Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding

Alejandro Tirachini a,⇑, David A. Hensher b,1, John M. Rose b,1

a Transport Engineering Division, Civil Engineering Department, Universidad de Chile, Santiago, Chile
b Institute of Transport and Logistics Studies (ITLS), The University of Sydney Business School, The University of Sydney, NSW 2006, Australia

A R T I C L E   I N F O

Article history:
Received 18 June 2013
Received in revised form 15 January 2014
Accepted 15 January 2014

Keywords:
Bus design
Congestion
Crowding
Fare
Quality of service
Walking

A B S T R A C T

The interplay between congestion and crowding externalities in the design of urban bus systems is identified and analysed. A multimodal social welfare maximisation model with spatially disaggregated demand is developed, in which users choose between travelling by bus, car or walking in a transport corridor. Optimisation variables are bus fare, congestion toll, bus frequency, bus size, fare collection system, bus boarding policy and the number of seats inside buses. We find that optimal bus frequency results from a trade-off between the level of congestion inside buses, i.e., passengers’ crowding, and the level of congestion outside buses, i.e., the effect of frequency on slowing down both buses and cars in mixed-traffic roads. A numerical application shows that optimal frequency is quite sensitive to the assumptions on crowding costs, impact of buses on traffic congestion, and overall congestion level. If crowding matters to users, buses should have as many seats as possible, up to a minimum area that must be left free of seats. If for any other reason planners decide to have buses with fewer seats than optimal (e.g., to increase bus capacity), frequency should be increased to compensate for the discomfort imposed on public transport users. Finally, the consideration of crowding externalities (on both seating and standing) imposes a sizeable increase in the optimal bus fare, and consequently, a reduction of the optimal bus subsidy.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

When deciding whether or not to undertake a trip by public transport, travellers are influenced by a number of characteristics or attributes of the public transport mode, including accessibility, waiting time, travel time, price, reliability, comfort and safety. Demand is sensitive to the overall quality of service, which in turn depends on the design of the system; hence understanding the economic nature of urban public transport operations is crucial as a means to ensure the efficiency of a public transport network and, ultimately, the sustainability of the entire transport system. From a transport planner’s perspective, the challenge associated with the design of public transport services lays in the myriad number of trade-offs that need to be considered at once, in order to establish an optimal service design. For example: increasing bus frequency reduces waiting time for users but increases the cost of operation; increasing the number of bus stops reduces users’ access...

⇑ Corresponding author. Tel.: +56 2 2978 4380.
E-mail addresses: alejandro.tirachini@ing.uchile.cl (A. Tirachini), david.hensher@sydney.edu.au (D.A. Hensher), john.rose@sydney.edu.au (J.M. Rose).
1 Tel.: +61 2 9351 0169.
time but increases bus riding time for all-stop services; and investing in a quicker fare collection technology and dedicated road infrastructure for buses reduces bus travel time (and consequently may reduce operating cost) but increases capital cost. In this paper we introduce two trade-offs within the microeconomic modelling framework of optimal supply levels and pricing of an urban bus route, that are crucial to the level of service offered to public transport users: (i) the interaction between road congestion and passenger crowding externalities when setting supply levels of public transport, and (ii) the decision on the number of seats that public transport vehicles should have.

First, the influences of congestion and crowding on optimal bus service frequency have been treated independently in the literature. On the one hand, Jara-Díaz and Gschwender (2003) show that passenger crowding externalities push optimal public transport frequency up, with a total cost minimisation model without road congestion. On the other hand, Tirachini and Hensher (2011) find that the existence of bus congestion in the form of queuing delays at bus stops pushes optimal frequency down, with a total cost minimisation model that ignores crowding externalities. Moreover, buses may also slow cars down in shared roads. Therefore, there are possible counter-effects of congestion and crowding on optimal frequency that need to be addressed simultaneously. In this paper we propose a multimodal social welfare maximisation model that includes both passenger crowding and mixed-traffic congestion externalities, and find that optimal bus frequency is the result of crowding and congestion acting as colliding forces.

Second, going beyond microeconomic models that optimise public transport frequency and/or vehicle size to set supply levels (e.g., Mohring, 1972; Jansson, 1980; Oldfield and Bly, 1988; Chang and Schonfeld, 1991), we look at the internal design of vehicles by including the number of bus seats as a decision variable that influences both comfort and capacity. Different configurations of vehicles regarding number of seats and space for standees are relevant for the level of crowding and standing externalities in public transport (Whelan and Crockett, 2009). This is a key insight from the estimation of crowding and standing disutilities that has not been given attention in the literature on the design and optimisation of public transport systems. Microeconomic models that have included the level of crowding as an influence on the value of in-vehicle time savings do not distinguish between passengers sitting and standing (Jara-Díaz and Gschwender, 2003; Tirachini et al., 2010); whereas Kraus (1991) applies a premium on the value of travel time savings for passengers standing, but his research is concerned with the marginal cost and pricing of services considering the discomfort of standing, rather than with the design of vehicles. Thus, even though crowding and discomfort externalities have been analysed in the literature, previous studies always assumed a given internal design or layout of the vehicles involved, i.e., a given bus or train capacity. In short, it is assumed that size implies capacity. However, reducing the number of seats increases bus capacity by allowing more standees; thus a seat implies a trade-off between comfort and capacity that is allowed for in our model.

The analytical approach consists of a social welfare maximisation model with disaggregated origin and destination demand, and multiple travel alternatives. The model is applied to a single transport corridor in Sydney, Australia, which allows us to obtain detailed measures of crowding levels section by section. In contrast to other social welfare maximisation models (e.g., De Borger et al., 1996; Proost and Van Dender, 2004; Wichiensin et al., 2007; Ahn, 2009; Parry and Small, 2009; Jansson, 2010; Basso et al., 2011), our approach provides a more comprehensive modelling of the bus mode, including bus frequency, bus size, fare collection system, bus boarding policy, number of bus seats and fare level as decision variables. We also show that the inclusion of a non-motorised mode (walking) as an alternative to choosing bus and car for short trips may have a significant role when the transport system is optimised in highly congested scenarios. Results are discussed for several scenarios with different demand levels and modelling assumptions.

We show that optimal bus frequency is quite sensitive to the assumptions regarding crowding costs, impact of buses on traffic congestion and the overall congestion level. In particular, if the planner takes into account that crowding matters to users, our numerical application shows that bus frequency should increase (for a given bus size) with demand even under heavy congestion, however that might not be the case if the crowding externality is not accounted for, in which case an increase of total demand might be met by a decrease of both frequency and number of seats per bus, at the expense of crowding passengers inside buses and making more passengers stand while travelling. Likewise, we find that buses should be designed with as many passenger seats as possible (up to a minimum area that must be left free of seats for an aisle, next to the driver and doors, for a wheelchair and other possible uses). If for any other reason planners decide to have buses with fewer seats than optimal (e.g., to increase bus capacity), frequency should be increased to compensate for the discomfort imposed on public transport users.

The remainder of the paper is organised as follows. The theoretical model is developed in Section 2, including assumptions and definitions (Section 2.1), demand and crowding modelling (Section 2.2), travel time and congestion (Section 2.3), internal bus layout (Section 2.4), and operator cost items (Section 2.5); the section concludes with the formulation of the social welfare maximisation problem. Section 3 presents the numerical application of the model to Sydney and discussion of results in several scenarios. Conclusions are provided in Section 4.

