PERGAMON

Transportation Research Part B 35 (2001) 789–813

TRANSPORTATION
RESEARCH
PART B

www.elsevier.com/locate/trb

Optimal timetables for public transportation

André de Palma [a,*], Robin Lindsey [b]

[a] *Department of Economics, Thema Transports and Réseaux, University of Cergy-Pontoise, 33 Bld du Port, F-95011 Cergy Pontoise Cedex, France*

[b] *Department of Economics, University of Alberta, Canada*

Received 6 August 1999; received in revised form 23 January 2000; accepted 7 February 2000

Abstract

This paper analyzes the optimal timetable for a given number of public transport vehicles on a single transit line when riders differ with respect to the times at which they prefer to travel and the schedule delay costs they incur from traveling earlier or later than desired. The problem of minimizing riders' total schedule delay costs is formulated in continuous time and first-order optimality conditions are identified. An explicit solution is derived for the "line" model in which preferred travel times are uniformly distributed in the population over part of the day and trips cannot be rescheduled between days. This solution is compared with the optimal schedule for the "circle" model in which preferred travel times are uniformly distributed over the full 24 h day and trips can be rescheduled between days.

Keywords: Bus usage; Scheduling; Public transportation; Social optimum; Trip timing

1. Introduction

This paper is about trip timing by users of public transportation and how a schedule or timetable for transit vehicles should be chosen to best serve demand. Individuals who travel either earlier or later than they would like to incur "schedule delays". These delays are unavoidable with public transport unless vehicles depart continuously around the clock, which is impossible.

Constructing a timetable is part of the overall transit planning process, which also includes network design, a choice of service frequency (i.e., time headways between vehicles) for each route, and allocations of vehicles and crews to routes. This is one of the most complex network

* Corresponding author. Tel.: +33-1-34-25-60-63; fax: +33-1-34-25-62-33.
E-mail address: thema@u-cergy.fr (A. de Palma).

PII: S0191-2615(00)00023-0

problems in operations research (Israeli and Ceder, 1996; Marqués et al., 1996). The current paper focuses on the timetabling sub-problem using an explicit traveller choice model.

The transit timetabling problem is related to the well known *p-median* problem in operations research (see for example Larson and Odoni, 1981; Labbé et al., 1995), where the objective is to locate p facilities (e.g., post offices) in order to minimize some measure of distance between the facilities and their users. In the transit vehicle timetabling problem, facilities are transit vehicles, users are travellers, and the distance separating them is measured in time rather than geographical location.

While the parallels between the transit timetabling problem and the *p-median* problem are close, there are three differences between the standard formulation of the *p-median* problem and the timetabling problem as treated here. First, users and facilities are typically assumed to be located at a discrete set of nodes. Travellers, however, may want to depart at any time, and it is more accurate to model their desired travel times as being distributed continuously as a function of time of day. Transit vehicles too can (subject to logistics constraints) be scheduled at any time, so that a continuum approach is appropriate for both transit supply and transit demand. Indeed, the continuum approach has been standard in location theory since Hotelling's (1929) seminal article. There is a variant of the *p-median* problem in which customers are assumed to be distributed continuously on a network. Following the terminology in Labbé et al. (1995, 3.3.3) it will be referred to here as the *S-continuous p-median* problem.

A second difference between the prototypical *p-median* problem and the timetabling problem is that travellers' schedule delay costs can be different for arriving early and arriving late. Someone traveling to an important meeting, for example, bears a much higher cost for arriving 10 min after the meeting starts than from getting there 10 min early. The timetabling problem therefore cannot be formulated using a distance metric with the property that the distance between two points is the same in either direction.

Finally, the costs of schedule delay vary with the purpose of the trip, family time pressures, income, and other individual and trip-specific characteristics. Designing a transit timetable on the basis of a representative or "average" traveller can result in aggregate costs well above the optimum, as will be shown.

A simple version of the timetabling problem is studied in this paper. There is a given number of individuals who travel by transit on a single link. Preferred travel times and unit schedule delay costs for arriving early or late vary from person to person. Service on the route is provided by a fixed number of vehicles. Vehicle capacity constraints are ignored, so that a vehicle can carry any number of passengers without congestion. Also ignored are logistics problems, such as how to get each vehicle to the starting point on the route at its designated time.

Two location models are considered. In the first, travellers' desired travel times are distributed over a segment of the day and rescheduling of trips between days is impossible. This is an example of the "line" model pioneered by Hotelling (1929) and widely used in location theory. It has been applied to modeling public transit systems by Alfa and Chen (1995) and Kraus and Yoshida (1999), to competition between private bus companies by Dodgson et al. (1993) and Ellis and Silva (1998), and to competition between passenger rail companies by Whelan et al. (1998). Using analytical methods Newell (1971) has solved for an optimal bus timetable using a line model. However, travellers in his model incur waiting time rather than schedule delay costs, and their arrival rate at the bus stop is exogenous.

In the second location model it is assumed that desired travel times are distributed around the clock and that rescheduling of trips between days is possible. This is an example of the "circle" model, also frequently used in location theory, and used to model competition in bus markets by Evans (1987) and Ireland (1991). One of the goals of this paper is to highlight the differences between the optimal bus timetables for these two models, and relatedly the importance of adopting the best model to describe the transit market in a given region.

The analysis for both the line model and the circle model proceeds in two steps. The first step is to determine for an arbitrary timetable of vehicles which individuals will travel on which vehicles. This can be thought of as a demand allocation problem. Given the assumption of no congestion on transit or other externalities, individuals' vehicle choices are optimal in this step. The second step is to determine the timetable that minimizes total schedule delay costs given the behaviour of individuals identified in step one.

The paper is organized as follows. Section 2 describes the line model in which travellers' preferred travel times are distributed on a finite time segment. Section 3 addresses the demand allocation and optimal timetabling problems when individuals differ in their preferred travel times but have identical schedule delay cost functions. An analytical solution is derived for the case in which the distribution of preferred travel times is uniform. Section 4 builds on Section 3 by allowing for individual differences in the costs of traveling earlier or later than desired. Some of the properties of the demand allocation and optimal timetabling problems identified in Section 3 do not generalize in the face of this heterogeneity. Section 5 analyzes the circle model. A summary and directions for further research are provided in Section 6.

2. The line model

Due to economies of vehicle size, public transit service is generally provided with large vehicles according to a timetable with sometimes substantial time headways between vehicles. Most transit users suffer a schedule delay even if the transit system is reliable and adheres perfectly to the timetable. This contrasts with travel by automobile, which can be initiated at any time. Schedule delays on auto trips are only incurred either inadvertently due to random travel times, or deliberately to avoid peak-period congestion – as in Vickrey (1969). Thus, the trade-off between travel time and schedule delay that is essential to modeling trip timing decisions in private transportation is not as fundamental in public transportation. To keep the analysis simple, congestion will be ignored. It is assumed that the travel speeds and station dwelling times of transit vehicles are independent of the passenger load (and the same for all vehicles), and further that passengers incur no disutility from crowding in vehicles, on platforms, at bus stops, etc.

To facilitate an analytical solution as far as possible, attention is focused on a single transit line with an origin node, a terminal node, and no intermediate stops. To be concrete, the model will be presented as one of an urban bus line, although it is also applicable to streetcars, light rail and subways. And though the model is limited to a single link in a network, it is unnecessary to assume that travellers on the link begin and end their trips at the same points. The only relevant characteristic of a trip is when the traveller wants to ride the link in question.

In the absence of congestion, travel time on the link is a constant, h. A vehicle that departs from the origin node at time t therefore arrives at the terminal node at time $t_a = t + h$. To simplify the notation it is assumed without loss of generality that $h = 0$ so that $t_a = t$. Departure time and arrival time are then the same, and it is possible to refer unambiguously to the "time" at which a trip is made.

Each individual is assumed to have a most preferred trip time, denoted t^*, and to incur a schedule delay cost if traveling at time $t \neq t^*$ instead. The schedule delay cost function $D(\cdot)$ is assumed to depend only on the difference between t and t^*, and to have the piecewise linear form

$$D(t - t^*) = \beta[t^* - t]^+ + \gamma[t - t^*]^+, \tag{1}$$

where β is the schedule delay cost per minute of arriving early (before t^*), γ is the schedule delay cost per minute of arriving late (after t^*), and $[x]^+ \equiv \mathrm{Max}[0, x]$. Eq. (1) has been widely used for the case $\beta = \gamma$ in the literature on automobile trip timing, as well as a number of studies of public transport. To allow for increasing marginal disutility from increasing earliness or lateness, a strictly convex function could be assumed instead (e.g. Ellis and Silva, 1998), although doing so would complicate the algebra.

The number of travellers, N, on the line is taken as given, i.e., independent of the schedule delay cost incurred on a trip. The distribution function of desired travel times, $F(\cdot)$, is assumed to be absolutely continuous with respect to the Lebesgue measure. (This assures that the total schedule delay cost function to be minimized is differentiable in the cases that are considered.) The distribution is assumed to have a support $[0, L]$ so that $F(0) = 0$ and $F(L) = N$. The time interval $[0, L]$ can be thought of as the period during which most people wish to travel, e.g., 6–10 a.m. for the morning commute. It is assumed that trips cannot be rescheduled to another day. Thus, an individual who wants to travel at 10 p.m. on Wednesday cannot (or will not) defer it until 7 a.m. on Thursday.

Until Section 4, it is assumed that individuals are identical except for their desired travel times. It will be convenient to refer to an individual with desired trip time t^* simply as "individual t^*".

