Journal of Transport Economics and Policy, Volume 37, Part 2, May 2003, pp. 261–277

From the Single Line Model to the Spatial Structure of Transit Services

Corridors or Direct?

Sergio R. Jara-Díaz and Antonio Gschwender

Address for correspondence: Professor Sergio R. Jara-Díaz, Universidad de Chile, Casilla 228-3, Santiago, Chile. Antonio Gschwender is also at the University of Chile. This research was partially financed by Fondecyt, Chile, grant 1010687, and the Millennium Nucleus ''Complex Engineering Systems''. The useful comments of one unknown referee are gratefully acknowledged.

Abstract

The microeconomic analysis of public transportation involves the optimisation of all resources, both users' and operators'. This has been applied to optimise frequency and fleet size for isolated transit lines, as well as to study optimal spacing for multiple lines serving a single destination. In this paper the spatial structure of transit services is analysed within the context of multiple origins and destinations. Direct services (no transfers) are compared against transit corridors for simple though illustrative spatially diversified demands. The best transit structure is shown to depend upon the demand volume, the relative time values and on network related parameters. Optimal fleet size preserves the ''square root rule''.

Date of receipt of final manuscript: January 2003

Introduction

In the microeconomic analysis of urban public transport (optimal pricing and design of services), two types of resources are considered as necessary: those provided by the operators, as vehicles, fuel, terminals, or labour, and those provided by the users, namely their time, usually divided into waiting, access, and in-vehicle time. In his pioneering work based on Vickrey's view (1955), Mohring (1972, 1976) constructed a microeconomic model to determine optimal frequency of buses serving a corridor with fixed demand. The main result was that frequency should be proportional to the square root of demand, and this happened only because all resources (operators' and users') were considered when finding the minimum cost operation.

Mohring (1972, 1976) then moved into a more complex model in which the number of stops in the corridor is also a variable, and where both the bus cycle time and users' in-vehicle time depend on the number of passengers boarding and alighting. Moreover, a probability of buses not stopping at a bus stop is also considered. The resulting total cost (including users' time) can not be minimised algebraically, and Mohring solves it numerically. Jansson (1980, 1984) simplifies this model by eliminating the number of bus stops as a variable and assuming buses always stop, but keeping the effect of passengers boarding and alighting on travel time. The analytical solution of the resulting cost minimisation problem yields a modified version of the "square root formula" for optimal frequency. This is then extended to two periods, regarding vehicle size as a new variable. The optimisation of isolated bus corridors has received further attention within the context of elastic demand. Oldfield and Bly (1988) consider congestion, making waiting time dependent on vehicle occupancy. Evans and Morrison (1997) include two new variables, namely accident risk and disruption or non-scheduled delay.

Isolated corridor analysis has played an important role in establishing a rigorous microeconomic approach to public transport analysis. Nevertheless, the spatial dimension is indeed important. As stated by Jansson (1979), public transport analysis has to include a network perspective. Along this line, Kocur and Hendrickson (1982), and Chang and Schonfeld (1991) considered an area served by parallel, equally spaced, bus lines. Both frequency and the spatial separation between lines were considered as optimisation variables. Bus cycle time and passenger travel time were regarded as independent of the number of users, which made it possible to obtain analytical solutions (though approximated). The main result (both articles) is that both the optimal interval between buses and the optimal spacing between lines are inversely proportional to the cubic root of demand. This result shows that when it is possible to act on the bus lines density, the optimal reaction to demand increases has two dimensions: increase lines density and increase bus frequency. Because of this, in this case (cubic root) optimal frequency grows less than in the isolated corridor case (square root).

When analysing single bus lines, all origin-destination (OD) pairs are included within that line, such that no transfers are needed. This is a big limitation indeed if one wants to explore the elements that determine the optimal spatial structure of a bus system serving a network in which passengers travel between many points in a two-dimensional space. But introducing trips with different origins and destinations within an area poses a great analytical difficulty, because bus lines can be organised in many different ways. One possible option to serve a demand pattern with non-aligned OD pairs is to design a set of bus corridors such that users can make the necessary transfers to reach the corresponding destinations. But it is also possible to design a set of routes that follow closely the spatial pattern of demand, with direct bus lines serving the main OD pairs, causing some overlapping. When demand is relatively low, such a "full coverage" may be neither in the interest of the bus company nor in that of the passengers because of the low frequencies that would very probably result. However, if demand is very large it may well happen that direct services can operate with sufficiently high frequencies and avoiding transfer time. This is of common occurrence in most Latin American capital cities, where public transport demand is very high and direct services have been observed to operate with fairly dense and connected networks.[1] Thus, in the case of a set of corridors, transfers induce a cost for the users, while direct services would reduce waiting, walking, and in vehicle times because of fewer transfers, but more buses could be required. It is then unclear which type of lines structure, direct or in corridors, is associated with the minimum total cost.

