CONGESTION PRICING, TRANSIT SUBSIDIES AND DEDICATED BUS LANES: EFFICIENT AND PRACTICAL SOLUTIONS TO CONGESTION

Leonardo J. Basso *,**
Cristián Angelo Guevara +
Antonio Gschwender *
Marcelo Fuster +

FEBRUARY 2010

ABSTRACT

We analyze urban congestion management policies through numerical analysis of a simple model that: allows users to choose between car, bus or an outside option (biking); consider congestion interactions between cars and buses; and allow for optimization of frequency, vehicle size, spacing between stops and percentage of capacity to be dedicated to bus lanes. We compare resulting service levels, social welfare and consumer surplus for a number of different policies and find that (i) dedicated bus lanes is a better stand-alone policy than transit subsidization or congestion pricing. The latter is marginally better than subsidization but has a negative impact in consumer surplus. (ii) efficient transit subsidies are quite large since in many cases first-best transit price is negative; establishing dedicated bus lanes or implementing congestion pricing render subsidies unnecessary for high demand levels. (iii) both subsidization and dedicated bus lanes would count with public support while congestion pricing would probably encounter strong opposition. (iv) transit subsidies and/or congestion pricing do not induce large changes on optimal bus size, frequency, circulation speeds and spacing between stops in mixed-traffic conditions: dedicated bus lanes do. (v) In all cases analyzed, revenues from congestion pricing are enough to cover transit subsidies; the optimal percentage of capacity that should be devoted for bus traffic is around one third.

* Civil Engineering Department, Universidad de Chile.
+ Facultad de Ingeniería, Universidad de los Andes
** Correspondence: lbasso@ing.uchile.cl, Casilla 228-3, Santiago, Chile

Acknowedgments: This project was partially funded by project FAI ICIV 001-08 and Instituto Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16)

1. INTRODUCTION

Congestion is, undoubtedly, the externality caused by urban transportation that has attracted the largest share of attention from economists and engineers. In the literature, two of the most popular ways to deal with congestion that have been suggested are congestion pricing and giving priority to public transportation. The first, because it follows the Pigouvian tradition of charging for the external costs produced by an agent's decision. The second, because public transportation provides a mean of transporting more people using less space, thus diminishing congestion.

Congestion pricing has been analyzed in a very large number of settings: with steady-state congestion models, with queuing models, taking into account network equilibrium, and considering all sort of second best issues such as the impossibility of pricing all links or having time-dependent charges. Many articles, books and review articles have been written on the topic of congestion pricing; references can be found in Small and Verhoef (2007) and Tsekeris and Voβ (2008). A particular feature of the results we would like to stress is that in most cases, if congestion pricing is implemented, travelers surplus will decrease since the full price consumers pay (time costs plus the tax) is larger than the time costs they pay without congestion pricing. Thus, total social welfare would be increased because tax collection dominates the travelers surplus reduction, making revenue recycling an important issue if political support is to be raised.

On the other hand, many authors have studied the optimal design of scheduled public transport services (Mohring, 1972, Jansson, 1984), seeking frequencies, vehicle sizes, and number and spacing of bus stops that minimize total costs. Although depending on the specific setting, the main result here is that, when one takes into consideration the resources supplied by operators (energy, crew, maintenance, administration, infrastructure, rolling stock and so on) and users (waiting, access and in-vehicle times), the efficient cost minimizing service requires subsidies. This happens because the sum of operators' and users' costs yields a total cost that grows less than proportionally with the demand, implying scale economies; this is sometimes known as *the Mohring effect* (for a review see e.g. Jara-Díaz and Gschwender, 2003; Jara-Díaz, 2007).

Now, as it is evident, in most cities people have a choice between using a car or public transportation, and these two modes share road capacity and thus interact with each other. This happens directly on the road, when vehicles are in motion, or when passengers are boarding and alighting in bus stops. In other words, buses delay cars and cars delay buses. Yet, as important as this may seem in practice, it has been very uncommon in the literature to consider congestion pricing and optimization of scheduled public transportation in a unique, joint model. Most of the congestion pricing literature deals with cases where only cars are considered, while the public transportation literature do not consider interactions either, nor the fact that buses may impose congestion on other buses if they are too many. Thus, we believe there is an important void that needs to be filled in order to better understand the full implications of different measures targeted at dealing with congestion in cities, such as congestion pricing, transit subsidies or dedicated bus lanes. Importantly as well, this should help to better assess what may be the level of public and political support for each of these policies; note that, depending on the country, some of these policies may have a real tough time being accepted and implemented. This is certainly known for congestion pricing, but it is also the case for transit subsidies in many developing countries.[1]

Some papers that do consider some of the features we are interested in are Mohring (1972), Small (1983), Viton (1983) and Huang (2000). Yet these models either do not fully consider the interactions between cars and buses, do not allow for optimization of some very relevant variables –such as the percentage of capacity dedicated to buses or bus size– or deals only with minimization of resources, without really considering demand effects. In this paper we propose a simple tractable optimization model that: (i) allows users to choose between car, public transportation or an outside option (biking) through a discrete choice model (ii) consider congestion interactions between cars and buses, including the effects of bus stops (iii) allow for optimization of frequency, vehicle size, spacing between stops and the percentage to be dedicated to bus lanes. We choose the best parameter values possible and numerically solve different optimization problems, each of which corresponds to a combination of alternative urban transport policies. Analyzing the results we can have first-idea of what would be the outcomes of these different policies, such as congestion pricing, allowing for transit subsidies (with or without a constraint on subsidies being covered with revenues from congestion pricing), dedicating a percentage of capacity only for buses, or any combination of these. The model is indeed parsimonious, with a number of simplifying assumption, yet we believe is an important first-step in a direction truly important for a deeper understanding of the impacts of transport policies.

