) (see (A16) and (A17) in Appendix) is:

\[
y^I = \left[ \int_{0}^{\infty} \left( \int_{0}^{\infty} x[G(p, F, \psi)](\phi \psi + \phi \psi) \partial \psi \right) dt \right] \]

\[
y^II = \left[ \int_{0}^{\infty} \left( \int_{0}^{\infty} (\phi \psi + \phi \psi) \partial \psi \right) dt + \frac{x\phi}{2F^2} \right] \tag{11b}
\]

4. CONTRASTING THE CASES WHERE TIMETABLES ARE USED AND NOT USED

It is obvious from the first-order conditions with respect to frequency that case I and II optima may be interpreted and contrasted by analysis of the two specific deficit terms. These terms, in turn, depend on the value of frequency delay and on the variation of this value with respect to its duration. In considering the two cases, we first analyse optimal price and deficit for various assumptions concerning the value of frequency delay, and then discuss the implications for peak and off-peak pricing.

4.1. The value of time for the frequency delay is zero

It is obvious from the first-order condition with respect to frequency (see (10) and (11)) that the deficit term disappears (since \( \phi^I = \phi^{II} = 0 \)), which means that the optimal price is above the operator's marginal cost but equal to the average variable cost. Although a zero value of time for the frequency delay is not a realistic assumption, its implications are interesting. It shows that, within our framework, the sole reason for a price below average variable cost is that there is a positive value attached to reductions in the frequency delay. It also shows that the deficit is zero, in spite of the fact that the operator's cost is decreasing (even when fixed costs are ignored). From the cost expression (1) we see that the average cost per departure, \( FC/X \), equals \( cb + FcYdX + FcYdX \), while the

\(^7\) By combining the first-order conditions with respect to price (9) and frequency (10), and by denoting the subsidy term \( y \) for both cases, we obtain: \( Cq - C\partial q = \partial\partial q + y \). That is, the difference between average and marginal operator's cost equals the sum of negative and positive external effects.
marginal cost, $F(\partial C/\partial X)$, equals $cb < FC/X$. The reason for the zero deficit is that the optimal price is the operator's marginal cost plus marginal effects on passengers.⁸

4.2. The value of time for the frequency delay is constant

For this situation ($\phi^{\text{II}}/\partial \tau = \phi^{\text{II}}/\partial \tau = 0$), the deficit terms in (11) can be rewritten as:⁹

$$y_t = \phi^{\text{I}} \int_0^t \left( \int_0^s x[G(p,F,\psi)d\psi]dt \right)$$

$$y^{\text{II}}_t = \chi \phi^{\text{II}}/2F^2 = X \phi^{\text{II}}/2F^2 = q \phi^{\text{II}}/2F$$

(Note that where passengers do not use timetables and the value of waiting time is independent of its length, the deficit per passenger, $y^{\text{II}}/q = \phi^{\text{II}}/2F$, is simply equal to the expected cost of waiting time per passenger, where the expected waiting time is half the interval between departures.)