---

2 Excluding non-motorised alternatives, the properties of the bimodal car-bus competition for user equilibrium and/or system optimum solutions are analysed in a number of contributions, e.g., Ahn (2009), Li et al. (2012), and Gonzales and Daganzo (2012, 2013). The influence of an un-congestible non-motorised alternative on the optimal public transport fare is analytically studied by Tirachini and Hensher (2012) with a three-mode model.
2. Model set up

2.1. Assumptions and definitions

We consider a linear bi-directional road of length $L$ and a single period of operation with directions denoted as 1 and 2. The road is divided into $P$ zones denoted as $i \in \{1, \ldots, P\}$, and the total demand $Y_{ij}$ per origin–destination pair $(i, j)$ is fixed. The distance between zone $i$ and zone $i+1$ is denoted as $L_i$ such that $L = \sum_{i=1}^{P-1} L_i$, as shown in Fig. 1. Users can choose to travel by car ($a$), bus ($b$) or to walk ($e$). Then, if $y_{ij}^m$ is the travel demand for mode $m$ between zones $i$ and $j$, it holds that:

$$Y_{ij} = \sum_m y_{ij}^m = y_{ij}^a + y_{ij}^b + y_{ij}^e$$

Let $f_{ia}$ be the traffic flow in section $i$, between zone $i$ and zone $i + 1$ (direction 1) and $f_{i a}^2$ be the traffic flow in section $i$, between zone $i + 1$ and zone $i$ (direction 2). The decision variables of the problem are denoted as follows:

- $f_b$: bus frequency (bus/h)
- $s_b$: bus length (m)
- $A$: fare collection technology and boarding policy (one-door or all-door boarding)
- $n_{\text{seat}}$: number of seats inside a bus
- $s_a$: car toll ($$/trip)
- $s_b$: bus fare ($$/trip)

It is assumed that there is only one bus stop per zone\(^4\) and that the travel distance between zones is the same for the three modes. Bus frequency is assumed to be continuous, whereas options on bus lengths are constrained by the size of commercial vehicles; four sizes are considered in the application of the model. These are mini (8 m long, 1 or 2 doors), standard (12 m long, 2 or 3 doors), rigid long (15 m long, 3 or 4 doors) and articulated (18 m long, 4 doors). The number of seats $n_{\text{seat}}$ can be freely chosen subject to lower and upper bounds, the former is given by a minimum number of seats per bus that is exogenously decided in order to provide a minimum level of service, whereas the latter is determined by a minimum area on a bus that must be clear of seats (i.e., aisle, doors, space for a wheelchair, area next to the driver). Following Tirachini and Hensher (2011), we consider four alternative fare collection technologies: on-board payment with (i) cash, (ii) magnetic strip (with contact) and (iii) smartcard (contactless), plus (iv) off-board payment (on the bus stop). The bus boarding and alighting policy can be chosen as well, two alternatives are available to implement in buses with more than one door: (a) simultaneous boarding and alighting, in which boarding is allowed at the front door only while alighting takes place at the back(s) doors, and (b) sequential boarding and alighting, in which boarding is allowed at all doors giving priority to passengers alighting to unload first. In principle, we assume that cars and buses share the right-of-way and that bus stops do not directly affect cars, an assumption that is revised in Section 3.3.

2.2. Demand modelling and crowding

Previous research has shown the influence of crowding and standing on increasing the value of travel time savings (e.g., Maunsell and Macdonald, 2007; Whelan and Crockett, 2009; Hensher et al., 2011; Wardman and Whelan, 2011). In this paper, mode choice models that include the proportion of available seats and the density of standees as attributes for buses are estimated. Data collected from a stated choice survey conducted in Sydney in 2009 is used to this end; the experimental design, study area, sample size and socioeconomic characteristics of respondents are described at length in Hensher et al. (2011), who estimate the crowding disutility as a function of the proportion of users seating (which affects the probability of getting a seat), and the total number of users standing, in order to estimate the willingness to pay to get a seat as a function of the number of people sitting and standing. In this paper, we use the density of standees per square metre, instead of

---

\(^3\) The model can be easily extended to more travel alternatives such as rail.

\(^4\) The location of bus stops is fixed in this model, which allows us to know the number of passengers that a bus carries in each segment of the route (between two consecutive zones). For a review of models that optimise the spacing of bus stops see Tirachini (2014).
the total number of standees, to represent the disutility of crowding, in order to have a common base for the application of the model with different internal bus layouts regarding allocation of space for seating and standing.

Let \( U_m^{ij} \) be the utility associated with travel by mode \( m \) in origin–destination (OD) pair \((i, j)\). In order to analyse differences in optimal bus service design due to alternative assumptions regarding user’s valuations of seating, standing and crowding levels inside buses, we propose three different models that incorporate attributes representing the number of passengers inside buses, we propose three different models that incorporate attributes representing the number of passengers

\[ U_m^{ij} = \sum \left( \beta_a^{M1} t_{ab}^{ij} + \beta_b^{M1} h_b + \beta_c^{M1} t_{cb}^{ij} + \beta_d^{M1} \right) \]

\[ U_m^{ij} = \sum \left( \beta_a^{M2} t_{ab}^{ij} + \beta_b^{M2} h_b + \beta_c^{M2} t_{cb}^{ij} + \beta_d^{M2} \right) \]

\[ U_m^{ij} = \sum \left( \beta_a^{M3} t_{ab}^{ij} + \beta_b^{M3} h_b + \beta_c^{M3} t_{cb}^{ij} + \beta_d^{M3} \right) \]

\[ U_m^{ij} = \sum \left( \beta_a^{M4} t_{ab}^{ij} + \beta_b^{M4} h_b + \beta_c^{M4} t_{cb}^{ij} + \beta_d^{M4} \right) \]

In (2)–(5), \( t_{ab}^{ij} \) is the access time at zone \( i \), \( h_b \) is the headway between two consecutive buses, \( t_{cb}^{ij} \) is the in-vehicle time between zones \( i \) and \( j \), \( \tau_{bc} \) is the bus fare, \( n_{den} \) is the density of standees per square metre, \( p_{seat} \) is the proportion of seats occupied. For each model M1–M4, the utility of the alternative modes (car and walk) have the same specification:

\[ U_c^{ij} = \beta_a^{M1} t_{ac}^{ij} + \beta_b^{M1} (c_{ij}^{seat} + \tau_{ac})/o_r \]

\[ U_w^{ij} = \beta_a^{M1} t_{aw}^{ij} + \beta_b^{M1} (c_{ij}^{seat} + \tau_{aw})/o_r \]

where \( c_{ij}^{seat} \) is the car running cost to travel between zones \( i \) and \( j \), \( \tau_{ac} \) is the road charge (decision variable) and \( o_r \) is the average car occupancy rate (therefore \( U_c^{ij} \) is the average utility of car users). In expressions (2)–(6), bus fare and car toll are independent of the origin and destination of trips, an assumption that is consistent with an area charging scheme: extensions to distance-based fares and tolls are discussed in Section 4 (Conclusions).

Assuming a multinomial logit model for the estimation of demand, the number of trips by mode \( m \) in OD pair \((i, j)\) is given by:

\[ y_m^{ij} = \sum e^{ij} / \sum e^{ij} \quad \forall i,j \]

where \( Y^{ij} \) is the total demand between zones \( i \) and \( j \). The estimation of parameters for models M1 to M4 is shown and discussed in Appendix A.1. In this framework, the consumer surplus \( B \) is given by the logsum formula:

\[ B = \sum \frac{Y^{ij}}{n} \ln \sum e^{ij} + B_0 \]

where \( I_a \) is the marginal utility of income,\(^6\) equal to minus the cost parameter \( \beta_a^{M} \) estimated with the choice models,\(^7\) and \( B_0 \) is a constant that has no effect on the solution of the problem, and therefore can be set to zero.

2.3. Travel time, congestion and bus stop delay

We assume that buses and cars share the right-of-way, which is subject to congestion. Furthermore, buses have to stop at bus stops to load and unload passengers. Bus stops are also subject to congestion in the form of queuing delays when the bus frequency is high and/or the dwell time is long. Taking direction 1 for illustration, we model travel time between zone \( i \) and

\(^5\) The assumed linearity in the headway attribute as a proxy of waiting time cost is more appropriate in high frequency services in which ex-ante schedules do not exist or are not relevant in the departure time decisions of users. A situation with low frequency in which passengers follow a schedule is better modelled with the inclusion of two or more headway parameters or with a non-linear formulation. The resulting frequencies in Table 3 are over 20 veh/h (headways of less than 3 min), which is consistent with the assumption of high frequency services.

\(^6\) The marginal utility of income is assumed constant, i.e., we ignore income effects on demand (Jara-Díaz and Videla, 1989; Jara-Díaz, 2007).

