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

Leonardo J. Basso, Universidad de Chile*
Cristián Angelo Guevara, Universidad de Los Andes
Antonio Gschwender, Universidad de Chile
Marcelo Fuster, Universidad de Los Andes

1. INTRODUCTION

In the literature, two of the most popular ways to deal with urban congestion that have been suggested are congestion pricing and giving priority to public transportation. Congestion pricing has been analyzed in a very large number of settings but, 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 (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

* Correspondence: lbasso@ing.uchile.cl, Casilla 228-3, Santiago, Chile

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).

Now, as it is evident, 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. 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.

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.[1]

[1] Some papers that do consider some of the features we are interested in are Mohring (1972), Small (1983), Viton (1983).

2. THE MODEL

We consider a road of infinite length with a capacity of Q vehicles/hour, where Y commuters per kilometer and hour would like to travel l kms (in the same direction). We model a representative kilometer of the road. 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]; number of equidistant bus stops per kilometer, p; the congestion toll for cars P_a [$/km]; the bus fare P_b [$/trip]; and the percentage of road capacity dedicated exclusively to bus services, η. Obviously, 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).

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; 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 order of the *B*s 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*. 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. Thus, the number of people using each mode is given by:

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

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

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

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:

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

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

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.

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-η). The time that a car takes to travel one kilometer will be given by:

$$t_a = \alpha\left(\frac{\frac{lY_a}{occ}}{(1-\eta)Q}\right) + \beta \tag{9}$$

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:

$$t_b = \left(\alpha\left(\frac{bf}{\eta Q}\right) + \beta\right) + \frac{Y_b t_{sb}}{f} + t_p p \tag{10}$$

On the right hand side of (10), 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 attempts to capture 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 $b(K)= K/100+1$, where K is the capacity of the bus, which nicely captures the fact that a bus is equivalent to something between 1.5 and 3 cars, depending on its size. In the second term in (10), t_{sb} is the average time that a passenger takes to board and alight the bus, thus this term captures delays for 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. 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:

$$t_a = \left(\alpha \left(\frac{bf + \frac{lY_a}{a}}{Q} \right) + \beta \right) + \left(\frac{\frac{Y_b t_{sb}}{f} + t_p p}{2} \right) \tag{11}$$

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

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

Finally, the cost of the bus system (in dollars per hour) is given by $C_b = (c_{b0} + c_{b1}K) f t_b$ where the term in brackets represents operational cost per bus and hour, which are larger for larger buses.

3. OBJETIVE FUNCTION, SCENARIOS AND PARAMETER VALUES

We 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. The common objective function –per hour social welfare–is:

$$SW = CS + P_b Y_b - C_b + P_a \frac{Y_a}{occ} l \tag{13}$$

Where CS is consumer surplus –easily obtainable as the sum of individual utilities– 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. We can now move to the description of the different scenarios we analyze. In all cases, the planner maximizes (13) 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. Thus, the planner must consider:

$$K = \frac{Y_b l}{f} \quad , \quad Y_a \geq 0 \quad , \quad Y_b \geq 0 \quad , \quad Y_a + Y_b \leq Y \tag{14}$$

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 (9) and (10) 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 (11) and (12) there. The scenarios are as follows:

Scenario 1: 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:

$$\text{Max}\, SW \;\; w.r.t \quad f, p, Y_a, Y_b \quad s.t. \quad P_a = 0 \,,\, P_b Y_b \geq C_b \,,\, eqs. \text{ (9), (10), (15)}$$

<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: the problem is as in scenario (1) but we no longer consider the constraint $P_b Y_b \geq C_b$.

<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. Given the policies at hand, this scenario will lead to the maximal social welfare level for mixed-traffic conditions as now we no longer consider the restriction $P_a = 0$.

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

What we intend to explore here, 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. Thus, the optimization problem now consider a constraint of the form $\frac{Y_a}{occ} P_a l + P_b Y_b \geq C_b$.

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. Given that in these cases there is on 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 (11) and (12).

As it is evident, the optimization model we propose to solve requires a large number of parameters. We have chosen to use parameters that represent reality as close as possible, in this case, they represent a morning-peak in Santiago, Chile, where monetary values correspond to 2006 US dollars. We omit the parameter values for space reasons, but area available upon request.

4. SIMULATION RESULTS

All eight scenarios were solved as optimization problems using the software Wolfram Mathematica; numerical results are omitted for space reasons but are available upon request. Figure 1 below summarizes the results of each scenario in terms of the value achieved of Consumer Surplus and Total Social Welfare, with respecto to the base case (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.

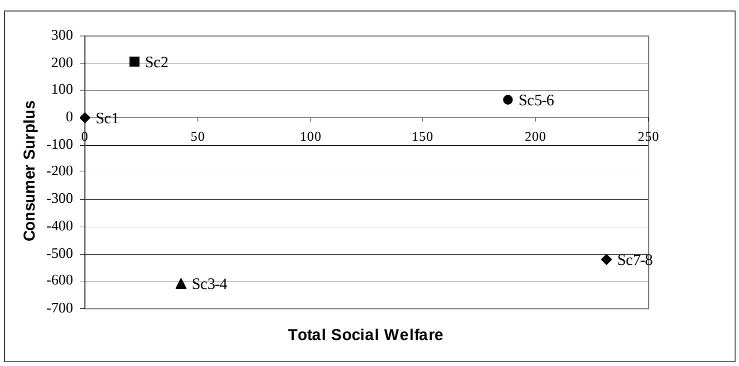

Figure 1: Differences in Consumer Surplus and Total Social Welfare (US$/km/hr) with respect to Scenario 1

Effects of Transit Subsidization

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 do 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.

How subsidization does alone compared to other policies also in isolation? 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. 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.

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.

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. 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. He hypothesized that dedicated bus lanes may be a tool equivalent to congestion pricing in achieving a change in modal split. To study this 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. 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. 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. This paper deals exactly with this issue, by proposing a simple tractable model that incorporate demand and engineering interactions. Numerical results of the model clearly indicate that this type of approach may be fruitful to better understand the full implications of different urban transport policies.

REFERENCES

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

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

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

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