The bus system is assumed to operate n buses at times $T_1, \ldots, T_n$, where $T_i < T_j$ for $i < j$. Each individual must therefore travel at T_i for some i. Each is assumed to minimize his/her trip cost. Given the same constant (zero) travel time for each bus, and no differences in fares (which are implicitly zero), this means that a traveller will choose the bus that has the lowest schedule delay cost. Individual t^* will choose bus 1 if $t^* \leqslant T_1$, and choose bus n if $t^* \geqslant T_n$. If neither of these inequalities holds, then $T_i \leqslant t^* \leqslant T_{i+1}$ for some $i = 1, \ldots, n-1$. As a convention assume that someone who is indifferent between two buses selects the earlier bus. Then individual t^* with $T_i \leqslant t^* \leqslant T_{i+1}$ chooses bus i if

$$\beta(t^* - T_i) \leqslant \gamma(T_{i+1} - t^*),$$

and chooses bus $i + 1$ otherwise. Let $t^*_{i,i+1}$, $i = 1, \ldots, n-1$, denote the desired travel time of an individual who is indifferent between bus i and bus $i + 1$. And define $t^*_{0,1} \equiv 0$ and $t^*_{n,n+1} \equiv L$. It then follows that

$$t^*_{i,i+1} = \begin{cases} (\beta T_i + \gamma T_{i+1})/(\beta + \gamma) & \text{for } i = 1, \ldots, n-1, \\ 0 & \text{for } i = 0, \\ L & \text{for } i = n. \end{cases} \tag{2}$$

All travellers with $t^* \in (t^*_{i-1,i}, t^*_{i,i+1}]$ choose bus i. Those with $t^* \in (t^*_{i-1,i}, T_i)$ travel *late*, those with $t^* \in (T_i, t^*_{i,i+1})$ travel *early*, and anyone with $t^* = T_i$ travels *on time*. If $\beta = \gamma$, then the "market boundary" $t^*_{i,i+1}$ between bus i and bus $i+1$ is located halfway between T_i and T_{i+1}. As the cost of arriving early (β) increases, the boundary shifts to the left and ridership on bus $i+1$ increases at the expense of bus i. Analogously, as the cost of arriving late (γ) increases, the boundary shifts to the right.

3. Optimal timetables with homogeneous schedule delay cost functions

In this section, it is assumed that travellers have the same schedule delay cost parameters, β and γ. Though all may differ in their desired travel times, they will be referred to as "homogeneous" individuals. Heterogeneity in β and γ will be considered in Section 4.

The bus system operators are assumed to choose $T_1, \dots, T_n$ to minimize the total schedule delay costs of the N travellers. As noted in the introduction, this problem is related to the well known $p-$*median* problem in operations research. Because the optimal bus timetable here involves n buses, and provides service to a continuous distribution of travellers with asymmetric early and late cost penalties ($\beta \neq \gamma$), it will be called the *asymmetric S-continuous n-median* problem.

Given the optimal timetable for each n, and values for the operating cost and fixed cost per bus, it is possible to solve numerically for the number of buses that minimizes the sum of travellers' schedule delay costs and bus system costs. This interesting and practically important problem is not addressed here; for a recent treatment see Kraus and Yoshida (1999).

3.1. Optimality conditions

Let $f(\cdot)$ denote the density function corresponding to the distribution of desired travel times, $F(\cdot)$. Given a timetable $T_1, \dots, T_n$, total schedule delay costs are

$$\begin{aligned} C(T_1, \dots, T_n) = {} & \gamma \int_0^{T_1} (T_1 - t) f(t)\,\mathrm{d}t + \beta \int_{T_1}^{t^*_{1,2}} (t - T_1) f(t)\,\mathrm{d}t + \cdots \\ & + \gamma \int_{t^*_{i-1,i}}^{T_i} (T_i - t) f(t)\,\mathrm{d}t + \beta \int_{T_i}^{t^*_{i,i+1}} (t - T_i) f(t)\,\mathrm{d}t + \cdots \\ & + \gamma \int_{t^*_{n-1,n}}^{T_n} (T_n - t) f(t)\,\mathrm{d}t + \beta \int_{T_n}^{L} (t - T_n) f(t)\,\mathrm{d}t. \end{aligned}$$

The first line in this sum specifies the schedule delay costs incurred by riders on bus 1, the second line specifies the costs of riders on bus i, and the last line the costs of riders on bus n.

The function $C(\cdot)$ is to be minimized by choice of the timetable $T_1, \dots, T_n$. This task is complicated by the fact that $C(\cdot)$ may not be a convex function of the T_i. To see this, suppose as shown in Fig. 1 that desired travel times are concentrated in three symmetric peak periods, centred about t_1, t_4 and t_7, respectively, with the early and late peaks of equal size and smaller than the central peak. Assume $n = 2$. One plausible solution, shown by the large black dots, is to schedule the first bus at t_1 and the second bus slightly to the right of t_4 at t_5. (Scheduling the second bus at t_5 rather than t_4 reduces schedule delay costs for travelers in the late peak without increasing aggregate

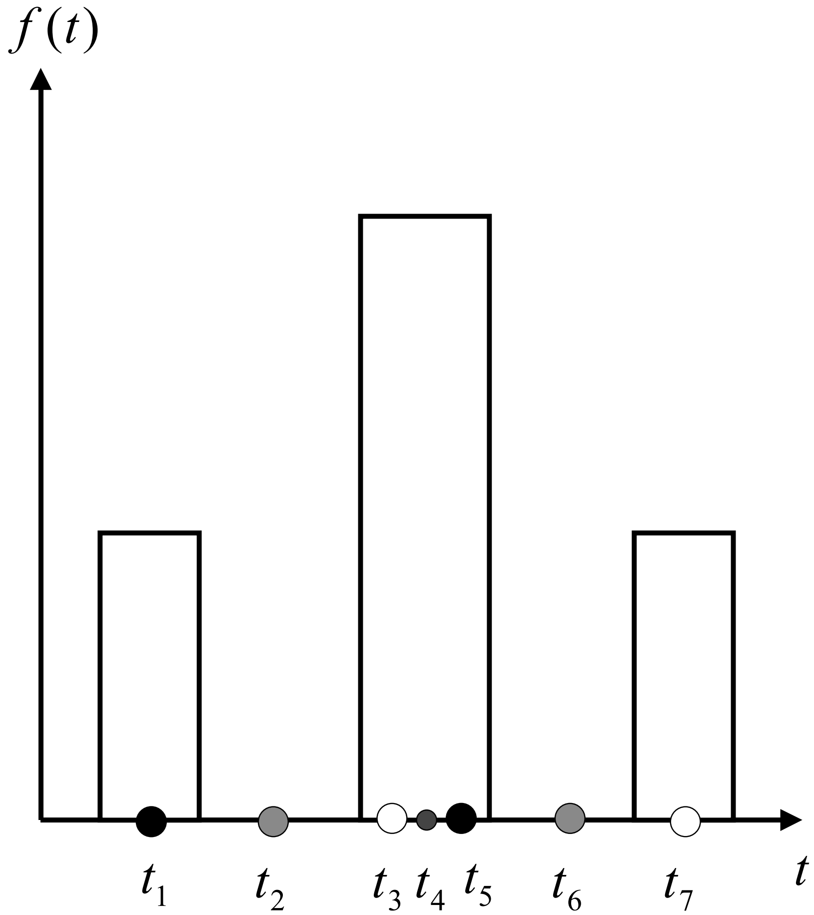

Fig. 1. A distribution of preferred travel times for which total schedule delay costs are not convex.

schedule delay costs of travelers in the central peak by very much.) Another possibility, shown by the open dots, is to schedule the first bus just before t_4 at t_3, and the second bus at t_7. This yields the same total cost as the first timetable. But scheduling the two buses at some convex combination of these two solutions, such as bus 1 at $t_2 = (1/2)t_1 + (1/2)t_3$ and bus 2 at $t_6 = (1/2)t_5 + (1/2)t_7$ (see the shaded dots), would result in schedule delay costs that are strictly higher.

Because the cost function is not convex, optimization methods such as Newton's and fastest descent are not guaranteed to find a global minimum. With this caveat in mind, the analysis will proceed on the assumption that the first-order conditions for minimization of $C(0)$ nevertheless do define a global minimum. This is the case when the density function $f(0)$ is uniform, as will be assumed in much of what follows. A proof that the first-order conditions define a global optimum is provided in Appendix A for the more general case of heterogeneous individuals, considered in Section 4.

The first-order condition for the timing of bus 1 is

$$\begin{aligned}\frac{\partial C(T_1,\dots,T_n)}{\partial T_1} &= \gamma\int_0^{T_1} f(t)\,\mathrm{d}t - \beta\int_{T_1}^{t^*_{1,2}} f(t)\,\mathrm{d}t + \left[\beta\left(t^*_{1,2}-T_1\right) - \gamma(T_2 - t^*_{1,2})\right] f(t^*_{1,2})\frac{\partial t^*_{1,2}}{\partial T_1} \\ &= 0. \end{aligned} \tag{3}$$

This condition stipulates that total schedule delay costs do not change if bus 1 is rescheduled slightly later. The three terms in Eq. (3) are interpreted as follows. The first term denotes the increase in late arrival costs for riders on bus 1 who prefer to travel earlier than T_1. The second term denotes the reduction in early arrival costs for riders on bus 1 who prefer to travel later than T_1. And the final term is the change in schedule delay costs for travellers with $t^* = t^*_{1,2}$ who switch from traveling late on bus 2 to traveling early on bus 1. By Eq. (2), the third term is zero, and Eq. (3) simplifies to

$$\gamma\int_0^{T_1} f(t)\,\mathrm{d}t = \beta\int_{T_1}^{t^*_{1,2}} f(t)\,\mathrm{d}t. \tag{4}$$

The first-order conditions for $T_2,\dots,T_n$, are similar

$$\gamma\int_{t^*_{i-1,i}}^{T_i} f(t)\,\mathrm{d}t = \beta\int_{T_i}^{t^*_{i,i+1}} f(t)\,\mathrm{d}t, \quad i = 2,\dots,n. \tag{5}$$

The system of n Eqs. (4) and (5) defines the optimal timetable. The solution of this system is the *homogeneous asymmetric S-continuous n-median* of the distribution $f(\cdot)$.

