Accounting for Travel Time Variability in the Optimal Pricing of Cars and Buses

Alejandro Tirachini 1
David A. Hensher 2
Michiel C.J. Bliemer 2

1 Corresponding author
Transport Engineering Division
Civil Engineering Department
Universidad de Chile
Tel: +56 2 2978 4380
alejandro.tirachini@ing.uchile.cl

2 Institute of Transport and Logistics Studies (ITLS)
The University of Sydney Business School,
The University of Sydney, NSW 2006, Australia

To appear in Transportation

Keywords: optimal pricing, travel time variability, reliability, headway variability, fare, toll

Abstract
A number of studies have shown that in addition to the common influences on mode, route and time of day of travel choices such as travel time and cost, travel time variability plays an increasingly important role, especially in the presence of traffic congestion on roads and crowding on public transport. The dominant focus of modelling and implementation of optimal pricing that incorporates trip time variability has been in the context of road pricing for cars. The main objective of this paper is to introduce a non-trivial extension to the existing literature on optimal pricing in a multimodal setting, building in the role of travel time variability as a source of disutility for car and bus users. We estimate the effect of variability in travel time and bus headway on optimal prices (i.e., tolls for cars and fares for buses) and optimal bus capacity (i.e., frequencies and size) accounting for crowding on buses, under a social welfare maximisation framework. Travel time variability is included by adopting the well-known mean-variance model, using an empirical relationship between the mean and standard deviation of travel times. We illustrate our model with an application to a highly congested corridor with cars and buses as travel alternatives in Sydney, Australia. There are three main findings that have immediate policy implications: (i) including travel time variability results in higher optimal car tolls and substantial increases in toll revenue, while optimal bus fares remain almost unchanged; (ii) when bus headways are variable, the inclusion of travel time variability as a source of disutility for users yields higher optimal bus frequencies; and (iii) including both travel time variability and crowding discomfort leads to higher optimal bus sizes.

Acknowledgements

Part of this research was developed while the first author was with the Institute of Transport and Logistics Studies in Sydney. This study is supported in part 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). We thank NSW Roads and Maritime Services (RMS) for providing the data to estimate the travel time variability functions in Sydney, and the referees for their extensive and invaluable comments.
1. Introduction

Travellers do not like wasting time in traffic or waiting at a bus stop. A major element of transport research has focussed on estimating and monetising the (average) time savings of infrastructure investment and the demand management measures targeted at reducing travel times. We are, however, increasingly aware that users are not only willing to pay for a shorter travel time, but also for a more reliable trip. Uncertain travel times cause users to arrive earlier or later than expected at their destination, and influences mode choice, route choice and departure time decisions; suggesting that the variability of travel time plays an important role in the generalised cost of travel. For example, in a survey of Dutch drivers, Verhoef et al. (1997) reported that 97.4 percent of respondents disliked driving in congested conditions, and when asked about the reasons for disliking congestion, the most important factors were time losses (4.14 points on a five-point scale), uncertainty (3.61), and unpleasant driving conditions (3.52). This is one of many pieces of evidence that points to the relevance of certainty and reliability of travel times for users (see Li et al., 2010, for a review).

Since travel time variability is often related to traffic congestion (e.g., Eliasson, 2007; Tu et al., 2007; Peer et al., 2012), a transport policy aimed at reducing the level of congestion, such as road pricing, has the potential of reducing travel times and increasing trip time reliability. This expected result has been empirically corroborated through the implementation of road pricing in Stockholm and London, which has resulted in reductions in both the mean and the standard deviation of travel times (Transport for London, 2007; Eliasson, 2009). Therefore, there is a case for incorporating the benefits from reducing travel time variability in the pricing of both car and public transport use.

Most analyses on the optimal pricing of urban passenger transport include mean travel times only, whereas the few studies that incorporate travel time variability, formally defined as the distribution of travel time over repeated trips with the same mode and route and time of day, focus on car tolling only (e.g., Li et al., 2008a; Jiang et al., 2011). In contrast, this paper investigates the optimal pricing structure of both cars and public transport (buses), as well as determining the optimal frequency and capacity of public transit, with an approach that explicitly accounts for travel time variability as a source of disutility for users.

When analysing car traffic, if there is a positive correlation between the mean ($\mu$) and standard deviation ($\sigma$) of travel time on a specific link (for a given mode and time of day), an increase in the number of cars in congested conditions increases both $\mu$ and $\sigma$. However, the analysis of public transit is not so simple; there are at least three basic differences with the case of cars that can worsen the consequences of unreliability associated with public transport. First, buses have to stop in order to transfer passengers, creating interactions between vehicles and passengers (in the boarding and alighting process), and among vehicles (e.g., queuing delays). The dwell time may also be variable, and such variability depends on several factors, including the scheduled headway, the number of passengers getting on and off, and the bus fare collection system (Dorbritz et al., 2009).

Second, the variability in travel times impacts not only in-vehicle time for users, but also waiting time, since unstable travel times yield schedule delays and headway variability, which in turn increases the waiting time of users (Welding, 1957) and influences activity scheduling decisions. Bus frequency (the inverse of headway) impacts both waiting time and in-vehicle travel time for users; therefore the overall impact of bus frequency on the full trip is more complex than the traffic flow-travel time relationship for car users.

Third, the unreliability and uncertainty of travel times associated with public transit also represents an extra cost for operators, who need to adjust the scheduling of services with larger slack times in the case of less reliable travel times (Furth, 2000). All of these considerations make the inclusion of buses in a multimodal analysis for the optimal pricing of travel time variability far from trivial.
In this paper, a multimodal social welfare maximisation model is formulated, that accounts for travel time and bus headway variability, specified using a mean-variance model (Jackson and Jucker, 1982; Senna, 1994), as defined in the next section. Although the main focus is on travel time variability, we also allow for levels of crowding in public transport (see Tirachini et al., 2013). For short trips we allow for the additional possibility of walking instead of using motorised modes. Demand is spatially disaggregated along a transport corridor. The design (decision) variables are car toll, bus fare, bus frequency and bus size. Besides presenting a methodology to optimise price levels for private and public transport, we also apply the model to an actual transport corridor in Sydney that is subject to congestion, and show the significant impact of including variability on the optimal prices and design of the public transport service. This leads to the final conclusion that including (travel time and headway) variability in the optimal design of the transport system is important from a social welfare point of view.

The remainder of the paper is organised as follows. Section 2 provides a literature review on the determinants of travel time variability and the valuation of travel time variability. Section 3 presents regression models for the relationship between the mean and standard deviation of travel time, estimated with data collected across 423 roads in Sydney. The reliability-sensitive social welfare maximisation approach is introduced in Section 4. In Section 5 the main results of the numerical application are discussed. Conclusions and directions for further research are summarised in Section 6.

2. Literature Review

In this section we provide a review of the literature on travel time variability. The majority of studies have focussed on car traffic (Section 2.1); however we find a growing number of studies that have investigated travel time variability in public transport (Section 2.2). Then we review studies on how travel time variability is valued by travellers (Section 2.3) and studies on the relationship between travel time variability and road pricing (Section 2.4).

2.1 Determinants of travel time variability: car traffic

Travel time variability (TTV) is related to random variations in travel time caused by factors that cannot be anticipated or foreseen by a traveller (Fosgerau et al., 2008; Tu, 2008). Tu (2008) divides the sources of TTV in two groups: demand fluctuations and supply fluctuations. Notable sources of variability in traffic demand include temporal effects (e.g., peak/off-peak, weekday/weekend), network effects (effect of traffic in one lane or road over travel times on other parallel or intersecting lanes/roads), and spatial and temporal differences in driving attitude. On the other hand, factors such as volatile or adverse weather conditions, traffic incidents and accidents, and traffic composition influence both demand and road capacity (Tu, 2008).

