Transport Reviews: A Transnational Transdisciplinary Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ttrv20

Multimodal Transport Pricing: First Best, Second Best and Extensions to Non-motorized Transport

Alejandro Tirachini [a] & David A. Hensher [a]

[a] Institute of Transport and Logistics Studies, The Business School, The University of Sydney, NSW 2006, Australia

Available online: 30 Nov 2011

To cite this article: Alejandro Tirachini & David A. Hensher (2012): Multimodal Transport Pricing: First Best, Second Best and Extensions to Non-motorized Transport, Transport Reviews: A Transnational Transdisciplinary Journal, 32:2, 181-202

To link to this article: http://dx.doi.org/10.1080/01441647.2011.635318

Transport Reviews, Vol. 32, No. 2, 181–202, March 2012

Multimodal Transport Pricing: First Best, Second Best and Extensions to Non-motorized Transport

ALEJANDRO TIRACHINI AND DAVID A. HENSHER

Institute of Transport and Logistics Studies, The Business School, The University of Sydney, NSW 2006, Australia

(*Received 20 February 2011; revised 19 October 2011; accepted 22 October 2011*)

ABSTRACT *In this paper, we examine the main concepts of transport pricing in an urban environment, focusing on the automobile, public transport and walking or cycling as travel alternatives. A review of the literature on the first-best and second-best pricing policies is provided, with an emphasis on public transport pricing, including the setting of frequency and vehicle capacity, the influence of bus congestion externalities and the interactions between transport pricing reforms and the broader tax system. A model is developed to analyse the impact of non-motorized transport on optimal public transport pricing policy, congestion interactions between cars and buses associated with the transfer of passengers at bus stops and the existence of a capacity constraint within the public transport mode.*

1. Introduction

There have been extensive efforts made to analyse the merits of road pricing as a tool to manage congestion and other transport externalities; however, the analysis of road pricing for private transport has received a disproportionate amount of attention relative to public transport and multimodal analysis. In this paper, we focus on multimodal pricing, with an emphasis on public transport and the influence of non-motorized transport on optimal pricing decisions.

First, we review the main concepts associated with the economics of public transport pricing. Second, we develop a multimodal pricing model incorporating automobile, public transport and non-motorized transport. This model extends the previous literature by identifying the role that non-motorized transport can play in the optimal setting of fares for public transport, an issue raised by Kerin (1992). We analyse how the optimal fare, frequency and vehicle size should be determined when the capacity constraint is binding for a public transport service, that is, when demand meets the capacity offered by the operator. The imbalance in the demand distribution throughout the day is often associated with a binding capacity constraint in high-demand peak periods, as observed in many public transport systems. In the framework, we include the cost of

Correspondence Address: Alejandro Tirachini, Institute of Transport and Logistics Studies, The Business School, The University of Sydney, NSW 2006, Australia. Email: alejandro.tirachini@sydney.edu.au

0144-1647 print/1464-5327 Online/12/020181-22

http://dx.doi.org/10.1080/01441647.2011.635318

externalities other than congestion, such as accidents, pollution and noise, and the toll collection cost, all of which increase the marginal cost of motorized transport compared with walking or cycling. The emphasis of this paper is not on the numerical value of optimal fares and subsidies as reported in the literature, but on the underlying economic principles.[1]

The paper is organized as follows. Section 2 reviews the basic concepts on optimal pricing in urban transport. Section 3 extends the result on fares to the analysis of several relevant outputs, such as frequency and capacity in bus transport, the need for subsidies and interactions with other sectors of the economy, and Section 4 develops a pricing model highlighting the main insights, with conclusions being provided in Section 5.

2. Setting Public Transport Fares: The First-Best and Second-Best Models

The analysis of transport pricing schemes usually distinguishes between the first-best (in which all prices match marginal costs) and the second-best policies. As reviewed by Quinet (2005), in the first-best world, there are no external effects, no public goods, firms are price-takers, there is no tax or taxes are optimal, there is no uncertainty or asymmetry in information, and there are no transaction costs and no redistribution issues. However, transport systems in the real world do not match these conditions, creating a *second-best* outcome.[2] Technological or acceptability constraints impose second-best situations within the transport sector, given the impossibility of taxing at marginal costs all modes or all locations in a network.

2.1 *The First-Best Pricing*

The principles of marginal cost pricing of private transport have a long history.[3] In the context of automobiles, it has been recognized that establishing a cost function for the study of demand and welfare must include travel time as a key factor. In the study of public transport pricing, Mohring (1972), Turvey and Mohring (1975) and Jansson (1979) were the first to recognize this. The addition of user time costs as an input in the social cost function of public transport proved to have remarkable consequences for the application of the marginal cost pricing rule. When an increase in demand is met by an increase in the frequency of service, the travel cost of all users decreases due to savings in waiting time (assumed to be inversely related to frequency), a phenomenon that is not observed when only operator cost is considered in the public transport cost function. Consequently, marginal cost lies below the average cost, which is the first-best argument for subsidizing public transport operation, as introduced by Mohring (1972) and Turvey and Mohring (1975). Intuitively, a lower fare will encourage more travellers to use public transport, which would be accompanied by an increase in the optimal frequency that produces benefit for all passengers (Jansson, 1993).

The first-best fare is the one that maximizes social welfare, defined as the sum of user and operator benefits. The unrestricted solution of this problem is a well-known result, that the optimal public transport fare equals total marginal cost (i.e. the summation of user and operator marginal costs) minus the average user cost (e.g. Else, 1985; Tisato, 1998). The principle of marginal cost pricing as a means to achieve economic efficiency applies to public transport services, but with the subtraction of what users already 'pay' when using the service, that is, their own time (Jara-Díaz, 2007).

Several refinements to these basic principles have been introduced. The contribution of Tabuchi (1993) highlights a renewed interest in the properties of the bimodal equilibrium between private and public transport under different pricing regimes. Instead of assuming static congestion for the automobile, Tabuchi assumed a dynamic bottleneck that arises when the flow of cars exceeds the capacity of the road (Vickrey, 1969; Arnott *et al.*, 1993). With a highly stylized model that ignores travel time as a cost for rail users and capacity constraints, Tabuchi showed that as demand grows, it is more attractive to have a rail-based alternative competing with cars, due to economies of scale in the former mode and congestion externality in the latter mode. Subsequently, Danielis and Marcucci (2002) extended Tabuchi's two-mode approach to include budget constraints on rail operations, and Huang (2002) introduced a stochastic (logit) modal choice model.

A different approach was presented by Kraus and Yoshida (2002), who adopted the highway bottleneck model of Vickrey (1969) for the modelling of rail commuting, assuming that users arrive at stations at the same time as trains do. They showed that the average user cost increases with demand, that is, the opposite result to the decreasing average user cost of all the Mohring'type models, a result explained in part because the length of the peak period is not fixed, such that as demand grows, the peak period enlarges (i.e. some passengers take earlier trains), which increases the schedule delay cost at the destination, given that the desired arrival time is fixed. Kraus and Yoshida (2002) provided an important insight into how the scheduling considerations of users affect average costs of travelling; however, their approach is less appropriate for modelling high-frequency services, in which it has been empirically observed that passengers arrive at stations or bus stops randomly at a more or less constant rate.[4] Therefore, the waiting time at stops exists even if the capacity constraint is not binding, and consequently, the economies of scale induced by increasing frequency should be accounted for.

Not only are additional benefits for users associated with a more frequent public transport service, but costs could also be incurred if providing extra bus kilometres has a negative effect on speeds for both buses and cars (Section 3.6). In this case, an increase in frequency can augment total average cost, and Mohring's (1972) scale economies argument for bus subsidies no longer applies, as shown in Mohring (1983). Nonetheless, there are a number of strategies that can be used to make bus transport more efficient in order to minimize or avoid the congestion related to high bus frequency. An example is the provision of faster fare collection systems at bus stops, as analysed by Tirachini and Hensher (2011), who, using an optimization model for congested bus corridors, showed that increasing total costs are observed for high-demand services if passengers are allowed to pay fares on boarding the buses, and frequency is over 120 veh/h, but decreasing total costs are still obtained even for higher frequencies when the fare payment is done before boarding buses.

2.2 *The Second-Best Pricing*

As widely recognized in the literature, several departures from ideal first-best conditions exist in reality. In the case of public transport pricing, the most evident and analysed situation is where buses or trains compete with underpriced cars, which imposes a second-best constraint on the determination of public transport fares. The classical argument is that if cars are underpriced, there is an excess of car travel; therefore, it would be welfare improving to reduce the public transport fare in order to attract some car users to trains or buses, reducing the level of congestion and other traffic externalities on the road network. This is a second economic rationale to subsidizing public transport, after the economies of scale (first-best) argument[5] (Preston, 2008; Parry and Small, 2009). As argued by Small (2008), from the first-best and second-best fare analysis, congestion charging could be seen as a way to reduce the financial needs of public transport, since an optimal road charge should decrease the subsidy required for public transport, even if the revenue from road pricing is not earmarked to public transport.