Proposition 1. *The solution to the homogeneous asymmetric S-continuous n-median problem over the segment* $[0,L]$, $T_1^0,\dots,T_n^0$, *satisfies the first-order conditions*:

$$\gamma\int_{t^*_{i-1,i}}^{T_i^0} f(t)\,\mathrm{d}t = \beta\int_{T_i^0}^{t^*_{i,i+1}} f(t)\,\mathrm{d}t, \quad i = 1,\dots,n,$$

where the market boundaries $t^*_{i,i+1}$, $i = 0,\dots,n$, *are given by Eq.* (2).

Note that the solution depends on the shape of the density function $f(\cdot)$ but not on the scale, i.e., on the total number of users, N. Note also that the solution depends only on the ratio β/γ and not on the scale of the schedule delay cost parameters. If $n = 1$, the optimal timetable for the one bus can be solved, implicitly or explicitly, from Eq. (4), which reduces to

$$F(T_1^0) = \frac{\beta}{\beta+\gamma}. \tag{6}$$

With two or more buses, a numerical solution is typically required.

3.2. A uniform distribution of desired travel times

Suppose that t^* is uniformly distributed: $f(t^*) = N/L$ for $0 \leqslant t^* \leqslant L$ and $f(t^*) = 0$ otherwise. Consider first the case of one bus. Given Eq. (6), this bus is optimally located on $[0,L]$ at

$$T^0_{1,n=1} = \frac{\beta}{\beta+\gamma}L. \tag{7}$$

The bus is scheduled earlier the larger the cost per minute of arriving late relative to early (γ/β). Average schedule delay cost per traveller is

$$\bar{c}^0_{n=1} = \frac{1}{2}\frac{\beta\gamma}{\beta+\gamma}L. \tag{8}$$

Note that the traveller with the earliest desired travel time, $t^* = 0$, incurs the maximum schedule delay cost of $\gamma T^0_{1,n=1} = (\beta\gamma/\beta+\gamma)L$. The traveller with the latest desired arrival time, $t^* = L$, incurs the same cost: $\beta(L - T^0_{1,n=1}) = (\beta\gamma/\beta+\gamma)L$. Meanwhile, the traveller with $t^* = T^0_{1,n=1}$ incurs no schedule delay cost. Given the uniform density of t^*, the average schedule delay cost is just half of the maximum, as recorded in Eq. (8).

Suppose now that there are n buses. The first-order conditions in Eqs. (4) and (5) become

$$\gamma\frac{N}{L}T_1 = \beta\frac{N}{L}(t^*_{1,2} - T_1), \tag{9}$$

$$\gamma\frac{N}{L}(T_i - t^*_{i-1,i}) = \beta\frac{N}{L}(t^*_{i,i+1} - T_i), \quad i = 2,\dots,n-1, \tag{10}$$

$$\gamma\frac{N}{L}(T_n - t^*_{n-1,n}) = \beta\frac{N}{L}(L - T_n). \tag{11}$$

Eqs. (9)–(11) have a unique solution which – as noted above – is a global maximum (see Appendix A). Substituting Eq. (2) into Eqs. (9)–(11) in turn, one obtains

$$T_2 = \frac{2\beta+\gamma}{\beta}T_1, \tag{12}$$

$$T_{i+1} = 2T_i - T_{i-1}, \quad i = 2,\dots,n-1, \tag{13}$$

$$T_n = \frac{(\beta+\gamma)L + \gamma T_{n-1}}{\beta+2\gamma}. \tag{14}$$

It follows by induction from Eqs. (12) and (13) that

$$T_i = \frac{i\beta + (i-1)\gamma}{\beta}T_1. \tag{15}$$

Eqs. (14) and (15) then give $T_1 = (\beta/\beta+\gamma)L/n$, which with Eq. (15) yields finally

$$T^0_{i,n} = \left(i - \frac{\gamma}{\beta+\gamma}\right)\frac{L}{n}, \quad i = 1,\dots,n-1.$$

Therefore the optimal timetable is periodic. Each bus serves a market of width L/n and carries N/n riders. Market boundaries are defined by the times:

$$t^*_{i,i+1} = i\frac{L}{n}, \quad i = 1,\dots,n-1.$$

Average schedule delay cost can be deduced immediately from Eq. (8) for one bus by substituting L/n in place of L

$$\overline{c}_n^0 = \frac{1}{2} \frac{\beta\gamma}{\beta+\gamma} \frac{L}{n}. \tag{16}$$

The solution is summarized in the following proposition.

Proposition 2. *Consider the homogeneous asymmetric S-continuous n-median problem over the segment* $[0, L]$ *when desired travel times are uniformly distributed. The optimal bus timetable is*

$$T_{i,n}^0 = \left(i - \frac{\gamma}{\beta+\gamma}\right)\frac{L}{n}, \quad i = 1, \ldots, n.$$

The market boundaries between buses are

$$t_{i,i+1}^* = i\frac{L}{n}, \quad i = 1, \ldots, n-1,$$

and average schedule delay cost is

$$\overline{c}_n^0 = \frac{1}{2} \frac{\beta\gamma}{\beta+\gamma} \frac{L}{n}.$$

The solution has several properties

1. Buses are spaced L/n apart, and each carries N/n passengers.
2. A fraction $(\beta/\beta+\gamma)$ of the riders on each bus arrives late, and a fraction $(\gamma/\beta+\gamma)$ arrives early.
3. The average cost for early arrivals and the average cost for late arrivals are both equal to $\overline{c}_n^0$.
4. The ratio of the total late schedule delay cost to the total early schedule delay cost is β/γ.
5. If $\beta = \gamma$, the buses are located at $(1/2)(L/n)$, $(3/2)(L/n), \ldots, (2n-1/2)L/n$. This is a standard result in location theory (e.g. Archibald et al., 1986).
6. In the limit $\beta/\gamma \downarrow 0$, $T_{i,n}^0 \downarrow (i-1)L/n$, $i = 1, \ldots, n$: as late arrival becomes infinitely costly relative to early arrival, buses are scheduled so that no one has to travel late. Analogously, as $\beta/\gamma \uparrow \infty$, $T_{i,n}^0 \uparrow i(L/n)$, $i = 1, \ldots, n$.

4. Optimal timetables with heterogeneous schedule delay costs

So far it has been assumed that while travellers differ in their preferred travel times, they have the same (i.e., homogeneous) schedule delay cost functions. In reality, scheduling costs depend on a number of individual-specific characteristics such as trip purpose, time pressures, occupation, age, sex, income and family size. As far as occupation, for example, individuals who work independently, such as writers, software developers and academics, typically have weaker preferences for when they start work than do workers who interact with others, such as stockbrokers, assembly line workers and clerical support staff. Independent workers also tend to incur lower relative penalties for starting "late" than starting "early" than do interactive workers.

Suppose then that there are K classes or groups of travellers, indexed by k, with class k comprising N^k individuals with schedule delay cost parameters β^k and γ^k, and a distribution of desired travel times characterized by the density function $g^k(\cdot)$. Estimates of the β^k, γ^k and $g^k(\cdot)$

can be obtained using revealed or stated preference methods, see for example (Nuzzolo and Russo, 1996, Tables 1 and 2; Whelan et al., 1998, Table 4). Section 4.1 following identifies the first-order conditions for an optimal timetable given arbitrary $g^k(\cdot)$ functions. Section 4.2 derives an explicit solution when the $g^k(\cdot)$ are uniform.