With the increasing availability of observed travel times, traffic flows and travel speeds on urban and inter-urban networks, analysts have been trying to explain the determinants of TTV based on empirical measurement of these traffic variables. There is no agreement on the dependent variable used as a measure of TTV, and several measures have been proposed to account for the degree of variability of travel time (Pu, 2011), including the standard deviation of travel time (May et al., 1989; Eliasson, 2007; Hellinga et al., 2012; Mahmassani et al., 2012; Peer et al., 2012), the difference between the 90th and 10th percentile of travel time (Eliasson, 2007; Tu et al., 2007), the coefficient of variation of travel time (May et al., 1989; Eliasson, 2006), the standard deviation and the variance of the delay\(^1\) (Mott MacDonald, 2008) and the probability that travel time is below a certain threshold (Asakura, 1998). In some cases, variability is analysed for whole sections or links (May et al., 1989; Eliasson, 2006; 2007; Peer et al., 2012), whereas other authors model variability per unit of road length (per

---

\(^1\) Delay defined as actual travel time minus free-flow travel time.
kilometre), as a way to have a distance-free measure (Tu et al., 2007; Mott MacDonald, 2008; Mahmassani et al., 2012).

The most common variable used to analyse travel time variability is the mean travel time or the mean delay. A majority of authors have found a positive correlation between travel time variability and mean travel time (May et al., 1989; Eliasson, 2007; Mott MacDonald, 2008; Hellinga et al., 2012; Peer et al., 2012), nevertheless the shape of the relationship varies from case to case. For example, using travel time data from a set of Dutch highways, Peer et al. (2012) show that TTV, measured as the standard deviation of travel time (σ), increases with the mean travel time (μ) and that the relationship is concave, i.e., the rate at which variability grows with the mean travel time decreases with travel time. Hellinga et al. (2012) found a similar result, explaining the standard deviation as a function of the mean travel time by using a logarithmic (concave) relationship. On the other hand, Mott MacDonald (2008) analyse TTV for different types of links on English motorways, finding that the shape of the relationship depends on the section or type of highway analysed; in particular, the relationship between μ and σ can be concave or convex, i.e., the coefficient of variation may be an increasing or decreasing function of travel time. In links with extreme congestion, Eliasson (2006) uses data from a number of urban roads in Stockholm containing traffic lights, and shows that the standard deviation divided by travel time might be a decreasing function of the travel delay. Eliasson (2007) finds that σ is higher in the “after AM peak” and “after PM peak” periods, which are interpreted as queue dissipation phases, and that a higher speed limit also increases σ. Li et al. (2008b) have shown a similar result theoretically derived from a stochastic bottleneck model.

Instead of analysing the relationship between variability and mean travel time, Tu (2008) relates variability directly to traffic flow. Using highway sections in China and the Netherlands, he found that the impact of inflow on TTV depends on the flow itself; there is a low demand range at which travel times are fairly constant and variability is low. However, when flow reaches a ‘critical transition inflow’, an increase in demand is associated with a rapid increase in TTV. This increased variability is maintained until flow reaches a ‘critical capacity inflow’, after which TTV can decrease with demand.

It is clear from this brief review of the published literature that there is no uniquely preferred measure of TTV, although the majority of the researchers seem to opt for the standard deviation of travel time. We follow this majority and consider the standard deviation of travel time, which is also appealing in adopting the mean-variance model (see Section 2.3).

2.2 Determinants of travel time variability: the case of public transport

Research on characterising TTV has mainly focused on cars. Nonetheless, public transport modes are also subjected to variations in (in-vehicle) travel times and headways (which translate into larger waiting times). The social cost of unreliability in public transport may be substantial; for example, Van Oort (2011) estimates a yearly cost of €12 million in The Hague, The Netherlands, due to unreliable buses and trams. Improving public transport reliability yields multiple benefits, including increased accessibility, additional ticket revenue and reductions in congestion and environmental externalities, if a modal shift from car to public transport is induced (Van Oort, 2011).

There are a number of studies that have analysed bus TTV based on empirical data (e.g., Abkowitz and Engelstein, 1983; Strathman and Hopper, 1993; Strathman et al., 1999; El-Geneidy et al., 2008; Mazloumi et al., 2010; Moghaddam et al., 2011). Common indicators proposed to assess the reliability of a public transport service include the standard deviation of travel time, the probability of on-time performance2, the travel time ratio (observed travel time/scheduled travel time), the average additional travel time per passenger (Van Oort, 2011) and measures to analyse the variability of headways. These studies usually find that travel time variability, however it is measured, increases with factors such as

---

2 Defined as a bus being between 1 min early to 5 min late at the destination point (Strathman and Hopper, 1993)
the length of a route, number of stops and signalised intersections, with longer headways and higher passenger activity (boarding and/or alighting), with part-time or unexperienced drivers, and that a deviation in travel time at an early stage on a route (including a late departure from the first stop) propagates further downstream as buses proceed. Mazloumi et al. (2010) found an almost linear relationship between the standard deviation and mean travel time for buses on an urban route in Melbourne. Moghaddam et al. (2011) also estimated a positive relationship between the standard deviation of travel time and the volume/capacity ratio as an indicator for congestion on the route. On-board fare collections systems, including cash payment, have been found to increase the standard deviation of boarding times (Dorbritz et al., 2009).

2.3 Estimation of users’ valuation of travel time variability

In this section we provide a brief summary of the main approaches that have been proposed to examine the users’ valuation of travel time variability (for in-depth reviews see Li et al., 2010; and Carrion and Levinson, 2012). The scheduling model and the mean-variance model are the two most common methods to deal with travel time reliability and departure decisions. The scheduling model (Small, 1982; Noland and Small, 1995; Bates et al., 2001) assumes that being early or late at a destination is an additional source of disutility for travellers, besides travel times and travel costs. The general form for the utility function \( U \) (for a given combination of mode, departure time, and route) in this model is:

\[
U = \delta C_t + \alpha T_t + \beta SDE_t + \gamma SDL_t + \vartheta D_t,
\]

where for each departure time \( t \), \( C_t \) is the monetary cost of travel, \( T_t \) is travel time, \( SDE_t \) and \( SDL_t \) are the schedule delay penalties for arriving early and late, and \( D_t \) is a dummy variable that is equal to one when arriving late at the destination, and zero otherwise. Behavioural parameters \( \delta, \alpha, \beta, \gamma, \) and \( \vartheta \) denote the (negative) marginal utilities of cost, travel time, minutes arriving early and late, and a fixed penalty for a late arrival, respectively. These parameters have been estimated by e.g., Small (1982), Bates et al. (2001) and Van Amelsfort et al. (2008).

The mean-variance approach (Jackson and Jucker, 1982; Senna, 1994; Lam and Small, 2001 among others) suggests that the variability of travel time is a cost by itself, no matter if travellers arrive early or late. Under these assumptions, expected utility (given the probabilities of different travel time outcomes) can be expressed as:

\[
U = \delta C_t + \alpha \mu_t + \rho \sigma_t,
\]

where \( \mu_t \) and \( \sigma_t \) are the mean and standard deviation of travel time when departing at time \( t \) and \( \rho \) is a behavioural (negative) parameter that describes the travellers’ sensitivity towards TTV. Small et al. (1999) amongst others, estimated values of reliability using the mean-variance model in Eqn. (2). Analogous to the value of travel time savings, also referred to as the value of time (VOT) and is equal to \( \frac{\alpha}{\delta} \) in Eqn (1), the value of reliability (VOR) is defined as \( \frac{\rho}{\alpha} \) in Eqn. (2). Another popular outcome of the mean-variance model is the reliability ratio (RR), defined as VOR/VOT = \( \frac{\rho}{\alpha} \) in Eqn. (2).

Recently, Fosgerau and Karlström (2010) have shown that the scheduling and mean-variance models are equivalent under certain conditions\(^4\). Empirical evidence suggests that, however, the valuation of

\(^3\) The scheduling model is also used in Small et al. (1999).

