

Efficiency and Substitutability of Transit Subsidies and Other Urban Transport Policies[†]

By LEONARDO J. BASSO AND HUGO E. SILVA*

This paper analyzes the efficiency of and the substitutability between three urban congestion management policies: transit subsidization, car congestion pricing, and dedicated bus lanes. The model features user heterogeneity, cross-congestion effects between cars and transit, intertemporal and total transport demand elasticities, and is simulated using data for London, UK and Santiago, Chile. We find that the substitutability between policies is large and, in particular, the marginal contribution of increased transit subsidies, as other policies are implemented first, diminishes rapidly. Bus lanes are an attractive way to increase frequencies and decrease fares without injecting public funds. (JEL L92, L98, O18, R42, R48)

Subsidies to public transport systems are large in the developed world; for example, they reach around 70 percent of operational cost on average for the largest 20 cities in the United States (Parry and Small 2009). Similar figures are found in other developed nations (see Kenworthy, Laube, and Vivier 2001; Elgar and Kennedy 2005). The reality is quite different in the developing world. In Latin America, for example, subsidies are zero everywhere, with the only exceptions of Buenos Aires (50 percent), Montevideo (10 percent), and São Paulo (5 percent); in Santiago, for the time being, subsidies have also been considered (see Corporación Andina de Fomento 2010 for figures).

Transit subsidies are indeed controversial and their existence is a matter of continuous debate, both in countries that have them as in countries that do not. The reasons that are advanced to support transit subsidization are manifold. First, there is the so-called Mohring effect: increased ridership induces an increase in frequency, which should diminish waiting times of all users (Mohring 1972); he shows that, if one minimizes costs taking into account the resources contributed by operators (fleet and its costs) and users (their time), the resulting cost function displays economies of scale implying the need for subsidies in the first-best (see also Jara-Díaz and Gschwender 2009). Alternatively, one can interpret the

*Basso: Universidad de Chile, Casilla 228-3, Santiago, Chile (e-mail: lbasso@ing.uchile.cl); Silva: VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands and the Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands (e-mail: h.silvamentalva@vu.nl). We are indebted to Vincent van den Berg, Bruno de Borger, Achim Czerny, Antonio Gschwender, Sergio Jara-Díaz, Stef Proost, Luis Rizzi, Kenneth Small, Harry van der Weijde, and two anonymous referees for helpful comments and suggestions that materially improved the paper. Financial support from FONDECYT-Chile, Grant 1130133, and grants Conicyt FB0-16 and Milenio P-05-004-F are gratefully acknowledged.

[†]Go to <http://dx.doi.org/10.1257/pol.6.4.1> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

Mohring effect as a positive externality, which requires a Pigouvian subsidy to be internalized. A second argument to favor of subsidization refers to unpriced negative externalities that car travel generates (e.g., congestion, pollution, and noise); since car travel is a substitute product of transit travel, when these negative externalities are not charged to car drivers, the second-best alternative is to reduce the price of transit, potentially implying the need for subsidies. A third argument is that the operators' costs themselves might display economies of scale (Boyd, Asher, and Wetzler 1978; Allport 1981).¹ Finally, there are equity considerations: usually, transit is used by poorer people and, therefore, subsidizing transit is a mean to achieve income redistribution.

Usual rebuttals against these arguments are that the cross price-elasticity between car usage and transit price is rather low (see Hensher 1998; Winston and Shirley 1998); that the public funds have a non-negligible cost so the funds necessary to cover transit deficits may produce welfare losses (Proost and Van Dender 2008); and that subsidization develops a negative and important effect on cost efficiency generated, among other things, because of inefficient use of labor and capital.² Indeed, an interested reader will find it extremely difficult to infer from the literature which approach is correct since, on one hand there are appealing economic arguments on both sides while, on the other, the empirical literature on optimal pricing of urban transport has delivered very different results. For example, Proost and Van Dender (2008) find that the optimal transit fare in the peak-period in Brussels may be close to zero, while the recent analysis by Parry and Small (2009) for London, Washington, DC, and Los Angeles shows that extending subsidies far beyond two-thirds of operating costs is in most cases (all but one) welfare improving. Winston and Shirley (1998, 59), on the other hand, find that for major US cities an efficient policy would sharply raise all bus fares and substantially cut frequency of service everywhere.