4.1. Optimality conditions

Define $f^k(\cdot) \equiv N^k g^k(\cdot)$. The choice of bus by class k is defined by market boundaries analogous to those given by Eq. (2)

$$t^{k*}_{i,i+1} = \begin{cases} (\beta^k T_i + \gamma^k T_{i+1})/(\beta^k + \gamma^k) & \text{for } i = 1, \ldots, n-1, \\ 0 & \text{for } i = 0, \\ L & \text{for } i = n. \end{cases} \qquad k = 1, \ldots, K. \tag{17}$$

For a given bus timetable, $T_1, \ldots, T_n$, total schedule delay costs for all users are

$$\begin{aligned} C(T_1, \ldots, T_n) = \sum_k \Bigg[& \gamma^k \int_0^{T_1} (T_1 - t) f^k(t)\, \mathrm{d}t + \beta^k \int_{T_1}^{t^{k*}_{1,2}} (t - T_1) f^k(t)\, \mathrm{d}t + \cdots \\ & + \gamma^k \int_{t^{k*}_{i-1,i}}^{T_i} (T_i - t) f^k(t)\, \mathrm{d}t + \beta^k \int_{T_i}^{t^{k*}_{i,i+1}} (t - T_i) f^k(t)\, \mathrm{d}t + \cdots \\ & + \gamma^k \int_{t^{k*}_{n-1,n}}^{T_n} (T_n - t) f^k(t)\, \mathrm{d}t + \beta^k \int_{T_n}^{L} (t - T_n) f^k(t)\, \mathrm{d}t \Bigg]. \end{aligned}$$

The first-order condition for the timing of bus 1 is

$$\begin{aligned} \frac{\partial C(T_1, \ldots, T_n)}{\partial T_1} = & \sum_k \left[\gamma^k \int_0^{T_1} f^k(t)\, \mathrm{d}t - \beta^k \int_{T_1}^{t^{k*}_{1,2}} f^k(t)\, \mathrm{d}t \right] \\ & + \sum_k \left\{ \left[\beta^k \left(t^{k*}_{1,2} - T_1 \right) - \gamma^k (T_2 - t^{k*}_{1,2}) \right] f^k(t^{k*}_{1,2}) \frac{\partial t^{k*}_{1,2}}{\partial T_1} \right\} = 0. \end{aligned} \tag{18}$$

Eq. (18) is interpreted in the same way as Eq. (3) for homogeneous travellers. As in Eq. (3), the term in square brackets on the second line is zero for all k by Eqs. (17), and (18) reduces to the analogue of Eq. (4)

$$\sum_k \left[\gamma^k \int_0^{T_1} f^k(t)\, \mathrm{d}t \right] = \sum_k \left[\beta^k \int_{T_1}^{t^{k*}_{1,2}} f^k(t)\, \mathrm{d}t \right]. \tag{19}$$

The first-order conditions for $T_2, \ldots, T_n$ are analogues of Eq. (5)

$$\sum_k \left[\gamma^k \int_{t^{k*}_{i-1,i}}^{T_i} f^k(t)\, \mathrm{d}t \right] = \sum_k \left[\beta^k \int_{T_i}^{t^{k*}_{i,i+1}} f^k(t)\, \mathrm{d}t \right], \quad i = 2, \ldots, n. \tag{20}$$

The solution to the system of n Eqs. (19) and (20), $T_1^0, \ldots, T_n^0$, will be called the *heterogeneous asymmetric S-continuous n-median* of the distributions $f^k(\cdot)$, $k = 1, \ldots, K$. The following proposition generalizes Proposition 1.

Proposition 3. *The solution to the heterogeneous asymmetric S-continuous n-median problem over the segment* $[0, L]$ *satisfies the first-order conditions*

$$\sum_k \left[\gamma^k \int_{t_{i-1,i}^{k*}}^{T_i} f^k(t)\,\mathrm{d}t\right] = \sum_k \left[\beta^k \int_{T_i}^{t_{i,i+1}^{k*}} f^k(t)\,\mathrm{d}t\right], \quad i = 1, \ldots, n,$$

where the market boundaries $t_{i,i+1}^{k*}$, $i = 0, \ldots, n$, $k = 1, \ldots, K$, *are given by Eq.* (17).

4.2. Uniform distributions of desired arrival times

It is now assumed that desired travel times in each class are uniformly distributed on $[0, L]$. With this assumption it is possible to derive explicitly the optimal bus timetable and to compare it with the corresponding optimal timetable for homogeneous travellers. To save on writing, define $\delta^k \equiv (\beta^k\gamma^k/\beta^k + \gamma^k)$, $\lambda^k \equiv N^k/N$ (the fraction of the population that belongs to class k), and the population-weighted mean values of the β^k, γ^k and δ^k

$$\begin{aligned} B &= \sum_k \lambda^k \beta^k, \\ \Gamma &= \sum_k \lambda^k \gamma^k, \\ \Delta &= \sum_k \lambda^k \delta^k. \end{aligned} \tag{21}$$

To compare the bus timetable with the timetable for homogeneous travellers, repeated use will be made of the following lemma which is proved in Appendix B.

Lemma 1. *Let* B, Γ *and* Δ *be as defined in Eq.* (21). *Then* $B\Gamma \geqslant (B + \Gamma)\Delta$, *with a strict inequality unless* γ^k/β^k *is the same for all classes of travellers.*

The optimal bus timetable can be derived using the same steps as in Section 3 for the timetable with homogeneous travellers, see Appendix C. The solution is given in

Proposition 4. *Consider the heterogeneous asymmetric S-continuous n-median problem over the segment* $[0, L]$ *when desired travel times are uniformly distributed for each class of traveller. The optimal bus timetable is*

$$T_{i,n}^0 = \frac{(i-1)B\Gamma + B\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta} L, \quad i = 1, \ldots, n. \tag{22}$$

Buses are spaced $(B\Gamma/(n-1)B\Gamma + (B+\Gamma)\Delta)L$ *apart. The market boundaries between buses for traveller class* k *are*

$$t_{i,i+1}^{k*} = \frac{(i-1)B\Gamma + B\Delta + \frac{\gamma^k}{\beta^k+\gamma^k}B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta}L, \quad i = 1,\dots,n-1, \quad k = 1,\dots,K.$$

And the average schedule delay cost for all travellers is

$$\bar{c}_n^0 = \frac{1}{2}\frac{B\Gamma\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta}L. \tag{23}$$

Recall that with homogeneous travellers, buses are optimally spaced L/n apart and carry equal loads of N/n. With heterogeneous travellers, buses are also uniformly spaced, but the spacing is greater. The difference in spacing is, by Proposition 4

$$\frac{B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta}L - \frac{1}{n}L \stackrel{s}{=} B\Gamma - (B+\Gamma)\Delta \geqslant 0, \tag{24}$$

where $\stackrel{s}{=}$ means identical in sign. By the Lemma, the right-hand side of Eq. (24) is strictly greater than zero unless all traveller classes have the same relative costs of early and late arrival. This result can be seen as a sort of product differentiation. Individuals with relatively strong aversion to being late (high γ^k/β^k) are accommodated by scheduling the first bus very early so that few travellers in this group are obliged to travel late. And individuals with relatively strong aversion to being early (small γ^k/β^k) are served by scheduling the last bus very late. Compared to the optimal timetable with homogeneous travellers, therefore, buses are run over a longer part of the day when travellers differ in their relative early and late arrival costs.

Recall too from Section 3 that with homogeneous travellers the first bus and last bus are run at times $T_{1,n}^0 = (\beta/\beta+\gamma)L/n$ and $T_{n,n}^0 = [n - (\gamma/\beta+\gamma)]L/n$. With one group $B = \beta$, $\Gamma = \gamma$, and these equations can be written $T_{1,n}^0 = (BB+\Gamma)L/n$ and $T_{n,n}^0 = [n - (B/B+\Gamma)]L/n$. Comparing these scheduled times with the scheduled times for heterogeneous travellers given in Eq. (22) one has

$$T_{1,n}^0 - \frac{B}{B+\Gamma}\frac{L}{n} = \left[\frac{nB\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta} - \frac{B}{B+\Gamma}\right]\frac{L}{n} \stackrel{s}{=} (B+\Gamma)\Delta - B\Gamma \leqslant 0, \tag{25}$$

and

$$T_{n,n}^0 - \left[n - \frac{\Gamma}{B+\Gamma}\right]\frac{L}{n} = \left[n\frac{(n-1)B\Gamma + B\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta} - \left(n - \frac{\Gamma}{B+\Gamma}\right)\right]\frac{L}{n} \stackrel{s}{=} B\Gamma - (B+\Gamma)\Delta \geqslant 0, \tag{26}$$

where the inequalities follow from the Lemma. These results are summarized in the following proposition.

Proposition 5. *Buses are more widely spaced in the optimal timetable with heterogeneous travellers than with homogeneous travellers. The first bus is scheduled earlier* (*see Eq.* (25)), *and the last bus is scheduled later* (*see Eq.* (26)). *The bus timetable is therefore extended at both the beginning and at the end of the day.*

Because the first bus is run earlier, and the last bus later, than with homogeneous travellers one might suspect that these buses cater to peripheral or "isolated" travellers and accordingly carry fewer than N/n passengers each. To see whether this is so, let N_i denote the number of passengers carried by bus i, $i = 1, \dots, n$. Using Proposition 4 it is readily shown that

$$N_1 = \sum_k \frac{N^k}{L} t_{1,2}^{k*} = \frac{B\Gamma + \left(\sum_k (\gamma^k/\beta^k + \gamma^k) N^k/N\right) B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta} N, \tag{27}$$

and

$$N_n = \sum_k \frac{N^k}{L} \left(L - t_{n-1,n}^{k*}\right) = \frac{B\Gamma + \Gamma\Delta - \left(\sum_k (\gamma^k/\beta^k + \gamma^k) N^k/N\right) B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta} N. \tag{28}$$

It then follows that

$$N_1 + N_n - \frac{2}{n} N \overset{s}{=} -(n-2)[B\Gamma - (B+\Gamma)\Delta].$$

By the Lemma, the right-hand side of this equation is strictly negative for $n \geqslant 3$ unless γ^k/β^k is the same for all groups. Therefore the first and last buses together do indeed carry a smaller than average load. Yet it is possible for either the first bus or the last bus to carry an above-average load. To see this, consider a numerical example in which there are two groups ($K = 2$) of equal size. Following the typology entertained at the beginning of Section 4, suppose group 1 consists of interactive workers and group 2 consists of independent workers. Small's (1982) empirical study of commuters is based on a sample of mainly interactive workers for which he obtains an estimate of $(\gamma/\beta) \cong 4$. Take this value for γ^1/β^1. Since Eqs. (27) and (28) depend only on the ratio of schedule delay cost parameters, one can assume without loss of generality that $\beta^1 = 1$. Given $\gamma^1/\beta^1 = 4$, this implies $\gamma^1 = 4$.