\(^4\) Namely that the scheduling utility function is linear (such as equation 1), there is no discontinuous penalty for being late (i.e., \( \vartheta = 0 \) in equation 1) and the travel time distribution is independent of the departure time. Fosgerau and Karlström (2010) also analysed the case in which the mean and standard deviation of travel time vary linearly with the departure time, and found that the equivalency between the two approaches (scheduling and mean-variance models) does not hold exactly but can be used as an approximation.
travel time variability from a scheduling model may be significantly smaller than that of a mean-variance model (Börjesson et al., 2012). A relationship between the mean and standard deviation of travel time is the simplest construct that can be obtained from empirical data (see also Section 3). Hence, in this paper we adopt the mean-variance approach as the mathematical conceptualisation of TTV in the utility functions for cars and buses in a mode choice context (instead of departure time choice).

2.4 Road pricing and travel time reliability

While most researchers looking at optimal road pricing strategies have focused on the minimising travel times in the network (e.g., Yang and Lam, 1996), maximising total toll revenues (e.g., Joksimovic et al., 2005) and minimising emissions (e.g., Johansson, 1997) or externalities in general, some have suggested looking at pricing strategies from a network reliability perspective (e.g., Brownstone and Small, 2005). Chan and Lam (2005) were among the first to look at the impact of road pricing on travel time reliability, and formulated a reliability-based static user equilibrium problem and optimised toll levels in order to optimise network travel time reliability based on the probability that the travel time is below a certain threshold.

Setting tolls for the optimisation of network travel time reliability in the context of a dynamic user equilibrium was first investigated by Li et al. (2007, 2008) using the standard deviation as a measure of travel time unreliability. Jiang et al. (2011) considers a multicriterion dynamic user equilibrium problem in which travel time, travel cost, and travel time reliability are included, and different vehicle types are considered (i.e. low and high occupancy vehicles). As far as we are aware, road pricing in the context of travel time reliability with respect to cars as well as public transport has not yet been considered in the literature. In this paper we will, for the first time, jointly consider both modes.

3. Empirical relationship between mean and standard deviation of travel time: the Sydney case

In order to estimate a relationship between the mean and standard deviation of travel times, we use a database of floating car data provided by the Roads and Maritime Services (RMS) office of the New South Wales Government in Australia. The data comprises measurements of travel time along 423 roads in Sydney, in which vehicles are equipped with a GPS device. For each road, a particular trip is repeated ten times over two weeks (from Monday to Friday, the first week in October 2011, the second week in March 2012) at the same time each day. Then, for each trip, a normalised mean and standard deviation of travel time [min/km] is calculated over ten observations. In this paper, only major urban roads are considered (highways are not accounted for). The total number of roads is 423. The scatter plot of mean versus standard deviation of travel times is shown in Figure 1.

Considering the mean travel time per unit distance (i.e., divided by the travel distance $L$), $\mu_L$ (min/km), and the standard deviation of travel time per unit distance, $\sigma_L$ (min/km), we conducted linear and nonlinear regressions as shown in Figure 1. Comparing model fits, the linear model with an adjusted $R^2$ of 0.682, slightly outperformed the power function with an adjusted $R^2$ of 0.668 (the constant was not significant and was therefore omitted). The linear model can be compared to Mahmassani et al. (2012), who estimate linear and non-linear (square root and quadratic) relationships between $\mu_L$ and $\sigma_L$ for three locations in the U.S. (Irvine, the Baltimore-Washington Corridor and New York City) using the traffic simulation model DYNASMART. They estimate regression models at four aggregation levels: network, O-D pair, path and link. For their path level model (the equivalent

---

5 We note that ten observations is a relatively small number for calculating stable standard deviations, but note that the estimates are obtained from 423 roads, and we acknowledge the real possibility that the estimates will vary as the number of observations increases; however, we anticipate that the estimates as the number increase would not be as different as might be expected from a random sample over higher numbers of observations. Although currently the available data is limited, new technologies enable larger data sets in the future.
to our regression models), the slope of their linear function lies between 0.25 and 0.53, which is comparable with our finding of 0.316. This number means that an increase of one minute in the mean travel time per kilometre implies an average increase of 19 seconds in the standard deviation of travel time.

![Figure 1: Mean and standard deviation of travel times, Sydney](image)

4. Social welfare maximisation approach

The analytical model used in this paper for the study of optimal TTV pricing of cars and public transport is an extension of the social welfare maximisation model developed by Tirachini (2012) for the analysis of optimal pricing and design of a bus route (which includes frequency and capacity), including congestion and crowding externalities, but ignoring TTV as a factor that influences users choices. Using the present model, we will investigate the impact of including TTV in transport models on the costs for each mode and the design of the bus service. To be more precise, we will determine the optimal car toll \( \tau_{\text{car}} \) (\$/trip), bus fare \( \tau_{\text{bus}} \) (\$/trip), the bus frequency \( f_{\text{bus}} \) (veh/h) and the bus capacity, defined in terms of the bus length \( s_{\text{bus}} \) (m). We first introduce the main variables and inputs into the model, then we describe the behavioural model, and we finish with the optimisation problem formulation.

4.1 Transport corridor definition

We consider a linear bi-directional corridor of length \( L \) and a single period of operation with two directions. The corridor is divided into \( P \) consecutively numbered zones and the total travel demand \( Y^0 \) (trips/hour) per origin-destination pair \((i, j)\) is fixed, where \( i, j \in \{1, \ldots, P\} \). The distance between two consecutives zones \( i \) and \( i+1 \) is denoted by \( L_i \), where both directions are assumed to have the same length. By definition, it holds that \( \sum_{i=1}^{P-1} L_i = L \). Users can choose to travel by car, bus, or decide to walk\(^6\). Denote the (endogenous) travel demand per mode \( m \) by \( y_m^\theta \), where \( m \in \{\text{car, bus, walk}\} \).

We now describe the traffic flows between the \( P \) zones. Let \( q_m^i \) denote the passenger flow in per hour on the road segment between zones \( i \) and \( i+1 \) using mode \( m \). For clarity reasons we will omit the

---

\(^6\) The analytical approach of this section is multimodal: any mode can be included. In the application (Section 5); however we only consider bus, car and walking because these are the main modes in our study area.
direction index from our variables, but note that each variable has a certain direction attached, i.e., from \( i \) to \( i+1 \) is direction 1, and from \( i+1 \) to \( i \) is direction 2. We assume there is exactly one bus stop per zone, and the travel distance between zones is the same for all modes. The passenger flows \( q^i_m \) can be easily determined from the travel demands \( \gamma^i_m \), as indicated in Figure 2. Hence, \( q^i_m \) can be computed as

\[
q^i_m = \sum_{j'=1}^{i} \sum_{j''=i+1}^{P} \gamma^i_m.
\] (3)

The car passenger flow can be converted to traffic flow in cars per hour by considering an average occupancy of a car, denoted by \( o_{\text{car}} \). The car traffic flow \( f^i_{\text{car}} \) (veh/h) can then be easily computed as \( f^i_{\text{car}} = q^i_m / o_{\text{car}} \). For public transport, the reasoning is the other way around. Given the frequency of the bus, denoted by \( f_{\text{bus}} \) (veh/h), we can calculate the occupancy of the bus at each road segment \( i \) as \( o_{\text{bus}}^i = q^i_m / f_{\text{bus}} \). We further note that the average bus headway (the average time between two consecutive buses) is by definition the inverse of the frequency, \( 1/ f_{\text{bus}} \). An important assumption that makes the model more realistic but at the same time more complex is that cars and buses share the same road infrastructure, such that cars and buses interact and influence each other’s travel times, which affects the TTV of both modes. This makes our model much more realistic than most models that assume buses following a strict schedule without any delays and without interactions with other road traffic. People who walk have separate infrastructure available (sidewalks) and are assumed to have fixed travel times. Further, we assume that there are separate bays for buses at the bus stops, such that bus stops do not directly affect cars.