\(^7\) Note that \( \beta_a^{M} \) is estimated with the choice of motorised modes only because walking is for free.
zone $i + 1$ by car ($t'_{ac1}$) and bus ($t'_{s1}$) as a function of traffic flow and bus frequency by using the well-known Bureau of Public Roads (BPR) formula:

$$t'_{ac1}(f_0, f_b) = t'_{s0} \left[ 1 + \alpha_0 \left( \frac{f_0 + \varphi(s_b)f_b}{K} \right)^{a_1} \right]$$

(10)

$$t'_{s1}(f_0, f_b) = t'_{s0} \left[ 1 + \alpha_0 \left( \frac{f_0 + \varphi(s_b)f_b}{K} \right)^{a_1} \right] + t'_s$$

(11)

where $t'_{ac0}$, $t'_{s0}$, $\alpha_0$ and $a_1$ are parameters ($t'_{ac0}$ and $t'_{s0}$ are the free-flow travel times by car and bus, respectively), $\varphi \geq 1$ is the passenger car equivalency factor of a bus, which depends on the bus length $s_b$, and $K$ is the capacity of the road. The travel time by bus includes the delay due to bus stops, $t'_s$, which consists of the acceleration and deceleration delay $t'_{ac1}$, the average queuing time $t'_{q1}$ and the dwell time $t'_{d1}$, i.e.,

$$t'_{s1} = t'_{ac1} + t'_{q1} + t'_{d1}$$

(12)

The delay in the process of accelerating and decelerating at bus stops is modelled by assuming uniform acceleration and deceleration; thus the extra stopping delay on top of the uniform travel time given by the running speed $v'_{b1}$, is expressed as (13):

$$t'_{ac1} = \frac{v'_{b1}}{2} \left( \frac{1}{a_0} + \frac{1}{a_1} \right)$$

(13)

The queuing time $t'_{q1}$ is a measure of the external congestion caused by a bus stop, observed when a bus arrives at a stop and all berths are occupied. This delay is commonly present in high frequency services, but it may also occur in poorly controlled low frequency services where buses tend to bunch. Following Fernández et al. (2000) and Tirachini and Hensher (2011), we use the bus stop simulator IRENE ( Gibson et al., 1989; Fernández and Planzer, 2002) to estimate the queuing delay $t_q$ (s/bus) as a function of the design of the bus stop (number and length of berths), bus length, bus frequency and average dwell time. Bus stops with one, two and three linear berths are considered. The regression model for the simulated data has the exponential form (14), as developed in Tirachini (2014) and explained in Appendix A.2:

$$t_q = 0.001[b_0 + b_1s_b + (b_{d1} + b_{d2}Z_2) + b_{d3}Z_3]t_d|e^{0.001}f_0|f_b|b_{d6} + b_{d7} + b_{d8} + b_{d9} + b_{d10} + b_{d11}]$$

(14)

where $s_b$ (m) is the bus length, $t_d$ (s/bus) is the dwell time, $f_b$ (veh/h) is the bus frequency and $b_0$, $b_1$, $b_{d1}$, $b_{d2}$, $b_{d3}$, $b_{d6}$, $b_{d7}$, $b_{d8}$, $b_{d9}$, $b_{d10}$ and $b_{d11}$ are estimated parameters; factors 0.001 are introduced for scaling of the parameters (see Appendix A.2 for further details). $Z_2$ and $Z_3$ are dummy variables defined as follows:

$$Z_2 = \begin{cases} 1 & \text{if bus stop has two berths} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_3 = \begin{cases} 1 & \text{if bus stop has three berths} \\ 0 & \text{otherwise} \end{cases}$$

Eq. (14) is defined for bus stops with one, two and three berths; therefore it allows the number of berths to be a variable. In the numerical application of Section 3 we assume the existence of two berths per bus stop. The estimation of the dwell time per stops requires the cases with boarding allowed at all doors ($TnBn$) and at the front door only ($TnB1$) to be addressed separately, since in $TnBn$ boarding and alighting is sequential at all doors, whereas in $TnB1$ boarding at the front door occurs simultaneously with alighting at the rear doors. These two cases are summarised in expression (15):

$$t'_{d} = \begin{cases} c_{oc} + p_b \beta_d \check{\lambda} + p_a \beta_a \check{\lambda}^- & \text{if boarding at all doors ($TnBn$)} \\ c_{oc} + \max(\beta_d \check{\lambda}, p_b \beta_d \check{\lambda}^-) & \text{if boarding at front door only ($TnB1$)} \end{cases}$$

(15)

where $c_{oc}$ is the time to open and close doors, $\beta_a$ and $\beta_d$ are the average alighting and boarding times per passenger, $\check{\lambda}$ and $\check{\lambda}^-$ are the number of passengers boarding and alighting a bus at the bus stop and factors $p_a$ and $p_b$ are the proportion of passengers boarding and alighting at the busiest door, which are given in Appendix A.3 (Table A3.1). Eqs. (13)–(15) conclude the derivation of the delay at bus stops (12).

2.4. The choice of bus size and internal layout

Using data from London, Jansson (1980) finds a linear relationship between bus running costs and bus size measured as the number of seats per bus, a relationship that has been used by Jansson and other authors to find the optimal size of buses in urban routes (e.g., Jara-Díaz and Gschwender, 2003; Tirachini and Hensher, 2011) under the implicit assumption that there is a unique relationship between bus size and capacity, measured as number of seats or total number of passengers that can be carried. However, the number of passengers that a bus can carry is not only given by the bus size, but also by
the internal layout of space allocated to seating and standing, as a passenger sitting takes up more space than a passenger standing. A standard value for the area needed for a passenger sitting is 0.5 m$^2$ (TRB, 2003), whereas, depending on crowding conditions, passengers standing may have a density of up to five or six passengers per square metre, and as such the minimum area required by a standee is approximately 0.17–0.20 m$^2$, i.e., less than half the space required for a person seated. Therefore, if the number of seats inside a bus can be varied, there is no one-to-one relationship between capacity and bus size, and the final capacity of a bus is the outcome of decisions on both the bus length and internal layout regarding seating and standing areas. In this context, capacity is not an absolute value, but rather a function of the number of seats and the maximum density of standees that is acceptable to have or provide, given by policy, demand and cultural constraints.

Several physical constraints need to be considered when deciding the number of seats, including minimum space for aisles, doors in front of the bus (next to the driver) and for a wheelchair, that must be clear of seats. Let $A(s_b)$ be the total area available in a bus for seating and standing, which is a function of the bus length $s_b$. If $P_s$ is the proportion of $A$ allocated to seating, the areas for seating $A_{\text{seat}}$ and standing $A_{\text{stand}}$ can be formulated as:

$$A_{\text{seat}}(P_s, s_b) = P_s A(s_b)$$

$$A_{\text{stand}}(P_s, s_b) = (1 - P_s) A(s_b)$$

If $A_{\text{seat}}$ is the area required by one bus seat (m$^2$), then the number of seats $n_{\text{seat}}$ per bus is

$$n_{\text{seat}} = \frac{A_{\text{seat}}}{A_{\text{seat}}}$$

For the estimation of in-vehicle time costs it is necessary to determine the proportion of seats being occupied $p_{\text{seat}}$ and the density of standees $n_{\text{den}}$ (if any) in each segment of a bus trip. Taking direction 1, if $\lambda^+$ and $\lambda^-$ are the number of passengers getting on and off a bus at stop $i$, the number of passengers $q^i$ on board a bus between stops $i$ and $i+1$ is calculated recursively:

$$q^0 = 0$$

$$q^i = q^{i-1} + \lambda^+ - \lambda^- \quad \forall i \in \{1, P - 1\}$$

Separating $q^i$ among passengers seating $q^i_{\text{seat}}$ and standing $q^i_{\text{stand}}$, we can obtain $p^i_{\text{seat}}$ and $n^i_{\text{den}}$ as follows:

$$p^i_{\text{seat}} = \frac{q^i_{\text{seat}}}{n_{\text{seat}}} = \frac{\min\{n_{\text{seat}}, q^i\}}{n_{\text{seat}}}$$

$$n^i_{\text{den}} = \frac{q^i_{\text{stand}}}{A_{\text{stand}}} = \frac{q^i - q^i_{\text{seat}}}{A_{\text{stand}}}$$

To calculate the area available for seating $A_{\text{seat}}$ and standing $A_{\text{stand}}$, we need an estimation of the area occupied by seats, standees, doors and other elements. The US Transportation Research Board recommends the following values (TRB, 2003):

We use Table 1 and the following assumptions in order to calculate seating and standing areas, feasible numbers of seats and total bus capacity:

(A1) Buses have transverse seating only, therefore 0.5 m$^2$ is the value used for the area occupied by passengers sitting.

(A2) The maximum density of standees $d_{\text{max}}$ is around 6.7 pax/m$^2$, equivalent to an area of 0.15 m$^2$ per standee in Table A4.1. However, given the Sydney context $d_{\text{max}}$ is set as 4 pax/m$^2$ in the Military Road application (Section 3).

(A3) Buses are 2.55 m wide (regardless of length).

---

\[ In some crowded public transport systems around the world (e.g., Moscow, Sao Paulo, Santiago de Chile, Tokyo), it is not unusual to operate at crush capacity, with 6 passengers standing per square metre in peak periods, however, such a high density of standees could not be acceptable in other regions.\]
The front area must be left clear of passengers, for the driver and front door. This area is 1.5 m long. Next to each rear door there has to be a 0.8 m² area clear of standees. The number of doors per bus is denoted as \( n_{\text{doors}} \).

Buses must have a 0.95 m² area reserved for wheelchairs.