Suppose first that the independent workers have relatively weak schedule preferences so that $\beta^2 = \gamma^2 = 1$. If $n = 3$, buses 1, 2 and 3 then carry respectively 31.2%, 34.4% and 34.4% of the passengers. Bus 3 carries more than a third of the total. If $n = 10$, bus 1 carries 9.2% of the total load and buses 2…10 carry 10.1% each.

Suppose alternatively that independent workers have relatively strong schedule preferences so that $\beta^2 = \gamma^2 = 4$. Then with $n = 3$, buses 1, 2 and 3 carry 34.4%, 34.4% and 31.2% respectively of the passengers. Now it is the first bus that carries more than a one-third share. With $n = 10$, buses 1…9 carry 10.1% of the load each, and bus 10 carries 9.2%.

One property of the optimal timetable with a homogeneous population, Property 6 in Section 3.2, continues to hold in a modified form with heterogeneous users. Pick any class k. In the limit $\beta^k \uparrow \infty$, $B \uparrow \infty$, while Γ and Δ remain finite. From Eq. (22) one then obtains $\lim_{\beta^k \uparrow \infty} T_{n,n}^0 = L$. Thus, if any class of traveller is infinitely averse to arriving early, the last bus is scheduled at the latest desired travel time so that no one has to arrive early. Similarly, $\lim_{\gamma^k \uparrow \infty} T_{1,n}^0 = 0$: if any class of traveller is infinitely averse to arriving late, the first bus is scheduled at the earliest desired travel time so that no one has to arrive late.

From the foregoing it is evident that heterogeneity in schedule delay costs should be taken into account in choosing a bus timetable. Indeed, one can ask: how much would travellers' aggregate schedule delay costs increase if heterogeneity were ignored in designing the timetable? To address this question it will be assumed that the timetable is inappropriately chosen on the basis of a "representative" traveller whose schedule delay cost parameters equal the population-weighted mean values; i.e. B, Γ and Δ. Hereafter this will be called the "representative traveller" approach.

Using the representative traveller approach the timetable is set as in Proposition 2 with B in place of β and Γ in place of γ:

$$T^R_{i,n} = \left(i - \frac{\Gamma}{B+\Gamma}\right)\frac{L}{n}, \quad i = 1,\dots,n,$$

where the superscript R denotes the representative traveller approach. Again by Proposition 2, average schedule delay costs are computed to be

$$\hat{c}^R_n = \frac{1}{2}\Delta\frac{L}{n}. \tag{29}$$

As is shown in Appendix D, average schedule delay costs will actually be

$$\overline{c}^R_n = \frac{1}{2}\left\{1 + \frac{1}{n}\frac{B\Gamma - (B+\Gamma)\Delta}{\Delta(B+\Gamma)}\right\}\Delta\frac{L}{n}. \tag{30}$$

Using Eqs. (23), (29) and (30), and after some algebra, one obtains:

Proposition 6. *Consider the heterogeneous asymmetric S-continuous n-median problem when desired travel times are uniformly distributed on $[0, L]$ for each class of traveller. If the representative traveller approach is used to choose the timetable, average schedule delay cost will be higher than at the true optimum by the fraction*

$$\frac{\overline{c}^R_n - \overline{c}^0_n}{\overline{c}^0_n} = \frac{n-1}{n^2}\left[\frac{(B\Gamma - (B+\Gamma)\Delta)^2}{B\Gamma(B+\Gamma)\Delta}\right] = \frac{n-1}{n^2}\left[\frac{\left(1 - \frac{(B+\Gamma)\Delta}{B\Gamma}\right)^2}{\frac{(B+\Gamma)\Delta}{B\Gamma}}\right]. \tag{31}$$

By the Lemma, this fraction is strictly positive for $n > 1$ unless γ^k/β^k is the same for all k. Average costs according to the representative traveller approach are lower than actual costs by the fraction

$$\frac{\hat{c}^R_n - \overline{c}^R_n}{\overline{c}^R_n} = \frac{\frac{1}{n}\frac{B\Gamma-(B+\Gamma)\Delta}{\Delta(B+\Gamma)}}{1 + \frac{1}{n}\frac{B\Gamma-(B+\Gamma)\Delta}{\Delta(B+\Gamma)}} \geqslant 0. \tag{32}$$

Again, this fraction is strictly positive unless γ^k/β^k is the same for all k.

To get an idea of the possible magnitude of the error introduced by the representative traveller approach, consider again the numerical example in the previous section with $\beta^1 = 1$ and $\gamma^1 = 4$. For group 2 take $\beta^2 = 1$ and assume $\gamma^2/\beta^2 = r(\gamma^1/\beta^1)$, where typically $r \ll 1$. And suppose there are two buses ($n = 2$). With this parameterization one obtains

$$\frac{(B+\Gamma)\Delta}{B\Gamma} = \frac{(6+4r)(1+9r)}{10(1+r)(1+4r)}.$$

If $\gamma^2/\beta^2 = \gamma^1/\beta^1$, then $r = 1$, $((B+\Gamma)\Delta/B\Gamma) = 1$, and by Eq. (31) use of the representative traveller approach results in no error in the choice of timetable, and therefore no increase in costs. If, contrarily, independent workers are equally averse to early and late arrival, then $\gamma^2 = \beta^2$, $r = 1/4$, and $((B + \Gamma)\Delta/B\Gamma) = 0.91$. With two buses, $((\overline{c}_n^R - \overline{c}_n^0)/\overline{c}_n^0) \cong 0.025$. The optimal timetable is not chosen, and average schedule delay costs are increased by about 2.5%. Suppose finally that $\gamma^2 = 0$, which is conceivable if independent workers do not like to rise early in the morning, but are otherwise indifferent about their work hours. One then has $r = 0$ and $((B + \Gamma)\Delta/B\Gamma) = 3/5$. With two buses, $(\overline{c}_n^R - \overline{c}_n^0/\overline{c}_n^0) = 1/6$. Use of the representative traveller approach results in costs that are one sixth higher than optimal. With five buses the increase is still above 10%. While such gross errors seem unlikely in practice, this exercise does highlight the importance of obtaining information about travellers' characteristics when designing a bus timetable.

5. The circle model

Up to now it has been assumed that the distribution of preferred travel times is limited to part of the day and that trips cannot be rescheduled from one day to another. This is consistent with the Hotelling (1929) line model. An alternative topology is the circle model developed by Salop (1979). In this model, desired travel times are spread around the full 24 h clock and rescheduling of trips between days is feasible. This means that an individual who prefers to travel on day d before the first bus on that day will no longer necessarily take that bus, but may instead take the last bus on day $d - 1$. Similarly, someone who prefers to travel after bus n on day d may choose to take the first bus on day $d + 1$.

The line model and the circle model are both idealizations. The main virtue of these models is analytical tractability, rather than realism. It can be debated which model provides the better approximation to the demand for bus travel (or demand for another public transportation mode) in a given market. Indeed, both models have been adopted in the literature on public transportation. To the extent that the two models "bracket" the reality of a given market, it is instructive to examine them both.

The only change that has to be made to the line model in converting it to a circle is in Eq. (17) defining the market boundaries between buses. Let L now denote the length of a day (24 h). In place of $t_{0,1}^{k*} = 0$, one obtains $t_{n,1}^{k*} = (\beta^k(T_n - L) + \gamma^k T_1/\beta^k + \gamma^k)$. And in place of $t_{n,n+1}^{k*} = L$, one has $t_{n,n+1}^{k*} = (\beta^k T_n + \gamma^k(T_1 + L)/\beta^k + \gamma^k)$. These two new relations are actually the same because bus $n + 1$ is bus 1 on the next day. Market boundaries for all buses are therefore described by a single equation

$$t_{i,i+1}^{k*} = \frac{\beta^k T_i + \gamma^k T_{i+1}}{\beta^k + \gamma^k}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, K, \tag{33}$$

where $T_0 = T_n - L$ and $T_{n+1} = T_1 + L$.

The market boundaries of the first and last buses are now determined by travellers' choices, rather than by end-of-the-day constraints. Though this change to the model may seem minor, it affects the optimal bus timetable with heterogeneous travellers in a qualitative way. To see this, suppose as earlier that each class of traveller has a uniform distribution of desired travel times. It then follows (see Appendix E) that, regardless of how parameters β^k and γ^k vary across classes, the optimal timetable is *periodic*: all buses are scheduled L/n apart and carry N/n passengers each.

Unlike with the line model, the bus timetable is no longer extended to accommodate a diverse population. This is because with rescheduling of trips between days now possible, a traveller can always choose between taking an early bus or taking a late bus (or possibly a bus on time).

Given the periodicity of the timetable, the average schedule delay cost of travellers in class k is given by the analogue of (16)

$$\overline{c}_n^{k0} = \frac{1}{2}\frac{\beta^k\gamma^k}{\beta^k+\gamma^k}\frac{L}{n}.$$

The average schedule delay cost of all travellers together is

$$\overline{c}_n^{0} = \sum_k \lambda^k \overline{c}_n^{k0} = \frac{1}{2}\left(\sum_k \lambda^k \frac{\beta^k\gamma^k}{\beta^k+\gamma^k}\right)\frac{L}{n} = \frac{1}{2}\Delta\frac{L}{n}. \tag{34}$$

Comparing (34) with the average schedule delay cost formula (23) for the line model one obtains immediately from the lemma.