The objective of this paper is twofold: first, to depart from the single line analysis in order to explore the Mohring–Jansson approach for relatively simple OD network structures; second, to discuss the potential advantages or disadvantages of a transfer-based transit system against direct services. The idea is to gain insight on the type of solutions for the optimal transit service spatial patterns, including the optimisation rules

[1]According to the last OD survey (2001) in Santiago, Chile, there are more than 4.3 million daily bus trips. They are served mostly with direct lines. During the peak hour, frequency is high, coverage is full, and many of the buses circulate at full capacity. The total fleet is 8,000 vehicles, organised in 350 lines with a typical average frequency of 8 buses per hour.

relating frequencies and fleet size. We will show that using the aforementioned approach it is possible to compare overall resources consumption (operators and users) for different line structures. To achieve this, we will start in section two with Jansson's model (1980, 1984) for a single corridor, and we will extend it in sections three and four to simple networks involving non-aligned OD structures and alternative spatial organisations of bus services. One of the main results is that optimal frequency keeps a square-root-like solution, involving case specific parameters that depend on relative time values, among other things. Most important, the microeconomic model permits the identification of the most efficient spatial transit pattern among those analysed. Finally, neither direct services nor transit corridors appear as a systematically superior alternative. The optimum depends, among other things, on the level of demand as expected.

The Single Line Model

Following Jansson (1980, 1984), let us consider a corridor served by one circular bus line of L kilometers long, operating at a frequency f with a fleet of B vehicles. This service is used by a total of Y passengers per hour, homogeneously distributed along the corridor where each travels a distance l. If T denotes time in motion of a vehicle within a cycle, and t is boarding or alighting time per passenger, then cycle time t_c is

$$t_c = T + 2t\frac{Y}{f}. \tag{1}$$

On the other hand, frequency is given by the ratio between fleet size and cycle time (B/t_c), which combined with (1) yields

$$B = fT + 2tY. \tag{2}$$

If c is the cost per bus-hour for the operator, and P_w and P_v are the values of waiting and travel time respectively, then the total value of the resources consumed (VRC) per hour is

$$VRC = Bc + P_w\frac{Y}{2f} + P_v\frac{l}{L}t_cY, \tag{3}$$

where the first term of the right-hand side of equation (3) corresponds to the operator expenses, and the second and third are users' waiting and travel time value respectively. Note that access time is not included in VRC

because route design is not a variable and, therefore, access cost is a constant that is not relevant to optimise the service. Note also that c has been assumed independent of vehicle size, a variable that is ignored in this formulation.

Using equations (1) and (2), we can write expression (3) as a function of B, that is,

$$VRC = Bc + P_w \frac{T}{2(B - 2tY)} Y + P_v \frac{l}{L} \left(T + \frac{2tTY}{B - 2tY} \right) Y. \tag{4}$$

This expression shows that, *ceteris paribus*, increasing the number of vehicles diminishes users' costs but increases operators' costs. Users' cost reduction occurs because increasing frequency diminishes both waiting and in-vehicle travel times, the latter because fewer individuals board and alight per bus.

Minimising VRC with respect to B yields the optimal fleet size B^*, given by

$$B^* = 2tY + \sqrt{\frac{TY}{c} \left(\frac{1}{2} P_w + P_v \; 2tY \frac{l}{L} \right)}, \tag{5}$$

which from (2), yields the optimal frequency

$$f^* = \sqrt{\frac{Y}{cT} \left(\frac{1}{2} P_w + P_v 2tY \frac{l}{L} \right)}, \tag{6}$$

known as the "square root formula" (Jansson, 1980, 1984). According to this result, optimal frequency increases proportionally to the square root of total demand if the second term in parenthesis is negligible relative to the first, but it can vary proportionally to demand if the contrary happens.

Finally, replacing optimal fleet size from (5) into (4), the minimum of VRC is obtained, that is, the cost function C:

$$C = 2tcY + 2\sqrt{cTY \left(\frac{P_w}{2} + P_v \; 2tY \frac{l}{L} \right)} + P_v TY \frac{l}{L}. \tag{7}$$

It is interesting to note that operators' cost corresponds to the first term plus half of the second (the square root), and the users' cost includes the square root plus the third term.

Extension to a Network

In this section we show how to use the modelling approach presented above to face the problem proposed and formulated in the introduction. Now we will deal with a network setting in which known passenger flows travel between non-aligned OD pairs. They have to be served with a bus fleet that can be spatially organised in different ways. Therefore, we have to compare total cost (operators and users) for each line structure.