Formal proofs that an alternative mode should be priced below the marginal cost when cars are priced at the average cost instead of at the marginal cost can be traced to Lévy-Lambert (1968), Marchand (1968) and Sherman (1971). The idea, linked to competitive neutrality, was extended by Glaister (1974), who found a second-best bus fare below marginal cost, not only in the peak period but also in the (congestion-free) off-peak period, the latter due to two effects—first, a low off-peak bus fare can attract peak car users, and second, peak bus users are attracted to travelling by bus during the off-peak periods, which relieves pressure in the peak periods, and therefore decreases the peak bus fare, which in turn attracts more car travellers to public transport. More recently, Parry and Small (2009) found that substantial gains in social welfare are obtained from diverting car drivers into public transport (second-best argument) in peak periods, whereas the case to subsidize fares due to the reduction of user costs (scale economies—first-best argument) is stronger in the off-peak periods.

In summary, we found that setting public transport fares below the average operator cost is supported by most of the formal analyses of pricing, resulting in the call for an 'optimal' subsidy regardless of it being based on the first-best or second-best grounds. Despite the rigorous analytical approaches and empirical evidence, the extant literature has a number of limitations, associated, in particular, with the omission of non-motorized modes such as walking and cycling, and the distortionary effect of bus subsidies, as identified by Kerin (1992). Some of these factors have been accounted for in more recent research, such as possible inefficiencies associated with subsidy (Section 2.3), the existence of tax distortions and their interaction with the transport system (Section 3.5), and the impact of bus congestion on travel times and operation costs (Section 3.6). In this paper, we identify the potential influence of non-motorized transport on optimal pricing decisions (Section 4).

2.3 *Issues that Arise When Subsidizing Public Transport*

Observed practice has shown a number of problems associated with public transport subsidies that stylized first-best and second-best models have ignored. The realization of the efficiency gains that optimal subsidies in theory yield in practice depends on several factors, such as the form of the subsidy (e.g., operating subsidy per passenger or passenger-kilometre vs. one-off grant), the structure of the service provider (private or public company) and the relationship between the provider and the subsidizing body (Else, 1985). Moreover, the authority may not have sufficient information on costs and demand to estimate the optimal level of subsidy (Frankena, 1983).

A potentially major problem is the inefficiency induced in the operation of public transport services by some types of subsidies (Bly *et al.*, 1980; Cervero, 1984; Pickrell, 1985). Recent research has shown that there are ways to contain the cost spiral in the presence of subsidy, through performance-based benchmarking and the use of service quality indicators in service contracts (Hensher and Prioni, 2002; Hensher and Stanley, 2003; Mazzulla and Eboli, 2006; Gatta and Marcucci, 2007), action taken by the regulator to enforce penalties for poor performance and the application of competitive tendering (Hensher and Houghton, 2004; Hensher and Wallis, 2005).

A related issue discussed by Preston (2008) comes from the distinction between capital and operating subsidies. One-off subsidies targeted specifically at capital investment may condition the decisions of policy-makers and operators towards over-investing in capital, for example, acquiring more sophisticated or newer vehicles instead of spending on the maintenance of the current fleet. Pre-defined rail-specific capital subsidies may also lead to unjustified rail investments in areas with low demand for public transport, with the second-round effect of inducing an unnecessarily large subsidy for operations. Therefore, the correct *ex ante* determination of capital and operating subsidies is crucial to ensure efficiency in the allocation of resources to public transport service provision.

In general, the way in which an 'optimal' subsidy is paid is crucial, and the business environment should be defined to minimize or eliminate potential money waste induced by ill-designed subsidies. The design of contracts to tackle this problem is a topic of ongoing research and continuous learning in public transport agencies around the world.

3. Results that Matter

3.1 *Optimal Frequency and Capacity*

Optimal values of frequency and capacity are obtained when the marginal social benefits are equal to their marginal cost. When the effect of frequency on waiting and dwell times for users is taken into account, the first-best scenario over a single route usually leads to some form of the square root rule for the optimal frequency, first introduced by Mohring (1972) and later extended by several authors including Jansson (1980), Jara-Díaz and Gschwender (2003, 2009) and Jara-Díaz *et al.* (2008). An important and sometimes forgotten outcome is that the square root rule does not necessarily mean that optimal frequency depends on the square root of demand; this is a result of the first Mohring model (Mohring, 1972). Subsequent extensions with more accurate representations of the user cost function have shown that even though the square root form is maintained, demand under the root appears to a degree higher than one, for example, the quadratic formulation in Jansson (1980) that included the boarding and alighting impact on travel time.[6] Therefore, even though the functional form for the optimal frequency is a square root when a single route is considered, it can vary with demand to a power higher than 0.5 (e.g., around 0.8 as numerically found by Tirachini and Hensher, 2011, with a model that includes bus congestion between stations).

A relevant issue for the economic analysis of pricing options is the determination of the optimal change in public transport frequency and capacity when road pricing is introduced. The answer is not straightforward; for instance, Jansson (2010) found that bus frequency, when car travel is underpriced, should be lower than that when marginal cost road pricing is in place due to the negative impact of frequency on the environment and excessive congestion derived from the greater-than-optimal car traffic. However, the bimodal rail–car analysis of Kraus (2003) concludes that both rail frequency and capacity should increase if cars are underpriced, assuming no congestion interaction and disregarding the environmental cost associated with rail, assumptions that are relaxed in the model given in Section 4. The existing literature does not offer unambiguous evidence for the direction of change in frequency and capacity of public transport after applying road pricing; indeed, the outcome seems to depend on the modelling assumptions. Bus (and rail) frequency should be increased with congestion pricing in situations where the expected modal switching (given the relevant cross-price elasticities) might lead to a shortage of service capacity, at least in peak periods. The anticipation of modal switching in London and Stockholm delivered increased buses in advance of the application of cordon pricing, which was used to show that the revenue raised from the congestion charge was being hypothecated back to the transport sector for the benefit of modal switchers.

3.2 *When the Capacity Constraint is Binding*

Capacity constraints play a role in the optimization of public transport fares. Transport capacity on a public transport route is given by the product of service frequency f and the capacity of vehicles K. This transport capacity sets the maximum flow that the service is able to accommodate in a given period of time. The first transport pricing study to consider capacity considerations is that of Glaister (1974), who found that the bus fare should include the shadow price of capacity, that is, the extra social benefit achieved if capacity is increased by one unit. Glaister did not provide an expression for the shadow price of capacity because capacity is not an optimization variable in his model, but Pedersen (2003) and Small and Verhoef (2007) derived an expression for the shadow price of capacity as a function of user and operator cost parameters, which is added to the optimal fare when the capacity constraint is binding.

Another argument to increase bus fares when the capacity constraint is binding was provided by Turvey and Mohring (1975), who argued for higher fares when buses run full (or close to full), as this increases the probability of passengers not being able to board the first bus that arrives at their stop and having to wait for one or more buses to continue their trip.

In summary, transport capacity appears to play a role in increasing both the first-best and second-best fares when the system is operating at capacity. Nevertheless, the fact that the capacity constraint is binding does not necessarily mean that the provided frequency and bus size are not optimal. This issue is analysed in Section 4.

3.3 *The Effect of Including Other Externalities Beyond Congestion*

When environmental externalities are included in the first-best pricing models, optimal prices increase for motorized modes, which would in turn reduce the first-best subsidy calculated for public transport (Kerin, 1992). However, the second-best analysis is different. Taking the case of fuel emissions, one bus is likely to pollute more than one car, but it can carry more people with a single vehicle, thus reversing the result of comparing vehicles only; that is, the marginal external cost of car users is usually higher than that of public transport riders;[7] therefore, it is expected that the fare premium on optimal prices associated with considering externalities other than congestion is greater for private transport than for public transport. On second-best grounds, this would tend to reduce the bus fare even more and consequently justify higher subsidies (Else, 1985).

The contribution of environmental and accident externalities to optimal fares relative to the congestion externality strongly depends on the specific application, in particular, on the degree of congestion observed. It is common that in peak periods in highly urbanized areas, the marginal cost of congestion is significantly higher than that of other externalities, whereas in the off-peak periods, the external costs of congestion, accidents and pollution have approximately the same order of magnitude, as reported by De Borger *et al.* (1996) for Belgium and Parry and Small (2009) for London and US cities. Therefore, we can conclude that ignoring externalities other than congestion should not have a substantial impact on fares in the peak period, but it does matter for off-peak travel.