![Transport corridor diagram](https://via.placeholder.com/150)

**Figure 2: Transport corridor diagram**

### 4.2 Mode choice model

Assuming a multinomial logit model\(^7\) for mode choice, the number of trips by mode \( m \) in OD pair \((i, j)\) is given by:

\[
y^i_j = Y^i_j \cdot e^{U^i_m^j} \sum_n e^{U^i_n^j}, \quad \forall i, j, m,
\] (4)

where \( U^i_m^j \) is the utility when travelling from \( i \) to \( j \) using mode \( m \). These utilities for each mode are determined using the following utility functions, which are of the form of the mean-variance model in Eqn. (2):

\[
U^i_{\text{car}} = \delta \left( \frac{c_{\text{car}}^i + \tau_{\text{car}}^i}{o_{\text{car}}} \right) + \alpha_{\text{car}} q^i_m + \rho_{\text{car}} \sigma_{\text{car}},
\] (5)

\(^7\) Although more complex models like nested or mixed logit have advantages over an MNL; the simplicity of the MNL is preferred for the computation of users benefit and social welfare in the optimisation model. Future research should investigate the differences that might result from more advanced choice models.
\[ U_{\text{bus}}^{ij} = M_{\text{bus}} + \delta \tau_{\text{bus}} + \alpha_{\text{bus}} t_{\text{bus}}^{ij} + \rho_{\text{bus}} \sigma_{\text{bus}}^{ij} + \beta_{a} \left( t_{\text{bus}}^{ij} + t_{\text{bus}}^{ij} \right) + \beta_{b} \left( h_{\text{bus}} + \frac{\rho_{\text{bus}}}{\alpha_{\text{bus}}} \sigma_{\text{bus}}^{ij} \right) + \beta_{d} r_{\text{bus}}^{ij} + \beta_{s} s_{\text{bus}}^{ij}, \] (6)

\[ U_{\text{walk}}^{ij} = M_{\text{walk}} + \alpha_{\text{walk}} t_{\text{walk}}^{ij}. \] (7)

Attributes in the utility for the car are the total costs per passenger (running costs \( c_{\text{bus}}^{ij} \) plus the car toll \( \tau_{\text{bus}} \) divided by the car occupancy), the travel time \( t_{\text{bus}}^{ij} \), and the standard deviation of the travel time, \( \sigma_{\text{bus}}^{ij} \). The utility function for the bus contains these same attributes, namely the travel costs per passenger (i.e., the bus fare) \( \tau_{\text{bus}} \), the (in-vehicle) travel time \( t_{\text{bus}}^{ij} \), and the standard deviation of the in-vehicle travel time \( \sigma_{\text{bus}}^{ij} \), plus the following additional attributes: access time \( t_{\text{bus}}^{ij} \), average headway \( h_{\text{bus}} \), the standard deviation of the headway \( \sigma_{\text{bus}}^{ij} \), and two attributes related to crowding, namely the weighted average density of standees per square metre, \( n_{\text{bus}}^{ij} \), and the weighted average proportion of seats occupied, \( p_{\text{bus}}^{ij} \), which both depend on the bus occupancy at each road segment, \( k \). The dummy indicator \( \delta \) (relative to car), a cost coefficient, \( \beta_{a} \), two coefficients for the standard deviation of travel time, \( \beta_{s} \), and several additional parameters for the bus related to the access time, \( \beta_{t} \), the average headway, \( \beta_{h} \), the average density of standees, and the average seat occupancy, \( \beta_{b} \).

Before we present the social welfare objective function, it is important to understand the relationships within the model, in particular with respect to the four decision variables. First of all, the travel times \( t_{\text{bus}}^{ij} \) and \( t_{\text{bus}}^{ij} \) are determined as the sum of the travel times of the consecutive road segments \( k \) (in case driving in direction 1, \( i, ..., j-1 \)), which depend on the number of vehicles (cars and buses, assuming they share the right-of-way) on each road segment and the size of the buses,

\[ t_{\text{bus}}^{ij} = \sum_{k=i}^{j-1} \left[ \frac{L_{k}}{v_{m}^{k}} \left( 1 + \gamma_{0}^{k} \left( f_{\text{car}}^{k} + \phi(s_{\text{bus}}) f_{\text{bus}}^{k} \right) \gamma_{1}^{k} \right) + \xi_{m}^{k} \left( \nu_{b} + \nu_{h}(s_{\text{bus}}) \sum_{l=k+1}^{j} v_{l}^{ij} + \nu_{a}(s_{\text{bus}}) \sum_{l=1}^{k} v_{l}^{ij} \right) \right], m \in \{\text{car, bus}\}. \] (8)

The (congested) car and bus travel times are assumed to be determined by the well-known Bureau of Public Roads (BPR) function, given a maximum speed \( v_{m}^{k} \), road capacity \( K_{k} \) measured in passenger car units (pcu), and road-specific parameters \( \gamma_{0}^{k} \) and \( \gamma_{1}^{k} \) for each road segment \( k \). The number of vehicles (measured in pcu) is equal to \( f_{\text{car}}^{k} = a_{\text{car}} / a_{\text{bus}} \) (see Section 4.1) and the number of buses \( f_{\text{bus}}^{k} \) multiplied with a pcu-value \( \phi(s_{\text{bus}}) \) that depends on the bus size. The dummy indicator \( \xi_{m}^{k} \) equals one if the mode \( m \) is a bus, and zero otherwise. This allows the addition of an extra delay for boarding and alighting a bus, known as dwell time, which includes a fixed time for acceleration/deceleration and time to open and close the doors, \( \nu_{b} \), a time for boarding that depends on the number of boarding passengers, and a time for alighting that depends on the number of alighting passengers, both multiplied with factors that depend on the size of the bus (i.e., number of doors), \( \nu_{h}(s_{\text{bus}}) \) and \( \nu_{a}(s_{\text{bus}}) \), respectively. The walking time \( t_{\text{walk}}^{ij} \) is simply determined by dividing the trip length by the fixed walking speed \( v_{\text{walk}} \), i.e., \( t_{\text{walk}}^{ij} = \sum_{k=i}^{j-1} L_{k} / v_{\text{walk}} \).

The standard deviation \( \sigma_{\text{bus}}^{ij} \) is determined from \( t_{\text{bus}}^{ij} \) based on an empirical relationship such as estimated in Figure 1 (taking the trip length into account). Bus travel time \( t_{\text{bus}}^{ij} \) is composed of running and dwell times (Eqn. (8)) and we assume that both parts are subject to variability. Since the bus shares the same road infrastructure as the car, the standard deviation of running times \( \left( \sigma_{\text{bus}}^{ij} \right) \) is calculated using the same empirical expression used for the car. Adopting results found in Dorbritz et al. (2009), we compute the standard deviation of the dwell time \( \left( \sigma_{d, \text{bus}}^{ij} \right) \) as an empirical function of
the mean dwell time. Under the assumption that running and dwell times are uncorrelated, the standard deviation of the bus travel time, $\sigma_{i,j,\text{bus}}$, can be easily determined as $\sigma_{i,j,\text{bus}} = \sqrt{\sigma_{i,j,\text{tr}}^2 + \sigma_{i,j,\text{bus}}^2}$.

The mean bus headway $h_{i,j,\text{bus}}$ is simply computed as the inverse of the frequency, i.e. $h_{i,j,\text{bus}} = 1/f_{i,j,\text{bus}}$. If bus headway is subject to variability, it is assumed that the standard deviation of the headway is equal to its mean, i.e. $\sigma_{i,j,\text{bus}} = h_{i,j,\text{bus}}$. This result stems from assuming that the arrival of buses at bus stops follows a Poisson distribution, as done in several models that consider random bus arrival times at bus stops (e.g., Delle Site and Filippi, 1998; Cominetti and Correa, 2001; Cepeda et al., 2006; Cortés et al., 2011). Implicit in (6) is that we ignore any correlation between headway and travel time in the specification of the bus utility function. The understanding and inclusion of such correlation is a matter of further research.