It has been shown (see K. Jansson (1991)) that in general Max $\{y^{\text{II}}/q\} = \phi^{\text{II}}/2F$, that is, $\phi^{\text{II}}/F^{\text{II}} \geq \phi^{\text{I}}/F^{\text{I}}$ for any demand function $x$ that is monotonically increasing in $t$. Thus, the larger the difference in values of frequency delay between the situation where passengers do and do not use timetables respectively, the larger the range of frequencies is for which the deficit per passenger is larger where timetables are not used than where they are. Empirical studies (see Algers, Colliander and Widlert (1985)) indicate that $\phi^{\text{II}}$ is on average 4 to 5 times larger than $\phi^{\text{I}}$. If the limiting frequency below which and above which passengers use timetables is assumed to be around 6 (10 minutes' headway), we conclude that for typical frequencies in urban public transport, the optimal deficit per passenger tends to be larger for frequent than for infrequent urban transport. The figure below illustrates the deficit per passenger as a function of the interval between departures, $1/F$, expressed in minutes. We assume that the value of frequency delay is a constant £10 per hour when timetables are not used (for intervals below 10 minutes) and £2.5 per hour when timetables are used (for intervals above 10 minutes). Under these conditions the figure shows that for intervals below 2.5 minutes, the deficit when timetables are not used is smaller than the minimum deficit when they are used, and that, for intervals above 40 minutes the deficit when timetables are used is larger than the maximum deficit when they are not used. In the more realistic case mentioned above, in which the value of frequency delay is zero also leads to the solution for the special case where virtually all passengers have the same ideal departure time (for example in rural areas where a service has the single objective of bringing all employees to a work place at 8.30 a.m.). $F$ now denotes frequency, in terms of number of departures at one point in time, but also the number of vehicles. Total demand, condensed to a point in time, $x$, corresponds to demand density in case II. Demand per departure, or vehicle, is $q = x(1/F)$. Omitting the expression for frequency delay (since its value is zero) in the objective function, the solution of case II provides optimal price, number of vehicles and deficit, where the latter is apparently equal to zero.

⁸ The assumption that the cost of frequency delay is zero also leads to the solution for the special case where virtually all passengers have the same ideal departure time (for example in rural areas where a service has the single objective of bringing all employees to a work place at 8.30 a.m.). $F$ now denotes frequency, in terms of number of departures at one point in time, but also the number of vehicles. Total demand, condensed to a point in time, $x$, corresponds to demand density in case II. Demand per departure, or vehicle, is $q = x(1/F)$. Omitting the expression for frequency delay (since its value is zero) in the objective function, the solution of case II provides optimal price, number of vehicles and deficit, where the latter is apparently equal to zero.

⁹ For this situation where the value of frequency delay is constant, (12b) corresponds to the result of Mohring (1972), Turvey and Mohring (1975) and J. O. Jansson (1979,1980), while observing the fact that (12a) corresponds to the result of Panzar. I am indebted to a referee.)

42
4.3. The value of time for frequency delay is a function of the delay itself
This is considered the most plausible situation, for which it is assumed that:

\[
\begin{align*}
\text{case I:} & \quad (a1) \quad T^{d1}/\partial \tau > 0 \quad (b1) \quad \phi^{d1}/\partial \tau < 0 \quad (c1) \quad 2T^{d1}/\partial \tau^2 < 0 \quad \text{(concavity)} \\
\text{case II:} & \quad (aII) \quad T^{dII}/\partial \tau > 0 \quad (bII) \quad \phi^{dII}/\partial \tau > 0 \quad (cII)\phi^2T^{dII}/\partial \tau^2 > 0 \quad \text{(convexity)}.
\end{align*}
\]

Conditions (a1) and (aII) mean that a long frequency delay is worse than a short frequency delay. Conditions (b1) and (c1) mean that the value of frequency delay decreases with its duration and that the effect on the cost of frequency delay is strictly concave in the delay. The reason is that for low frequencies people may more easily find alternative use for a long period of time than for a short period. For high frequencies, when people prefer to go to the stop spontaneously and wait there, conditions (bII) and (cII) say that a long wait is perceived as more uncomfortable per minute than a short wait, that is, that passengers find waiting worse the longer they have to wait.\(^\text{10}\)

\(^{10}\) That the cost of frequency delay is convex in the delay where timetables are not used and concave where they are is supported by the empirical study by Alger, Collander and Widlert (1985).
From (12a) and (12b) it is evident that the deficit term $y^{II}$ is larger if $\phi^{III}/\tau > 0$ than if $\phi^{III}/\tau = 0$, and that the deficit term $y^{I}$ is smaller if $\phi^{I}/\tau < 0$ than if $\phi^{III}/\tau = 0$. Thus, for the situation where the cost of frequency delay is convex in the delay where timetables are not used and concave where they are, we conclude that the range of frequencies for which the deficit per passenger is larger where timetables are not used than where they are, is wider than for the situation where the value of time for the frequency constraint is constant.\footnote{In Section 2 we mentioned in a note a third situation, for which $t$ equals $1/2F - t$. Since the third situation means that the integration range over departure times is smaller, it is easy to show that the subsidy is higher compared to the two situations that we have considered. We may then also conclude that the larger the proportion of passengers to whom the third situation applies, the larger the range of frequencies for which case II deficits per passenger exceed case I deficits.}