Using (A3)-(A6), the total area \( A \) (m²) available for seating and standing is:

\[
A = A_{\text{sit}} + A_{\text{stand}} = 2.55(s_b - 1.5) - 0.8(n_{\text{doors}} - 1) - 0.95
\]

and the capacity of a bus (maximum number of passengers that can be accommodated) is:

\[
K(s_b, P_s, n_{\text{doors}}) = \left[ \frac{P_s \sum_{\text{seat}} + (1 - P_s)d_{\text{max}}}{a_{\text{seat}}} \right] A
\]

### 2.5. Constraints

(C1) An aisle is provided in the centre of the bus, with a minimum width of 0.5 m. This aisle does not necessarily have to cover the full length of the bus as the back row may have a seat in the middle (where the aisle ends). Therefore, assuming that 1.5 m is left at the front and 0.7 m is used for a seat at the back, the minimum area that has to be reserved for the aisle is \( A_{\text{stand}} = 0.5(s_b - 2.2) \). Then, the number of seats is upper bounded by:

\[
n_{\text{seat}} \leq n_{\text{seat}}^\text{max} = \frac{A - A_{\text{stand}}}{a_{\text{seat}}}
\]

(C2) A minimum number of seats must be provided, i.e., the proportion \( P_s \) of \( A \) allocated to seating has a lower bound \( P_s^\text{min} \), which is arbitrarily decided (e.g., \( P_s^\text{min} = 0.3 \) meaning that at least 30% of the available area must be reserved for passengers sitting). Therefore,

\[
n_{\text{seat}} \geq n_{\text{seat}}^\text{min} = \frac{P_s^\text{min} A}{a_{\text{seat}}}
\]

### 2.6. Bus operator cost and problem formulation

In this section, we formulate bus operator costs. Let operator cost be divided into four components:

\[
\begin{align*}
&c_1: \text{Station infrastructure cost ($/station-h),} \\
&c_2: \text{Personnel costs (crew) and vehicle capital costs ($/bus-h),} \\
&c_3: \text{Running costs (fuel consumption, lubricants, tyres, maintenance, etc.) ($/bus-km),} \\
&c_4: \text{Implementation cost related to the fare payment technology (e.g., software requirements) ($/h).}
\end{align*}
\]

The station cost \( c_1 \) (Eq. (27)), consists of two components: the station infrastructure cost which depends on the bus length i.e., \( c_{10}(s_b) \), and the cost of fare vending machines and fare collection readers (if validation is undertaken at the station and not on bus), \( c_{11}(A) \), where the dependency on \( A \) denotes the fare payment method.

\[
c_1(s_b, A) = c_{10}(s_b) + c_{11}(A)
\]

Second, the cost per bus-hour \( c_2 \) also has two elements: the personnel cost (wages) and the capital cost of a vehicle, which includes the cost of the fare collection readers (validation devices) installed in buses. If \( c_{20}(s_b) \) is the cost associated to bus size and \( c_{21}(A, s_b) \) is the cost of the fare collection readers (which in turn depends on bus size \( s_b \) as well, because for on-board payment methods, fare collection devices are installed at each boarding door, see Appendix A.4), the total cost per bus-hour \( c_2 \) is simply expressed as:

\[
c_2(s_b, A) = c_{20}(s_b) + c_{21}(A, s_b)
\]

The third component of operator cost is the running cost per veh-km \( c_3(s_b) \), which includes fuel consumption, lubricants, tyres, maintenance, etc., and depends on the size of the bus (the capital cost of garages, which depends on bus and fleet size, is ignored). Finally, \( c_4(A) \) accounts for the cost of software and implementation of the alternative fare collection technologies and boarding and alighting policies. The estimation of parameters for the operator cost components is given in Appendix A.4.

With this, the total operator cost \( C_0 \) can be defined as:

\[
C_0(s_b, A, F) = c_1(s_b, A)S + c_2(s_b, A)F + c_3(s_b)VF + c_4(A)
\]

where \( S \) is the number of bus stops, \( F \) is the fleet size and \( V \) is the operating speed (commercial speed including running and stops). The fleet size requirement is given by \( F = f_sT_c \), in which \( T_c \) is the cycle or round-trip time (given by the summation of bus travel time (11) at all sections and both directions, plus a scheduled slack time at termini if required). Rewriting \( T_c \)
as $2L/V$, we obtain that the third term in (18) does not depend on the operating speed and passenger demand. Therefore, the final expression for bus operator cost is given by (30).

$$C_b(f_b, s_b, A) = c_1(s_b, A)S + c_2(s_b, A)f_bT_c(f_b, s_b, A) + 2c_3(s_b)Lf_b + c_4(A)$$ (30)

Importantly, we are assuming that the number of seats inside a bus (and consequently, the number of passengers) has no effect on the bus capital cost, which is only determined by the bus size and arrangements regarding fare collection readers (i.e., the cost of seats if assumed negligible relative to the cost of the bus). After obtaining an expression for the operator cost (30), we can formulate the social welfare maximisation problem as follows:

$$\text{Max} \quad SW = \sum_{ij} y_{ij} \ln \frac{e^{y_{ij}}}{n} + \sum_{ij} y_{ij} f_s + \sum_{ij} y_{ij} f_b - C_o$$ (31)

Subject to

$$\max \{y_{i1}, y_{i2}\} \leq k_f s_b$$ (32a)
$$n_{\text{min}} \leq n_{\text{seat}} \leq n_{\text{max}}$$ (32b)
$$f_{\text{min}} \leq f_b \leq f_{\text{max}}$$ (32c)
$$s_b \in \{s_{b1}, \ldots, s_{b4}\}$$ (32d)
$$A \in \{A_1, \ldots, A_4\}$$ (32e)

$$y_{ij} = \frac{e^{y_{ij}}}{\sum_n e^{y_{ij}}} \quad \forall i, j$$ (32f)

Inequality (32a) is a capacity constraint that ensures that the bus transport capacity ($k_f s_b$) is large enough to accommodate the maximum bus load ($y_{i1}$ and $y_{i2}$ are the demands on section $i$ in directions 1 and 2, respectively; see Fig. 1); $k$ is a design factor introduced to have spare capacity to absorb random variations in demand (for example, $k = 0.9$ assumes a system in which capacity is 90% of the maximum average demand). (32b) states that the number of seats is constrained by minimum and maximum values (Eqs. (25) and (26)). Frequencies are also constrained by a minimum policy frequency $f_{\text{min}}$ (set to have a minimum level of service, if desired) and the maximum feasible frequency $f_{\text{max}}$ as given in expression (32c). Expressions (32d) and (32e) establish that bus size $s_b$ and the boarding and alighting policy and fare collection technology $A$ are taken from available choices. Finally, a set of equilibrium constraints is necessary because modal choice depends on travel times, which in turn depend on modal choice; inducing a fixed-point problem that is solved by iterating between modal choice and travel times until convergence is reached, using Eq. (32f).

Contrary to simpler models in which it is possible to analytically find the optimal level of variables such that bus frequency, size and fare (e.g., Mohring, 1972; Jansson, 1980; Chang and Schonfeld, 1991; Tirachini et al., 2010), the constrained optimisation (31) and (32) requires a numerical approach. The problem is solved using the Sequential Quadratic Programming method implemented in the constrained optimisation toolbox of MATLAB v7.12.0. The solution procedure applied considers bus frequency as a continuous variable, while the number of seats, car toll and bus fare are discrete (fare and toll are constrained to be a multiple of 5 cents).

3. Application

3.1. Physical setting and input parameters

In order to illustrate the effects of explicitly accounting for the crowding discomfort in the design of buses and the pricing structure of urban transport systems, we apply the social welfare maximisation model with demand and supply data from a specific transport corridor, Military Road in North Sydney, Australia, shown in Fig. 2. The section modelled comprises 3.44 km of road which is divided in 12 zones (therefore the average zone length is 286 m). The origin–destination matrix for car and bus trips is obtained from a traffic simulation study undertaken in this corridor by the Roads and Traffic Authority (RTA). In order to add walking trips to the matrix we use Sydney’s Household Travel Survey (TDC, 2010) to obtain the city’s modal split by trip distance; 66.7% of trips shorter that 1 km are made on foot, a Fig. that drops to 24.7% for trips between 1 and 2 km, and 5.7% for trips between 2 and 5 km (considering car, bus and walk only). Then we amplify each cell (bus + car trips) by the respective percentage of walking trips according to the distance between origin and destination. The matrix obtained

---

9 This corridor is chosen because of the availability of origin-destination demand data at the level of small zones. The estimation of taste parameters for utility functions (2)-(5) is done with data collected in adjacent area in Sydney (the CBD and the North West); we assume that the estimated parameters are also applicable to the Military Road area.
with this procedure is presented in Fig. 3, with a total of 19,231 trips in the morning peak (7.30–8.30 am), from which 54.3% are from east to west, towards the CBD (Direction 2 in Fig. 1).