The crowding attributes are computed as follows. Let $\eta(s_{\text{bus}})$ be the total number of available seats and let $\varphi(s_{\text{bus}})$ the total amount of square metres standing space, respectively, both depending on the bus size. For each road segment $k$, the density per standee is zero if $\alpha^k_{i,j,\text{bus}} \leq \eta(s_{\text{bus}})$, i.e. everyone can sit, and is equal to $(\alpha^k_{i,j,\text{bus}} - \eta(s_{\text{bus}}))/\varphi(s_{\text{bus}})$ otherwise, and the proportion of occupied seats is 1 if $\alpha^k_{i,j,\text{bus}} \geq \eta(s_{\text{bus}})$, and is equal to $\alpha^k_{i,j,\text{bus}}/\eta(s_{\text{bus}})$ otherwise. The trip density and trip proportion of occupied seats is a weighted sum over all road segments, where the weight is a function of the mean and standard deviation of the travel times inside the vehicle, i.e. for direction 1 we obtain

$$h_{i,j,\text{bus}} = \sum_{k=1}^{k_{\text{end}}} \left[ \frac{P_{i,j,\text{bus}}}{\sigma_{i,j,\text{bus}}} \right] \cdot \max \left\{ \frac{\alpha^k_{i,j,\text{bus}} - \eta(s_{\text{bus}})}{\varphi(s_{\text{bus}})} \right\},$$

$$P_{i,j,\text{bus}} = \sum_{k=1}^{k_{\text{end}}} \left[ \frac{f_{i,j,\text{bus}}}{\alpha_{i,j,\text{bus}}} \right] \cdot \min \left\{ \frac{\alpha^k_{i,j,\text{bus}}}{\eta(s_{\text{bus}})} \right\}.$$  

We have tried to put as much as possible realism into the mode choice model, however, this also leads to a more complex model. In particular the interdependent relationships in the model make it more challenging to solve. The mode choice (through the utility functions) depends (among others) on travel times, travel time variability and bus occupancy, resulting in numbers of car and bus passengers, which in the end again influence the travel times, travel time variability, and bus occupancy. Therefore, for each evaluation of the design inputs (car toll, bus fare, bus frequency, and bus size), we need to iteratively find a solution to what the literature calls a fixed point problem.

### 4.3 Social welfare design problem

In this subsection we will mathematically formulate the social welfare (SW), which in economics is defined as the sum of consumer and producer surplus (in $)$:

$$SW = \sum_{g} \left( \frac{Y^g}{\delta} \right) \ln \sum_{m} e^{\ln \left( \frac{Y^g}{\delta} \right) + \ln \left( \frac{Y^g}{\delta} \right) \tau_{\text{car}} + \frac{Y^g}{\delta} \tau_{\text{bus}}} - \left( c_1(s_{\text{bus}})P + c_2(s_{\text{bus}})f_{\text{bus}} \left( t_{\text{bus}} + t_{\text{bus}}^i + t_i \right) + 2c_3(s_{\text{bus}})f_{\text{bus}} L \right)$$

The consumer surplus consists of the well-known logsum of the utility functions, multiplied by the number of consumers (trips), divided by the (negative) cost coefficient in the utility functions in order to monetise the value. The producer surplus consists of the total car toll revenues and the total bus fare revenues, minus the bus operator costs. These costs are split up in bus stop/station costs (where $P$ is the number of stops and $c_1(s_{\text{bus}})$ is the infrastructure cost per station that depends on the bus size), personnel and vehicle capital costs (where $f_{\text{bus}}(t_{\text{bus}} + t_{\text{bus}}^i + t_i)$ is the total travel time by all buses spent in both directions, considering a slack time $t_i$ at termini, and $c_3(s_{\text{bus}})$ are the hourly costs depending
on the bus size), and running costs (where $2f_{\text{bus}}L$ is the total distance spent by all buses in both directions and $c_i(s_{\text{bus}})$ are the costs per kilometre).

Hence, our social welfare optimisation problem can be mathematically formulated as follows:

$$
\begin{align*}
\max_{\tau_{\text{car}}, \tau_{\text{bus}}, f_{\text{bus}}, s_{\text{bus}}} & \quad SW \\
\text{s.t.} & \quad \tau_{\text{car}}, \tau_{\text{bus}} \geq 0, \\
& \quad \omega(s_{\text{bus}}) f_{\text{bus}} \geq \max_i q_i, \\
& \quad f_{\text{min}} \leq f_{\text{bus}} \leq f_{\text{max}}, \\
& \quad s_{\text{bus}} \in S, \\
\end{align*}
$$

(12) - (16)

Non-negativity constraints (13) are set to avoid negative car tolls or bus fares, constraint (14) guarantees that the bus capacity (in both directions) is sufficient to accommodate the maximum number of bus passengers, where $\omega(s_{\text{bus}})$ is the capacity (crush capacity plus some spare capacity) of a bus of size $s_{\text{bus}}$, constraint (15) ensures that the bus frequency falls within possible upper and lower limits defined by policies, and constraint (16) states that only certain discrete bus sizes can be chosen, dependent on the sizes available on the market denoted by set $S$. Finally, of course the results need to be consistent with mode choice behaviour and resulting travel times and bus occupancies described in the previous two subsections.

The above constrained optimisation is solved using the optimisation toolbox of Matlab. The solution procedure implemented considers bus frequency as a continuous variable while the bus length, car toll and bus fare are discrete (fare and toll are constrained to be a multiple of 5 cents).

5. Numerical application

5.1 Physical setting, input parameters and assumptions

The social welfare maximisation model is applied with demand and supply data from Military Road in North Sydney, shown in Figure 3. This road is well-known locally as very congested in the peak hours, and is a main bus corridor. The section modelled comprises $L = 3.44$ km of a two-lane road in both directions which is divided into 12 zones (therefore the average zone length $L_i$ is 286 metres). The origin-destination matrix of trips $Y^{ij}$ for all modes combined (car, bus and walking) for the morning peak (7.30 to 8.30am) is shown in Figure 4.

All input parameters related to the bus size are summarised in Table 1, where the pcu values have been based on Basso and Silva (2010). Boarding and alighting times have been derived based on off-board fare collection (Tirachini, 2012). Buses only stop when signalled. Cost parameters are been adopted from Tirachini (2012). The behavioural coefficients in the utility functions of the mode choice model are listed in Table 2, which are based on model estimations (Tirachini, 2012) from a static choice survey conducted in Sydney (Hensher et al., 2011). Coefficients are shown for a model without considering crowding, and a model taken crowding into account (see values between brackets). The mode specific constants $M$ have been calibrated based on the current Sydney model split (for trips < 5 km) of 62.5 percent car, 31.6 percent walk, and 5.9 percent bus (TDC, 2010), using a current bus frequency of 16 buses/hour in the morning peak, a bus fare of $2.10, and no car toll. Bus access and egress times, $t_{\text{bus}}^i$ and $t_{\text{bus}}^j$, are assumed to be 5 minutes. Bus slack time at termini is also assumed to be 5 minutes. The calculation of the number of seats and floor space for standees in Table 1 include a number of technical restrictions (i.e., doors must be clear of seats), see Tirachini (2012) for details.

Note that since we did not have data available to estimate the coefficients for the standard deviation parameters, $\rho_m$, we will assume a series of different values in our numerical analyses.