Our results show, among other things, that: (i) dedicated bus lanes is a better stand-alone policy than transit subsidization or congestion pricing. The latter is marginally better than subsidization but has a negative impact in consumer surplus. (ii) efficient transit subsidies are quite large since in many cases first-best transit price is negative; establishing dedicated bus lanes or implementing congestion pricing render subsidies unnecessary for high demand levels. (iii) both subsidization and dedicated bus lanes would count with public support while congestion pricing would probably encounter strong opposition. (iv) transit subsidies and/or congestion pricing do not induce large changes on optimal bus size, frequency, circulation speeds and spacing between stops in mixed-traffic conditions: dedicated bus lanes do. (v) In all cases analyzed, revenues from congestion pricing are enough to cover transit subsidies; the optimal percentage of capacity that should be devoted for bus traffic is around one third.

The plan of the paper is as follows. In Section 2 we explain in detail the model we use and how it captures what we consider to be the essential features of the problem. Section 3 explains the scenarios we analyze, specifying objective functions and constraints, and shows the parameter values used. Section 4 contain our numerical results and the comparative analyses between scenarios and urban transport policies. Section 5 summarizes our results and concludes.

---

[1] In the case of Chile, the discussion about whether to subsidize the new metropolitan transits system, Transantiago, achieved a large public notoriety, spiced up by the rather poor results of the system in its first months. In the end, subsidies were approved but these are only aimed at covering a fare reduction that students receive, something that before was covered with the fare revenues collected from the rest of the users. The spirit of Transantiago is that it has to be self-financed.

2. THE MODEL

We consider a road of infinite length with a capacity of Q vehicles/hour, where Y commuters per kilometer and hour travel l kms in the same direction. In order to avoid dealing with border conditions, we shall model a representative kilometer of the road; in this sense, a circular road model would be equivalent. All travel commuters choose one of three modes –car, bus or bicycle– in a utility maximization framework. For the two motorized modes we consider congestion externalities caused both by their interaction while in motion, as well as congestion caused by the existence of bus stops. The variables that the planner can (potentially) adjust in order to maximize social welfare are: bus frequency, f [bus/h] and bus capacity, K [passengers/bus]; the bus fare P_b [\$/trip]; number of equidistant bus stops per kilometer, p; the congestion toll for cars P_a [\$/km]; and the percentage of road capacity dedicated exclusively to bus services, η. Obvisouly, as these variables change, utility levels are affected and, consequently, so will be the modal split. The possible policies we consider are: congestion pricing, transit subsidies and dedicated bus lanes. Then, the scenarios we analyze are made of combinations of these policies and, therefore, some of the variables may not be available to the planner in some of the scenarios (for example, in some of the scenarios we will not allow the planner to use dedicated bus lanes). An optimum point, then, for a given scenario is reached when the welfare is maximized and each user has chosen the mode that maximizes her utility.

We can now move unto the specific functional forms we consider. Let us start by the modal utilities. The utility a commuter perceives, for traveling by automobile (a), bus (b) or cycling (c) are respectively:

$$U_a = Inc + B_a\theta - l\left(SVT \cdot t_a + \frac{(P_a + c_a)}{occ}\right) - \frac{g}{occ} \tag{1}$$

$$U_b = Inc + B_b\theta - P_b - SVT \cdot \left(lt_b + \frac{\gamma_E}{2f} + \frac{\gamma_{AC}}{2p\nu_{AC}}\right) \tag{2}$$

$$U_c = Inc - SVT \cdot lt_c \tag{3}$$

In each case, the utility of using a mode corresponds to the benefits of undertaking the trip, given here by the daily income Inc –which without loss of generality we normalize to zero– plus a modal constant which we will discuss further momentarily, minus generalized costs. These generalized costs include car tolls and bus fares, P_a and P_b; and in vehicle travel times, t_a , t_b and t_c, which are multiplied by the Subjective Value of Time, SVT, and the travel distance l. In (1) we also consider operational cost per kilometer, c_a, and parking costs, g, which are shared by the occupants of a car (occ). In (2) on the other hand, we also consider (average) waiting time, given by $1/2f$, and (average) walking time to and from the bus-stop, given by $1/2p\nu_{AC}$, where ν_{AC} [km/h] is walking speed. γ_{AC} and γ_E are the ratios between the in-vehicle SVT and waiting SVT and walking SVT respectively.

Something that is key to capture is the fact that people, even if facing the exact same alternatives, do different things. This users' heterogeneity can be addressed in a number of ways, such as differences in income, differences in values of time (these two may be related through the marginal utility of income, see e.g. Jara-Diaz, 2007) and so on. Here, we have chosen a simpler framework which, perhaps at the expense of some realism, increases tractability: we assume that all commuters share the same value of time and income but differ in their valuation of some other attributes such as safety, comfort, social status and so on. The level of these other attributes are modal specific and captured by B_i in equations (1)-(3). In this way θ is an idiosyncratic term that varies across the population and that accounts for the importance each person assigns to the other attributes. We assume that $B_a > B_b > B_c$, that θ is uniformly distributed in [0;1] and, without further loss of generality, that $B_c = 0$; in other words, that ceteris paribus, people prefer the car over the bus because, for example, of higher comfort or status, but not everyone with the same intensity. Note though, that the values of Bs and θ are intertwined, that is, one can always allow θ to achieve a larger value and then adjust Bs adequately so that the modal split does not change. Also, the order of the Bs may change if one consider different attributes. For example, B may be the amount of pollutants per person each mode emits (which would reverse the order between modes) and θ may be *green consciousness*. Finally, Small (1983) found that socio-economic variables play an important role in mode choice; for example, families with young children are more likely, *ceteris paribus*, to choose a car; these type of effects may also be captured in our modeling.