3.4 *Dedicated Bus Lanes*

The study of private and public transport pricing options is different if modes share the right of way or run on segregated roads. Mohring (1983) analysed the convenience of having reserved lanes for buses and found that travel cost savings of providing dedicated road infrastructure for buses are small when marginal cost pricing is in place, but considerable benefits are obtained when toll and fare constraints are present (second-best scenarios), to the point that the travel cost in a situation with exclusive bus lanes, toll and bus fare constraints is only slightly higher than that when first-best pricing is implemented for mixed-traffic (bus–automobile) roads. Berglas *et al.* (1984) showed that if travel cost decreases with road width, and the cost of separating the right of way for buses and cars is zero, the mixed-traffic operation is never superior and is more likely to be inferior than providing exclusive lanes for buses and cars, given that a bus passenger has a lower contribution to congestion than a car user.

The superiority of providing exclusive bus lanes was supported by Basso and Silva (2010), who using data from Santiago de Chile found that the provision of one-bus lane on a corridor increases social welfare with respect to any scenario in which buses and cars share the right of way (even when optimal pricing is applied in mixed traffic but not for exclusive bus lanes). The optimal operation with dedicated bus lanes is translated into a lower requirement on the number of buses, a lower bus fare and higher frequency, providing large benefits for bus users.

In summary, implementing dedicated bus road infrastructure to reduce travel costs is shown as being slightly worse (Mohring, 1983) or better (Basso and Silva, 2010) than providing marginal cost pricing on mixed-traffic conditions; however, there is the added advantage that bus lanes as a transport policy tool are likely to be more politically and sociably acceptable than imposing marginal cost pricing (Mohring, 1983), a fact that is evident when comparing the number of cities in which marginal cost pricing has been implemented in contrast to cities with dedicated bus lanes. A limitation of all economic models incorporating bus lanes is that they abstract from the extra cost of reserved bus lanes produced by diversions and extra delays at intersections, as some movements need to be prohibited for cars. This consideration is likely to reduce the welfare gain estimates of segregated bus lanes; however, it is unlikely to change the conclusions obtained by the authors.

3.5 *Interactions with Other Sectors of the Economy*

The previous analyses and results are based on partial equilibrium models that abstract from the interaction between transport and other sectors of the economy. This is a significant issue because the findings of a partial equilibrium model establish, for example, the need to subsidize public transport, but say nothing as to how that subsidy should be financed and what its repercussions are on the wider fiscal system. In order to answer these questions, one needs a general equilibrium model to estimate the impact of transport pricing reforms on the government budget, the labour market, land use, firms and so on.[8]

As to how to fund public transport subsidies was first analytically addressed by Dodgson and Topham (1987), who investigated the efficiency of raising the subsidy for public transport through an increase in the tax on other goods. The convenience of such a subsidy strongly depends on the marginal cost of public funds (MCF), which measures the welfare loss for society in raising additional revenue to finance public spending through the application of distortionary taxes (Browning, 1976; Kleven and Kreiner, 2006). The MCF depends on what tax instrument is used to increase government revenue (e.g., uniform lump sum tax and income tax), and hence the welfare analysis of transport pricing policies depends on the source of the money required to cover financial deficits or investments (Proost *et al.*, 2007; Calthrop *et al.*, 2010), or how the revenue is allocated if there is a surplus.

What is the impact of wider fiscal considerations on optimal fares and subsidies? It is expected that estimated public transport subsidies would decrease, given that when there is no account as to how the subsidies are funded, their cost is misrepresented in the social welfare analysis (Kerin, 1992). This issue can be analysed in a simple (but not complete) way that avoids dealing with general equilibrium models, by simply including the marginal cost of public funds in partial equilibrium models, as implemented, for example, by Proost and Van Dender (2008), who found that road prices and public transport fares increase in the presence of costly public funds, as the benefits of generating revenue to be used elsewhere in the economy (or the benefits from reducing the subsidy for public transport) are taken into account. A similar conclusion was reached by Parry and Small (2009), who suggested that fiscal considerations would decrease optimal public transport subsidies in the USA, but not to the point of jeopardizing their need on second-best grounds.

In summary, approaching the problem by including the MCF in the net revenues of a transport intervention in a partial equilibrium model is useful as a first approximation to assist in answering the question as to how tax distortions affect, and are affected by, reforms in the transport sector, but a fuller understanding requires a general equilibrium model. For example, the impacts of reduced congestion on other markets (as shown by Parry and Bento, 2001, for the income tax), derived from a public transport subsidy, are not going to be captured with an approach that only considers the MCF as representing the rest of the economy (Calthrop *et al.*, 2010).

3.6 *Effects of Bus Congestion and Congestion Interactions*

Traditional first-best models consider that the travel time of buses is fixed in between bus stops, that is, there are no time delays caused by cars and buses themselves (e.g., Mohring, 1972; Jansson, 1979). This assumption is plausible for services in which the frequency is relatively low, with no noticeable bus interaction due to bunching or queuing delays behind bus stops. Nevertheless, as frequency grows, it is more likely that buses will arrive at bus stops when there are other buses transferring passengers, with queuing delays arising at bus stops. This is a relevant issue for pricing analysis, since frequency-induced congestion increases bus travel time for users and operators, in contrast to the economies of scale effect on reducing waiting times (Kerin, 1992).

A technical problem when bus congestion is included in formal pricing analysis is that the bus congestion technology has not been realistically understood and defined because of the myriad number of factors that influence how buses interact with each other, with other modes and with passengers. In theoretical models, the authors typically apply bus flow–delay functions borrowed from car traffic models, such as the linear function implemented by Ahn (2009) and the Bureau of Public Roads (BPR) function used by Fernández *et al.* (2005) and Wichiensin *et al.* (2007). These are not necessarily good representations of the interaction between buses and cars, bus stops and passengers. Moreover, these functions do not explicitly account for the fact that buses have to stop at bus stops, which is sometimes implicitly internalized by applying to buses a large passenger car-equivalency factor, for example, to assume that a standard bus is equivalent to four or five cars (Parry and Small, 2009).

More recently, there have been improvements in the characterization of bus congestion technology. Basso and Silva (2010) proposed a non-linear function for bus frequency that accounts for the delay that cars experience when buses stop at a bus stop, in such a way that the mean delay transferred to cars is small when bus frequency is low and equals bus dwell time when bus frequency is high. Another example is that given by Tirachini and Hensher (2011), who estimated bus queuing delay functions at bus stops as a function not only of the bus frequency but also of the number of passengers being transferred and the fare payment technology used by passengers (e.g., cash, magnetic strip, contactless card and off-board payment). The travel time functions are consequently more realistic than traffic-borrowed formulae for the analysis of congestion induced by bus stops.

We are far from a realistic characterization of the phenomenon of congestion when urban buses are involved. The inclusion of engineering or simulation models that deal with bus dynamics at bus stops (Fernández and Tyler, 2005; Fernández, 2010) into economic pricing analysis is a possible way forward to improve our understanding of bus and car delays in mixed systems and its implications for pricing policy.

3.7 *Non-motorized Transport*

Second-best pricing models that consider only two modes—cars and transit (bus or rail)—have found that subsidies for public transport are desirable, with fares offered below marginal cost, due to the underpricing of cars. However, as argued by Kerin (1992), this approach overlooks the existence of other modes, notably walking and cycling, that play a crucial role in urban transport systems, especially for short trips. Disregarding non-motorized transport is a growing concern because low bus fares not only deter some drivers from using their cars but also divert walkers and cyclists into trains or buses, which is not necessarily a desirable outcome. As such, a pricing model that also includes non-motorized transport seems to be desirable in order to estimate the impact of these modes on (possibly decreasing) optimal subsidies for public transport. To the best of our knowledge, this issue has not been formally analysed in the first-best or second-best pricing models. It is addressed in Section 4.

4. A Three-mode Pricing Model

4.1 *Introduction*

We propose a simple model that incorporates non-motorized transport and revisits the impact of a capacity constraint on optimal public transport pricing. Even though there are no analytical models that address the issue of the influence of non-motorized transport on urban transport pricing policy, we do find that walking and cycling are considered as travelling alternatives in applied models (e.g., Safirova *et al.*, 2006; Proost and Van Dender, 2008), but no attempt is made to identify as to how the design of the pricing instrument would change by considering or ignoring walking and cycling.