In isolated corridor models like the one synthesised in the previous section, total resources consumed are optimised with respect to the number of vehicles (B). Now we have OD flows in a network, and it is necessary first to assign optimally a parametrically given fleet size among lines, in order to optimise B as a second step. This can be done for different line structures, finding the corresponding minimum total cost and then comparing across structures. Thus, the problem can be formulated in two stages that have to be solved for each lines structure. These structures will be assumed to cover the same space (links) such that passenger access time is constant across designs (transfers will be assumed to increase waiting time only). Thus, regarding users' costs, only waiting and in-vehicle time are relevant for optimisation and comparison.

In the **first stage** we consider a given fleet size B of vehicles that have to be distributed in quantities B_i among the different lines. Waiting times (t_{wj}) and travel times (t_{vj}) corresponding to each O-D pair j can be calculated as a function of the number of vehicles assigned to each line as well as other parameters of the problem, that is,

$$t_{wj} = t_{wj}(\vec{B}), \qquad t_{vj} = t_{vj}(\vec{B}), \tag{8}$$

where $\vec{B}$ is the vector of fleets B_i. From (8), average travel and waiting times (over all OD pairs) can be calculated as a function of all B_i and other parameters.

$$\bar{t}_w = \frac{\sum_j t_{wj}(\vec{B})Y_j}{Y} = \bar{t}_w(\vec{B}), \quad \bar{t}_v = \frac{\sum_j t_{vj}(\vec{B})Y_j}{Y} = \bar{t}_v(\vec{B}), \tag{9}$$

where Y_j is flow (passengers/hour) on OD pair j and Y is total flow. Then, taking into account restriction $B = \sum_i B_i$, we can minimise VRC over all B_i with VRC given by

$$VRC = cB + P_w\bar{t}_w(\vec{B})Y + P_v\bar{t}_v(\vec{B})Y, \tag{10}$$

with c, P_w and P_v defined as above. As B is exogenous on this stage, we obtain

$$B_i^* = B_i^*(B). \tag{11}$$

Replacing (11) in (9), we obtain the optimal waiting and travel times as a function of B, that is,

$$\bar{t}_w^* = t_w[B^*(\vec{B})] = \bar{t}_w^*(B), \qquad \bar{t}_v^* = t_v[\vec{B}^*(B)] = \bar{t}_v^*(B). \tag{12}$$

For the **second stage**, these results are replaced in (10), obtaining VRC as a function of B:

$$VRC = cB + P_w\bar{t}_w^*(B)Y + P_v\bar{t}_v^*(B)Y. \tag{13}$$

Minimising with respect to B yields the optimal number of vehicles $B^*(c, P_w, P_v, Y)$. Finally, replacing the result in (13) the minimum cost C_l for each line structure can be obtained:

$$\begin{aligned} C_l &= cB^*(c, P_w, P_v, Y) + P_w\bar{t}_w^*[B^*(c, P_w, P_v, Y)]Y + P_v\bar{t}_v^*[B^*(c, P_w, P_v, Y)]Y \\ &= C_l(c, P_w, P_v, Y) \end{aligned} \tag{14}$$

from which the structure with minimum overall cost can be identified.

Application

The cases

As mentioned above, a given network and OD flow pattern can be served with different line structures. Each structure implies different amounts of resources consumed. In what follows, we will name **direct lines** a structure of bus services in which there is at least one line serving each OD pair directly, that is, a system in which transfers are unnecessary. Note that this structure does not imply as many lines as OD pairs, because one line can serve many aligned OD pairs. On the other hand, we will name **corridor lines** a structure of bus services in which lines do not overlap. Thus, if the origin lies along one corridor and the destination lies along a different one, a transfer is mandatory. As is evident, these two structures just defined are extreme cases, which will be shown to be helpful for analysis and discussion.

In this section we will define some OD structures and networks that could be taken as aggregate representations of a street network serving urban zones. In so doing, we will privilege those assumptions that permit a clean analytical solution. Figure 1 shows a simple network with four arcs (*a*-*b*, *b*-*d*, *e*-*b*, *b*-*c*) on which passengers have to travel from different origins to different destinations.

The three cases present increasingly complex OD structures where the total number of passengers/hour that enters the system is Y, which is equally distributed across OD pairs. Thus, there are $Y/2$ passengers/hour on each OD pair in case one, $Y/4$ in case two and $Y/8$ in case three. Note that this demand distribution keeps the spirit of the single line cases, where a total demand Y results from passengers that board the system in different places but with the same trip length, implying the same number of passengers on each OD pair. For each case, direct and corridor line structures are also shown.[2] The relevant efficiency question is which structure is better from the viewpoint of total resources consumption (users and operators). To answer this, we will construct the cost function for each structure in each case, applying the two stages method introduced in the previous section, following Jansson's (1980, 1984) approach. Details will be shown only for the first network, as the remaining algebraic derivations are similar.

First stage: fleet distribution

Let T_0 be the round trip time between consecutive nodes required by a vehicle (not including boarding and alighting), and let t be the time each passenger has to assign to either board or alight. Let f_i be service frequency of line i. As discussed previously, in the first stage we have to distribute B among the lines in order to calculate the average travel and waiting times as a function of B. Let us analyse the first case.