Proposition 7. *Consider the heterogeneous asymmetric S-continuous n-median problem with uniformly distributed desired travel times for each class of traveller. Average schedule delay costs are lower for the optimal timetable on the circle than for the optimal timetable on the line unless* γ^k/β^k *is the same for all k.*

Average schedule delay costs are lower in the circle model because travellers can reschedule trips between days, whereas they cannot in the line model.

If the representative traveller approach were used, the same periodic timetable (except perhaps for its phase) would be chosen. Average schedule delay costs would be computed to be

$$\hat{c}_n^{R} = \frac{1}{2}\frac{B\Gamma}{B+\Gamma}\frac{L}{n}. \tag{35}$$

Given Eqs. (34) and (35) one obtains:

Proposition 8. *Consider the heterogeneous asymmetric S-continuous n-median problem when desired travel times are uniformly distributed around the circle for each class of traveller. If the representative traveller approach is used, an optimal timetable will be chosen and* $\overline{c}_n^R = \overline{c}_n^0$. *But average schedule delay costs will be overestimated by a fraction*

$$\frac{\hat{c}_n^R - \overline{c}_n^0}{\overline{c}_n^0} = \frac{B\Gamma - (B+\Gamma)\Delta}{(B+\Gamma)\Delta} \geqslant 0.$$

By the Lemma, this fraction is strictly positive unless γ^k/β^k *is the same for all k.*

Costs are overestimated using the representative traveller approach because it overlooks the ability of travellers to choose between travelling early and late according to their individual preferences. Using the two-group parameterization of Section 4.2, the proportional error is about

10% if $r = 1/4$, and as large as 2/3 if $r = 0$. The miscalculation in travel costs through use of a representative traveller can thus be appreciable. Because costs are overestimated with the approach, a transit authority could be induced to overinvest in bus capacity.

The brief analysis in this section illustrates that, as is well known in location theory, the properties of location problems vary according to whether they are set on a line or on a circle.

6. Conclusions

This paper uses the location theory framework to study the optimal timetable for transit vehicles on a single transit line. Each transit rider is assumed to have an ideal time of day to travel, and to incur a schedule delay cost from traveling earlier or later. Two location models are considered: the line model and the circle model. In the line model, travellers' preferred travel times are distributed over part of the day, and rescheduling of trips between days is impossible. In the circle model, desired travel times are distributed around the clock, and rescheduling of trips between days is possible.

The modelling approach used here is similar to Vickrey (1969) model of automobile trip timing in its use of schedule delay cost functions. It differs in that congestion is ignored (i.e., travel time is constant), and more fundamentally in that – unlike drivers – transit users cannot start a trip whenever they want to, but rather must choose a departure time according to the bus timetable. The model is also related to the *p-median* problem of optimal facility location. It differs from the standard formulation in treating the locations of both users (travellers) and facilities (transit vehicles) as continuous rather than discrete, and in allowing unit schedule delay costs to differ for early and late arrival and from traveller to traveller.

While the formulation of the optimal timetable problem here is simple, finding a solution can be difficult because the total schedule delay cost function to be minimized is not necessarily a convex function of the choice variables (the departure times of the buses). The solution is characterized by a set of necessary first-order conditions (see Propositions 1 and 3). In general, numerical methods are required to solve these equations. But an explicit analytical solution can be found when the distribution of preferred travel times is uniform. Section 3 describes the solution when users have identical schedule delay cost functions (Proposition 2), and Section 4 for the case where their costs differ (Proposition 4). The two solutions are compared in Section 4.2 (Proposition 5).

The optimal timetable is also derived for the circle model and compared with the timetable for the line model (Proposition 7). Finally, the use of a representative traveller approach to compute the timetable is considered, and biases in the calculation for the line and circle models are identified (Propositions 6 and 8).

The model omits a number of important features of transit systems that should be included in future work:

1. The number of transit vehicles has been treated as exogenous. Given a solution to the optimal bus timetable for any fixed number of vehicles, and data on operating and maintenance costs per bus, the optimal number of buses could be determined by minimizing the sum of total schedule delay and system costs.

2. Scheduling of transit vehicles may be constrained by network considerations, such as the time required to do the back-haul on a round trip, or to move vehicles between routes.
3. The number of travellers is taken as given, thereby suppressing individual trip frequency and mode choice decisions. While this assumption might be defended on the grounds that fare elasticities of demand are low in the short run, fare elasticities typically increase over time and demand can be relatively sensitive to time headways (Berechman, 1993) which are determined by the choice of timetable.
4. Travellers are assumed to know the bus timetable, so that they can reach a bus stop just as a bus arrives and do not have to wait. An alternative assumption, adopted by Evans (1991) and Ellis and Silva (1998), is that travellers do not know the timetable. In practice, of course, some individuals know the timetable and others do not. The decision whether to be informed depends on how much schedule delay and waiting time can be reduced, which in turn depends on how frequently trips are made and on headways between buses (Jolliffe and Hutchinson, 1975; Tisato, 1991).
5. Buses are assumed to adhere perfectly to the timetable. In reality, travel times are random and buses can arrive late or be cancelled. Stochasticity in bus arrival times and its impact on riders' departure time choice are considered by Sumi et al. (1990).
6. Vehicle capacity constraints should be incorporated, as in (Alfa and Chen, 1995; Kraus and Yoshida, 1999; Lam et al., 1999b).
7. Service on a given route can be provided by more than one type of vehicle. Transit operators face a trade-off between the scale economies of larger vehicles, and the more frequent service made possible by smaller vehicles such as jitneys, which have been successfully integrated in various cities (Klein et al., 1997).
8. Congestion aboard vehicles and at bus stops (or on railway station or subway platforms) has been ignored. Empirical evidence on the importance of congestion on light rail systems is documented in Lam et al. (1999a).

Acknowledgements

We are grateful to Frank Haight and five anonymous referees for most helpful comments. We would also like to thank Cedric Fontan, Fabrice Marchal and Omar Mekkaoui for research assistance, and PREDIT, Ministry of Transport, France, for financial support.

Appendix A. Sufficiency of first-order conditions with heterogeneous travellers and uniform distributions of desired travel times

Because interchanging any two buses in a timetable leaves total schedule delay costs unchanged the objective function is clearly not globally convex. The first-order conditions are nevertheless sufficient for a cost minimum if the total schedule delay cost function is convex for $T_1, \dots, T_n$ such that $0 \leqslant T_1 \leqslant T_2 \leqslant \cdots \leqslant T_n \leqslant L$. Convexity will be established by showing that the matrix of second-order partial derivatives is positive definite given this ordering of buses.

The first-order derivative of total schedule delay costs with respect to T_1 is given in Eq. (18)

$$\frac{\partial C}{\partial T_1} = \sum_k \left[\gamma^k \int_0^{T_1} f^k(t)\,\mathrm{d}t - \beta^k \int_{T_1}^{t_{1,2}^{k*}} f^k(t)\,\mathrm{d}t\right] + \sum_k \left\{\left[\beta^k\left(t_{1,2}^{k*} - T_1\right) - \gamma^k(T_2 - t_{1,2}^{k*})\right] f^k(t_{1,2}^{k*}) \frac{\partial t_{1,2}^{k*}}{\partial T_1}\right\}, \tag{A.1}$$

where the term in square brackets on the second line is zero. Given uniform distributions for each traveller class ($f^k(t) = f^k$) Eq. (A.1) reduces to

$$\frac{\partial C}{\partial T_1} = \sum_k f^k \left[\gamma^k T_1 - \beta^k (t_{1,2}^{k*} - T_1)\right].$$

The non-zero second-order derivatives are

$$\frac{\partial^2 C}{\partial T_1^2} = \sum_k f^k\left(\gamma^k + \delta^k\right), \quad \text{and} \quad \frac{\partial^2 C}{\partial T_1 \partial T_2} = -\sum_k f^k \delta^k. \tag{A.2}$$

For $i = 2, \ldots, n-1$ the first-order derivatives are

$$\begin{aligned}\frac{\partial C}{\partial T_i} = & \sum_k \left[\gamma^k \int_{t_{i-1,i}^{k*}}^{T_i} f^k(t)\,\mathrm{d}t - \beta^k \int_{T_i}^{t_{i,i+1}^{k*}} f^k(t)\,\mathrm{d}t\right] \\ & + \sum_k \left[\beta^k\left(t_{i-1,i}^{k*} - T_{i-1}\right) - \gamma^k(T_i - t_{i-1,i}^{k*})\right] f^k(t_{i-1,i}^{k*}) \frac{\partial t_{i-1,i}^{k*}}{\partial T_i} \\ & + \sum_k \left[\beta^k\left(t_{i,i+1}^{k*} - T_i\right) - \gamma^k(T_{i+1} - t_{i,i+1}^{k*})\right] f^k(t_{i,i+1}^{k*}) \frac{\partial t_{i,i+1}^{k*}}{\partial T_i}.\end{aligned} \tag{A.3}$$

By Eq. (17) the last two lines are zero. Given uniform distributions Eq. (A.3) reduces to

$$\frac{\partial C}{\partial T_i} = \sum_k f^k \left[\gamma^k(T_i - t_{i-1,i}^{k*}) - \beta^k(t_{i,i+1}^{k*} - T_i)\right].$$