The road has two lanes per direction, BPR functions (10) and (11) are assumed to represent travel times with commonly used parameter values $a_0 = 0.15$ and $a_1 = 4$, and a capacity $K_r = 2000$ veh/h obtained by assuming a 60% for effective green time ratio at signalised intersections. Speed at free flow is 50 km/h. With these assumptions, the average car speed is 26.3 km/h in direction 1 (outbound) and 21.5 km/h in the direction 2 (inbound), similar to the measured average speed of 22 km/h on this road (RTA, 2011, which only reports average speed in the inbound direction in the morning peak). The bus equivalency factors $\varphi(s_b)$ are 1.65 for small buses (8 m), 2.19 for standard buses (12 m), 2.60 for rigid long buses (15 m) and 3.00 for articulated buses (18 m).

Users can choose between travelling by car, bus or to walk; other alternatives like switching time period or changing origin and/or destination are not considered. The car operating cost is 14 cents/km (fuel consumption) and the average car occupancy 1.45 pax/car (TDC, 2010), which we assume remains unchanged after pricing reforms (the sensitivity of car occupancy to raising tolls is ignored). Walking speed is assumed to be 4 km/h.
The parameter estimates for the utility functions (2)–(7) is presented in Table 2. The estimation of parameters is explained and discussed in the Appendix A.1, including goodness-of-fit tests and t-ratios. The constraints for the minimum and maximum number of seats per bus are explained in Section 2.4.

3.2. Base results

Results with the current OD matrix (Fig. 3) for demand models M1 to M4 are shown in Table 3. First, the solution regarding bus size, frequency, fare, toll and number of seats is similar for M1 and M2, and for M3 and M4. In the case of M1 (no crowding or standing externality internalised) it is optimal to operate with mini buses (8 m long) at a frequency of 21.7 veh/h and to charge a fare of 10 cents, whereas in M2 (with standing disutility) the optimal solution has a slightly greater frequency of 23.7 veh/h. The similarity of results is because at this level of bus demand almost all passengers are sitting, as shown by the maximum occupancy rate (over number of seats), which is 1.08 for M1 and 0.98 for M2 (tenth row in Table 3), therefore, due to the absence of standees, both models have similar optimal outputs. A different result is obtained if we assume that the proportion of bus riders sitting is also a source of disutility, either in a linear (M3) or quadratic (M4) form; in these cases the optimal solution comprises bigger (12 m) and more frequent buses (between 25.0 and 26.1 veh/h). Importantly, the external marginal cost of crowding reflects an increase in the optimal fare of 30 cents per passenger; as in M1 and M2 the marginal cost of carrying an extra passenger is only given by the extra boarding and alighting time, whereas...
for M3 and M4 the optimal fare also accounts for the discomfort caused by a passenger that reduces the number of free seats on a bus. This strong increase in optimal bus fare points to the large effect of including a crowded seating disutility when setting optimal public transport prices.

Next, regarding the optimal number of seats, in all cases the optimal result is having the maximum number of seats that is technically possible (24 for 8 m-long buses, 39 for 12 m-long buses), constrained by the minimum area required free of seats. Out of the available area for seating and standing, 80% is allocated to seating and 20% to standing. The greater frequency and bus size of models M3 and M4 considerably reduces the average occupancy rate (as a function of the number of seats) from over 50% in M1 and M2, to 30% in M3 and M4 (the supply of seats per hour is almost doubled from 521 in M1 to 1017 in M3). The key role of the number of seats as a decision variable is unveiled in Section 3.3, when an increase in road congestion challenges the optimality of having the maximum number of seats if the planner ignores the external cost of crowding.

The outputs regarding number of doors, bus boarding policy and fare collection technique are described as follows. First, in all cases it is optimal to have the maximum number of doors given by the bus size, i.e., 2 doors for 8 m buses and 3 doors for 12 m buses, as the more doors are in place the shorter are the boarding and alighting times. Second, sequential boarding and alighting at all doors (TnBn system) is more efficient than operating with boarding at the front door only (TnB1). Third, the optimal fare collection technology is off-board with M1 and on-board with a magnetic strip with M2–M4.

The consideration of a crowded seating disutility has a strong effect on the financial state of the public transport provider and the subsidy required to run the system: in M1 and M2 with an optimal fare of 10 cents it is required a subsidy that needs to cover 83% of the operator cost, whereas if the optimal fare of M3 is charged (40 cents) the required subsidy is halved. In all cases the toll revenue is more than enough to cover the bus operator deficit (ignoring toll collection costs).

Since the model includes disaggregated origin–destination (OD) demand information, we are able to compute marginal external costs of crowding and congestion per OD pair. The marginal external costs are computed by deriving travel time costs with respect to modal demand in order to account for the effect of an extra traveller resulting in increasing travel time on the road (i.e., the congestion externality captured by BPR functions 10 and 11) and in increasing the proportion of seats occupied or the density of standees (crowding externality captured in utility functions, Eqs. (3)–(5)). In all cases, external time costs (min/trip) are divided by the marginal utility of cost ($/trip) to have the correct ($/trip) measure. For illustration, marginal external cost curves for trips from five origins are depicted in Fig. 4 for model M3, in which, for example, “crowding 1–12” shows the marginal external crowding cost (MECrC, between 10 and 63 cents of dollar) of bus trips starting in origin 1 with destination in all zones between 2 and 12, whereas “crowding 12–1” shows the MECrC for trips with origin in zone 12 and destinations between 1 and 11 (between 16 and 74 cents in the dollar). The two MECrC curves grow approximately linearly as a trip gets longer, and the value is slightly larger in Direction 2 (from 12 to 1) given the larger demand in this direction (54.3% of total). Using the distance in kilometres between stops as explanatory variable, the following relationship between MECrC and trip length $L_{km}$ can be estimated:

$$\text{MECrC} = \alpha + \beta L_{km},$$

where $\alpha$ and $\beta$ are estimated coefficients.

10 Note that bus utility in M1 is indifferent to the number of seats inside buses, therefore as long as the capacity constraint is not binding, any number of seats would produce the same level of social welfare. In Table 3 the capacity constraint is inactive for buses with the maximum number of seats $n_{\text{max}}$, therefore $n_{\text{max}}$ is arbitrarily chosen for M1.

11 This result ignores that the time to open and close doors may increase with the number of doors, because drivers may spend more time to check that all doors are clear of passengers if more doors are provided in a bus.

12 The current operation of Sydney buses has a minimum fare of $2.10 for a single ticket, which in our model would produce profits, however the current system has to be subsidised. This divergence is explained by a number of elements, including the likely existence of a large amount of fixed costs that is not considered in this application, and that we are only modelling the morning peak period (in which demand is the highest) in an area of high demand density.
Eqs. (33a) and (33b) imply that for trips starting in sections 1 and 12, MECrC grows by between 16 and 18 cents of dollar for every kilometre that a person is inside a bus. On the other hand, the marginal external congestion cost (MECoC) is larger than MECrC as shown in Fig. 4. The relatively larger level of the curve “Congestion 12–1” is derived from the large travel demand from section 12 to sections 11 and 10, whilst in comparison, a person starting his or her trip in section 11 (curve “crowding 11–1”) has a much lower MECoC.

In the next sections, we analyse how the bus service and pricing levels (fare and toll) should be adapted when faced with an increase in transport demand (e.g., through a future urban densification around the corridor), assuming that it is not possible to increase road capacity ($K_r = 2000$ veh/h). The idea is to analyse the evolution of key design variables when the system is stressed and severe congestion arises. The trips by origin and destination of Fig. 3 are uniformly scaled in five steps, up to a total demand of 28,850 trips/h (50% higher than the current number of trips). The main results for modal split, bus service design, pricing, crowding and congestion are discussed. Model M4 is not shown because it produces similar results to those of M3.