4.2 *Model Assumptions*

Consider a single origin–destination pair and three modes: automobile (a), public transport (b), which could be a bus or a rail-based mode, and a non-motorized mode (e), which could be walking or cycling. At this point, we need to distinguish between non-motorized modes as being complementary or an alternative to motorized modes; walking is commonly an access and/or egress mode in a trip chain that includes driving or riding a bus or train, in which case the modes are complementary. In our model, we assume walking or cycling as the (linehaul) mode, that is, as an alternative to choosing a motorized mode (walking and cycling as an access mode is included in the motorized alternatives). The competitiveness of walking and cycling is mainly associated with trip distance and factors such as steepness, weather and availability of safe walking and cycling facilities. In all situations, walking as a substitute to motorized modes typically declines as distance increases. For example, in Sydney, we found that 35.4% of total trips are shorter than 2 km and 24.9% of total trips are between 2 and 5 km; on these distance ranges, 65.8% of trips shorter than 1 km are walking-only trips, a fraction that is 23.6% for trips between 1 and 2 km and 5.4% for trips between 2 and 5 km (TDC, 2010). Donoso *et al.* (2006) reported a similar pattern for Santiago de Chile. Then, there is a (location-specific) distance range in which walking is an alternative for motorized modes.

In the model, the decision variables are optimal prices for both automobile and public transport, and frequency and size (capacity) of buses or trains. We consider only one period of operation,[9] which allows us to find a closed-form formula for the optimal prices of automobile and public transport to shed light on the impact of non-motorized transport and capacity constraints. Road capacity is fixed and tax distortions are ignored.

We follow much of the notation of Small and Verhoef (2007). Ignoring income effects, the joint demand for the three modes can be obtained from the benefit function $B(q_{\mathrm{a}}, q_b, q_e)$, which expresses the consumers' willingness to pay for a particular combination $\{q_{\mathrm{a}}, q_{\mathrm{b}}, q_e\}$ of travel by automobile, public transport and non-motorized mode. The inverse demand function d_i for mode i is given by

$$d_i(q_{\mathrm{a}}, q_{\mathrm{b}}, q_{\mathrm{e}}) = \frac{\partial B(q_{\mathrm{a}}, q_{\mathrm{b}}, q_{\mathrm{e}})}{\partial q_i}, \quad i \in \{a, b, e\}. \tag{1}$$

Let C_i and c_i be the total and average cost functions of mode i, respectively, including both time and operation costs, that is

$$C_i = q_i c_i. \tag{2}$$

Let $c_{\mathrm{a}}(q_{\mathrm{a}}, q_{\mathrm{b}}, f_{\mathrm{b}}, K_{\mathrm{b}})$ and $c_{\mathrm{b}}(q_{\mathrm{a}}, q_{\mathrm{b}}, f_{\mathrm{b}}, K_{\mathrm{b}})$ be the average costs of car and bus travel, respectively. We assume that these cost functions depend on car demand q_{a}, bus frequency f_{b} and bus capacity K_{b} (related to bus size), and the activity of buses at bus stops, which is given by f_{b}, q_{b} and K_{b} if dwell time increases with crowding. The relationship between car demand q_{a} and car flow f_{a} is $f_{\mathrm{a}} = v_{\mathrm{a}} q_{\mathrm{a}}$, where v_{a} is the inverse of the average occupancy rate per car.[10] Bus cost c_{b} includes user cost c_{u} (access, waiting and in-vehicle time costs) and operator cost c_{o} (which depends on bus frequency and size); hence

$$c_{\mathrm{b}} = c_{\mathrm{u}} + c_{\mathrm{o}}. \tag{3}$$

We assume that the travel time associated with walking or cycling is fixed and independent of demand or flow of any mode, that is, the non-motorized mode is uncongestible.

In equilibrium, the marginal benefit is equal to the generalized price, $c_{\mathrm{a}} + \tau_{\mathrm{a}}$ and $c_{\mathrm{u}} + \tau_{\mathrm{b}}$ for cars and public transport (Equation (4)), respectively, where τ_{a} is the road use charge for the automobile and τ_{b} is the fare for public transport:

$$\frac{\partial B}{\partial q_{\mathrm{a}}} = c_{\mathrm{a}} + \tau_{\mathrm{a}}, \quad \frac{\partial B}{\partial q_i} = c_{\mathrm{u}} + \tau_{\mathrm{b}}. \tag{4}$$

We assume that the public transport mode is a bus that shares the right of way with cars, resulting in congestion dependence between the two modes. The case of trains or buses running on segregated busways is a particular case of the above, derived after assuming congestion independence between modes, as usually assumed by researchers who address the rail–car pricing problem (Tabuchi, 1993; Arnott and Yan, 2000; Pels and Verhoef, 2007).

The social welfare function SW (5) is, maximized subject to a capacity constraint for public transport vehicles, given by expression (6), which states that the transport capacity $f_b K_b$ must be sufficient to carry total demand q_b:

$$\text{Max SW} = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e. \quad (5)$$

$$\text{subject to } q_b \leq f_b K_b. \quad (6)$$

4.3 *The First-Best Pricing*

To solve the constrained maximization problems (5) and (6), we set the Lagrange function L given by

$$L = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda[f_b K_b - q_b], \quad (7)$$

where λ is the Lagrange multiplier associated with constraint (6), that is, the marginal social benefit of increasing bus transport capacity by one unit. After applying the first-order conditions (see the Appendix), we find that

$$\tau_a = q_a \frac{\partial c_a}{\partial q_a} + q_b \frac{\partial c_b}{\partial q_a} \quad (8)$$

$$\tau_e = 0. \quad (9)$$

Equation (8) is the well-known Pigouvian tax for cars, including here the marginal cost on bus cost due to car demand (the second term), whereas Equation (9) shows that the price for walking or cycling is zero (the uncongestible mode).

The solution for the optimal bus fare, frequency and capacity depends on whether or not the capacity constraint (6) is binding.

Case 1: Capacity constraint is not binding

In this case, $\lambda = 0$ and the optimal fare is obtained as

$$\tau_b = c_o + q_a \frac{\partial c_a}{\partial q_b} + q_b \frac{\partial c_b}{\partial q_b}. \quad (10)$$

The optimal frequency and capacity are obtained by solving the following system of equations:

$$q_a \frac{\partial c_a}{\partial f_b} + q_b \frac{\partial c_b}{\partial f_b} = 0 \quad (11a)$$

$$q_a \frac{\partial c_a}{\partial K_b} + q_b \frac{\partial c_b}{\partial K_b} = 0. \quad (11b)$$

Case 2: Capacity constraint is binding

In this case, constraint (6) is active, that is, $q_b = f_b K_b$, and the Lagrange multiplier is $\lambda \neq 0$. From Equation (A3),

$$\tau_b = c_o + q_a \frac{\partial c_a}{\partial q_b} + q_b \frac{\partial c_b}{\partial q_b} + \lambda. \tag{12}$$

From Equation (A5), the marginal welfare benefit of capacity can be expressed as

$$\lambda = \frac{1}{K_b}\left[q_a \frac{\partial c_a}{\partial f_b} + q_b \frac{\partial c_b}{\partial f_b}\right], \tag{13}$$

and using that $1/K_b = f_b/q_b$, we finally obtain

$$\tau_b = c_o + q_a \frac{\partial c_a}{\partial q_b} + q_b \frac{\partial c_b}{\partial q_b} + f_b\left[\frac{q_a}{q_b}\frac{\partial c_a}{\partial f_b} + \frac{\partial c_b}{\partial f_b}\right]. \tag{14}$$

Equation (14) shows the effect of the capacity constraint on the optimal bus fare. A similar result was obtained by Pedersen (2003) in a model with no car–bus interactions. When the capacity constraint is binding, one possibility is to increase the frequency to satisfy constraint (6) to a higher-than-optimal value. In that case, the term within the brackets in Equation (14) is positive and represents the impact on car and bus marginal costs of the increased frequency necessary to deal with a demand that the optimal frequency (solution of Equation (11a)) cannot meet. Nevertheless, note that frequency and capacity can be optimal and the capacity constraint can indeed be binding, if, for example, there is no extra benefit of providing excess capacity (no crowding or comfort costs), and therefore, once the frequency has been optimized, the vehicle size is obtained as the minimum that satisfies Equation (6). In this case, expression (14) is valid, but the capacity-related term within the brackets is zero because the frequency is optimal (solution of Equation (11a)), and then Equation (14) is reduced to the optimal fare (11) with no capacity constraints. An important outcome of Equation (14) is that what really matters when setting optimal fares is not if the capacity constraint is binding, but whether or not the operator provides the optimal transport capacity.

4.4 *The Second-Best Pricing*

We can solve the same problem assuming that there is no road price for cars, that is, $\tau_a = 0$. The Lagrange function is

$$\begin{aligned} L &= B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda[f_b K_b - q_b] \\ &\quad + \gamma_a\left(c_a - \frac{\partial B}{\partial q_a}\right) + \gamma_b\left(c_u + \tau_b - \frac{\partial B}{\partial q_b}\right) + \gamma_e\left(c_e - \frac{\partial B}{\partial q_e}\right). \end{aligned} \tag{15}$$

The first-order conditions are given in the Appendix.