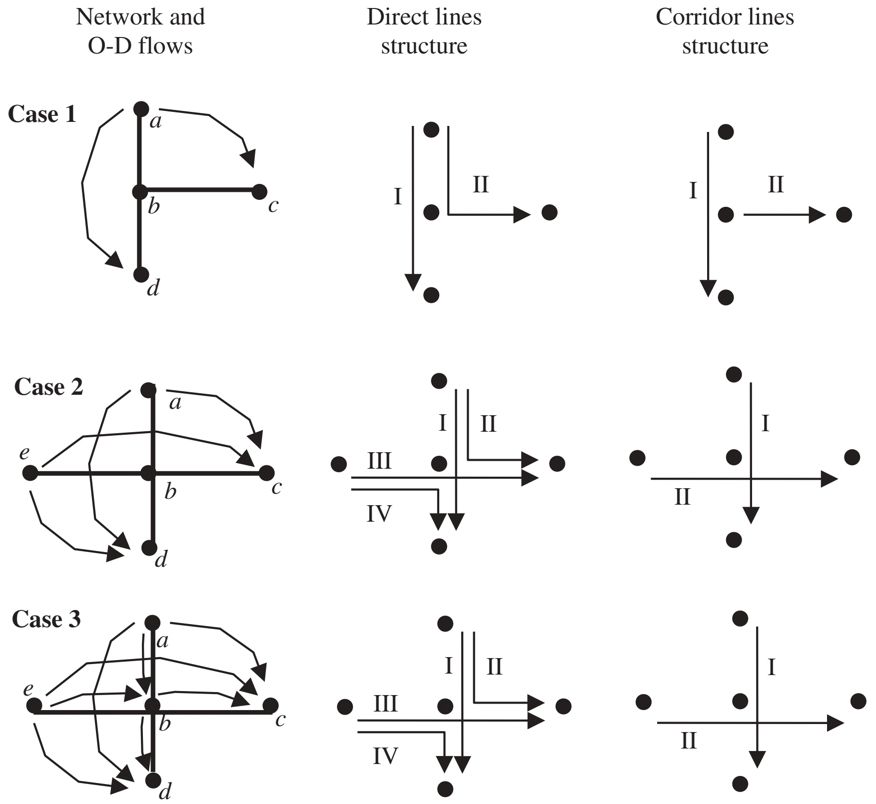
In **Case 1** (which could be looked at as a centre-to-periphery case), the **direct lines structure** presents a symmetric case for the assignment of demand to lines. The optimal distribution of the fleet is to assign $B/2$ buses to each line. Cycle time on each line is then given by

$$t_{ci} = 2T_0 + 2t\frac{Y/2}{f_i}, \tag{15}$$

where $2T_0$ corresponds to the vehicle cycle time with no passengers boarding or alighting. In the second term, the ratio between $Y/2$ (total number of passengers/hour) and the line frequency yields the passengers that board and alight each vehicle during a cycle. Therefore, this number times $2t$ is the total time each vehicle remains at the bus stops on each cycle.

[2]Note that in all three cases cyclical single line services passing through node b many times could accommodate all OD flows, but cannot be taken as cases of interest in this network context.

Figure 1
Direct and Corridor Lines Structures



On the other hand, frequency is also the ratio between the number of vehicles and cycle time. As we have $B/2$ vehicles on each line, using (15) we obtain

$$f_i = \frac{B/2}{2T_0 + 2t(Y/2/f_i)}, \tag{16}$$

from which

$$f_i = f_{\mathrm{I}} = f_{\mathrm{II}} = \frac{B - 2tY}{4T_0}. \tag{17}$$

As frequency is the same for both lines, passengers in both OD pairs will experience the same average waiting time, given by half the interval between vehicles (the inverse of frequency). Thus, the average waiting time for each user, as a function of B and other parameters, is

$$\bar{t}_w^* = \frac{1}{2f_i} = \frac{2T_0}{B - 2tY}. \tag{18}$$

In-vehicle time is also equal for all users, given by

$$\bar{t}_v = T_0 + \frac{t}{2}\frac{Y/2}{f_i}. \tag{19}$$

The first term corresponds to vehicle in motion, that is, the time it takes to travel along the two consecutive arcs. The second term is the time the passenger has to wait on average while others alight at the destination, taken as half the total alighting time. Note that waiting after boarding should not be included because this was already accounted for in total waiting time. Replacing frequency from (17) in (19) we get the average travel time as a function of B and other parameters:

$$\bar{t}_v^* = T_0 + \frac{T_0 t Y}{B - 2tY}. \tag{20}$$

For the **corridor lines structure** in this same Case 1, demand assignment is no longer symmetric and the trivial distribution of vehicles disappears. Cycle times for each line now are