4.4. Peak and off-peak price
Irrespective of whether the value of frequency delay is zero, constant or variable, we have found that the optimal price increases with demand per departure. The operator’s variable cost is higher for peak than for off-peak due to vehicle-related costs for the peak. This does not imply, however, that optimal peak price generally exceeds optimal off-peak price. This is true with certainty only if optimum belongs to either case I or II for both peak and off-peak.

If $\phi^{II} < \phi^{III}$ and even more so if $\phi^{II}/\tau < 0$ and/or $\phi^{III}/\tau > 0$, the peak optimum may very well belong to case II and the off-peak optimum to case I. Even though the marginal operator costs are higher during peak (due to vehicle-related costs) than off-peak (effect 1), the marginal passenger costs may be smaller during peak (effect 2). The latter effect is due to the higher value of frequency delay in case II, which may imply a lower occupancy rate (per vehicle) in peak than in off-peak. Since effect 2 may very well be larger than effect 1, we conclude that the optimal off-peak price may exceed the optimal peak price.

5. EXTENSIONS

Without going into detail, let us briefly consider the situation where standing is not permitted, and then the potential impacts on investment and information of the analysis in this paper.

5.1. Standing not permitted
On certain modes and/or in some areas, standing passengers are not permitted. This may apply to trains and long-distance buses and definitely airlines. In our framework this situation can be taken into account directly by letting the value of riding time or the cost approach infinity when the number of passengers exceeds the number of seats. An alternative way is to use an occupancy constraint and compute a shadow price for one more seat. When the constraint is binding, the shadow price corresponds to an additional
negative external effect. Since the constraint cost is influenced by frequency (if passengers who are left behind wait for the next departure), the constraint, when binding, would also give rise to an additional positive external effect through increased frequency.\footnote{An alternative way to treat this situation is to internalise a probability cost of being left behind. Assume such a probability cost $T^p$ equal to $T^p[X[p, F], F]$ as part of the generalised cost. The first-order condition with respect to price would generate an additional negative external effect, $X(\partial T^p/\partial X)$. The first-order condition with respect to frequency would generate an additional positive external effect, $F(\partial T^p/\partial F)$. This probability cost is observed by Turvey and Mohring (1975) and dealt with in more detail by Panzar (1979).}

5.2. Impact on investment and information decisions
Although it is not really a part of the theme of this paper, we cannot refrain from mentioning a potentially important impact if our approach is applied also to investment and information decisions. When timetables are not used frequency delay constitutes a large part of the cost of travel time, appearing both at first boarding and at transfers between services. Also, whether or not it is worthwhile to use timetables will affect the benefit of having several services to choose from, the benefit of express services and the passengers’ choice of routes and stops. It is also obvious that introduction of easily available information for the passengers about departure times would in many situations radically reduce the cost of frequency delay and facilitate their choice of the most favourable service for each specific situation. This means that application of dual passenger behaviour and the fact that passengers place different values on frequency delay in different situations have a considerable potential impact on design and investments and on the resources spent on information.

6. CONCLUSIONS

We have found that the optimal price equals the operator’s marginal cost plus a negative external effect in terms of riding time and crowding costs, and that this price increases less with journey length the more boarding affects the run time of the service. The optimal price also equals the operator’s average variable cost minus a positive external effect. The intuitive reason for the subsequent financial deficit is that a subsidised price will encourage more passengers, which in turn raises the optimal frequency and benefits all passengers. Note that this corresponds to the principles behind optimal road pricing. More cars on the streets cause negative external effects, which causes optimal price to exceed average costs. Note also that this argument for an optimal deficit in public transport holds even if car trips are priced at full marginal social cost.