We can simplify the differential notation as follows:

$$\frac{\partial^2 B}{\partial q_i \partial q_j} = \frac{\partial^2 B}{\partial q_j \partial q_i} \equiv B_{ij}. \tag{16a}$$

$$\frac{\partial c_i}{\partial q_j} \equiv c_{ij}, \tag{16b}$$

where B_{ij} is the derivative of the inverse demand d_i in Equation (1) with respect to q_j, that is, it measures a marginal change in willingness to pay for mode i due to a marginal change in the amount of travel on mode j. If there is no substitution between two modes then $B_{ij} = 0$. If all modes are substitutes (e.g., an increase in bus fare would increase the amount of car and non-motorized travel), then $B_{ij} < 0$ $\forall\, i, j$. On the other hand, $B_{ij} > 0$ would mean that i and j are complements. Moreover, as Kraus (2003) discussed, following standard microeconomics for utility maximizing consumers, it should hold that (assuming that trip demand is independent of income) $B_{ii} \leq 0$ and $B_{ii}B_{jj} > B_{ij}B_{ji}$ for any modes i and j.

Case 1: Capacity constraint is not binding

After some algebraic manipulation, we obtain the second-best bus fare $\tau_{\mathrm{b}}^{\mathrm{SB}}$ as

$$\tau_{\mathrm{b}}^{\mathrm{SB}} = \tau_{\mathrm{b}} - (q_{\mathrm{a}}c_{aa} + q_{\mathrm{b}}c_{ba})\frac{c_{ab} - B_{ab} + (B_{ae}B_{be}/B_{ee})}{c_{aa} - B_{aa} + B_{ae}^2/B_{ee}}, \tag{17}$$

where τ_{b} is the first-best fare (10). Unlike the first-best pricing rule, under the second-best rule, the non-motorized mode plays a role through the substitution parameters B_{ae}, B_{be} and B_{ee}. Note that if car is an uncongestible mode and does not interact with buses, then the second-best correction is zero (the second term on the right-hand side of Equation (17)), and consequently the second-best fare is equal to the first-best fare, $\tau_{\mathrm{b}}^{\mathrm{SB}} = \tau_{\mathrm{b}}$, analogous to a two-link road pricing analysis when one link is uncongestible (e.g., Knight, 1924; Verhoef *et al.*, 1996).

Two new results can be derived from Equation (17). First, if we assume that there is no substitution between modes a and e and b and e, then $B_{ae} = B_{be} = 0$, and Equation (17) is reduced to

$$\tau_{b0}^{\mathrm{SB}} = \tau_{\mathrm{b}} - (q_{\mathrm{a}}c_{aa} + q_{\mathrm{b}}c_{ba})\frac{c_{ab} - B_{ab}}{c_{aa} - B_{aa}}, \tag{18}$$

which is the second-best bus fare considering only two modes, as obtained by Small and Verhoef (2007), for the case in which there is no congestion interaction between modes, that is, $c_{ab} = c_{ba} = 0$, and by Ahn (2009), who considered that bus demand does not affect car travel time, that is, $c_{ab} = 0$. If $c_{ab} = 0$, the second-best bus fare equals the first-best price ($\tau_{b0}^{\mathrm{SB}} = \tau_{\mathrm{b}}$) when there is no cross-demand elasticity between car and bus, that is, when $B_{ab} = 0$, and therefore a low bus fare has no effect on mode shifting, as noted by Small and Verhoef (2007) and Ahn (2009). Nevertheless, when delays related to bus passenger activities affect cars ($c_{ab} \neq 0$), the second-best fare (18) is not reduced to the first-best fare (10) even if $B_{ab} = 0$ (noting that this does not mean that the second-best fare decreases with c_{ab} because c_{ab} increases the first-best fare τ_{b} in Equation (18), as shown in Equation (10)).

Second, Equation (17) can be used to formally assess Kerin's (1992) claim that second-best fares obtained by considering car and public transport only are likely to be lower than the optimal fares if the analysis is extended to walking and cycling. A comparison between Equations (17) and (18), assuming for illustrative purposes that demand and congestion levels are the same, indicates that the second-best bus fare will be larger when considering non-motorized transport if

$$\tau_{\mathrm{b}}^{\mathrm{SB}} > \tau_{b0}^{\mathrm{SB}} \Leftrightarrow \frac{B_{be}}{B_{ae}} > \frac{c_{ab} - B_{ab}}{c_{aa} - B_{aa}}, \tag{19}$$

that is, the larger the value of B_{be} and the lower B_{ae} and B_{ab} (in absolute values), the more likely is Equation (17) to be greater than Equation (18). The intuition behind this result is that *if* the modal substitution between public transport and non-motorized modes (B_{be}) is strong relative to the substitution between car and public transport (B_{ab}) and car and non-motorized modes (B_{ab}), a lower public transport fare attracts more passengers who would otherwise be walking or cycling than driving. But note that the change could go the other way as well ($\tau_{\mathrm{b}}^{\mathrm{SB}} < \tau_{b0}^{\mathrm{SB}}$) if the modal substitution between automobile and non-motorized transport is stronger than that between public transport and non-motorized modes (low value of B_{be}/B_{ae}). Certainly, the result depends on trip distance, since for long trips cycling and walking are unlikely to be an option (as discussed in Section 4.2 when we analysed modal split per trip distance), which means $B_{ae} = B_{be} = 0$ and the analysis can be reduced to motorized modes only.

Optimal frequency and bus capacity are the solution of

$$(q_{\mathrm{a}} - \gamma_{\mathrm{a}})\frac{\partial c_{\mathrm{a}}}{\partial f_{\mathrm{b}}} + q_{\mathrm{b}}\frac{\partial c_{\mathrm{b}}}{\partial f_{\mathrm{b}}} = 0, \tag{20a}$$

$$(q_{\mathrm{a}} - \gamma_{\mathrm{a}})\frac{\partial c_{\mathrm{a}}}{\partial K_{\mathrm{b}}} + q_{\mathrm{b}}\frac{\partial c_{\mathrm{b}}}{\partial K_{\mathrm{b}}} = 0 \tag{20b}$$

with $\gamma_{\mathrm{a}} = (q_{\mathrm{a}}c_{aa} + q_{\mathrm{b}}c_{ba})/(c_{aa} - B_{aa} + B_{ae}^2/B_{ee})$.

That is, in the second-best case, the congestion externality of buses to cars is less internalized because $\gamma_{\mathrm{a}} > 0$, as commented by Ahn (2009), the intuition being that due to the underpricing of cars, the negative effect of buses on car travel time should be weighted less. If there is no congestion interaction, that is, $\partial c_{\mathrm{a}}/\partial f_{\mathrm{b}} = 0$, the rules for the first-best and second-best frequencies and bus capacities are the same (Equations (11) and (20)), then a higher bus demand q_{b} in the first-best case (due to the pricing of cars) would make the first-best frequency higher than the second-best one, which is not a straightforward result with cross-congestion, due to the presence of γ_{a} in Equation (20a).

Case 2: Capacity constraint is binding

Analogously to the first-best case, the second-best bus fare is obtained as

$$\tau_{\mathrm{b}}^{\mathrm{SB}} = \tau_{\mathrm{b}} - (q_{\mathrm{a}}c_{aa} + q_{\mathrm{b}}c_{ba})\frac{c_{ab} - B_{ab} + B_{ae}B_{be}/B_{ee}}{c_{aa} - B_{aa} + B_{ae}^2/B_{ee}} + f_{\mathrm{b}}\left(\frac{q_{\mathrm{a}} - \gamma_{\mathrm{a}}}{q_{\mathrm{b}}}\frac{\partial c_{\mathrm{a}}}{\partial f_{\mathrm{b}}} + \frac{\partial c_{\mathrm{b}}}{\partial f_{\mathrm{b}}}\right). \tag{21}$$

4.5 *Extensions: Other External Costs and Collection Costs*

In this section, we extend the preceding approach by including more cost components, namely toll collection costs and external costs such as accidents, pollution and noise. The toll collection and operator costs are usually disregarded from the formal analysis of pricing policies, even though current road pricing schemes show that they are not negligible; operating costs account for 7% of the revenues in Singapore, 25% in Stockholm and 48% in London (May *et al.*, 2010), mostly influenced by the choice of technology for charging and enforcement.[11] A simple way to include operating costs OC is proposed in

$$\mathrm{OC}(q_\mathrm{a}) = oc_0 + oc_1 q_\mathrm{a}, \tag{22}$$

where oc_0 is a fixed cost and oc_1 is the marginal cost per transaction. The fare collection cost for public transport is partially included in the bus or rail operator cost c_o (Equation (4)), which may include the fixed collection cost due to software requirements plus fare payment devices at stations or vehicles. The cost per transaction, if not negligible, can be incorporated in the same way as Equation (22).