With these assumptions and equations (1) to (3), it is easy to show that under mild conditions, there exist threshold values of θ characterized by $0 < \theta^b < \theta^a < 1$, which define a modal split where people with value of θ between 0 and θ^b choose cycling, people with value of θ between θ^b and θ^a choose bus, while the remainder choose car – we refer readers to Basso and Zhang (2008) for the proof, as this is analogous to the one there for the choice of peak and off-peak travel. Thus, the number of people using each mode is given by:

$$Y_a = Y\left(1-\theta^a\right) \tag{4}$$

$$Y_b = Y\left(\theta^a - \theta^b\right) \tag{5}$$

$$Y_c = Y - Y_a - Y_b = Y\left(\theta^b - 0\right) \tag{6}$$

where the values of the thresholds are obtained by searching for the indifferent types, i.e. by equating the utilities. This process straightforwardly leads to:

$$\theta^a = \frac{\dfrac{l\left(P_a + c_a\right) + g}{occ} - P_b - SVT\left(l\left(t_b - t_a\right) + \dfrac{\gamma_E}{2f} + \dfrac{\gamma_{AC}}{2p\nu_{AC}} \right)}{B_a - B_b} \tag{7}$$

$$\theta^b = \frac{P_b + SVT\left(l(t_b - t_c) + \frac{\gamma_E}{2f} + \frac{\gamma_{AC}}{2p\nu_{AC}} \right)}{B_b} \tag{8}$$

Note that, by replacing the threshold values (7) and (8) in equations (4) to (6), one obtains the number of consumers per mode as functions of the variables that the planner chooses, such as the congestion toll for cars P_a and the bus fare P_b, or the frequency f of the transit system. These, however, are not real demand functions because travel times t_a and t_b do depend on the numbers of users of each mode because of congestion effects. Hence, what we have, more than a system of demands, is a fixed-point system that implicitly defines the modal split as a function of optimization variables. Yet, what it is possible to do to overcome this difficulty is to describe the optimization problem in terms of inverse demands, that is, the problem of finding a modal split Y_a and Y_b plus the other three variables, rather than one of finding prices plus the other variables. The two problems are indeed equivalent, but using inverse demands is computationally simpler because they have explicit expressions, easily obtained by replacing (7) and (8) in equations (4) – (6), and then solving for P_a and P_b. We get:

$$P_a = \frac{a\left(B_a(Y - Y_a) + B_b Y_b\right)}{Yl} - \frac{aSVT(t_a - t_c)l + c_a l + g}{l} \tag{9}$$

$$P_b = \frac{B_b(Y - Y_a - Y_b)}{Y} - SVT\left(l(t_b - t_c) + \frac{\gamma_E}{2f} + \frac{\gamma_{AC} l}{2p\nu_{AC}} \right) \tag{10}$$

We can now move on to the central issue of in-vehicle travel time functions, i.e. t_a, t_b and t_c. Ideally, one would like to use functions that capture, as close to reality as possible, the effects that the distance between stops, number and size of buses and cars, and number of lanes has on the average speed of cars and buses. Yet, we are not aware of any model that proposes this in a manner that can be interacted with the microeconomic framework and, therefore, we have opted for choosing simple linear forms, which capture the effects we desire yet may be unrealistic if second order effects are strong.

Suppose first that buses and cars are physically separated, such that buses can use a proportion η of the capacity Q, while cars use $(1-\eta)$. The time that a car takes to travel one kilometer will be given by:

$$t_a = \alpha \left(\frac{\frac{lY_a}{occ}}{(1-\eta)Q} \right) + \beta \tag{11}$$

where the figure in the numerator corresponds to the flow of cars –since l is the distance of each trip, and occ is the number of people per car– and thus show congestion effects. α and β are parameters. Obviously, since in this case buses and cars are not really interacting with each other, there are no cross-congestion effects.

On the other hand, the time it takes a bus to travel one kilometer when it has exclusive use of a proportion η of the road capacity is:

$$t_b = \left(\alpha \left(\frac{bf}{\eta Q} \right) + \beta \right) + \frac{Y_b t_{sb}}{f} + t_p p \tag{12}$$

On the right hand side of (12), the first term in brackets represent travel time while the vehicle is in motion: buses, like cars, can suffer from congestion. The flow of buses is multiplied by an equivalence factor *b* that captures the differences in size and maneuverability between cars and buses, factor that has usually been assumed to be constant (e.g. Mohring 1979). Here, however, we let this parameter be given by

$$b(K) = \frac{K}{100} + 1 \tag{13}$$

where *K* is the capacity of the bus. For values of *K* that go from 50 to 200, (13) nicely captures the fact that a bus is equivalent to something between 1.5 and 3 cars, thresholds habitually used in the literature for this factor. In the second term in (12), t_{sb} is the average time that a passenger takes to board and alight the bus, thus the product of this term and the demand for bus Y_b , devided by the frequency *f*, captures the delays caused by bus stops operations. Finally, the third term captures the fact that, in order to load and unload the bus at a bus stop, the driver has to slow down the bus before stopping and then speed up the vehicle, which causes further delays at a rate of *p* seconds per stop.

Let us consider now mixed-traffic conditions, where buses and cars share the road. Here, we would like to capture not only that buses and cars causes congestion to each other, but also that bus stop operations can cause delays to car users. As mentioned above, we are not aware of a travel time function that, grounded on real data, delivers the effects that the distance between stops, number, size and load factor of buses has on average speed of cars and buses. Hence what we do, is to simply consider that a fraction of the extra-time that a bus requires for bus top operations is also incurred by cars. We set this fraction to one half (since it may be possible for the car to surpass a bus), and thus obtain the travel time for cars for mixed-traffic conditions as:

$$t_a = \left(\alpha \left(\frac{bf + \frac{lY_a}{occ}}{Q} \right) + \beta \right) + \left(\frac{\frac{Y_b t_{sb}}{f} + t_p p}{2} \right) \tag{14}$$

Note that in the first term, the capacity is now shared (there is no η) and a bus is treated as *b* cars, according to (13).

Buses on the other hand, still use time for boarding-alighting operations and acceleration from bus stops, but now they also suffer from congestion caused by cars. Their travel time function in mixed-traffic conditions is then:

$$t_b = \left(\alpha \left(\frac{bf + \frac{lY_a}{occ}}{Q} \right) + \beta \right) + \frac{Y_b t_{sb}}{f} + t_p p \tag{15}$$

Finally, before moving to the planner's objective function and the different policies she may implement, we need to specify one final function and a technical constraint. The cost of the bus system (in dollars per hour) is given by

$$C_b = \left(c_{b0} + c_{b1}K\right) f t_b \tag{16}$$

where the term in brackets represents operational cost per bus and hour, which are larger for larger buses ($c_{b1} > 0$). And since bus capacity may also be optimized, a constraint that ensures that buses are large enough to carry the demand has to be imposed, that is:

$$1 \geq \frac{Y_b l}{f} \Big/ K \tag{17}$$

i.e. the load factor has to be smaller than 100%.