---

8 As the model estimated in Tirachini (2012) only includes mean travel times, the valuation of uncertainty is embedded in the parameters estimated. The reliability parameters are introduced on top of the average travel
Remaining input parameters are car running costs of $0.14/km for determining $c_{ijc}$, an average car occupancy $o_{car} = 1.45$ (TDC, 2010), which we assume remains unchanged after pricing reforms, (free-flow) speeds $v_{car} = v_{bus} = 50$ km/h for all road segments $i$ and $v_{walk} = 4$. Further, for the travel time functions we assume the values of $\gamma_0 = 0.15$, $\gamma_1 = 4$, and $K = 2000$ pcu/h for the entire corridor (assuming an effective green-time on signalised intersections of 60 percent on a two-lane road).
Table 1: Bus attributes and their values per bus type

<table>
<thead>
<tr>
<th>Bus attribute</th>
<th>mini</th>
<th>standard</th>
<th>rigid long</th>
<th>articulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>length (m)</td>
<td>8</td>
<td>12</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>number of doors (--)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>pcu value $\varphi(s_{bus})$ (--)</td>
<td>1.65</td>
<td>2.19</td>
<td>2.60</td>
<td>3.00</td>
</tr>
<tr>
<td>boarding time $\psi(s_{bus})$ (sec)</td>
<td>0.88</td>
<td>0.63</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>alighting time $\psi(s_{bus})$ (sec)</td>
<td>0.88</td>
<td>0.63</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>floor space $\psi(s_{bus})$ (m$^2$)</td>
<td>3</td>
<td>5</td>
<td>6.2</td>
<td>7.7</td>
</tr>
<tr>
<td>number of seats $\eta(s_{bus})$ (--)</td>
<td>24</td>
<td>40</td>
<td>50</td>
<td>62</td>
</tr>
<tr>
<td>stop/station cost $c_1(s_{bus})$ ($/h$)</td>
<td>4.4</td>
<td>6.5</td>
<td>8.7</td>
<td>10.9</td>
</tr>
<tr>
<td>driver/capital cost $c_2(s_{bus})$ ($/h$)</td>
<td>42.7</td>
<td>49.5</td>
<td>54.5</td>
<td>59.6</td>
</tr>
<tr>
<td>operating cost $c_3(s_{bus})$ ($/km$)</td>
<td>0.9</td>
<td>1.3</td>
<td>1.4</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table 2: Parameter values of the utility functions in the mode choice model

<table>
<thead>
<tr>
<th>Behavioural coefficients</th>
<th>Car</th>
<th>Bus</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost $\delta$</td>
<td>-0.110 (-0.111)</td>
<td>-0.110 (-0.111)</td>
<td>--</td>
</tr>
<tr>
<td>$M_m$</td>
<td>--</td>
<td>-1.922 (-1.976)</td>
<td>-0.101 (-0.109)</td>
</tr>
<tr>
<td>Travel time $\alpha_m$</td>
<td>-0.019 (-0.021)</td>
<td>-0.019 (-0.021)</td>
<td>-0.035 (-0.035)</td>
</tr>
<tr>
<td>Standard deviation $\rho_m$</td>
<td>variable</td>
<td>variable</td>
<td>--</td>
</tr>
<tr>
<td>Access/egress time $\beta_a$</td>
<td>--</td>
<td>-0.021 (-0.019)</td>
<td>--</td>
</tr>
<tr>
<td>Headway $\beta_h$</td>
<td>--</td>
<td>-0.009 (-0.010)</td>
<td>--</td>
</tr>
<tr>
<td>Density $\beta_d$</td>
<td>--</td>
<td>(-0.003)</td>
<td>--</td>
</tr>
<tr>
<td>Occupied seats $\beta_s$</td>
<td>--</td>
<td>(-0.012)</td>
<td>--</td>
</tr>
</tbody>
</table>

For computing the standard deviations, we assume the linear empirical relationship found in Figure 1 for $\sigma^2_{\text{car}}$ and $\sigma^2_{\text{bus}}$, while for $\sigma^2_{\text{bus}}$ we take 50 percent of the mean dwell time (consistent with Dorbritz et al. (2009) who found that the standard deviation is between 36 and 58 percent of the mean boarding time in Zürich, Switzerland).

With these inputs, 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).

We would like to examine what the impact on optimal prices and bus design is under different assumptions on travel time variability. Since we do not know the current reliability ratio for Sydney, we consider four reliability ratios for the car mode, namely, $\rho_{\text{car}}/\alpha_{\text{car}} = 0.5, 1.0, 1.5$ and 2.0, in the range of the values estimated in the literature (two recent reviews are Li et al., 2010; and Carrion and Levinson, 2012). The reliability ratio of buses is usually expected to be larger than that of cars because of the discrete nature of bus departures. For example, Bates et al. (2001) suggest reliability ratios of around 1.3 for cars and “somewhat higher” for public transport, whereas De Jong et al. (2009) suggest reliability ratios of 0.8 for cars and 1.4 for public transport. We assume that the reliability ratio of buses is 50 percent larger than that of cars in each case, i.e., $\rho_{\text{bus}}/\alpha_{\text{bus}} = 1.5(\rho_{\text{car}}/\alpha_{\text{car}}) = 0.75, 1.5, 2.25$ and 3.0.

Our model is capable of including four sources of externalities, namely (i) congestion, (ii) crowding, (iii) TTV, and (iv) headway variability. In our analyses, we consider the following four scenarios:
Scenario S1: inclusion of congestion and TTV
Scenario S2: inclusion of congestion, TTV, and headway variability
Scenario S3: inclusion of congestion, TTV and crowding
Scenario S4: inclusion of congestion, TTV, crowding and headway variability

The following subsections present the results. First (Section 5.2), we will analyse the effect of the sensitivity to TTV on prices and bus design in detail, assuming all sources of externalities are included (scenario S4). Then in Section 5.3, we investigate the effect of crowding and headway variability on prices and bus design. Section 5.4 examines the effect of different cost sensitivities (i.e., different income groups), and in Section 5.5 we look at the common situation of no road pricing in addition to a public transport budget constraint. Finally, Section 5.6 discusses the impact of an increased (future) travel demand.

5.2 Sensitivity to travel time variability

Results with the current OD matrix (Figure 4) are shown in Table 3, in which we have optimised the social welfare by setting optimal values for our optimisation variables (shaded in the table) for the four different reliability ratios. First, we study the sensitivity of the solution on optimal pricing (\( \tau_{\text{car}} \) and \( \tau_{\text{bus}} \)) and bus service design (\( f_{\text{bus}} \) and \( s_{\text{bus}} \)) to the increasing values of TTV (given by the reliability ratio). We consider the scenario in which four sources of externalities are included, namely congestion, crowding, TTV, and headway variability (scenario S4).