External costs EC other than congestion (see Section 3.4) can be expressed as follows:

$$\mathrm{EC}(q_\mathrm{a}, q_\mathrm{b}, f_\mathrm{b}, K_\mathrm{b}) = v_\mathrm{a} q_\mathrm{a} \mathrm{EC}_\mathrm{a}(q_\mathrm{a}, q_\mathrm{b}, f_\mathrm{b}, K_\mathrm{b}) + f_\mathrm{b} \mathrm{EC}_\mathrm{b}(q_\mathrm{a}, q_\mathrm{b}, f_\mathrm{b}, K_\mathrm{b}), \tag{23}$$

where EC_a and EC_b are the external cost rates per vehicle for car and public transport (assuming the external costs of walking or cycling to be zero), and the car flow is $f_\mathrm{a} = v_\mathrm{a} q_\mathrm{a}$, where v_a is the inverse of the average occupancy rate per car. Expressions (22) and (23) can be subtracted from the social welfare formula (5) to derive the first-best and second-best pricing results. Denoting $\mathrm{EC}_{ij} \equiv \partial \mathrm{EC}_i / \partial q_j$, the results for the first-best prices are

$$\tau_\mathrm{a} = q_\mathrm{a} c_{aa} + q_\mathrm{b} c_{ba} + oc_1 + v_\mathrm{a} q_\mathrm{a} \mathrm{EC}_{aa} + v_\mathrm{a} \mathrm{EC}_\mathrm{a} + f_\mathrm{b} \mathrm{EC}_{ba}, \tag{24}$$

$$\tau_\mathrm{b} = c_o + q_\mathrm{a} c_{ab} + q_\mathrm{b} c_{bb} + v_\mathrm{a} q_\mathrm{a} \mathrm{EC}_{ab} + f_\mathrm{b} \mathrm{EC}_{bb}. \tag{25}$$

Since external costs other than congestion are assumed to be positive for car and bus users, it is likely that the result of expressions (24) and (25) will be greater than the optimal prices when considering only congestion externalities (Equations (8) and (10)), and therefore, the internalization of accidents, noise or pollution costs would increase the generalized cost of motorized transport modes compared with non-motorized modes (although the final result depends on the sensitivity of demands q_a and q_b to price) and reduce the amount of subsidy for public transport on first-best grounds. The second-best analysis can be undertaken in the same fashion. Regarding the toll collection costs, only the marginal cost per transaction oc_1 shows up in the optimal toll (24); however, the fixed cost of collection oc_0 in Equation (22) is accounted for in the calculation of social welfare; furthermore, oc_0 may be so high that the total collection cost is larger than the welfare gain from the internalization of the external cost, in which case tolling is not welfare improving unless a more cost-effective way of collecting tolls is implemented.

5. Summary and Conclusions

In this paper, we have reviewed some of the main issues associated with pricing in urban transport, with special attention being paid to the pricing of public transport services. The focus was on the economic fundamentals of pricing policies and their implication for a number of variables and outputs and not on the actual value of road prices, fares and subsidies obtained in the literature.

We have developed a model to analyse some issues that have been partially treated in the literature. It has been shown analytically that the effect of considering non-motorized transport alternatives on optimal public transport fares depends on the demand substitution between modes; the stronger is the demand substitution between public transport and non-motorized modes, relative to the substitution between automobile and public transport and automobile and non-motorized modes, the more likely it is that a higher optimal public transport fare would result when considering walking or cycling with respect to fare setting.

We revisited the role of a capacity constraint in public transport service provision, which suggested that a capacity constraint plays a role in optimal pricing only when the transport capacity cannot be set at its optimal level. We also presented a way to include externalities other than congestion and toll collection costs into the analysis of optimal pricing under the first-best and second-best rules, which showed that the internalization of externalities other than congestion is likely to increase optimal fares and road charges, therefore increasing the generalized price of motorized transport modes relative to a non-motorized alternative.

Future research needs to recognize the extended set of influences on optimal pricing of public transport and the implications this has for identifying optimal subsidy levels. For example, although travel time variability and modal reliability are known to have a significant role in influencing the quality of service and hence demand, the relationship between service reliability and pricing on public transport needs to be better understood. Optimal investment in public transport infrastructure, particularly in the case of buses, has also received little attention in the literature on the economics of public transport, despite being a topic of growing relevance nowadays as high-performance urban bus systems (Bus Rapid Transit or BTR) are spreading quickly in cities around the world (Wright and Hook, 2007; Hensher and Golob, 2008). The examination of high-performance–high-demand bus-based systems also highlights the importance of analysing bus congestion and crowding costs in the context of pricing decisions and public transport service design. Alternative modes such as motorcycles have not been included in the analysis and could be accommodated in future developments with a special qualification, as in this case the estimation of congestion, and particularly accident externalities, is harder to establish in a multimodal setting in which motorcycles may share the right of way with cars, buses, trucks and bicycles.

Other areas worthy of further research include building in preference heterogeneity for the fuller set of attributes, recognizing the wider economic impacts of transport pricing, and the distributive justice implications of road pricing. Distributional concerns emphasize the need for a general equilibrium approach to identify who is affected by the taxes levied to fund transport projects or subsidy for public transport (Dodgson and Topham, 1987; Proost *et al.*, 2007). In ongoing research, Hensher is developing a new referendum model to identify the impact of alternative road pricing schemes on community acceptability; the modelling approach conditions preference heterogeneity by the believability of subjective evidence judgements. The political economy of road pricing suggests that the use of toll revenues to subsidize public transport is crucial for the *ex ante* public acceptability of road pricing reforms (Marcucci *et al.*, 2005; De Borger and Proost, 2012).

Acknowledgements

This work has received financial support from Conicyt (Chile) through its programme of PhD scholarships. The detailed comments of two anonymous referees have greatly improved the final version of the paper.

Notes

1. For numerical comparisons on fares and subsidies, among several studies, see Proost and Van Dender (2008) and Parry and Small (2009).
2. See Rouwendal and Verhoef (2006) or Small and Verhoef (2007) for a more detailed discussion on second-best issues.
3. The history of road pricing and the evolution of the research on this topic have been extensively reviewed by Marcucci (2001) and Lindsey (2006). For a detailed survey on road pricing issues, see Tsekeris and Voß (2009).
4. A headway of 10 min or shorter is usually taken as the one that makes most passengers to ignore a timetable.
5. Other arguments in favour of subsidizing public transport include pursuing distributional or social objectives and option values, which are not treated in this paper (see Kerin, 1992; Preston, 2008).
6. Another extension is including route density as a decision variable, in which case, the optimal frequency results in a cubic root of a function of demand (e.g., Kuah and Perl, 1988; Chang and Schonfeld, 1991; Small, 2004).
7. As empirically found for pollution and accidents, but not for noise (De Borger *et al.*, 1996). The Milan Ecopass scheme is a pioneer in the application of differentiated charges based on the emission standard of vehicles (Rotaris *et al.*, 2010).
8. For an extended discussion on the advantages of general equilibrium models, see Calthrop *et al.* (2010).
9. Examples of multiperiod analyses are Glaister (1974), Glaister and Lewis (1978), De Borger *et al.* (1996) and Proost and Van Dender (2008).
10. We assume that the occupancy rate does not change with pricing reforms, that is, we ignore the possibility of car-pooling if road price increases.
11. In the case of London, other authors presented higher estimates of operating costs. Prud'homme and Bocarejo (2005) estimated that in 2003 the London congestion charging scheme's operating costs were 85% of toll revenue and that net revenue would not be enough to cover the annualized capital cost. Mackie (2005), Santos and Shaffer (2004) and Santos (2005) were more optimistic; they concluded that the operating cost was, respectively, 75%, 72% and 53−60% of the net revenue.

References

Ahn, K. (2009) Road pricing and bus service policies, *Journal of Transport Economics and Policy*, 43(1), pp. 25–53.

Arnott, R. and Yan, A. (2000) The two-mode problem: second-best pricing and capacity, *Review of Urban & Regional Development Studies*, 12(3), pp. 170–199.

Arnott, R., Palma, A. D. and Lindsey, R. (1993) A structural model of peak-period congestion: a traffic bottleneck with elastic demand, *The American Economic Review*, 83(1), pp. 161–179.

Basso, L. J. and Silva, H. E. (2010) A microeconomic analysis of congestion management policies. 5th Kuhmo Nectar Conference in Transport Economics Valencia, Spain, 8−9 July.

Berglas, E., Fresko, D. and Pines, D. (1984) Right of way and congestion toll, *Journal of Transport Economics and Policy*, 18(2), pp. 165–187.

Bly, P. H., Webster, F. V. and Pounds, S. (1980) Effects of subsidies on urban public transport, *Transportation*, 9(4), pp. 311–331.

Browning, E. K. (1976) The marginal cost of public funds, *The Journal of Political Economy*, 84(2), pp. 283–298.