3. OBJETIVE FUNCTION, SCENARIOS AND PARAMETER VALUES

We consider that the planner seeks to maximize a social welfare function given by the (un-weighted) sum of consumer surplus plus government revenues minus operational transit costs. What changes from one scenario to the next are the policies that the planner chooses to –or can– implement. These policies are congestion pricing, transit subsidies and dedicated bus lanes and, therefore, the scenarios we analyze –which as we show below correspond to different constraints imposed on the optimization problem– are made of combinations of these policies. We start then by explicitly writing the common objective function. The first thing needed is consumer surplus; given that the modal split is univocally related to θ values, we can calculate it as:

$$CS = Y \left(\int_{\theta^a}^{\bar{\theta}} U_a d\theta + \int_{\theta^b}^{\theta^a} U_b d\theta + \int_{\underline{\theta}}^{\theta^b} U_c d\theta \right) \tag{18}$$

where U_a, U_b and U_c are given by equations (1) to (3) and directly depend on the optimization variables f, K, Y_a, Y_b and η, the latter when relevant. Note that the utilities depend on travel time functions, which in turn depend on optimization variables. The expression one obtains after replacing both the utilities and the travel time functions in (18) is not particularly informative hence we omit it. It is also important to recognize that the travel time functions to be used depend on whether exclusive lanes are considered as a policy or not: One should use (11) and (12) for the case of segregated circulation, and (14) and (15) for the case of mixed traffic. Still, in (18) the integration limits are not clearly dependent on optimization variables. Using (4) to (6) however, one can easily write the limits as functions of Y_a and Y_b, which are optimization variables:

$$\theta^a = 1 - \frac{Y_a \overline{\theta}}{Y} \qquad \theta^b = \frac{Y - Y_a - Y_b}{Y} \tag{19}$$

Having consumer surplus at hand, the (per hour) social welfare function is easily finished by adding the financial result of the bus system and congestion pricing revenues:

$$SW = CS + P_b Y_b - C_b + P_a \frac{Y_a}{occ} l \tag{20}$$

where the second term on the left hand side is (per hour) transit revenue, the third term is transit cost, and the fourth term is congestion pricing revenue. Importantly, note that we have not included costs related to the implementation of congestion pricing, which may be sizeable (see e.g. Prud'homme and Bocarejo, 2005).

We can now move to the description of the different scenarios we analyze. In all cases, the planner maximizes (20) and must consider at least three technical constraints: First, there is the constraint of minimal bus size, given by $K \geq Y_b l f^{-1}$. Yet, since having idle capacity only decreases the value of the objective function, buses will always be chosen to meet demand so the constraint binds. On the other hand, the number of commuters in each mode cannot be negative. In summary, the planner must consider:

$$K = \frac{Y_b l}{f} \quad , \quad Y_a \geq 0 \quad , \quad Y_b \geq 0 \quad , \quad Y_a + Y_b \leq Y \tag{17}$$

We can now move to the description of the eight different scenarios we analyze. The first four consider mixed-traffic conditions –thus we use equations (11) and (12) for travel times there– while the last four consider that the percentage of exclusive capacity for buses, η, is also optimized –thus we use equations (14) and (15) there. It may be important to remind the reader here that even though from a mathematical point of view, the problems will be described as ones of choosing frequency, bus stop spacing and demand levels, these are absolutely equivalent to the –perhaps more natural– problems of choosing frequency, bus stop spacing and bus fare and congestion toll. The scenarios are as follows:

<u>Scenario 1</u>*: Self-financing transit, no congestion pricing, mixed-traffic*

The first scenario we consider corresponds to the current situation in many cities around the world and therefore we refer to it as our base case. It features self-financing for the bus system (through fares only), absence of congestion pricing and road capacity shared by buses and cars. The problem solved by the planner in this scenario is then:

$$
\begin{aligned}
&\text{Max}\, SW \quad w.r.t \quad f, p, Y_a, Y_b \\
&s.t. \quad eqs.\ (11), (12), (17) \\
&\qquad P_a = 0\ ,\ P_b\, Y_b \geq C_b
\end{aligned}
$$

<u>Scenario 2</u>: Transit subsidies, no congestion pricing, mixed-traffic

In this second case we consider a transit subsidization policy, which here takes the form of no longer asking the transit fare to cover transit costs. In this sense, the subsides are *optimal* given the rest of the system environment; a (worse) alternative would be to fix an *a priori* amount for the subsidy. We also note that the way in which the money for subsidies is raised is not considered in the optimization problem:

$$
\begin{aligned}
&\text{Max}\, SW \quad w.r.t \quad f, p, Y_a, Y_b \\
&s.t. \quad eqs.\ (11), (12), (17) \\
&\qquad P_a = 0
\end{aligned}
$$

<u>Scenario 3</u>: Transit subsidies, congestion pricing, mixed-traffic

The third scenario, in addition to transit subsidies, considers congestion pricing which, according to the model above, consists of a per-kilometer charge. Here we do not necessarily link transit subsidies to congestion pricing revenues. Given the policies at hand, this scenario will lead to the maximal social welfare level for mixed-traffic conditions.

$$
\begin{aligned}
&\text{Max}\, SW \quad w.r.t \quad f, p, Y_a, Y_b \\
&s.t. \quad eqs.\ (11), (12), (17)
\end{aligned}
$$

<u>Scenario 4</u>: Transit subsidies paid for by congestion pricing revenues, mixed-traffic

In the final scenario with mixed-traffic conditions, what we intend to explore by comparison with scenario 3 is whether optimal transit subsidies can be covered by optimal congestion pricing plus optimal bus fare. In other words, whether imposing a urban transport sector *self-financing constraint* leads to welfare losses or not. The optimization problem is now:

$$
\begin{aligned}
&\text{Max}\, SW \quad w.r.t \quad f, p, Y_a, Y_b \\
&s.t. \quad eqs.\ (11), (12), (17) \\
&\qquad \frac{Y_a}{occ} P_a\, l + P_b Y_b \geq C_b
\end{aligned}
$$

The next four scenarios, 5 to 8, are similar to scenarios 1 through 4 but now we consider cases where buses circulate on dedicated lanes that use a share η of the total capacity. We optimize η but allow it to take the values 1/3 or 2/3 only, that is, the decision is whether to give one or two lanes exclusively for buses. Note that a continuous optimization of η is much more problematic that continuous optimization of other variables that are also discrete in nature, since η can only take on two values. Indeed, even if mathematically one obtains for example that it would be optimal to give 15% of the capacity to buses, a bus circulating in half-a-lane does not make much sense.