The non-zero second-order derivatives are

$$\frac{\partial^2 C}{\partial T_i^2} = 2\sum_k f^k \delta^k, \quad \frac{\partial^2 C}{\partial T_{i-1} \partial T_i} = -\sum_k f^k \delta^k, \quad \text{and} \quad \frac{\partial^2 C}{\partial T_{i+1} \partial T_i} = -\sum_k f^k \delta^k. \tag{A.4}$$

Finally, for $i = n$ the first-order derivative is

$$\begin{aligned}\frac{\partial C}{\partial T_n} = & \sum_k \left[\gamma^k \int_{t_{n-1,n}^{k*}}^{T_i} f^k(t)\,\mathrm{d}t - \beta^k \int_{T_n}^{L} f^k(t)\,\mathrm{d}t\right] \\ & + \sum_k \left[\beta^k\left(t_{n-1,n}^{k*} - T_{n-1}\right) - \gamma^k(T_n - t_{n-1,n}^{k*})\right] f^k(t_{n-1,n}^{k*}) \frac{\partial t_{n-1,n}^{k*}}{\partial T_n}.\end{aligned} \tag{A.5}$$

By Eq. (17) the second line is zero, and Eq. (A.5) reduces to

$$\frac{\partial C}{\partial T_n} = \sum_k f^k \left[\gamma^k(T_n - t_{n-1,n}^{k*}) - \beta^k(L - T_n)\right].$$

The non-zero second-order derivatives are

$$\frac{\partial^2 C}{\partial T_n^2} = \sum_k f^k(\delta^k + \beta^k), \quad \text{and} \quad \frac{\partial^2 C}{\partial T_{n-1}\partial T_n} = -\sum_k f^k \delta^k. \tag{A.6}$$

Now given $f^k = \frac{N^k}{L}$ and $\lambda^k = N^k/N$, $f^k = \lambda^k N/L$. It follows from Eq. (21) that $\sum_k f^k \beta^k = N/LB$, $\sum_k f^k \gamma^k = N/L\Gamma$ and $\sum_k f^k \delta^k = N/L\Delta$. Given Eqs. (A.2), (A.4) and (A.6), the matrix of second-order derivatives of C can then be written

$$\left[\frac{\partial^2 C}{\partial T_i \partial T_j}\right] = \frac{N}{L}\begin{bmatrix} \Gamma + \Delta & -\Delta & & & \\ -\Delta & 2\Delta & -\Delta & & \\ & -\Delta & 2\Delta & \ddots & \\ & & \ddots & \ddots & -\Delta \\ & & & -\Delta & B + \Delta \end{bmatrix}.$$

Let D_i denote the principal minor of order i. $[(\partial^2 C/\partial T_i \partial T_j)]$ is positive definite if $D_i > 0,\ i = 1, \ldots, n$.

Lemma 2.

$$D_i = \delta^{i-1}[i\gamma + \delta] > 0, \quad i = 1, \ldots, n-1.$$

Proof (*By induction*). $D_1 = \Gamma + \Delta > 0$. By inspection of the matrix

$$\begin{aligned} D_{i+1} &= 2\delta D_i - \delta^2 D_{i-1} = \delta(2D_i - \delta D_{i-1}) = \delta(2\delta^{i-1}[i\gamma + \delta] - \delta\delta^{i-2}[(i-1)\gamma + \delta]) \\ &= \delta^i[(i+1)\gamma + \delta]. \qquad \square \end{aligned}$$

Finally

$$\begin{aligned} D_n &= (\beta + \delta)D_{n-1} - \delta^2 D_{n-2} = (\beta + \delta)\delta^{n-2}[(n-1)\gamma + \delta] - \delta^2\delta^{n-3}[(n-2)\gamma + \delta] \\ &= \beta\delta^{n-2}[(n-1)\gamma + \delta] + \delta^{n-1}[(n-1)\gamma + \delta - (n-2)\gamma - \delta] = \beta\delta^{n-2}[(n-1)\gamma + \delta] + \delta^{n-1}\gamma > 0. \end{aligned}$$

Appendix B. Proof of Lemma

By the definitions of B, Γ and Δ in (21)

$$\begin{aligned} B\Gamma - (B + \Gamma)\Delta &= \left(\sum_k \lambda^k \beta^k\right)\left(\sum_k \lambda^k \gamma^k\right) - \left(\sum_k \lambda^k \beta^k + \sum_k \lambda^k \gamma^k\right)\sum_k \lambda^k \frac{\beta^k \gamma^k}{\beta^k + \gamma^k} \\ &= \sum_k \lambda^k \frac{\beta^k \beta^k}{\beta^k + \gamma^k}\left(\sum_k \lambda^k \gamma^k\right) - \left(\sum_k \lambda^k \beta^k\right)\sum_k \lambda^k \frac{\beta^k \gamma^k}{\beta^k + \gamma^k}. \end{aligned} \tag{B.1}$$

By Jensen's inequality

$$\sum_k \lambda^k \frac{\beta^k \beta^k}{\beta^k + \gamma^k} \geqslant \frac{\left(\sum_k \lambda^k \beta^k\right)\left(\sum_k \lambda^k \beta^k\right)}{\left(\sum_k \lambda^k \beta^k\right) + \left(\sum_k \lambda^k \gamma^k\right)} = \frac{BB}{B + \Gamma},$$

with a strict inequality unless γ^k/β^k is the same for all k. Again, by Jensen's inequality

$$\sum_k \lambda^k \frac{\beta^k \gamma^k}{\beta^k + \gamma^k} \leqslant \frac{\left(\sum_k \lambda^k \beta^k\right)\left(\sum_k \lambda^k \gamma^k\right)}{\left(\sum_k \lambda^k \beta^k\right) + \left(\sum_k \lambda^k \gamma^k\right)} = \frac{B\Gamma}{B + \Gamma},$$

with a strict inequality unless γ^k/β^k is the same for all k. We therefore have on the right-hand side of (B.1)

$$\sum_k \lambda^k \frac{\beta^k \beta^k}{\beta^k + \gamma^k}\left(\sum_k \lambda^k \gamma^k\right) - \left(\sum_k \lambda^k \beta^k\right)\sum_k \lambda^k \frac{\beta^k \gamma^k}{\beta^k + \gamma^k} \geqslant \frac{BB}{B + \Gamma}\Gamma - B\frac{B\Gamma}{B + \Gamma} = 0,$$

with a strict inequality unless γ^k/β^k is the same for all classes of travellers. □

Appendix C. Optimal timetable with heterogeneous travellers and uniform distributions of desired travel times on the line

With uniform distributions of desired travel times, the first-order conditions Eqs. (19) and (20) become:

$$\sum_k \gamma^k \frac{N^k}{L} T_1 = \sum_k \beta^k \frac{N^k}{L}(t^{k*}_{1,2} - T_1), \tag{C.1}$$

$$\sum_k \gamma^k \frac{N^k}{L}(T_i - t^{k*}_{i-1,i}) = \sum_k \beta^k \frac{N^k}{L}(t^{k*}_{i,i+1} - T_i), \quad i = 2, \ldots, n-1, \tag{C.2}$$

$$\sum_k \gamma^k \frac{N^k}{L}(T_n - t^{k*}_{n-1,n}) = \sum_k \beta^k \frac{N^k}{L}(L - T_n). \tag{C.3}$$

Multiplying through by L/N, using $\lambda^k = N^k/N$, and substituting (17), these conditions can be written

$$\sum_k \lambda^k \gamma^k T_1 = \sum_k \lambda^k \beta^k \left(\frac{\beta^k}{\beta^k + \gamma^k} T_1 + \frac{\gamma^k}{\beta^k + \gamma^k} T_2 - T_1\right),$$

$$\sum_k \lambda^k \gamma^k \left(T_i - \left(\frac{\beta^k}{\beta^k + \gamma^k} T_{i-1} + \frac{\gamma^k}{\beta^k + \gamma^k} T_i\right)\right)$$
$$= \sum_k \lambda^k \beta^k \left(\frac{\beta^k}{\beta^k + \gamma^k} T_i + \frac{\gamma^k}{\beta^k + \gamma^k} T_{i+1} - T_i\right), \quad i = 2, \ldots, n-1,$$

$$\sum_k \lambda^k \gamma^k \left(T_n - \left(\frac{\beta^k}{\beta^k + \gamma^k} T_{n-1} + \frac{\gamma^k}{\beta^k + \gamma^k} T_n\right)\right) = \sum_k \lambda^k \beta^k (L - T_n).$$

These equations in turn reduce to

$$\Gamma T_1 = \Delta(T_2 - T_1),$$

$$\Delta(T_i - T_{i-1}) = \Delta(T_{i+1} - T_i), \quad i = 2, \ldots, n-1,$$

$$\Delta(T_n - T_{n-1}) = B(L - T_n).$$

Solving these equations sequentially yields

$$T_2 = \frac{\Gamma + \Delta}{\Delta} T_1, \tag{C.4}$$

$$T_{i+1} = 2T_i - T_{i-1}, \quad i = 2, \ldots, n-1, \tag{C.5}$$

$$T_n = \frac{B}{B+\Delta} L + \frac{\Delta}{B+\Delta} T_{n-1}. \tag{C.6}$$

It follows by induction from Eqs. (C.4) and (C.5) that

$$T_i = \frac{(i-1)\Gamma + \Delta}{\Delta} T_1. \tag{C.7}$$

Eqs. (C.6) and (C.7) then give

$$T_1 = \frac{B\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta} L,$$

which with (C.7) yields finally the timing of each bus

$$T_i = \frac{(i-1)B\Gamma + B\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta} L, \quad i = 1, \ldots, n,$$

and the market boundaries between buses

$$t_{i,i+1}^{k*} = \frac{(i-1)B\Gamma + B\Delta + \frac{\gamma^k}{\beta^k+\gamma^k} B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta} L, \quad i = 1, \ldots, n-1, \quad k = 1, \ldots, K.$$