### 3.3. Optimal bus frequency: The trade-off between congestion and crowding

The evolution of the optimal bus frequency is presented in Fig. 5. It is evident that regardless of the demand model considered, frequency does not vary monotonically with demand, in particular optimal frequency can decrease as demand grows, although the reasons for this result are not the same across the models. Focusing on M2 (standing disutility) first, we observe that frequency is increasing up to 32 veh/h for 1900 pax/h, but drops to 22 veh/h for 2100 pax/h; this is because up at 1900 pax/h the optimal bus is mini (8 m) whereas at 2100 pax/h it becomes optimal to operate with standard 12 m buses with a higher capacity. Similarly for M3, a discrete increase in bus length (from 12 to 15 m) also explains the drop in frequency from 26 to 24 veh/h with 1550 pax/h. However, if bus size remains unaltered, frequency is always an increasing function of demand if we assume that crowding and standing disutilities matter (case of M2 up to 1900 pax/h and M3 beyond 1550 pax/h) which is in line with all total cost minimisation models that optimise bus frequency either assuming a fixed bus size, or that bus size can be freely adjusted to meet demand once frequency has been optimised (e.g., Mohring, 1972; Jansson, 1980; Jara-Díaz and Gschwender, 2003).

What happens with the optimal frequency in the model that is insensitive to crowding as a source of disutility (M1) is even more noteworthy. In this case the optimal bus size does not change across the whole demand range (mini buses) and in spite of that, frequency slightly decreases from 21.7 to 20.8 veh/h as demand increases from 1370 to 2250 pax/h. This is because of congestion on the road: as total demand grows so does the number of people that use the congestible road facility (the actual speed drop for cars and buses is shown in Fig. 6 for both directions, D1 and D2), and given that bus frequency adds to traffic congestion, the model tries to reduce the number of buses on the street at the expense of increasing crowding levels inside buses, which in M1 is welfare improving because crowding comes at no comfort loss. This is a clear sign of the relevance of including crowding in the optimisation of transport corridors that have cross-congestion between buses and cars.

The examination of optimal frequencies does not provide a full picture of the transport supply being provided by the bus operator because different optimal bus sizes are chosen in Fig. 5. The total seat supply (frequency times number of seats per

---

13 This frequency reduction is not necessarily in opposition of traditional bus optimisation models that predict bus frequency to increase with demand, such as Mohring (1972). In Mohring’s “square root rule”, frequency decreases with the cycle time, which in this case is increasing with demand because of road congestion.
bus) and seat plus stand supply (frequency times bus capacity) are shown in Figs. 7 and 8. It is clear that the optimal capacity that a planner would choose is quite sensitive to the characterisation of the crowding and standing disutilities. M1 is insensitive to the number of seats chosen as long as the bus capacity constraint (32a) is not binding; therefore we have shown the maximum number of seats per bus such that (32a) is not active, which passes from 24 seats per bus when demand is 1370 pax/h, to 11 seats per bus when demand is 2250 pax/h, as reflected in Fig. 7 with a total seat supply decreasing for M1. In other words, when confronted with an increase in demand, part of the (optimal) increase in supply is provided simply

Fig. 6. Average speed M1.

Fig. 7. Optimal seat supply.

Fig. 8. Optimal seating + standing supply.
by reducing the number of seats in order to increase the number of passengers that can be accommodated in a bus, at no crowding cost in M1.\textsuperscript{14} A completely different outcome is obtained if crowding and standing matter, in which case the number of seats is kept at the maximum possible given constraint (32b) and total seat capacity is increasing for the whole demand range on M2 and M3, as shown in Fig. 7.

The fact that planners or bus operators would choose to reduce the number of seats per bus if crowding and standing disutilities are not explicitly accounted for (M1) does not mean that the total transport capacity (seat plus stand) is decreasing; as Fig. 8 shows that with M1, total capacity is actually increasing, due to the increase in bus capacity coupled with a slightly decreasing (almost flat) bus frequency (Fig. 5).

In order to show that M1’s optimal frequency drop (Fig. 5) is actually due to the congestion interaction with cars, an alternative scenario in which there is an exclusive bus lane is modelled, and therefore, bus frequency does not influence car travel time, which is only given by traffic flow (cars remain in two lanes). As depicted in Fig. 9, when buses do not affect cars, optimal frequency has an increasing tendency along the demand range.

Finally, it is worth mentioning that in all scenarios, bus frequencies are low enough not to cause any queuing delay at bus stops (Eq. (14)), which are assumed to have two berths each.

3.4. Optimal pricing and modal split

The change in optimal toll and bus fare is analysed in this section. The optimal toll is largely insensitive to the specification of crowding in the bus utility functions, therefore only one value is presented in Fig. 10, which shows the increase in the optimal toll as total demand (and consequently road congestion) grows. On the other hand, the optimal bus fare slightly increases for M2 and M3, up to 50 cents per ticket, whereas in M1 the fare is maintained at 10 cents.

Next, modal shares are analysed. The predicted modal splits are almost identical under the four models, and that. The resulting modal split with optimised bus design and pricing structure is almost identical under all demand models (Table 3), thus only M2 is shown for illustration in Fig. 11. Compared to the observed modal split (62.5% car, 31.6% walk and 5.9%), more people decide to walk (33%) and ride a bus (7%), reducing the car modal split to 60%. The relatively modest increase in bus modal share (from 5.9% to 7%) after optimisation is due to the large and negative value of the bus modal constant (around \( \frac{1}{C_0} \approx 2.1 \) in Table 2), which has been calibrated to predict observed modal shares in the base situation (very unbalanced in favour of the car mode).

The worsening of road congestion\textsuperscript{15} as total demand grows encourages walking. This results in a car modal share dropping from 60% to 54%, the (assumed uncongestable) alternative of walking increases from 33% to 38%, and the bus choice grows from 7% to 8% of all trips (due to increased frequency and price difference between toll and fare, see Fig. 10). Therefore, if transport demand grows in the future and road capacity is held constant, the model predicts walking to become more relevant as a travel alternative, which in this example is supported by the fact that trips are relatively short (the corridor is 3.4 km long). In fact, Fig. 12 displays the modal split per trip length for trips starting in Zone 1; it is clear that there is a loss of competitiveness of walking as an alternative to motorised modes as trip length increases.

3.5. The case with increased bus-induced congestion

In the previous scenario it was assumed that passenger car equivalency factor for buses \( \phi(s_b) \) is solely given by bus size, from 1.65 for mini buses (8 m) to 3 for articulated buses (18 m). However, some authors such as Parry and Small (2009)

\[ \text{Fig. 9. Optimal bus frequency on shared and dedicated right-of-way.} \]
assume that, in mixed traffic, buses should be given a greater weight in the congestion functions (10) and (11), given that their stops to load and unload passengers have an effect on the capacity of lanes and impose delays on other modes including cars (Koshy and Arasan, 2005; Zhao et al., 2007). We find that when doubling the passenger car equivalency factor (to between 3.3 and 6) optimal bus frequency is reduced, and that the impact is stronger if no crowding externalities are explicitly modelled (M1, Fig. 13) than when the crowding disutility is accounted for, in which case the crowding externality dominates over the congestion externality (M2, Fig. 14). In Fig. 13, the increase in frequency for a bus demand beyond 1800 pax/h is because the minimum number of seats \( n_{\text{min}} \) has been reached, the capacity constraint (26) is binding and therefore the operator has no option but increasing the bus frequency to meet demand.
3.6. The relationship between the number of seats and optimal frequency

In this section, we study the sensitivity of the optimal bus frequency to alternative bus layouts regarding number of seats. As previously discussed, in all scenarios in which the crowding externality is considered (M2, M3 and M4), the optimal bus design comprises having as many seats as possible, given an optimal bus size, in order to reduce the crowding effects of seating and reduce the number of standees. In this context, we study what happens if the number of seats is exogenously chosen to be lower than the maximum (and therefore the bus capacity is increased); Fig. 15 shows that for both M2 (mini buses, 8 m) and M3 (standard buses, 12 m) frequency should be increased as a response to the users’ discomfort of having fewer seats. In other words, the number of seats inside a bus does have an effect on the optimal design of a public transport system if the planner acknowledges that users dislike crowding.

Fig. 13. Optimal bus frequency M1, double equivalency factor for buses.

Fig. 14. Optimal bus frequency M2, double equivalency factor for buses.

Fig. 15. Optimal bus frequency for suboptimal numbers of seats.
3.7. The second best scenario

The preceding analysis was undertaken by assuming that a congestion toll on cars is in place, as shown in Fig. 10. In this section, the second best case in which there is no car toll is investigated.\footnote{The principles behind first best and second best pricing are extensively analysed in \textcite{Small2007} and \textcite{Tirachini2012} among others.} We limit the analysis to a graphical comparison of relevant optimisation outputs between the first best and second best scenarios.