$$t_{c\mathrm{I}} = 2T_0 + 2t\frac{Y}{f_{\mathrm{I}}}, \quad t_{c\mathrm{II}} = T_0 + 2t\frac{Y/2}{f_{\mathrm{II}}}, \tag{21}$$

respectively, because corridor I is twice as long as corridor II, and all passengers have to board at a, but half of them have to transfer at b. As frequency is the ratio between the number of vehicles and cycle time, from (21), the following expressions obtain for frequency on each line as a function of the corresponding number of vehicles

$$f_{\mathrm{I}} = \frac{B_{\mathrm{I}} - 2tY}{2T_0}, \qquad f_{\mathrm{II}} = \frac{B_{\mathrm{II}} - tY}{T_0}. \tag{22}$$

For each OD pair (a-d and a-c) average waiting times are

$$t_{w(a\text{-}d)} = \frac{1}{2f_{\mathrm{I}}}, \qquad t_{w(a\text{-}c)} = \frac{1}{2f_{\mathrm{I}}} + \frac{1}{2f_{\mathrm{II}}}. \tag{23}$$

The second term for trips on the a-c pair is due to the necessary transfer in b. As both O-D pairs have the same number of passengers by construction, then the average waiting time as a function of frequencies is

$$\bar{t}_w = \frac{1}{2}\left[\frac{1}{f_{\mathrm{I}}} + \frac{1}{2f_{\mathrm{II}}}\right]. \tag{24}$$

On the other hand, average in-vehicle travel time for each pair is

$$t_{v(a-d)} = T_0 + t\frac{Y/2}{f_{\mathrm{I}}} + \frac{t}{2}\frac{Y/2}{f_{\mathrm{I}}} = T_0 + \frac{3}{4}\frac{tY}{f_{\mathrm{I}}},$$

$$t_{v(a-c)} = T_0 + \frac{t}{2}\frac{Y/2}{f_{\mathrm{I}}} + \frac{t}{2}\frac{Y/2}{f_{\mathrm{II}}} = T_0 + \frac{tY}{4}\left[\frac{1}{f_{\mathrm{I}}} + \frac{1}{f_{\mathrm{II}}}\right]. \tag{25}$$

In both cases, the third term of the expression in the middle corresponds to the average time a passenger has to stay in the vehicle while other passengers alight at their destination. The second term represents in-vehicle time while others transship at *b*. For those that travel between *a* and *d*, this corresponds to the whole alighting time of those that change vehicle in order to go to *c*, while the former remain in the vehicle. For those that go to *c*, instead, in-vehicle time at the interchange is half the alighting time because they are part of the group. Then, average in-vehicle travel time is

$$\bar{t}_v = \frac{1}{2}\left[2T_0 + \frac{tY}{f_{\mathrm{I}}} + \frac{tY}{4f_{\mathrm{II}}}\right]. \tag{26}$$

By replacing equations (22) in (24) and (26) the overall average waiting and travel times can be calculated as a function of the number of vehicles assigned to each line. But $B_{\mathrm{II}} = B - B_{\mathrm{I}}$ and, therefore, average times can be expressed as a function of B_{I} only:

$$\begin{aligned} \bar{t}_w &= \frac{1}{2}\left[\frac{2T_0}{B_{\mathrm{I}} - 2tY} + \frac{T_0}{2(B - B_{\mathrm{I}} - tY)}\right], \\ \bar{t}_v &= \frac{1}{2}\left[2T_0 + \frac{2T_0tY}{B_{\mathrm{I}} - 2tY} + \frac{T_0tY}{4(B - B_{\mathrm{I}} - tY)}\right]. \end{aligned} \tag{27}$$

Replacing these results into the expenditure expression (10) and minimising over B_{I}, the optimal fleet distribution is obtained as

$$\begin{aligned} B_{\mathrm{I}}^* &= \frac{\sqrt{\frac{P_w}{P_v} + tY}}{\frac{1}{2}\sqrt{\frac{P_w}{P_v} + \frac{tY}{2}} + \sqrt{\frac{P_w}{P_v} + tY}}\,B + \frac{tY\left[\sqrt{\frac{P_w}{P_v} + \frac{tY}{2}} - \sqrt{\frac{P_w}{P_v} + tY}\right]}{\frac{1}{2}\sqrt{\frac{P_w}{P_v} + \frac{tY}{2}} + \sqrt{\frac{P_w}{P_v} + tY}} \\ &= \alpha B + \gamma, \end{aligned} \tag{28}$$

where the parameters α and γ depend on the ratio between waiting and travel time prices, and also on the product tY. The parameters have a peculiar expression, such that if the square roots values were similar, γ would vanish and α would be 2/3. This suggests a numerical sensitivity analysis using actual values for the variables involved. After a number of experiments (see appendix), we concluded that exact calculations are all in the neighbourhood of

$$B_{\mathrm{I}}^* \approx \tfrac{2}{3}B, \qquad B_{\mathrm{II}}^* \approx \tfrac{1}{3}B, \tag{29}$$

which we adopted for comparison between line structures. Replacing these values in (27) we obtain the overall waiting and travel times as a function of B:

$$\bar{t}_w^* = \frac{\frac{3}{2}T_0}{\frac{2}{3}B - 2tY}, \qquad \bar{t}_v^* = T_0 + \frac{\frac{5}{4}T_0 tY}{\frac{2}{3}B - 2tY}. \tag{30}$$