The spacing between buses is

$$T_i - T_{i-1} = \frac{B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta} L. \tag{C.8}$$

The average schedule delay cost for all travellers is a weighted average of the costs for travellers in each time interval between buses. Given (C.8) the average cost for group k for $t^* \in [T_i^0, T_{i+1}^0]$, $i = 1, \ldots, n-1$ is

$$\overline{c}_{i,n}^k = \frac{1}{2} \frac{\beta^k \gamma^k}{\beta^k + \gamma^k} \frac{B\Gamma}{(n-1)B\Gamma + (B+\Gamma)\Delta} L, \quad i = 1, \ldots, n-1.$$

For $t^* \in [0, T_1^0]$,

$$\overline{c}_{0,n}^k = \frac{1}{2} \gamma^k T_1^0 = \frac{\gamma^k}{2} \frac{B\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta} L,$$

and for $t^* \in [T_n^0, L]$,

$$\bar{c}_{n,n}^k = \frac{1}{2}\beta^k(L - T_n^0) = \frac{\beta^k}{2}\frac{\Gamma\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta}L.$$

Average schedule costs are

$$\bar{c}_n^0 = \frac{1}{L}\sum_k \lambda^k\left\{\bar{c}_{0,n}^k T_1^0 + \bar{c}_{i,n}^k(T_n^0 - T_1^0) + \bar{c}_{n,n}^k(L - T_n^0)\right\},$$

which reduces after substitution to

$$\bar{c}_n^0 = \frac{1}{2}\frac{B\Gamma\Delta}{(n-1)B\Gamma + (B+\Gamma)\Delta}L.$$

Appendix D. Average schedule delay costs on the line using the representative traveller approach

Using the representative traveller approach the timetable chosen is

$$T_{i,n}^R = \left(i - \frac{\Gamma}{B+\Gamma}\right)\frac{L}{n}, \quad i = 1,\ldots,n.$$

Buses are spaced L/n apart. Average schedule delay costs for this timetable are calculated using the same procedure as for the optimal timetable. The average cost for group k individuals with $t^* \in [T_{i,n}^R, T_{i+1,n}^R]$ is

$$\bar{c}_{i,n}^{kR} = \frac{1}{2}\frac{\beta^k\gamma^k}{\beta^k + \gamma^k}\frac{L}{n}, \quad i = 1,\ldots,n-1.$$

For $t^* \in [0, T_{1,n}^R]$,

$$\bar{c}_{0,n}^{kR} = \frac{1}{2}\gamma^k T_{1,n}^R = \frac{\gamma^k}{2}\frac{B}{B+\Gamma}\frac{L}{n}.$$

For $t^* \in [T_{n,n}^R, L]$,

$$\bar{c}_{n,n}^{kR} = \frac{1}{2}\beta^k(L - T_{n,n}^R) = \frac{\beta^k}{2}\frac{\Gamma}{B+\Gamma}\frac{L}{n}.$$

Average aggregate schedule costs are

$$\bar{c}_n^R = \frac{1}{L}\sum_k \lambda^k\left\{\bar{c}_{0,n}^{kR} T_{1,n}^R + \bar{c}_{i,n}^{kR}(T_{n,n}^R - T_{1,n}^R) + \bar{c}_{n,n}^{kR}(L - T_{n,n}^R)\right\},$$

which reduces after substitution to

$$\bar{c}_n^R = \frac{1}{2}\left\{1 + \frac{1}{n}\frac{B\Gamma - (B+\Gamma)\Delta}{\Delta(B+\Gamma)}\right\}\Delta\frac{L}{n}.$$

Appendix E. Optimal timetable with heterogeneous travellers and uniform distributions of desired travel times on the circle

The optimal timetable is derived using the same logic as for the line model, but with fewer steps. The first-order conditions for the T_i are

$$\sum_k \gamma^k \frac{N^k}{L}(T_i - t^{k*}_{i-1,i}) = \sum_k \beta^k \frac{N^k}{L}(t^{k*}_{i,i+1} - T_i), \quad i = 1, \ldots, n. \tag{E.1}$$

Substituting for the $t^{k*}_{i,i+1}$ with Eq. (33), (E.1) reduces to

$$\Delta(T_i - T_{i-1}) = \Delta(T_{i+1} - T_i), \quad i = 1, \ldots, n$$

or

$$T_{i+1} = 2T_i - T_{i-1}, \quad i = 1, \ldots, n. \tag{E.2}$$

By induction one gets

$$T_i = iT_1 + (i-1)T_0, \tag{E.3}$$

where T_0, the timing of the last bus on the previous day, is arbitrary. Using the relation $T_n = T_0 + L$, one obtains finally

$$T_i = T_0 + i\frac{L}{n}.$$

The timetable is periodic with buses spaced L/n apart.

References

Alfa, A.S., Chen, M., 1995. Temporal distribution of public transport demand during the peak period. European Journal of Operational Research 83, 137–153.

Archibald, G.C., Eaton, B.C., Lipsey, R.G., 1986. Address models of value. In: Stiglitz, J.E., Mathewson, F.G. (Eds.), New Developments in the Analysis of Market Structure. MIT Press, Cambridge, pp. 3–47.

Berechman, J., 1993. Public Transit Economics and Deregulation Policy. North-Holland, Amsterdam.

Dodgson, J.S., Katsoulayos, Y., Newton, C.R., 1993. An application of the economic modelling approach to the investigation of predation. Journal of Transport Economics and Policy 27, 153–170.

Ellis, C.J., Silva, E.C.D., 1998. British bus deregulation: Competition and demand coordination. Journal of Urban Economics 43 (3), 336–361.

Evans, A., 1987. A theoretical comparison of competition and other economic regimes for bus services. Journal of Transport Economics and Policy 21, 7–36.

Hotelling, H., 1929. Stability in competition. Economic Journal 39, 41–57.

Ireland, N.J., 1991. A product differentiation model of bus deregulation. Journal of Transport Economics and Policy 25 (2), 153–162.

Bianco, L., Ceder A, 1996. Multi-objective approach for designing transit routes with frequencies. In: Toth, P. (Ed.), Advanced Methods in Transportation Analysis. Springer-Verlag, Berlin, pp. 157–182.

Jolliffe, J.K., Hutchinson, T.P., 1975. A behavioural explanation of the association between bus and passenger arrivals at a bus stop. Transportation Science 9, 248–282.

Klein, D.B., Moore, A., Reja, B., 1997. Curb Rights: A Foundation for Free Enterprise in Urban Transit. Brookings Institution Press, Washington, DC.

Kraus, M., Yoshida, Y., 1999. A model of a welfare-maximizing urban mass transit authority. Working Paper. Department of Economics, Boston College.

Labbé, M., Peeters, D., Thisse, J.-F., 1995. Location on networks. In: Ball, M.O., Magnanti, T.L., Monma, C.L., Nemhauser, G.L. (Eds.), Handbooks in Operations Research and Management Science, vol. 8, Network Routing, Elsevier, Amsterdam, pp. 551–624.

Lam, W.H.K., Cheung, C-Y, Lam, C.F., 1999a. A study of crowding effects at the Hong Kong light rail transit stations. Transportation Research A 33A, 401–415.

Lam, W.H.K., Gao, Z.Y, Chan, K.S., Yang, H., 1999b. A stochastic user equilibrium assignment model for congested transit networks. Transportation Research B 33B, 351–368.

Larson, R.C., Odoni, R., 1981. Urban Operations Research. Prentice-Hall, Englewood Cliffs, NJ.

Marqués, A., Torregrosa, M., Camarena, A., Darby-Dowman, K., Moody, S., Little, J., 1996. Flexible dynamic scheduling: A major improvement for public transport. In: Stephanedes, Y.J., Filippi, F. (Eds.), Applications of Advanced Technologies in Transportation Engineering. Proceedings of the Fourth International Conference. American Society of Civil Engineers, New York, pp. 134–138.

Newell, G.F., 1971. Dispatching policies for a transportation route. Transportation Science 5 (1), 91–105.

Nuzzolo, A., Russo, F., 1996. Stochastic assignment models for transit low frequency services: Some theoretical and operative aspects. In: Bianco, L., Toth, P. (Eds.), Advanced Methods in Transportation Analysis. Springer, Berlin, pp. 321–339.

Salop, S., 1979. Monopolistic competition with outside goods. Bell Journal of Economics and Management Science 10, 141–156.

Small, K., 1982. The scheduling of consumer activities: Work trips. American Economic Review 72, 467–479.

Sumi, T., Matsumoto, Y., Miyaki, Y., 1990. Departure time and route choice of commuters on mass transit systems. Transportation Research B 24B, 247–262.

Tisato, P., 1991. User costs in public transport: A cost minimisation approach. International Journal of Transport Economics 18 (2), 167–193.

Vickrey, W.S., 1969. Congestion theory and transport investment. American Economic Review (Papers and Proceedings) 59, 251–260.

Whelan, G., Preston, J., Wardman, M., Nash, C., 1998. The privatisation of passenger rail services in Great Britain: An evaluation of competition for the market and competition in the market. Paper presented at the Eighth World Conference on Transportation Research, Antwerp, Belgium.