Given that in these cases there is one extra optimization variable, one could think that compared one to one –for example scenario 2 vs. scenario 6–, welfare will be larger in the latter. This is not directly true however because the travel time functions are now different: now we should use equations (14) and (15). Also, something that seems to be interesting to analyze is how a policy of dedicated lanes –and nothing else– compares to pricing policies (subsidies and taxes) under mixed-traffic conditions. The new scenarios are then:

Scenario 5: Self-financing transit, no congestion pricing, dedicated bus lanes

$$\begin{aligned} \text{Max}\, SW \quad & w.r.t \quad f, p, Y_a, Y_b, \eta \\ & s.t. \quad eq.\ \text{(14), (15), (17)} \\ & P_a = 0\ ,\ P_b Y_b \geq C_b\ ,\ \eta = 1/2\ or\ \eta = 1/3 \end{aligned}$$

Scenario 6: Transit subsidies, no congestion pricing, dedicated bus lanes
Scenario 7: Transit subsidies, congestion pricing, dedicated bus lanes
Scenario 8: Transit subsidies paid for by congestion pricing revenues, dedicated bus lanes

Where we have omitted the optimization problems of the last three scenarios to save on space, and since they are quite obvious by now. As it is evident, the optimization model we propose to solve requires a large number of parameters. And while we do realize that our model is a simplifying abstraction, we have chosen to use parameters that represent reality as close as possible. In this sense, almost all parameters have been obtained or calculated from data that represent a morning-peak in Santiago, Chile, where monetary values correspond to 2006 US dollars.[2] The exception are the B_i parameters, since we did not have direct data for these and obtaining values from calibrated demand models was not simple; what we did then is to chose the values of *B*s such that the modal split in scenario 1 gave us something *reasonable*, meaning that all modes have a positive share, the largest being car, closely followed by bus with biking having a quite small share.

4. SIMULATION RESULTS

All eight scenarios were solved as optimization problems using the software Wolfram Mathematica and considering the parameters in Table 1; numerical results are presented in

[2] Data comes from DFT (2007), Fernández (2008), Ortega (2007), SECTRA (2005) and SEISTU (2001). Further details are available upon request.

Table 2 and analyzed in detail below. In order to increase the insights we can obtain, we have also created Figure 1 that summarizes the results of each scenario in terms of the value achieved of Consumer Surplus and Total Social Welfare, with respect to the base case Scenario 1 (which features no subsidies, no congestion pricing and no dedicated bus lanes). The idea with this is to jointly assess the social goodness of each policy and the level of public support that each policy may find. For example, a policy that produces and increase in social welfare but a decrease in consumer surplus is a policy that may find stronger opposition unless government revenues are *recycled* in some clear and known way.

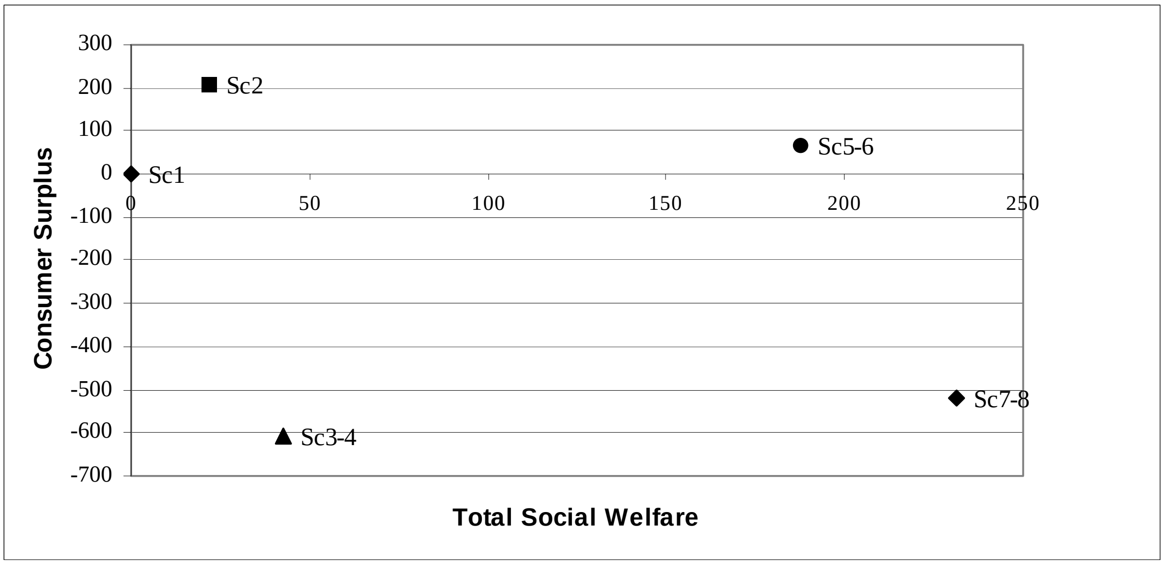

Figure 1: Differences in Consumer Surplus and Total Social Welfare (US$/km/hr) with respect to Scenario 1

Effects of Transit Subsidization

A look at Table 2 and Figure 1, comparing scenarios 1 and 2, shows that subsidizing transit is a policy that works in that it increases total social welfare and consumer surplus. It also changes the modal split importantly: the bus takes all former bicycle users and some car users as well. It is, however, a policy quite expensive for the government that, actually, lead to negative prices. In terms of service variables, when buses are analyzed in isolation, transit subsidies should increase frequency and bus size (see e.g. Jara-Díaz and Gschwender, 2003). Both of this happen here, yet the change in the optimal bus size is marginal, from 104 to 107 passengers. More sizeable is the change in frequency, from 27 to 35 buses/hour, which reduces waiting times. Since there are now less people using cars, circulation speeds increase for both modes, although not by much (0.5 km/hour). The optimal spacing between bus stops does not change.