<table>
<thead>
<tr>
<th>Reliability ratio car (( \rho_{\text{car}} / \alpha_{\text{car}} ))</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal bus fare ($)</td>
<td>0.35</td>
<td>0.35</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>Optimal car toll ($)</td>
<td>1.35</td>
<td>1.55</td>
<td>1.75</td>
<td>1.90</td>
<td>2.05</td>
</tr>
<tr>
<td>Optimal bus size (m)</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Optimal bus frequency (veh/h)</td>
<td>26</td>
<td>32</td>
<td>33</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td>Bus headway (min)</td>
<td>2.3</td>
<td>1.9</td>
<td>1.8</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Bus fleet size (--)</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Social welfare ($)</td>
<td>67,653</td>
<td>66,162</td>
<td>64,731</td>
<td>63,345</td>
<td>61,996</td>
</tr>
<tr>
<td>Consumer surplus ($)</td>
<td>57,420</td>
<td>54,547</td>
<td>51,664</td>
<td>49,277</td>
<td>46,934</td>
</tr>
<tr>
<td>Bus operator profit ($)</td>
<td>-437</td>
<td>-542</td>
<td>-557</td>
<td>-629</td>
<td>-692</td>
</tr>
<tr>
<td>Subsidy/bus operator cost (--)</td>
<td>0.47</td>
<td>0.53</td>
<td>0.50</td>
<td>0.53</td>
<td>0.56</td>
</tr>
<tr>
<td>Maximum bus load factor (--)</td>
<td>0.90</td>
<td>0.77</td>
<td>0.70</td>
<td>0.63</td>
<td>0.59</td>
</tr>
<tr>
<td>Average bus load factor (--)</td>
<td>0.48</td>
<td>0.42</td>
<td>0.38</td>
<td>0.34</td>
<td>0.32</td>
</tr>
<tr>
<td>Toll revenue ($)</td>
<td>10,669</td>
<td>12,157</td>
<td>13,624</td>
<td>14,697</td>
<td>15,754</td>
</tr>
<tr>
<td>Car speed direction 1 (km/h)</td>
<td>26.3</td>
<td>26.4</td>
<td>26.4</td>
<td>26.5</td>
<td>26.5</td>
</tr>
<tr>
<td>Car speed direction 2 (km/h)</td>
<td>21.6</td>
<td>21.7</td>
<td>21.9</td>
<td>22.0</td>
<td>22.1</td>
</tr>
<tr>
<td>Bus speed direction 1 (km/h)</td>
<td>20.6</td>
<td>20.9</td>
<td>21.2</td>
<td>21.3</td>
<td>21.4</td>
</tr>
<tr>
<td>Bus speed direction 2 (km/h)</td>
<td>17.5</td>
<td>17.8</td>
<td>18.1</td>
<td>18.3</td>
<td>18.5</td>
</tr>
<tr>
<td>Modal split bus (%)</td>
<td>7.3</td>
<td>7.2</td>
<td>7.2</td>
<td>7.2</td>
<td>7.1</td>
</tr>
<tr>
<td>Modal split car (%)</td>
<td>59.6</td>
<td>59.1</td>
<td>58.7</td>
<td>58.3</td>
<td>57.9</td>
</tr>
<tr>
<td>Modal split walk (%)</td>
<td>33.1</td>
<td>33.6</td>
<td>34.1</td>
<td>34.5</td>
<td>34.9</td>
</tr>
</tbody>
</table>

A scenario with travel time variability without headway variability is a proxy for a system in which headway variability can be greatly controlled, for example, through real-time control strategies such as vehicle holding, traffic light pre-emption, bus stop skipping and deadheading. These strategies can be applied to keep intervals at a desired level, at the expense of a likely increase in the variability of running times.
As the sensitivity of users to travel time variability increases (higher reliability ratios), the optimal car toll increases approximately linearly, from $1.35 when $\rho_{car}/\alpha_{car} = 0$ (i.e., users are assumed insensitive to travel time variability) to $2.05$ when $\rho_{car}/\alpha_{car} = 2$ (currently the car toll is $0$). On the other hand, the optimal bus fare remains almost constant, on either $0.35$ or $0.40$ (currently $2.10$). That is, even though both car and bus users contribute to increases in travel time (and headway) variability, the contribution of car users is much higher and that is reflected in the socially optimal bimodal pricing structure. Due to the low bus demand on this corridor (5.9 percent, see Section 5.1), the optimal bus size is constant at 8 meters (mini bus). With these optimal settings, Table 3 shows a reduction in the number of car trips and an increase in bus and walking trips.

The larger the reliability ratio, the larger the optimal frequency, which ranges from 26 to 40 buses per hour (currently 16 buses per hour). As the reliability ratio grows, it is optimal to have a greater frequency in order to reduce not only the mean headway but also its standard deviation. The increase in optimal frequency is attached to an increase in the number of buses required for the service, which grows from 13 to 19, at the expense of increasing both operating and capital costs\(^{10}\). The consumer surplus and total social welfare are lower if users are more sensitive to travel time variability. The optimal public transport subsidy (first best) amounts to between 47 and 56 percent of the total operator cost, but the total loss is more than compensated by the toll revenue (without accounting for toll collection costs).

The average car speed is slightly higher than the current speed of 21.5 km/h in the peak direction (direction 2, towards Zone 1 in Figure 3) whereas bus speed including stops for boarding and alighting is between 17.5 and 18.5 km/h in the same direction. The modal speeds are the result of the number of buses and cars on the road. With larger reliability ratios there are less cars given the decreasing modal split, but due to increasing bus frequencies, there are more buses on the road. The net effect is a slight increase in speed for both modes.

When compared to the case of a zero reliability ratio, total toll revenues increase by 28 percent if the value of reliability for cars is equal to the value of travel time savings (i.e., $\rho_{car}/\alpha_{car} = 1$) and by 48 percent if the value of reliability is double the value of travel time savings (i.e., $\rho_{car}/\alpha_{car} = 2$). This is an indication of the substantial effects that including TTV into an optimal (first best) transport pricing scheme may have on toll revenue. Note that the increase in revenue happens in spite of the reduction in the total number of cars trips induced by a higher reliability-sensitive toll. As the reliability ratio increases, more people walk because walking is most reliable as it is assumed not affected by congestion. Maximum and average load factors (number of passengers over number of seats) also decrease due to the reduction of optimal headway\(^{11}\). The low occupancy levels are a result of including crowding parameters in the bus utility function (Tirachini, 2012).

5.3 Sensitivity to crowding and/or headway reliability

The previous analysis (scenario S4) was carried out assuming that crowding is a source of disutility for users, and that bus headways are subject to variability. In this section, we compare solutions on optimal pricing for cars and buses for scenarios in which neither are considered (S1), only headway variability is considered (S2), or only crowding is considered (S3).

The main results on the comparison of scenarios S1 to S4 are summarized next. Figure 5 shows that the optimal bus frequency is quite sensitive to the assumptions on the sources of disutility for users. If only congestion and TTV (for both cars and buses) are taken into account (S1), the optimal bus

\(^{10}\) The fleet size calculated in Table 3 includes that 5 percent of the fleet is kept in reserve at garages.

\(^{11}\) Load factors are calculated with total bus demand and aggregated bus frequency in the period under analysis, i.e., unbalanced load factors due to bunching effects are not included in the calculation of modal utility and demand.
frequencies remain almost constant regardless the assumed reliability ratio. In this scenario, the optimal bus size is 8 metres for all reliability ratios. However, when taking crowding into account (S3), the optimal bus frequencies decrease somewhat (from 27 to 21 buses per hour), however the optimal bus size increases to 12 metres (for reliability ratios of 0.75 and 1.5) and 15 metres (for ratios 2.25 and 3.0). In other words, when taking crowding into account but not headway variability, there will be less but larger buses.

In the two scenarios with headway variability (S2 and S4), the optimal frequency is steadily increased as the reliability ratio increases, a result previously observed in Table 3. In these scenarios, increasing frequency has the extra benefit of reducing the cost associated with headway variability. Increasing the frequency turns out to be more beneficial to do rather than increasing bus size (the optimal bus size is constant at 8 m). Finally, note that the optimal frequency when accounting for crowding externalities (S4) is larger than when crowding externalities are not considered (S2), an outcome also obtained in models that ignore travel time variability (Jara-Díaz and Gschwender, 2003; Tirachini et al., 2014).

The optimal prices for the different scenarios under different assumptions of the reliability ratio are depicted in Figure 6. All scenarios (S1 to S4) produce more or less the same optimal car tolls, hence only a single line is plotted for car tolls. This means that optimal car toll is given by congestion and travel time variability externalities, and is mostly insensitive to bus-specific attributes (passenger crowding and headway variability). On the other hand, including headway variability has no noticeable impact on the optimal fare (hence S1 produces the same results as S2, and S3 yields the same outcomes as S4), but including crowding increases the optimal bus fare for all reliability ratios. Regardless of the scenario considered, similar to Table 3, the optimal bus fare is insensitive to the reliability ratio, in contrast to the optimal car toll which steadily increases. In summary, bus frequency is sensitive to both crowding externalities and variability of travel times and headways, whereas optimal fare is only sensitive to crowding; the marginal cost of an extra passenger worsens crowding conditions but has a negligible effect in bus TTV.