Emphasising dual passenger behaviour with respect to frequency and a variable value of frequency delay, this analysis has shown that: (i) there may be one low-frequency/low-deficit per passenger optimum, and another high-frequency/high-deficit per passenger optimum; (ii) infrequent urban, rural and inter-regional services, and extraordinarily frequent urban services, have small optimal deficits; (iii) typical frequent urban services have large optimal deficits; (iv) for most urban public transport, it should be endogenously
determined whether price should be high and deficit low or vice versa; (v) it may be that the optimal off-peak price exceeds the optimal peak price.

We have also noted that the application of dual behaviour and valuation with respect to frequency potentially has a considerable impact on design, investment and information criteria in the public transport sector.

APPENDIX

Derivations of first-order conditions with respect to price and frequency

The objective function, \( w[S[G(p, F)] + \pi] \), is maximised with respect to price and frequency, by setting partial derivatives equal to zero:

\[
\frac{\partial w}{\partial p} = \left( \frac{\delta S}{\delta G} \frac{\delta G}{\delta T} \frac{\delta T}{\delta X} \right) \frac{\partial X}{\partial p} = (p - m) \frac{\partial X}{\partial p} = 0
\]

\[
\frac{\partial w}{\partial F} = \left( \frac{\delta S}{\delta G} \frac{\delta G}{\delta T} \frac{\delta T}{\delta X} \right) \frac{\partial X}{\partial F} + \frac{\delta S}{\delta G} \frac{\delta G}{\delta T} \frac{\delta T}{\delta F} + \frac{\delta S}{\delta G} \frac{\delta G}{\delta X} \frac{\partial X}{\partial F} + \delta \pi = (p - m) \frac{\partial X}{\partial F} + \frac{\delta S}{\delta G} \frac{\delta G}{\delta T} \frac{\delta T}{\delta F} + \frac{\delta S}{\delta G} \frac{\delta G}{\delta X} \frac{\partial X}{\partial F} = 0
\]

where, in order to distinguish between various partial derivatives, \( \delta \) is used for partial derivatives going via demand, \( \delta \) is used for direct partial derivatives and \( \delta \) is used for total partial derivatives, and \( m \) is interpreted as marginal social cost. Once the first-order condition with respect to price (or demand) is derived, we can directly insert \( (p - m) \) from this differentiation, so that we only have to calculate the direct partial derivatives when differentiating with respect to frequency.

Differentiating the objective function:

\[
w = F \int_{0}^{t_{F}} s[G[p, F, t]] dt + pX[p, F] - FC[X[p, F]]
\]

with respect to \( p \) yields the following first-order condition with respect to price:

\[
\frac{\partial w}{\partial p} = F \left[ \int_{0}^{t_{F}} \frac{\partial s}{\partial G} \frac{\partial G}{\partial p} \frac{\partial X}{\partial p} + X + \frac{\partial X}{\partial p} \right] - F \frac{\partial C}{\partial X} \frac{\partial X}{\partial p} = 0
\]

(A2)

where \( \frac{\partial G}{\partial p} = 1 + \frac{\partial T}{\partial X} \frac{\partial X}{\partial p} \equiv 1 + \frac{\partial \phi}{\partial R} \frac{\partial X}{\partial p} \frac{h}{\sigma F} + \phi[R] \frac{\partial X}{\partial p} \frac{b}{F} \) is independent of \( t \).

Using (2), (5), (7) and factoring out \( X/\partial p \) we obtain:

\[
p^* = X \phi[R] \frac{b}{F} + X \frac{\partial \phi[R]}{\partial R} \frac{h}{\sigma F} + F \frac{\partial C}{\partial X} \frac{1}{\partial X} \equiv X \frac{\partial T}{\partial X} + F \frac{\partial C}{\partial X}
\]

(A3)

where \( h = bX/F + \gamma \gamma \) and \( R = X/F \sigma \).