An important result is that transit subsidization does not seem to be necessary if either congestion pricing or dedicated bus lanes are in place: in scenario 3 where both congestion pricing and subsidies are allowed, the bus system (optimally) generate positive profits, implying that subsidies are not needed. In this sense, scenario 3 can be taken as representing a congestion pricing policy alone. On the other hand, in scenario 5, where dedicates bus lanes are used, the transit system self-finances as well, so that the transit budget constraint is not binding. That is why when we formally allow for subsidies in addition to exclusive lanes, in scenario 6, results do not change. As we discuss below, however, if total demand is smaller (results not shown), transit does not self finance even in the presence of congestion pricing and dedicated bus lanes, and therefore subsidies improve the situation.

Given that subsidization does not seem to be needed when any of the other two policies are at work (and demand is large), it becomes important to see how subsidization does alone compared to other policies also in isolation. In this sense, Figure 1 shows that subsidies produce the smallest increase in social welfare as compared to congestion pricing alone (scenario 3) or dedicated bus lanes alone (scenario 5). In turn, transit subsidization is the policy that produces the largest consumer surplus –obviously due to large negative prices– and, therefore, could be the one with the largest public support. Clearly, if positive prices were imposed (not shown) both social welfare and consumer surplus would decrease.

Effects of Congestion Pricing

The scenarios that consider congestion pricing are scenarios 3, 4, 7 and 8. The optimal congestion tax is calculated at 0.18 US$/km. The first obvious and expected result of congestion pricing is that it induces a change in the modal split, moving commuters from cars to the transit system. It actually moves more people from cars to bus than transit subsidies. Table 2 also reveals that congestion pricing induces: (i) larger speeds –due to decreased car usage– but these are not sizeable (about 1km/hour), (ii) larger bus frequency comparable to subsidization, but without increasing the fleet as much. Bus size and distance between bus stops are again not notably changed.

Importantly, the use of a congestion pricing policy induces a large increase in the bus fare which now is not only positive (recall it was negative under subsidization) but it generate revenues that more than cover transit costs. It is because of this that Scenarios 3 and 4, and 7 and 8 are identical: the question of whether congestion pricing revenues are enough to cover transit subsidies is irrelevant if subsidies are not needed at the optimal situation! Note, however, that the financial result of the transit system would be negative again if one considers a smaller total demand (not shown), i.e. subsidies would be required for smaller demand levels. However, the clear effect of congestion pricing on reducing importantly the size of transit subsidies remains and, if all scenarios are simulated again but with half the total demand, it is always the case that optimal congestion pricing revenues cover optimal transit subsidies; hence, even with half the demand, Scenarios 3 and 4, and 7 and 8 would coincide.

Figure 1 shows some important other insights. First, it can be noted that congestion pricing applied over the base case (from Scenario 1 to 3) produces an increase in social welfare but, at the same time, a decrease in consumer surplus. Thus, the usual result that consumer surplus decreases with congestion pricing remains. Note that in the light of the well know

Downs Thomson Paradox (Mogridge, 1990), this result may seem strange. According to this paradox, if car users are induced to switch to public transport, then this would imply benefits in generalized costs for everyone, as there will be less congestion for cars while the transit system would be more frequent. There are two issues here that explain the differences: first, that in our case we allow the bus system to be congestible, both in mixed traffic conditions and separate circulation; second, that the Downs Thompson paradox looks at costs while here we are taking into account differences in preferences through consumer surplus.

As mentioned before, if congestion pricing is applied then transit subsidies become redundant; congestion pricing though achieves a better result in terms of social welfare. Congestion pricing on top of a policy of dedicated bus lanes (move from Scenario 6 to scenario 7) do increase welfare by a large amount but, again, reduces consumer surplus. These reductions in consumer surplus together with increases in social welfare imply that it may be difficult, politically, to pursue congestion pricing policies without earmarking the revenues and having the public to believe that revenue recycling will indeed occur. It should also be recalled that our analysis considers that toll collection costs are zero, which is obviously not true in reality. The incorporation of collection costs in the model would decrease the welfare benefits of congestion pricing policies.

Mixed Traffic vs Dedicated lanes

What one expects from a policy that assigns part of road capacity to dedicated bus lanes is that bus speed should increase considerably, given that buses are no longer trapped in car congestion. Car speed may increase as well, because cars may now avoid conflict with buses (including bus stops operations), but decreased capacity for cars may have the opposite effect.

The first important result to note when looking at dedicated bus lanes policies is that the optimal number of lanes to be assigned to buses is one in all four cases (scenarios 5 to 8). Comparing scenarios 1 and 5 in order to see what would achieve the bus lanes policy lanes on its own, we see that indeed buses can now go more than three times faster, while cars decrease their speed by two km/hour. This large change in speeds induces a sizeable increase in bus frequency (about 70%) while decreasing the bus size from 100 to 80 people. It is interesting to note that the increase in frequency does not require an increase in bus fleet: the fleet needed is actually 80% smaller than in the base case. Higher bus speeds also induce a larger separation between bus stops, something that neither transit subsidies nor congestion pricing made. All in all, dedicated bus lanes induce sizeable changes in service levels, something that under mixed traffic conditions do not happen. As a result of all these changes, bus demand increases importantly with respect to mixed-traffic conditions. The optimal bus fare is ten cents higher than in the base case which, together with decreased costs lead to a positive financial result for the transit system.