Calthrop, E., De Borger, B. and Proost, S. (2010) Cost-benefit analysis of transport investments in distorted economies, *Transportation Research Part B*, 44(7), pp. 850–869.

Cervero, R. (1984) Cost and performance impacts of transit subsidy programs, *Transportation Research Part A*, 18(5–6), pp. 407–413.

Chang, S. K. and Schonfeld, P. M. (1991) Multiple period optimization of bus transit systems, *Transportation Research Part B*, 25(6), pp. 453–478.

Danielis, R. and Marcucci, E. (2002) Bottleneck road congestion pricing with a competing railroad service, *Transportation Research Part E*, 38(5), pp. 379–388.

De Borger, B. and Proost, S. (2012) A political economy model of road pricing, *Journal of Urban Economics*, 71(1), pp. 79–92.

De Borger, B., Mayeres, I., Proost, S. and Wouters, S. (1996) Optimal pricing of urban passenger transport: A simulation exercise for Belgium, *Journal of Transport Economics and Policy*, 30(1), pp. 31–54.

Dodgson, J. S. and Topham, N. (1987) Benefit-cost rules for urban transit subsidies: an integration of allocational, distributional and public finance issues, *Journal of Transport Economics and Policy*, 21(1), pp. 57–71.

Donoso, P., Martínez, F. and Zegras, C. (2006) The Kyoto Protocol and sustainable cities: the potential use of the clean development mechanism in structuring cities for "carbon-efficient" transport, *Transportation Research Record*, 1983, pp. 158–166.

Else, P. K. (1985) Optimal pricing and subsidies for scheduled transport services, *Journal of Transport Economics and Policy*, 19(3), pp. 263–279.

Fernández, R. (2010) Modelling public transport stops by microscopic simulation, *Transportation Research Part C*, 18(6), pp. 856–868.

Fernández, R. and Tyler, N. (2005) Effect of passenger-bus-traffic interactions on bus stop operations, *Transportation Planning and Technology*, 28(4), pp. 273–292.

Fernández, J. E., de Cea, J. and de Grange, L. (2005) Production costs, congestion, scope and scale economies in urban bus transportation corridors, *Transportation Research Part A*, 39(5), pp. 383–403.

Frankena, M. W. (1983) The efficiency of public transport objectives and subsidy formulas, *Journal of Transport Economics and Policy*, 17(1), pp. 67–76.

Gatta, V. and Marcucci, E. (2007) Quality and public transport service contracts, *European Transport/Trasporti Europei*, (36), pp. 92–106.

Glaister, S. (1974) Generalised consumer surplus and public transport pricing, *The Economic Journal*, 84(336), pp. 849–867.

Glaister, S. and Lewis, D. (1978) An integrated fares policy for transport in London, *Journal of Public Economics*, 9(3), pp. 341–355.

Hensher, D. A. and Golob, T. F. (2008) Bus rapid transit systems - a comparative assessment, *Transportation*, 35(4), pp. 501–518.

Hensher, D. A. and Houghton, E. (2004) Performance based quality contracts for the bus sector: Delivering social and commercial value for money, *Transportation Research Part B*, 38(2), pp. 123–146.

Hensher, D. and Prioni, P. (2002) A service quality index for area-wide contract performance assessment, *Journal of Transport Economics and Policy*, 36(1), pp. 93–113.

Hensher, D. A. and Stanley, J. (2003) Performance-based quality contracts in bus service provision, *Transportation Research Part A*, 37(6), pp. 519–538.

Hensher, D. A. and Wallis, I. (2005) Competitive tendering as a contracting mechanism for subsidising transportation: The bus experience, *Journal of Transport Economics and Policy*, 39(3), pp. 295–322.

Huang, H.-J. (2002) Pricing and logit-based mode choice models of a transit and highway system with elastic demand, *European Journal of Operational Research*, 140(3), pp. 562–570.

Jansson, J. O. (1979) Marginal cost pricing of scheduled transport services, *Journal of Transport Economics and Policy*, 13(3), pp. 268–294.

Jansson, J. O. (1980) A simple bus line model for optimization of service frequency and bus size, *Journal of Transport Economics and Policy*, 14(1), pp. 53–80.

Jansson, K. (1993) Optimal public transport price and service frequency, *Journal of Transport Economics and Policy*, 27(1), pp. 33–50.

Jansson, K. (2010) Public transport policy with and without road pricing. 5th Kuhmo-Nectar Conference on Transport Economics, Valencia, Spain, 8–9 July.

Jara-Díaz, S. R. (2007) *Transport Economic Theory* (Oxford: Elsevier).

Jara-Díaz, S. R. and Gschwender, A. (2003) Towards a general microeconomic model for the operation of public transport, *Transport Reviews*, 23(4), pp. 453–469.

Jara-Díaz, S. R. and Gschwender, A. (2009) The effect of financial constraints on the optimal design of public transport services, *Transportation*, 36(1), pp. 65–75.

Jara-Díaz, S. R., Tirachini, A. and Cortés, C. E. (2008) Modeling public transport corridors with aggregate and disaggregate demand, *Journal of Transport Geography*, 16(6), pp. 430–435.

Kerin, P. D. (1992) Efficient bus fares, *Transport Reviews: A Transnational Transdisciplinary Journal*, 12(1), pp. 33–47.

Kleven, H. J. and Kreiner, C. T. (2006) The marginal cost of public funds: Hours of work versus labor force participation, *Journal of Public Economics*, 90(10–11), pp. 1955–1973.

Knight, F. H. (1924) Some fallacies in the interpretation of social cost, *The Quarterly Journal of Economics*, 38(4), pp. 582–606.

Kraus, M. (2003) A new look at the two-mode problem, *Journal of Urban Economics*, 54(3), pp. 511–530.

Kraus, M. and Yoshida, Y. (2002) The commuter's time-of-use decision and optimal pricing and service in urban mass transit, *Journal of Urban Economics*, 51(1), pp. 170–195.

Kuah, G. K. and Perl, J. (1988) Optimization of feeder bus routes and bus-stop spacing, *Journal of Transportation Engineering*, 114(3), pp. 341–354.

Lévy-Lambert, H. (1968) Tarification des services a qualité variable- application aux péages de circulation, *Econometrica*, 36(3/4), pp. 564–574.

Lindsey, R. (2006) Do economists reach a conclusion on road pricing? The intellectual history of an idea, *Econ Journal Watch*, 3(2), pp. 292–379.

Mackie, P. J. (2005) The London congestion charge: a tentative economic appraisal. A comment on the paper by Prud'homme and Bocajero, *Transport Policy*, 12(3), pp. 288–290.

Marchand, M. (1968) A Note on optimal tolls in an imperfect environment, *Econometrica*, 36(3/4), pp. 575–581.

Marcucci, E. (2001) Road pricing: Old beliefs, present awareness and future research paths, *International Journal of Transport Economics*, XXVIII(1), pp. 49–80.

Marcucci, E., Marini, M. and Ticchi, D. (2005) Road pricing as a citizen-candidate game, *European Transport /Trasporti Europei*, (31), pp. 28–45.

May, A., Koh, A., Blackledge, D. and Fioretto, M. (2010) Overcoming the barriers to implementing urban road user charging schemes, *European Transport Research Review*, 2(1), pp. 53–68.

Mazzulla, G. and Eboli, L. (2006) A Service Quality experimental measure for public transport, *European Transport/Trasporti Europei*, (34), pp. 42–53.

Mohring, H. (1972) Optimization and scale economies in urban bus transportation, *American Economic Review*, 62(4), pp. 591–604.

Mohring, H. (1983) Minibuses in urban transportation, *Journal of Urban Economics*, 14(3), pp. 293–317.

Parry, I. W. H. and Bento, A. (2001) Revenue recycling and the welfare effects of road pricing, *The Scandinavian Journal of Economics*, 103(4), pp. 645–671.

Parry, I. W. H. and Small, K. A. (2009) Should urban transit subsidies be reduced?, *American Economic Review*, 99(3), pp. 700–724.

Pedersen, P. A. (2003) On the optimal fare policies in urban transportation, *Transportation Research Part B*, 37(5), pp. 423–435.

Pels, E. and Verhoef, E. (2007) Infrastructure pricing and competition between modes in urban transport, *Environment and Planning A*, 39(9), pp. 2119–2138.

Pickrell, D. (1985) Rising deficits and the uses of transit subsidies in the United States, *Journal of Transport Economics and Policy*, 19(3), pp. 281–298.

Preston, J. M. (2008) Public transport subsidisation, in: S. Ison and T. Rye (Eds) *The Implementation and Effectiveness of Transport Demand Management Measures: An International Perspective*, Chap. 10, pp. 189–209 (Hampshire: Ashgate Publishing Limited).