The first-order condition with respect to frequency is:
\[
\frac{\partial w}{\partial F} = \int_0^V s[G[p, F, \ell, \delta]] dt - F(1/F^2)s[G[p, F, 1/F]] - C + F \frac{\partial C}{\partial X} \frac{\partial X}{\partial F} + F \int_0^V \frac{\partial s}{\partial G} \frac{\partial G}{\partial F} dt = 0 \quad (A4)
\]

Inserting
\[
F \int_0^V \frac{\partial s}{\partial G} \frac{\partial G}{\partial F} dt = F \int_0^V \frac{\partial s}{\partial G} \frac{\partial T}{\partial F} dt + X \frac{\partial \phi[R]}{\partial R} \frac{X}{\sigma F^2} h + X \phi[R] \frac{X}{F^2}
\]

and replacing the first and second terms on the right hand side in (A4) with \( z \) and \( Z \) respectively, we obtain:
\[
\frac{\partial w}{\partial F} = z - Z + Y + \frac{X}{F} \left( \frac{X}{\sigma F^2} \frac{h}{F} + X \phi[R] \frac{\partial C}{\partial X} \right) - C = 0 \quad (A5)
\]

After inserting the first-order condition with respect to price, profit per departure, \( \pi \), can be written:
\[
\pi = pX/F - C = -(z - Z + Y) \quad (A6)
\]

where
\[
z = \int_0^V s[G[p, F, t]] dt = \int_0^V \left( \sum_{G[p, F, t]} x(p) dp \right) dt \quad (A7)
\]

\[
Z = F(1/F^2)s[G[p, F, 1/F]] = \int_0^V \left( \sum_{G[p, F, t]} x(p) dp \right) dt \quad (A8)
\]

\[
Y = F \int_0^V \frac{\partial s}{\partial G} \frac{\partial T}{\partial F} dt = -F \int_0^V x[G[p, F, t]](\partial \phi/\partial t)(dt/\partial F) + \phi_t(\delta t/\partial F) dt \quad (A9)
\]

Observing that \( dt/\partial F = -1/F^2 \), \( \phi_t/\partial t \equiv - \phi_t t \), \( \phi_t/\partial t \equiv G/\partial t \equiv -G/\partial t \), we achieve:
\[
Y = -(1/F) \int_0^V x[G[p, F, t]](\partial G/\partial t) dt \quad (A10)
\]

Since \( G \) is differentiable and \( G/\partial t \) is continuous in the interval \([0 < t \leq 1/F]\), we may substitute to obtain:
\[
Y = (1/F) \int_{G[p, F, 0]}^{G[p, F, 0]} x(p) dp = \int_0^V \left( \int_{G[p, F, 0]}^{G[p, F, 0]} x(p) dp \right) dt \quad (A11)
\]

We can then calculate \( y = z - Z + Y \) as follows:
\[
y = \int_0^{\text{UF}} \left( \int_{G[p,F,t]}^{G_{\text{max}}(p,F,t)} x[p] \, dp \right) \, dt - \int_0^{\text{UF}} \left( \int_{G[p,F,\text{UF}]}^{G_{\text{max}}(p,F,t)} x[p] \, dp \right) \, dt + \int_0^{\text{UF}} \left( \int_{G[p,F,\text{UF}]}^{G[p,F,0]} x[p] \, dp \right) \, dt
\]
\[
= \int_0^{\text{UF}} \left( \int_{G[p,F,0]}^{G[p,F,t]} x[p] \, dp \right) \, dt \tag{A12}
\]

Profit per departure, \( p^* \), and optimal price, \( p^* \), may then be written:
\[
p^* = \frac{FC}{X} - \frac{FX}{C} = \int_0^{\text{UF}} \left( \int_{G[p,F,t]}^{G_{\text{max}}(p,F,t)} x[p] \, dp \right) \, dt = -\frac{1}{y} \tag{A13}
\]
\[
p^* = \frac{FC}{X} - \frac{FX}{C} \int_0^{\text{UF}} \left( \int_{G[p,F,0]}^{G[p,F,t]} x[p] \, dp \right) \, dt \equiv FC/X - FY/X \tag{A14}
\]