Now, despite the fact that bus fare increased with respect to the base case and that car speed decreased, dedicated bus lanes actually increase consumer surplus, as can be seen in Figure 1, which may lead to think that it is a policy that would count with public support. This happen because transit service levels (with the exception of access time) improved so strongly, that they dominate the other effects. Furthermore, consumer surplus increases in a context in which the financial result of the transit system is positive, implying an even larger increase in social welfare: in fact, dedicated lanes is the policy that, by itself, achieves the largest increase of social welfare, and by a large amount. Note that these profits cannot be used to decrease the fare as this would be welfare reducing: instead they can be earmarked and recycled to push even further the political support for the policy.

As discussed before, the addition of dedicated bus lanes when a subsidization policy is in place (scenario 2 compared to scenario 6) has the effect of rendering subsidies unnecessary for these level of demand. Social welfare will increase substantially yet consumer surplus will decrease slightly, since there is a move from negative to positive prices. For smaller demands, dedicated lanes –just as congestion pricing– decrease the size of the subsidy importantly. This happens because increased speed helps to make the bus system more attractive, so that it is less necessary to use fare. Next, if from mixed-traffic conditions with congestion pricing (scenario 3, where subsidies are not needed), dedicated bus lanes are added (move to scenario 7), then both consumer surplus and social welfare increase; the former slightly, the later by a large amount. Hence implementing dedicated bus lanes seems to be a policy that, from a social welfare point of view, can improve any existent situation.

Congestion Pricing vs. Dedicated Lanes

Mohring (1979) argued that bus speed was one of the most important attributes of the system and that as such, it should be one the central objectives of planners, if they want to increase bus patronage. This is why he considered that dedicating lanes exclusively to bus traffic can be a quite successful policy. Furthermore, he argued that dedicated bus lanes may be a tool equivalent to congestion pricing in achieving a change in modal split. Though to this point, most of the comparison between congestion pricing and dedicated lanes have been made, it is good to spell out clearly what our results show. The scenarios one has to compare are scenarios 3 (pure congestion pricing in mixed-traffic conditions) and scenario 5, where the only policy is dedicated bus lanes (there is no congestion tax and buses have to self-finance). Figure 1 gives us a clear picture: dedicated bus lanes achieve both larger consumer surplus and larger total social welfare, and in both cases differences are large. It also induces a larger bus patronage which may impact other transportation externalities which were not explicitly included in our model, such as air pollution. The implications of all this are that bus lanes may find less opposition from the public than a congestion pricing policy, while achieving a larger positive impact.

5. SUMMARY AND CONCLUSIONS

People have a choice between using a car or public transportation, and these two modes share road capacity and thus interact with each other. This happens directly on the road, when vehicles are in motion, or when passengers are boarding and alighting in bus stops. In other words, buses delay cars and car delay buses. Yet, as important as this may seem in practice, it has been very uncommon in the literature to consider congestion pricing and optimization of scheduled public transportation in a unique, joint model. Most of the congestion pricing literature deals with cases where only cars are considered, while the public transportation literature do not consider interactions either, nor the fact that buses may impose congestion on other buses if there are too many.

This paper deals exactly with this issue, by proposing a simple tractable model that: (i) allows users to choose between car, public transportation or an outside option (biking) through a discrete choice model (ii) consider congestion interactions between cars and buses (iii) allow for optimization of frequency, vehicle size, spacing between stops and the number of lanes to be dedicated to buses if applicable. Analyses of numerical results of the model allowed us to better see the full implications of different measures targeted at dealing with congestion in cities –such as congestion pricing, transit subsidies or dedicated bus lanes–, as well as to explore what may be the level of public support for each of these policies.

Our results show, among other things, that: (i) dedicated bus lanes is the best stand-alone policy possible, as it achieves –and by far– larger social welfare than transit subsidization or congestion pricing. Congestion pricing is marginally better than subsidization in terms of social welfare but has a negative impact in consumer surplus. (ii) without congestion pricing in place, efficient transit subsidies are quite large since in many cases the efficient transit price is negative; establishing dedicated bus lanes or implementing congestion pricing render subsidies unnecessary for high demand levels, while decreasing substantially the amount of subsidy required importantly for smaller demands. (iii) in terms of the support that each policy may raise, both subsidization and dedicated bus lanes should count with public support although, probably, it is easier from a political standpoint, to implement dedicated bus lanes than raise political support for subsidies. Congestion pricing on the other hand, is a policy that will probably encounter strong public opposition. To reverse this, an important level of tax recycling would need to be promised. (iv) in mixed traffic conditions, bus size, frequency, circulation speeds and spacing between bus stops do not change much when moving from the base case to subsidies or congestion pricing. It is only dedicated bus lanes that actually induce sizeable changes in these variables; (v) in all the cases and settings we analyzed, revenues from congestion pricing are enough to cover transit subsidies when required; (vi) for the cases we analyzed, the optimal percentage of capacity that should be devoted for bus traffic is around one third.

Indeed, this is far to be that last word in terms of comparisons of impacts of different congestion management policies. We are absolutely aware of the simplifying assumptions we made to favor tractably. But we do believe that many of our qualitative results will go through in more complicated, and hence realistic, models, as in fact they did when making sensitivity analysis. In the aftermath, the main significance of this research is in showing that an integrated micro-economic framework for the analysis of different urban-transportation policies under multimodality and cross-congestion enable novel analyses of traditional transportation policies.

In our view future research on this topic needs to test whether our results hold or not when different assumptions are made, and should explore other issues which we did not start to address. Some of the improvements we are currently working on, but which are far from what needs to be done to have a complete picture– are: modeling the modal split using a random utility model calibrated with real data, which will allow us to assess the distributional impacts of different policies; modeling two periods, peak and off-peak, to capture the fact that the same bus fleet is used both in periods of high demand and low demand; test whether different congestion models will have a qualitative effect on the results; and, finally, incorporate effects of bus stop congestion.

REFERENCES

Basso, L., Zhang, A., (2008) Sequential Peak-load Pricing: The Case of Airports and Airlines. **Canadian Journal of Economics 41(3)**, 1087-1119.