Let us now see **Case 2**, in which both structures originate completely symmetric systems, as shown in the figure. This makes the first stage very simple, and fleet distribution is direct. Following similar procedures as for Case 1, we obtain closed expressions for overall average waiting and travel times. For the **direct lines structure** the results are

$$\bar{t}_w^* = \frac{4T_0}{B - 2tY}, \qquad \bar{t}_v^* = T_0 + \frac{T_0 tY}{B - 2tY}, \tag{31}$$

and for the **corridor lines structure**

$$\bar{t}_w^* = \frac{2T_0}{\frac{2}{3}B - 2tY}, \qquad \bar{t}_v^* = T_0 + \frac{\frac{3}{2}T_0 tY}{\frac{2}{3}B - 2tY}. \tag{32}$$

Case 3 is characterised by trips originating and ending at the intersection b of corridors a-d and e-c. Again, both lines structures generate symmetric systems, which permits the derivation of closed expressions for the average waiting and travel times. For the **direct lines structure** we obtain:

$$\bar{t}_w^* = \frac{3T_0}{B - 2tY}, \qquad \bar{t}_v^* = \tfrac{3}{4}T_0 + \frac{\frac{9}{8}T_0 tY}{B - 2tY}, \tag{33}$$

Table 1
Waiting and Travel Time Parameters

Case	*Structure*	Ψ	δ	φ_w	φ_v	$1/\delta$	φ_w/δ	$\varphi_v\delta$
1	Direct	1	1	2	1	1	2	1
1	Corridors	1	2/3	3/2	5/4	1.5	2.25	1.875
2	Direct	1	1	4	1	1	4	1
2	Corridors	1	2/3	2	3/2	1.5	3	2.25
3	Direct	3/4	1	3	9/8	1	3	1.125
3	Corridors	3/4	4/5	2	3/2	1.25	2.5	1.875

and for the **corridor lines structure**

$$\bar{t}_w^* = \frac{2T_0}{\frac{4}{5}B - 2tY}, \qquad \bar{t}_v^* = \frac{3}{4}T_0 + \frac{\frac{3}{2}T_0 tY}{\frac{4}{5}B - 2tY}. \tag{34}$$

The results for all cases and line structures, synthesised by equations (18), (20), (30–34), show that overall average waiting and travel times have the same form:

$$\bar{t}_w^* = \frac{\varphi_w T_0}{\delta B - 2tY}, \qquad \bar{t}_v^* = \psi T_0 + \frac{\varphi_v T_0 tY}{\delta B - 2tY}, \tag{35}$$

where the parameters δ, Ψ, φ_w and φ_v take different values for each case. These are shown in Table 1 where the values for the corridor lines structure in Case 1 are approximated (equations (28) to (30)). Note that Ψ is the only parameter that has the same value for both structures in each case.

Second stage: fleet optimisation and total costs comparison

The general forms found for average travel and waiting times in the first stage, allows us to obtain a general expression for VRC in (13) by simply replacing (35)

$$VRC = cB + P_w \frac{\varphi_w T_0}{\delta B - 2tY} Y + P_v \left[\Psi T_0 + \frac{\varphi_v T_0 tY}{\delta B - 2tY} \right] Y. \tag{36}$$

This can be minimised with respect to B in order to find the optimal number of vehicles, which happens to depend on the parameters that represent line structure, that is,

$$B^* = 2tY\frac{1}{\delta} + \sqrt{\frac{T_0 Y}{c}\left(P_w \frac{\varphi_w}{\delta} + P_v \frac{\varphi_v}{\delta} tY\right)}. \tag{37}$$

Finally, replacing B^* in (36) we obtain a general expression for the minimum VRC for each line structure l:

$$C_l = 2ctY\frac{1}{\delta^l} + 2\sqrt{cT_0 Y\left(P_w \frac{\varphi_w^l}{\delta^l} + P_v \frac{\varphi_v^l}{\delta^l} tY\right)} + P_v \Psi T_0 Y. \tag{38}$$

Note that both B^* and C_l follow the form obtained by Jansson (1980) in equations (5) and (7) for an isolated corridor. The "square root formula" survives a network oriented, line structure analysis. This is an interesting departure from the single line case.