The optimal bus frequency for different scenarios and reliability ratios is depicted in Figure 5.
5.4 Sensitivity to different marginal utilities of income

The present model assumes a fixed cost parameter $\delta$, which means that a single marginal utility of income is used for all income groups in the sample. It is worth analysing whether previous results regarding optimal pricing and bus service design hold when different cost sensitivities are explicitly considered. We analyse this issue by comparing the base case (scenario S4) against two extreme cases: (i) a low cost sensitivity in which $\delta$ is halved, representing a high income situation, and (ii) a high cost sensitivity in which $\delta$ is doubled, representing a low income situation. Results for optimal car toll, fare and bus frequency are shown in Figures 7, ranging from low cost sensitivity to high cost sensitivity.

As expected, the lower the sensitivity to cost (high income groups), the higher the optimal fare and car toll. Since the optimal prices are higher with higher reliability ratios, it holds that optimal prices are significantly higher for populations with a high income and a high sensitivity to travel time. Low cost sensitivities also results in higher optimal bus frequencies, while the optimal bus size does not change.
5.5 No road pricing and a public transport budget constraint

We can solve the problem of bus optimisation for the second best case in which there exists no road pricing instrument. As in Ahn (2009) and Tirachini (2012), when a budget constraint is not imposed on the bus operator, the second-best bus fare is negative in this scenario. Therefore, we consider the more realistic case in which there is a budget constraint that forces the bus operator to be self-sustained. In this case, optimal bus fare increases to values in the range between $0.70 (low reliability ratio) and $0.95 per trip (high reliability ratio), i.e., the optimal fare is roughly doubled relative to the first best case (see Table 3). Bus modal split reduces from 7.2 percent in Table 3 to values in the range 6.1-6.6 percent in the second best case, and optimal frequency also decreases compared to the base case.

5.6 Increased travel demand

In this section, we analyse how the bus service and pricing levels change when faced with an increase in travel demand (e.g., future urban densification around the corridor). We assume that it is not possible to increase road capacity, which is a realistic assumption in our corridor. The idea is to analyse the impact on the optimal design variables when the transportation system is stressed and severe congestion results. The trips in the OD matrix in Figure 4 are uniformly increased up to a total demand of 28,850 trips/h (i.e., an increase of 50 percent). Results of the optimal pricing structures and bus headways are shown in Figure 8. Considering scenario S4, two cases regarding variability are compared: (i) no variability cost (\( \rho_{car} = \rho_{bus} = 0 \)), hence the reliability ratio is zero, and (ii) an average fixed reliability ratio of 1.5 for buses (which means a ratio of 1.0 for cars). It is observed that the optimal car toll steadily increases as total demand grows, whereas the optimal bus fare remains almost constant (Figure 8(a)). The optimal bus frequency when TTV and headway variability are accounted for is always higher than when no variability is considered (Figure 8(b)).
6. Conclusions and policy implications

In this paper, we have studied the optimal pricing structure of both cars and buses when travel times and bus headways are subject to variability, in which travellers’ value reductions in both the mean and standard deviation of travel time. We have proposed a methodology in which we maximise social welfare by determining optimal values for the car toll, bus fare, bus frequency and bus size. As we have argued in the paper, this optimisation is far from trivial due to the complex relationships between the above mentioned design variables and variability in the travel times and bus headways.

In order to analyse the impact of including variability in the design process, we considered an important transport corridor in Sydney as a case study in which people can travel by car, bus, or can walk. The behavioural processes are modelled with a high level of detail, including running times and costs, dwelling times, travel time variability, bus headway variability, and crowding (seat availability and density of standees).

First, we looked at the impact of travel time variability on the optimal prices. We find that as the sensitivity of users to travel time variability increases, the optimal car toll increases approximately linearly, whereas the optimal bus fare remains almost constant in most cases. Even though both car and bus users contribute to increases in travel time (and headway) variability, the contribution of car users is much higher, and that is reflected in the socially optimal car tolls and bus fares. This result was obtained in a number of different scenarios, including alternative assumptions regarding crowding externalities and travel time and headway variability associated with the bus mode, and for different levels of total transport demand. Lower cost sensitivity (i.e., high income) leads to a significant increase in the car toll, while the bus fare is only slightly affected. While a positive relationship between travel time variability and optimal car tolls has been found in the literature, as far as we are aware this study shows for the first time a relationship between travel time variability, optimal bus fares and car tolls in an integrated fashion.

Second, we examined the impact of travel time variability on optimal bus frequency. If the bus headway is variable, then including travel time variability impacts upon the optimal bus frequencies. The higher the reliability ratio (that is, sensitivity to such variability), the higher the optimal frequencies. Taking crowding into account also leads to higher optimal frequencies, although in some scenarios the bus capacity is also increased by using longer buses instead of merely increasing frequencies.

Our study has a number of relevant policy implications in a multimodal network context. In recent years, empirical studies have increasingly shown that travel time variability in congested road environments has an important behavioural role in travel decisions (especially commuter trips that are repeated on a regular basis), that is additional to the actual travel time of a specific trip. Nonetheless, only a few governments formally recognise the important role of crowding and trip time variability and introduce these additional sources of user benefit into guidelines for project appraisal. For example, in The Netherlands there is a mark up of 25% on travel time savings due to reductions in travel time variability under congested road conditions (CPB, 2004), while in Australia, Transport for New South Wales released in 2013 a new version of its guidelines (TfNSW, 2013) which provides multipliers as mark ups for crowding levels and reliability of time relative to mean travel time. On the other hand, the road pricing experiences of London and Stockholm have shown that the application of urban tolls can reduce the variability of travel times (Transport for London, 2007; Eliasson, 2009), resulting in a significant user benefit that deserves quantification in project appraisal. This paper recognises this source of user benefit in a broader multimodal context; specifically showing through scenario analysis, the effect of travel time variability on the design of the transport system including public and private transport; notably the effect on optimal car tolls and bus frequencies, and to a lesser extent on bus fare and bus size. Car drivers will accept higher tolls if they not only improve the travel time but also travel time variability; and as shown in our case study this can potentially lead to
sizeable increases in total toll revenues. Therefore, the findings of our study reinforce the need of including travel time variability (and crowding) in formal cost-benefit analysis of pricing reforms, road investments and public transport service design problems from the early stages of project appraisal.

Several extensions and venues of further research originate from the topic of this article. First, the complex interrelationships that are present in public transport service provision, in particular the possible correlations between headway (which influences waiting time and scheduling delay), dwell time and in-vehicle time, and the correlations between crowding and reliability, are topics open for further enquiry. In the bus context (in contrast to car travel), an amount of the variability in travel time can be attributed to crowding at stops and in the vehicle, given it is a source of delay in entering and departing a bus. Second, the fact that travel time variability may decrease as a function of the inflow when there is severe congestion (Dion et al., 2004; Tu, 2008) and that different types of roads may have different travel time variability functions, were not considered in this study and should be assessed in future research efforts. Third, the unreliability of low frequency public transport systems that work with timetables and have a scheduling delay attached should be also included. Fourth, as previously mentioned, the analysis of a full-scale city-wide scenario in which more modal alternatives are in place (e.g., rail, bicycle) is a natural extension of the one-corridor one-period analysis undertaken in this paper, in order to uncover the implications on optimal toll system design (location of toll points/gates) and pricing levels of all private and public transport alternatives, when accounting or ignoring travel time variability as a source of disutility for users.

References


Pu, W., 2011. Analytic relationships between travel time reliability measures. Transportation Research Record 2254, 122-130.