DFT (2007) WEBTAG: Transport Analysis Guidance. United Kingdom Department for Transport (http://www.webtag.org.uk).

Fernández, R., (2008) Recolección de Datos y Calibración de Modelos de Tiempo de Servicio de Pasajeros en Buses y Metro de Transantiago. Proceedings XV Congreso Panamericano de Ingeniería de Tránsito y Transporte, Cartagena de Indias, Colombia.

Huang. H-J. (2000) Fares and Tolls in a Competitive System with Transit and Highway: the case with two groups of commuters. **Transportation Research Part E 36**. 267-284

Jansson. J.O. (1984) **Transport System Optimization and Pricing**. Wiley

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

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

Mohring. H. (1972) Optimization and Scale Economies in Urban Bus Transportation. **American Economic Review 62**. 591–604.

Mohring. H. (1979) The Benefits of Reserved Bus Lanes. Mass Transit Subsidies and Marginal Cost Pricing in Alleviating Traffic Congestion. In **Current Issues in Urban Economics**. Mieskowosky. P. y Straszheim M. editors.

Mogridge, M (1990) **Travel in towns: jam yesterday, jam today and jam tomorrow?** Macmillan Press, London

Ortega, M., (2008) *Efectos de la Omisión del Costo de los Usuarios en el Diseño de Servicios de Transporte Público Urbano*. Tesis de Magíster del Departamento de Ingeniería Civil de la Universidad de Chile.

Prud'homme, R., Bocarejo J.P., (2005) The London Congestion Charge: A Tentative Economic Appraisal. **Transport Policy 12(3)**, 279-287.

SECTRA (2005) **Análisis y Actualización del Modelo Estraus**. Secretaria Interministerial de Planificación de Transporte, Chile.

SEISTU (2001) **Análisis, Desarrollo y Evaluación de Proyectos Urbanos**. Estudio para la Secretaria Interministerial de Planificación de Transporte, Primera Etapa.

Small. K. (1983) Bus Priority and Congestion Pricing on Urban Highways. In **Research in Transportation Economics**. Keeler. T. Ed. JAI.

Small. KA. and Verhoef. ET. (2007) **The Economics of Urban Transportation**. London. Routledge.

Tsekeris. T. and Voβ. S. (2008) Design and Evaluation of Road Pricing: state-of-the-art and methodological advances. **Netnomics**. forthcoming.

Viton. P.. (1983) Pareto Optimal Urban Transportation Equilibria. In **Research in Transportation Economics.** Keeler. T. Ed. JAI.

TABLES

Table 1: Parameter values used in numerical analysis

Parameter	Description	Value	Units
α	Parameter of in-vehicle travel time (slope)	0.08	[]
β	Parameter of in-vehicle travel time (intercept)	0.0125	[]
SVT	In-vehicle subjective value of travel time savings	1	[US$/h]
γ_{AC}	Ratio between walking (access) value of time and SVT	2	[]
γ_E	Ratio between waiting value of time and SVT	3	[]
v_{AC}	Walking speed	3.6	[km/h]
a	Occupancy rate of cars	1.8	[passengers/car]
B_a	Car modal constant (maximum, for θ=1)	30	[US$]
B_b	Bus modal constant (maximum, for θ=1)	15	[US$]
c_a	Car operational cost per kilometer.	0.3804	[US$/car]
c_{b0}	Fixed operational cost per bus per hour	20.9930	[US$/h]
c_{b1}	Variable operational cost unit of bus capacity per hour	0.2095	[US$/h]
g	Fixed car cost (daily parking)	5	[US$/car]
l	Travel distance	12	[km]
Q	Road capacity	5400	[veh/h]
t_c	Travel time in bicycle	0.0833	[h/km]
t_p	Delays by breaking and accelerating to reach and leave a bus-stop	0.0028	[h/stop]
t_{sb}	Time to board or alight a passenger	0.0007	[h/passenger]
Y	Total demand per kilometer and hour.	900	[pax/km-h]

Table 2: Results of numerical simulations of scenarios

	Scenario 1	Scenario 2	Scenario 3 - 4	Scenario 5 - 6	Scenario 7 - 8
Car demand [pax/km]	622,84	585,56	548,82	571,22	505,81
Bus demand [pax/km]	237,47	314,44	319,52	328,78	394,19
Bicycle demand [pax/km]	39,69	0,00	31,66	0,00	0,00
Total Operational Cost of Cars [US$/hr/km]	3309,63	3111,52	2916,30	3035,33	2687,76
Total Operational Cost of Buses [US$/hr/km]	100,82	126,80	124,11	50,34	60,30
Financial Result of Bus system [US$/hr/km]	0,00	-185,02	0,03	119,11	139,87
Congestion pricing revenues [US$/hr/km]	0,00	0,00	649,62	0,00	610,23
Congestion Pricing Fee [US$/km]	0,00	0,00	0,18	0,00	0,18
Bus fare [US$/ride]	0,42	-0,19	0,39	0,52	0,51
Lanes for bus corridor				1,00	1,00
Car speed [km/hr]	12,41	12,95	13,59	10,30	11,44
Bus speed [km/hr]	11,59	12,05	12,61	35,79	34,58
Bicycle speed [km/hr]	12,00	12,00	12,00	12,00	12,00
Bus Frequency [buses/hr]	27,24	35,11	36,26	46,44	52,12
Bus capacity [pax]	104,63	107,48	105,73	84,96	90,76
Bus fleet [buses/km]	2,35	2,91	2,88	1,30	1,51
Car-bus equivalence [cars]	2,05	2,07	2,06	1,85	1,91
Number of stops per Km.	1,90	1,90	1,91	2,38	2,40
Consumer Surplus [US$/hr/km]	8500,93	8707,65	7893,86	8569,46	7982,39
Total Social Welfare [US$/hr/km]	8500,93	8522,63	8543,51	8688,56	8732,49
Results compared to Scenario 1					
Consumer Surplus [US$/hr/km]	0,00	206,73	-607,07	68,53	-518,54
Total Social Welfare [US$/hr/km]	0,00	21,71	42,58	187,63	231,57