The second best bus fare is negative across the demand range tested and for all utility specifications (M1–M4), i.e., the optimal decrease in bus fare to face a zero toll policy is larger than the optimal first best bus fare (between 10 and 40 cents).\footnote{A negative second best bus fare is also obtained by \textcite{Ahn2009}.} More generally, this result points to the potential optimality of very low public transport fares and large subsidies if car use is substantially under-priced (common in peak periods), in line with earlier findings (\textcite{Glaister1978, Proost2008, Ahn2009, Parry2009}).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig16.png}
\caption{Optimal toll minus bus fare, first best and second best scenarios.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig17.png}
\caption{Optimal bus frequency, first best and second best scenarios.}
\end{figure}

\textit{Fig. 16} shows the difference between optimal toll and fare in the first best and second best scenarios for demand models M1, M2 and M3 (therefore, in the second best scenarios the curves are equal to the absolute value of the negative bus fare). The difference between toll and fare is lower in the second best scenario, as also found by \textcite{Ahn2009}. In \textcite{Ahn2009}, the second best bus fare does not decrease sufficiently to maintain the difference between fare and toll in the first best scenario because such a low bus fare would produce a greater than socially optimal amount of total trips; whereas in our framework the amount of total trips is fixed but the amount of motorised trips is not, and hence a low (negative in this case) bus fare

\footnote{\textcite{Basso2013} show that the level of optimal bus subsidies might be influenced by the provision of dedicated bus lanes if congestion pricing is implemented.}
Estimation of parameters, MNL models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access time ( t_a )</td>
<td>-0.016</td>
<td>-0.017</td>
<td>-0.017</td>
<td>-0.017</td>
</tr>
<tr>
<td>Headway ( h_b )</td>
<td>(-1.22)</td>
<td>(-1.33)</td>
<td>(-1.33)</td>
<td>(-1.33)</td>
</tr>
<tr>
<td>Travel time bus ( \mu_{b} )</td>
<td>-0.019</td>
<td>-0.013</td>
<td>-0.004</td>
<td>-0.006</td>
</tr>
<tr>
<td>Egress time ( \beta_r )</td>
<td>-0.055</td>
<td>-0.058</td>
<td>-0.059</td>
<td>-0.059</td>
</tr>
<tr>
<td>MSC train ( \alpha_{t} )</td>
<td>-3.393</td>
<td>-3.455</td>
<td>-3.473</td>
<td>-3.476</td>
</tr>
<tr>
<td>MSC bus ( \beta_{b} )</td>
<td>-4.131</td>
<td>-4.275</td>
<td>-4.313</td>
<td>-4.315</td>
</tr>
<tr>
<td>MSC metro ( \alpha_{m} )</td>
<td>-2.526</td>
<td>-2.444</td>
<td>-2.465</td>
<td>-2.460</td>
</tr>
<tr>
<td>( r_{m} \times \text{den stand} ) &amp; -0.004</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td>( r_{m} \times \text{prop seat} ) &amp; -0.013</td>
<td>(-1.70)</td>
<td>-0.013</td>
<td>(-1.70)</td>
<td></td>
</tr>
<tr>
<td>Specification tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-1283.4</td>
<td>-1273.1</td>
<td>-1271.7</td>
<td>-1271.7</td>
</tr>
<tr>
<td>Adjusted ( R^2 ) (relative to ASCs)</td>
<td>0.107</td>
<td>0.113</td>
<td>0.113</td>
<td>0.113</td>
</tr>
<tr>
<td>Likelihood ratio test with respect to M1</td>
<td>20.53(&gt;X^2_{1,0.001}=10.83)</td>
<td>23.39(&gt;X^2_{2,0.001}=13.82)</td>
<td>23.34(&gt;X^2_{2,0.001}=13.82)</td>
<td></td>
</tr>
<tr>
<td>Likelihood ratio test with respect to M2</td>
<td>2.86(&lt;X^2_{1,0.05}=3.84)</td>
<td>0.0005</td>
<td>(-2.41)</td>
<td>(-2.46)</td>
</tr>
</tbody>
</table>

Note: \( t \)-tests in bracket below parameter estimates. Time in minutes, cost in $ (AUD).

In the scenario “M3 – Second best”, versus 38.1% in the scenario “M3 – Second best – zero profit”).

4. Conclusions

In this paper we have introduced a social welfare maximisation model with disaggregated origin destination demand and multiple travel alternatives, with the aim of optimising the design of urban bus routes including pricing decisions for both bus and car. The influence of bus crowding is highlighted as we analyse its impact on both the design of the bus service and the congestion level on the road. The consideration of crowding externalities as increasing the discomfort of public transport users pushes towards having bigger and more frequent buses (Jara-Díaz and Gschwender, 2003), which in turn may worsen both bus and traffic congestion on shared roads. The number of seats in buses is introduced as a decision variable for the first time in a microeconomic public transport model; the number of seats is the result of the trade-off between passengers’ comfort (that drives the number of seats up) and vehicle capacity (which might be increased by removing seats). The model is applied to the Military Road corridor in North Sydney and results are discussed in several scenarios with different demand levels and modelling assumptions.

A number of results stand out. The consideration of crowding externalities (at both seating and standing) imposes a higher optimal bus fare, and consequently, a reduction of the optimal bus subsidy. Optimal bus frequency results from a trade-off
between the level of congestion inside buses, i.e., passengers’ crowding, and the level of congestion outside buses, i.e., the effect of frequency on slowing down both buses and cars in mixed-traffic. In particular, optimal bus frequency is quite sensitive to the assumptions regarding crowding costs, the impact of buses on traffic congestion and the overall congestion level, as the crowding externality puts pressure on operators to provide more frequent services, which in turn may add to both bus and car congestion. We find that if crowding matters, bus frequency should increase (for a given bus size) with demand even under heavy congestion, however that might not be the case if the crowding externality is not accounted for, in which case an increase of total demand might be met by a decrease of both frequency and number of seats per bus, at the expense of crowding passengers inside buses and making more passengers stand while travelling.

We find that the existence of a crowding externality implies that buses should have as many seats as technically possible, up to a minimum area that must be left free of seats. If for any other reason planners decide to have buses with fewer seats than optimal (e.g., to increase bus capacity), bus frequency (and the number of buses itself) should be increased to compensate for discomfort imposed on public transport users. Future research should test the optimality of providing the maximum number of seats in the following two cases: (i) in routes with high public transport demand, in which an active capacity constraint may push the number of seats down, and (ii) if the standing disutility is low relative to sitting, in which case, in principle, it might be optimal to provide less seats than the maximum feasible, if the benefit of an increased capacity for the operator is larger than the comfort disbenefit for users.

Regarding the relevance of non-motorised modes in urban mobility, in a corridor of 3.4 km, an increase in total transport demand worsens traffic congestion which increases the choice of walking relative to its motorised alternatives (with optimised bus service, fare and toll). This suggests that at least for short trips, improving the travel conditions of non-motorised modes is a wise strategy to tackle worsening congestion problems in cities.

Including bus riders sitting as a source of disutility produces important changes in some variables (optimal bus fares, bus size, frequency, subsidy). However, optimising the transport corridors under different assumptions on the relevance of crowding externalities to users produced similar values of other outcomes such as social welfare, consumer surplus and

### Table A2.1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>-2.952</td>
<td>0.887</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.061</td>
<td>0.020</td>
</tr>
<tr>
<td>$b_{21}$</td>
<td>2.185</td>
<td>0.530</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>-1.903</td>
<td>0.495</td>
</tr>
<tr>
<td>$b_{33}$</td>
<td>-2.044</td>
<td>0.510</td>
</tr>
<tr>
<td>$b_{13}$</td>
<td>23.089</td>
<td>0.723</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.361</td>
<td>0.046</td>
</tr>
<tr>
<td>$b_{32}$</td>
<td>1.807</td>
<td>0.091</td>
</tr>
<tr>
<td>$b_{45}$</td>
<td>-0.374</td>
<td>0.093</td>
</tr>
<tr>
<td>$b_{56}$</td>
<td>-0.627</td>
<td>0.087</td>
</tr>
<tr>
<td>$k_1$</td>
<td>0.921</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>265</td>
<td></td>
</tr>
</tbody>
</table>

### Table A3.1

<table>
<thead>
<tr>
<th>Number of doors</th>
<th>TnBn $p_b$ ($p_a$) (%)</th>
<th>TnB1 $p_a$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>43</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>38</td>
</tr>
</tbody>
</table>

### Table A3.2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Time (s/pax)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boarding time cash</td>
<td>10.74</td>
</tr>
<tr>
<td>Boarding time magnetic strip</td>
<td>2.94</td>
</tr>
<tr>
<td>Boarding time contactless card</td>
<td>2.05</td>
</tr>
<tr>
<td>Boarding time off-board payment</td>
<td>1.46</td>
</tr>
<tr>
<td>Alighting time</td>
<td>1.46</td>
</tr>
</tbody>
</table>
modal splits across models M1–M4. More research is needed to assess if this finding holds in other contexts, for example, with a superior bus modal share (over 20%).