We can also rewrite the term for total deficit (difference between revenue and cost), \( y \), in (A13) by substitution, as follows:
\[
y = \int_0^{\text{UF}} \left( \int_{G[p,F,0]}^{G[p,F,t]} x[p] \, dp \right) \, dt \equiv \int_0^{\text{UF}} \left( \int_0^{G[p,F,t]} x[G(p,F,\psi)](\partial G/\partial \psi) \, d\psi \right) \, dt \tag{A15}
\]

where \( G/\psi = - (\phi \hat{\tau} / \partial \tau) \hat{\tau} - \phi \hat{\tau} = (\phi \hat{\tau} / \partial \psi) \psi - \phi \hat{\tau} \), yielding:
\[
y = \int_0^{\text{UF}} \left( \int_0^{G[p,F,0]} x[G(p,F,\psi)](- (\phi \hat{\tau} / \partial \psi) \psi + \phi \hat{\tau}) \, d\psi \right) \, dt \tag{A16}
\]

Where timetables are not used (case II), demand is independent of \( t \), implying the following subsidy term:
\[
y = x[G(p, F)] \int_0^{\text{UF}} \left( \int_0^{t'} (- (\partial \phi \hat{\tau} / \partial \psi) \psi + \phi \hat{\tau}) \, d\psi \right) \, dt
\]
\[
= x[\int_0^{\text{UF}} \left( \int_0^{t'} (- (\partial \phi \hat{\tau} / \partial \psi) \psi) \, d\psi \right) \, dt + x\phi \hat{\tau} \int_0^{\text{UF}} \, dt
\]
\[
= x[\int_0^{\text{UF}} \left( \int_0^{t'} (- (\phi \hat{\tau} / \partial \psi) \psi) \, d\psi \right) \, dt + x\phi \hat{\tau} / 2F^2] \tag{A17}
\]

Demand, consumers’ surplus and the terms \( z, Z, Y \) and \( y \) are visualised in Figure A1. For the sake of simplicity, all functions are illustrated as linear. The line BC is the demand \( x \) as a function of generalised cost \( G \), for passengers at time \( t = 0 \), that is, \( x[G(p, F, 0)] \). The line EU is the demand for passengers at \( t = 1/F \), that is, \( x[G(p, F, 1/F)] \). MBC is the consumers’ surplus for passengers at \( t = 0 \), that is, \( s[G(p, F, 0)] \). DEU is the corresponding consumers’ surplus for passengers at \( t = 1/F \), that is, \( s[G(p, F, 1/F)] \). The demand for one departure, \( q \), is the area MCUSD. \( z \) is the consumers’ surplus MCBEDU. \( Z \) is the consumers’ surplus for the passengers at \( t = 1/F \), multiplied by the interval \( 1/F \), that is, the volume BILUED. \( Y \) is the difference between consumers’ surplus at \( t = 0 \) and \( t = 1/F \), multiplied by the interval \( 1/F \), that is, the volume MCLIDUNQ. The total deficit, \( y \), corresponds to the volume MCUNQD. This is the aggregate cost saving in terms of frequency delay of not having the maximum frequency delay, or, in other words,
the difference between the total consumers' surplus and the consumers' surplus for the passengers who have the maximum $(1/F)$ frequency delay.

Where the value of time is constant, (A16) and (A17) yield the following expressions for the deficit per passenger, $y/q$, for the situations where passengers use (case I) and do not use (case II) timetables:

\[
\frac{y^I}{q} = \phi^{ui} \int_{0}^{t_f} x[G(p, F, \psi)] d\psi \int_{0}^{t_f} x[G(p, F, t)] \equiv \phi^{ui} A^I \quad (A18)
\]

\[
\frac{y^{II}}{q} = \phi^{ui} \frac{1}{2F} \equiv \phi^{ui} A^{II} \quad (A19)
\]

It is shown in Jansson (1991) that $\text{Max} \{A^I\} = A^{II}$. 

49
REFERENCES