Proost, S. and Van Dender, K. (2008) Optimal urban transport pricing in the presence of congestion, *economies of density and costly public funds, Transportation Research Part A*, 42(9), pp. 1220–1230.

Proost, S., De Borger, B. and Koskenoja, P. (2007) Public finance aspects of transport charging and investments, in: A. De Palma, R. Lindsey and S. Proost (Eds) *Investment and the Use of Tax and Toll Revenues in the Transport Sector, Research in Transportation Economics* 19, pp. 59–80 (Oxford: Elsevier).

Prud'homme, R. and Bocarejo, J. P. (2005) The London congestion charge: a tentative economic appraisal, *Transport Policy*, 12(3), pp. 279–287.

Quinet, E. (2005) Alternative pricing doctrines, *Research in Transportation Economics*, 14, pp. 19–47.

Rotaris, L., Danielis, R., Marcucci, E. and Massiani, J. (2010) The urban road pricing scheme to curb pollution in Milan, Italy: Description, impacts and preliminary cost-benefit analysis assessment, *Transportation Research Part A*, 44(5), pp. 359–375.

Rouwendal, J. and Verhoef, E. T. (2006) Basic economic principles of road pricing: from theory to applications, *Transport Policy*, 13(2), pp. 106–114.

Safirova, E., Houde, S., Lipman, D. A., Harrington, W. and Baglino, A. (2006) Congestion pricing, long-term economic and land-use effects, RFF Discussion Paper 06-37, Washington DC.

Santos, G. (2005) Urban congestion charging: A comparison between London and Singapore, *Transport Reviews*, 25(5), pp. 511–534.
Santos, G. and Shaffer, B. (2004) Preliminary result of the London congestion charging scheme, *Public Works Management and Policy*, 9(2), pp. 164–181.
Sherman, R. (1971) Congestion interdependence and urban transit fares, *Econometrica*, 39(3), pp. 565–576.
Small, K. A. (2004) Road pricing and public transport, *Research in Transportation Economics*, 9(1), pp. 133–158.
Small K. A. (2008) Urban transportation policy: A guide and road map, Unraveling the Urban Enigma: City Prospects, City Policies, Conference and Book, Wharton School, University of Pennsylvania, 3 May 2007.
Small, K. A. and Verhoef, E. T. (2007) *The Economics of Urban Transportation* (Abingdon: Routledge, Taylor & Francis Group).
Tabuchi, T. (1993) Bottleneck congestion and modal split, *Journal of Urban Economics*, 34(3), pp. 414–431.
TDC. (2010) *2008/09 Household Travel Survey*. Summary Report, NSW: Transport Data Centre.
Tirachini, A. and Hensher, D. A. (2011) Bus congestion, *optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors, Transportation Research Part B*, 45(5), pp. 828–844.
Tisato, P. (1998) Optimal bus subsidy and cross subsidy with a logit choice model, *Journal of Transport Economics and Policy*, 32(3), pp. 331–350.
Tsekeris, T. and Voß, S. (2009) Design and evaluation of road pricing: state-of-the-art and methodological advances, *Netconomics*, 10(1), pp. 5–52.
Turvey, R. and Mohring, H. (1975) Optimal bus fares, *Journal of Transport Economics and Policy*, 9(3), pp. 280–286.
Verhoef, E., Nijkamp, P. and Rietveld, P. (1996) Second-best congestion pricing: the case of an untolled alternative, *Journal of Urban Economics* 40*(3)*, pp. 279–302.
Vickrey, W. S. (1969) Congestion theory and transport investment, *The American Economic Review*, 59(2), pp. 251–260.
Wichiensin, M., Bell, M. G. H. and Yang, H. (2007) Impact of congestion charging on the transit market: An inter-modal equilibrium model, *Transportation Research Part A*, 41(7), pp. 703–713.
Wright, L. and Hook, W. (2007) *Bus rapid transit planning guide* , 3rd edn (New York: Institute for Transportation and Development Policy).

Appendix

The first-order conditions for first best (Section 4.3):

$$L = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda[f_b K_b - q_b], \qquad (A1)$$

$$\frac{\partial L}{\partial q_a} = \frac{\partial B}{\partial q_a} - c_a - q_a \frac{\partial c_a}{\partial q_a} - q_b \frac{\partial c_b}{\partial q_a} = 0, \qquad (A2)$$

$$\frac{\partial L}{\partial q_b} = \frac{\partial B}{\partial q_b} - q_a \frac{\partial c_a}{\partial q_b} - c_b - q_b \frac{\partial c_b}{\partial q_b} - \lambda = 0, \qquad (A3)$$

$$\frac{\partial L}{\partial q_e} = \frac{\partial B}{\partial q_e} - c_e = 0, \qquad (A4)$$

$$\frac{\partial L}{\partial f_b} = -q_a \frac{\partial c_a}{\partial f_b} - q_b \frac{\partial c_b}{\partial f_b} + \lambda K_b = 0, \qquad (A5)$$

$$\frac{\partial L}{\partial K_b} = -q_a \frac{\partial c_a}{\partial K_b} - q_b \frac{\partial c_b}{\partial K_b} + \lambda f_b = 0, \qquad (A6)$$

$$\lambda[f_b K_b - q_b] = 0. \qquad (A7)$$

Recalling the equilibrium condition (4), (A2) and (A4) yield results (8) and (9).

The first-order conditions for second best (Section 4.4):

$$L = B(q_a,q_b,q_e) - q_a c_a(q_a,q_b,f_b,K_b) \\ - q_b c_b(q_a,q_b,f_b,K_b) - q_e c_e + \lambda[f_b K_b - q_b] + \gamma_a\left(c_a - \frac{\partial B}{\partial q_a}\right) + \gamma_b\left(c_u + \tau_b - \frac{\partial B}{\partial q_b}\right) \\ + \gamma_e\left(c_e - \frac{\partial B}{\partial q_e}\right), \quad (A8)$$

$$\frac{\partial L}{\partial q_a} = \frac{\partial B}{\partial q_a} - c_a - q_a\frac{\partial c_a}{\partial q_a} - q_b\frac{\partial c_b}{\partial q_a} + \gamma_a\left(\frac{\partial c_a}{\partial q_a} - \frac{\partial^2 B}{\partial q_a^2}\right) + \gamma_b\left(\frac{\partial c_u}{\partial q_a} - \frac{\partial^2 B}{\partial q_a \partial q_b}\right) - \gamma_e\frac{\partial^2 B}{\partial q_a \partial q_e} = 0, \quad (A9)$$

$$\frac{\partial L}{\partial q_b} = \frac{\partial B}{\partial q_b} - q_a\frac{\partial c_a}{\partial q_b} - c_b - q_b\frac{\partial c_b}{\partial q_b} - \lambda + \gamma_a\left(\frac{\partial c_a}{\partial q_b} - \frac{\partial^2 B}{\partial q_b \partial q_a}\right) + \gamma_b\left(\frac{\partial c_u}{\partial q_b} - \frac{\partial^2 B}{\partial q_b^2}\right) - \gamma_e\frac{\partial^2 B}{\partial q_b \partial q_e} = 0, \quad (A10)$$

$$\frac{\partial L}{\partial q_e} = \frac{\partial B}{\partial q_e} - c_e - \gamma_a\frac{\partial^2 B}{\partial q_e \partial q_a} - \gamma_b\frac{\partial^2 B}{\partial q_e \partial q_b} - \gamma_e\frac{\partial^2 B}{\partial q_e^2} = 0, \quad (A11)$$

$$\frac{\partial L}{\partial f_b} = -q_a\frac{\partial c_a}{\partial f_b} - q_b\frac{\partial c_b}{\partial f_b} + \lambda K_b + \gamma_a\frac{\partial c_a}{\partial f_b} + \gamma_b\frac{\partial c_u}{\partial f_b} = 0, \quad (A12)$$

$$\frac{\partial L}{\partial K_b} = -q_a\frac{\partial c_a}{\partial K_b} - q_b\frac{\partial c_b}{\partial K_b} + \lambda f_b + \gamma_a\frac{\partial c_a}{\partial K_b} + \gamma_b\frac{\partial c_u}{\partial K_b} = 0, \quad (A13)$$

$$\frac{\partial L}{\partial \gamma_a} = c_a - \frac{\partial B}{\partial q_a} = 0, \quad (A14)$$

$$\frac{\partial L}{\partial \gamma_b} = c_u + \tau_b - \frac{\partial B}{\partial q_b} = 0, \quad (A15)$$

$$\frac{\partial L}{\partial \gamma_e} = c_e - \frac{\partial B}{\partial q_e} = 0, \quad (A16)$$

$$\frac{\partial L}{\partial \tau_b} = \gamma_b = 0, \quad (A17)$$

$$\lambda[f_b K_b - q_b] = 0. \quad (A18)$$