The analytical framework presented in the paper can be extended in several ways. Departure time decisions can be included by introducing a multi-period framework with modal and time-of-day substitution (e.g., Glaister and Lewis, 1978; De Borger and Wouters, 1998; Proost and Van Dender, 2008). Extensions to larger corridors or networks would make more apparent the need to introduce distance-based tolling and more complex public transport fare structures (by section or origin destination pair), extensions that can be added in the present framework by setting fares per kilometre. Note that a differentiation in price by distance is more attractive if destination choice is included as well; otherwise the model would overestimate the revenue from long-distance travellers, some of whom would try to travel shorter distances. Our model also ignores issues of income distribution (Dodgson and Topham, 1987; Mayeres and Proost, 1997), heterogeneity in the value of time savings (Verhoef and Small, 2004) and that road capacity may also be endogenous (De Borger and Wouters, 1998; Arnott and Yan, 2000), all dimensions that are promising venues of further research for the analysis of marginal cost pricing and optimal public transport design including congestion and crowding externalities. Finally, regarding the number of seats as a variable, the bus operator cost structure could include differences in capital and maintenance costs due to alternative standing and seating configurations.

Acknowledgements

Part of this research was developed while the first author was associated with ITLS-Sydney. This study has been supported by the Australian Research Council Discovery Program Grant DP120100201 titled: ‘Valuation of Service Reliability and Crowding under Risk and Uncertainty: Neglected Drivers of Demand for Public Transport’ and by the Complex Engineering Systems Institute, Chile (Grants ICM P-05-004-F, CONICYT FBO16). This paper is also a contribution to the research program of the Volvo Research and Education Foundation Bus Rapid Transit Centre of Excellence. The authors are grateful to Stef Proost, Bruno de Borger, two anonymous referees and participants of the TML-KUL seminar (Leuven, May 2012), the Kuhmo-Nectar Conference on Transportation Economics (Berlin, June 2012) and the Conference on Advanced Systems for Public Transport – CASPT12 (Santiago de Chile, July 2012) for helpful comments. Remaining errors are the authors’ responsibility alone.

Appendix A. Appendix

A.1. Estimation of demand models M1–M4

The estimation of parameters for commuting and specification tests are presented in Table A1.1 (n = 1932 observations): Focusing on the goodness-of-fit measures, the log-likelihood and adjusted $\hat{R}^2$ statistics relative to a model with alternative specific constants (ASCs) only, demonstrate that the three crowding models (M2–M4) outperform the model with no crowding (M1), but the difference in overall fitness amongst the crowding models is not significant. In fact, M2, M3 and M4 have the same adjusted $\hat{R}^2$ value, and a likelihood ratio test indicates that M2, M3 and M4 are significantly superior than M1 at the 99.9% confidence level. Models M2 and M3 are not statistically different at 95% confidence level, nevertheless from a behavioural perspective, the alternative crowding cost specifications do provide differences on the estimation of value of travel time savings.

Parameters for the utility functions (2)–(7) are taken from Table A1.1, with the exceptions of the time parameter for walking and the mode specific constants, which are estimated as follows. First, walking as a travel alternative was not considered in the survey of the main stated choice experiment from 2009 in Sydney, described in Section 2.2 (Hensher et al., 2011); therefore a reasonable value for the disutility of travel time while walking has to be supplemented. To this end, a secondary intra-CBD model described in an internal 2009 report by Hensher and Rose is used, in which walking was an alternative to

<table>
<thead>
<tr>
<th>Bus size (m)</th>
<th>Bus cost ($/bus-h)</th>
<th>Driver cost ($/bus-h)</th>
<th>Station cost ($/station-h)</th>
<th>Operating cost ($/bus-km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>5.1</td>
<td>37.6</td>
<td>4.4</td>
<td>0.9</td>
</tr>
<tr>
<td>12</td>
<td>11.9</td>
<td>37.6</td>
<td>6.5</td>
<td>1.3</td>
</tr>
<tr>
<td>15</td>
<td>16.9</td>
<td>37.6</td>
<td>8.7</td>
<td>1.4</td>
</tr>
<tr>
<td>18</td>
<td>22.0</td>
<td>37.6</td>
<td>10.9</td>
<td>1.6</td>
</tr>
</tbody>
</table>
public transport modes and taxi for short CBD trips; in the intra-CBD model, it is found that the time parameter of walking ($b_{ve}$) is 1.86 times greater than the in-vehicle time parameter for bus ($b_{vb}$). Thus, we assume a constant value of $b_{ve}$ across models, equal to 1.86 times $b_{vb}$ on M1 (because the latter is an average value of $b_{vb}$ for all crowding conditions); therefore, $b_{ve} = 1.86 \times -0.019 = -0.035$.

Second, mode specific constants for demand models M1–M4 are calibrated to represent the current Sydney modal split of trips shorter than 5 km: 62.5% car, 31.6% walk, and 5.9% bus (TDC, 2010). The current bus frequency of 16 bus/h in the morning peak is used, with a fare of $2.10 and no car toll. The car specific constant is fixed at zero. With these two considerations for the time parameter for walking and the mode specific constants, the estimated parameters used in Section 3 are presented in Table 2.

### A.2. Estimation of queuing delay function

To estimate the queuing delay of buses we use the bus stop simulator IRENE (Gibson et al., 1989; Fernández and Planzer, 2002). A total of 265 simulations were run encompassing all bus sizes (8-, 12-, 15- and 18-m long buses) and bus stop with one, two and three linear berths, for a range of frequencies from 20 to 220 bus/h and dwell times between 10 and 65 s. Buses are assumed to arrive at a constant rate at stops (no bus bunching) and bus stops are isolated from traffic lights. Estimated parameters are presented in Table A2.1. For more details see Tirachini (2014).

### A.3. Dwell time estimation

When there are multiple doors to board and alight, passengers can choose a door to get on and off buses, and the spatial dispersion of their decision will determine how long the boarding and alighting times are. It seems unreasonable to suppose that passengers will distribute uniformly across doors if middle or back doors have closer access to more seats than, say, the front door. We assume that the middle doors would attract a number of passengers that is 50% higher than that of the front or back doors (for example, for buses with two doors, the rear door is placed towards the centre of the bus, and is therefore assumed to attract 60% of the boarding demand, leaving 40% boarding through the front door, next to the driver). The same assumption is made regarding alighting. With this, the proportions $p_b$ and $p_a$ of passengers boarding and alighting at the busiest door (necessary for estimation of dwell times in Eq. (15)), is given in Table A3.1. Boarding and alighting times per fare collection system and boarding policy are taken from Tirachini (2012) (see Table A3.2).

### A.4. Operator cost

The station infrastructure cost depends on the amenities provided, quality of shelter and overall design, ranging from $15,000 for a simple shelter to $150,000 or more for stations with passenger enclosure, at-level boarding, retail services and detailed passenger information (FTA, 2009). In this paper, we assume that the cost increases linearly with bus length: $50,000 (8 m bus), $75,000 (12 m bus), $100,000 (15 m bus) and $125,000 (18 m bus), values that are amplified by 25% if off-board payment is provided.

There are two vending machines per station, and four fare collection readers in case of off-board payment. Fare collection costs are taken from Wright and Hook (2007), the cost of a fare collection reader is $750 (coins), $1750 (magnetic strip) and $2500 (contactless card), while the cost per vending machine is $10,000 (magnetic strip) and $15,000 (smart card). For on-board payment methods, two fare collection readers per boarding door are considered. The cost of software is $100,000 for coin payment, $300,000 for magnetic strip and $500,000 for contactless card. Bus driving cost is $29.9 (Hensher, 2010) and running costs are 0.6, 0.9, 1.0 and 1.1 $/veh-km for 8, 12, 15 and 18 m long buses, respectively (ATC, 2006), values that are increased by 21% to account for overhead costs. The cost of buses is $160,000 (8 m), $370,000 (12 m), $520,000 (15 m) and $700,000 (18 m). The estimated parameters in Tables A4.1 and A4.2 are adjusted to 2011 Australian Dollars assuming 20 years of asset life for buses, 15 years for stations and 5 years for software, card readers and vending machines; one year is equivalent to 2947 peak hours of operation for a typical urban bus service in Australia.

---

21 This figure is in the order of the values estimated by Jovicic and Hansen (2003) for Copenhagen (1.36 and 2.32 for the ratio $b_{ve}/b_{oa}$ for purposes commuting and education, respectively, considering walking and cycling altogether as a non-motorised mode, and trips up to 30 min long).