Now we have the elements to face the second challenge, namely which line structure is most appropriate to serve a spatially segregated demand. When comparing total cost C_l for each line structure for each case, the last term in (38) cancels out, as Ψ is equal for both structures. This corresponds to the cost of "in motion" travel time. The other two terms depend upon $1/\delta$ and on the ratios φ_w/δ and φ_v/δ, whose values are shown in Table 1. In **Case 1**, these three terms are lower for the direct lines structure, which means that it is this structure with no transfers that is most convenient for such a network and such a demand pattern. The comparison is less clear for **Cases 2** and **3**, in which both $1/\delta$ and φ_v/δ are lower for the direct lines structure, while φ_w/δ is larger. Therefore, in both cases we need to know the values of the rest of the parameters in order to choose the most efficient lines structure.

For a general analysis of the three cases, it is convenient to note that the first term in (38) is practically negligible compared to the second (see appendix). It is the square root term that dominates for comparison. Therefore, **the condition in all three cases for the direct lines structure to be the most efficient is**

$$2\sqrt{cT_0 Y\left(P_w \frac{\varphi_w^D}{\delta^D} + P_v \frac{\varphi_v^D}{\delta^D} tY\right)} < 2\sqrt{cT_0 Y\left(P_w \frac{\varphi_w^C}{\delta^C} + P_v \frac{\varphi_v^C}{\delta^C} tY\right)}, \qquad (39)$$

where superscripts D and C stand for direct and corridor respectively. Some algebraic manipulation turns the condition into

$$P_w\left(\frac{\varphi_w^D}{\delta^D} - \frac{\varphi_w^C}{\delta^C}\right) < P_v tY\left(\frac{\varphi_v^C}{\delta^C} - \frac{\varphi_v^D}{\delta^D}\right). \qquad (40)$$

If the left-hand side parenthesis is positive, it can divide the right-hand side without changing the sign of the inequality. The sign will change if that term is negative. Therefore, the conditions for the direct line structure to be more efficient than corridors are

$$\frac{P_w}{P_v tY} < \left(\frac{\varphi_v^C}{\delta^C} - \frac{\varphi_v^D}{\delta^D}\right) \Big/ \left(\frac{\varphi_w^D}{\delta^D} - \frac{\varphi_w^C}{\delta^C}\right) \quad \text{if } \frac{\varphi_w^D}{\delta^D} > \frac{\varphi_w^C}{\delta^C}, \qquad (41)$$

$$\frac{P_w}{P_v tY} > \left(\frac{\varphi_v^C}{\delta^C} - \frac{\varphi_v^D}{\delta^D}\right) \Big/ \left(\frac{\varphi_w^D}{\delta^D} - \frac{\varphi_w^C}{\delta^C}\right) \quad \text{if } \frac{\varphi_w^D}{\delta^D} < \frac{\varphi_w^C}{\delta^C}. \qquad (42)$$

This particular form of comparison is attractive because the left-hand side involves demand related information, while the right-hand side expression has only lines structure specific parameters. For Case 1, Table 1 shows that the ratio between φ_w and δ is larger for the corridors, which makes condition (42) the relevant one; replacing all parameter values a negative result is obtained for the right-hand ratio, which is always lower than $P_w/P_v tY$ (positive). This corroborates the result previously obtained: direct lines should be preferred.

On the other hand, for Cases 2 and 3 condition (41) holds, because φ_w/δ is larger for direct lines. Replacing the values of the right-hand side ratios from Table 1, we obtain 1.25 and 1.5, for Cases 2 and 3 respectively. In each case, this value could be larger or smaller than the left-hand side term. Thus, the direct lines structure is more likely to be the best the larger the product tY (as expected) and the smaller the ratio between P_w and P_v (that is, the smaller the relative value of waiting time). Note that the difference between Cases 2 and 3 is the presence of a destination in the node where corridors intersect in the latter case. Thus, everything else constant, the right-hand side values show that it is (marginally) less likely that direct lines are the answer in Case 2 when compared with Case 3. A similar analysis can be done regarding fleet size (equation (37)), which yields the same conclusions.

Conclusions

Optimising transit services poses many questions: fleet size, lines frequency, route structure, and vehicle capacity, are among the many variables involved. The literature has emphasised frequency and fleet size, based upon single bus lines analysis. However, urban demand structure presents many non-aligned OD pairs even at an aggregated level, and this introduces new aspects to understand and discuss. One such aspect is the spatial organisation of the transit fleet, which can follow demand patterns by means of direct services, or can rely upon passenger transfers among services organised in corridors, to mention two extremes only as other structures are possible. We have analysed simple OD structures in order to expand the single line microeconomic framework and to investigate the elements that determine the possible relative advantages of corridors against direct services. The most interesting result is that the classical square root expression for optimal frequency survives the expansion of the single line case to bus networks.

Our results show that optimal transit patterns depend upon many aspects, some related to the relative values of waiting and in-vehicle time, and some related to the particular spatial shape of demand and its level. Although it seems that there is no universal answer, large demand values tend to favour direct services. It is important to emphasise that the advantage of the direct lines structure is, in principle, a matter of avoiding transfers, diminishing users' costs (transfer time for some, travel time for others). On the other hand, corridors let users board any arriving vehicle. If total fleet size was constant, this type of analysis would be sufficient. However, fleet size is a variable whose optimal number depends on the lines structure, as shown. What would be the best spatial structure of services if users' costs were not taken into account? Clearly, in that case direct services would never be an undoubtedly superior solution (at most, a tie). On the other hand, the analysis presented here has treated demand parametrically, but users are sensitive to transfers indeed, which makes the corridor structure sensitive to other type of variables such as climate, bus stations design, pricing policy, and so on. Elastic demand analysis might favour direct lines slightly, but this is something that should be investigated further.

It has not been our objective to replicate actual urban public transport networks. Effective, sophisticated and huge public transport models have been built with that purpose. We have extended the microeconomic framework for the one line case in order to understand the economic underpinnings of bus services in a spatial setting. It has been proved to be a useful approach to verify that the basic conclusions of the single line case do survive the spatial expansion, which has other economic implications (for example optimal pricing; see Jansson, 1984). Can these results be generalised? We believe that this could be empirically explored, just as other economic models have helped understanding of the basics of production and consumption, setting the foundations of demand or industry structure studies. An alternative would be to use large scale models to analyse the relation between optimal fleet size and both the level and pattern of demand. The normative analysis of public transport systems will gain enormously from this potential verification.

References

Chang, S. K. and P. M. Schonfeld, (1991): "Multiple Period Optimization of Bus Transit Systems." *Transportation Research*, 25B, 453–78.

Evans, A. W. and A. D. Morrison, (1997): "Incorporating Accident Risk and Disruption in Economic Models of Public Transport." *Journal of Transport Economics and Policy*, 31, 117–46.

Jansson, J. O. (1979): "Marginal Cost Pricing of Scheduled Transport Services." *Journal of Transport Economics and Policy*, 13, 268–94.

Jansson, J. O. (1980): "A Simple Bus Line Model for Optimization of Service Frequency and Bus Size." *Journal of Transport Economics and Policy*, 14, 53–80.
Jansson, J. O. (1984): *Transport System Optimization and Pricing*. John Wiley & Sons.
Jara-Díaz, S. R. and A. Gschwender (1997): "Tarifas óptimas en transporte público programado (Optimal fares in scheduled public transport)." *Actas del VIII Congreso Chileno de Ingeniería de Transporte*, 265–78.
Kocur, G. and C. Hendrickson, (1982): "Design of Local Bus Service with Demand Equilibration." *Transportation Science*, 16, 149–70.
Mohring, H. (1972): "Optimization and Scale Economies in Urban Bus Transportation." *American Economic Review*, 62, 591–604.
Mohring, H. (1976): *Transportation Economics*. Cambridge, Mass.: Ballinger.
Oldfield, R. H. and P. H. Bly, (1988): "An Analytic Investigation of Optimal Bus Size". *Transportation Research*, 22B, 319–37.
Vickrey, W. (1955): "Some Implications of Marginal Cost Pricing for Public Utilities." *American Economic Review*, 45, 605–20.

Appendix

Exact Calculation of α and γ in equation (28), with $t = 2.5$ Seconds.

Y [*pass/hr*]	P_w/P_v	α	γ
2,000	3	0.686	−0.079
2,000	1	0.704	−0.154
1,000	3	0.678	−0.023
1,000	1	0.692	−0.052
200	3	0.669	−0.001
200	1	0.674	−0.003

If $B = 30$ vehicles, the calculation of B_I with the exact values of α and γ differs in less than 1 vehicle form the approximation with $\alpha = 2/3$ and $\gamma = 0$.

Numerical Comparison of Both First Terms of the Cost Function (38).

T_0 [*hr*]	Y [*pass/hr*]	P_v [*US$/hr*]	P_w [*US$/hr*]	x [*US$/hr*]	z [*US$/hr*]	z/x
3	2,000	4	12	24.8	2,517.4	102
3	2,000	2	6	24.8	1,780.1	72
1.5	1,000	2	6	12.4	847.2	68
0.5	500	2	6	6.2	336.8	54
0.5	250	2	6	3.1	234.9	76
0.2	200	2	6	2.5	132.5	53

Calculations were made considering $t = 2.5$ seconds, $c = 8.9$ US$/hr (Jara-Díaz and Gschwender, 1997), $\delta = 1$, $\Psi = 1$, $\varphi_w = 2$ and $\varphi_v = 1$; x corresponds to the first term in the cost function (38) and z to the second one (with the square root).