What we provide here is a new assessment of the efficiency and desirability of transit subsidies. For this, we use a transport mode choice model that considers substitution between private and public transport, intertemporal and total transport demand elasticities, and enables consumer surplus calculations in a theoretically sound manner. Our model also considers three main new features: mutual congestion, optimization of the transit design variables, and policy interactions. Regarding the first of these, our network model captures congestion interactions between cars and transit while vehicles are in motion, but also at bus stops. Bus stop congestion—caused both by buses and passengers boarding—can be a heavy problem at peak times, affecting the traffic of all vehicles on the road. Second, we take a somewhat longer-run view than others by letting the transit system design to adapt to new conditions; therefore, vehicle size, frequency and design of bus stops are optimization variables. And third—perhaps our most important contribution—we analyze the performance of alternative measures, such as congestion pricing or dedicates bus

¹ These type of economies—decreasing average cost as traffic density increases over a fixed network and route structure—are sometimes called economies of density in the transportation literature. See, e.g., Basso and Jara-Díaz (2006).

² This latter negative effect has been found empirically by Winston and Shirley (1998), Savage (2004), and documented by De Borger, Kerstens, and Costa (2002) on the basis of a review of several articles.

lanes, using welfare measures for the comparisons. The latter enables us to analyze the marginal benefit of an increase in subsidies, as in Parry and Small (2009), but recognizing that subsidies might operate together with other urban policies; we can, therefore, uncover the complementarity, or lack thereof, between them. Also, the welfare analyses allow us to assess the distributional impacts of the different transport policies in order to forecast support, and to assess which combination of urban transport policies might be best and what would be a reasonable implementation path. In order to analyze the robustness of results to different situations regarding consumer preferences, income, and costs, we simulate our model using data from a large city from the developed world—London, UK—and another large city from the developing world—Santiago, Chile.

Table 1 compares our work with the different papers that we build upon, and helps to highlight the novelties of our model and analyses. The table shows two important differences with previous literature, in addition to what has already been mentioned regarding the cross congestion modeling and the transit design optimization. The first difference is the set of urban policies analyzed. Our paper considers three—transit subsidies, congestion pricing, and bus lanes—and analyze them both in isolation and in different combinations.³ Many other authors do not separate, for example, the pricing of buses from car tolls; they simply consider optimal (social marginal cost) pricing. However, reality shows that in many cities there are transit subsidies but no congestion pricing, which begs the question of how good this is with respect to using both. In addition, it is quite obvious that if total demand was inelastic and subsidies caused no cost to society, the only thing that would matter is the price difference (e.g., Basso and Jara-Díaz 2012). Yet, total demand has an elasticity and, most importantly, public funds do have a marginal cost. When this is the case, congestion pricing not only helps to solve an externality but, by generating revenue, it allows for diminishing distortionary taxes elsewhere, while transit subsidies work in the opposite way. Therefore, the marginal cost of public funds—rarely considered in the previous literature—affects the amount of subsidy that is optimal. Regarding bus lanes, many authors have recognized that the congestion caused by cars on buses is one of the main impediments for transit to diminish its social marginal cost (e.g., De Borger et al. 1996). It is then only natural to wonder what would be the outcome of a bus lane policy and the effects on optimal car tolls and transit subsidies.

The second main difference with earlier literature is the welfare and distributional analyses. Indeed, in many cases authors provide results on how prices or traffic for each mode will change, but do not provide a welfare measure of the performance of the policy (e.g., the percentage change in total welfare) nor the change in consumer surplus for different groups. This complicates or precludes using the results for public policy, as it makes it difficult to compare the different policies or combinations of them. In this paper we enable these comparisons

³Of course, there is a large literature that analyzes each mode in isolation, without explicitly considering substitution. See Small and Verhoef (2007) for a review of congestion pricing and Jara-Díaz (2007) for transit economics. In some of these papers, some of the aspects we consider—such as delays to riders due to loading and unloading—are considered.

TABLE 1—RELATED LITERATURE

Paper	Peak and off-peak periods	Total demand elasticity	Transit design optimization	Cross-congestion effects	Marginal cost of public funds	Transport policies analyzed	Welfare analyses and comparisons	Distributional analyses
Basso and Silva (2014)	✓	✓	✓	✓	✓	Transit subsidies, congestion pricing, bus lanes	✓	✓
Glaister and Lewis (1978)	✓	✓	×	×	×	Transit fares	×	×
Mohring (1979)	×	×	✓	(b)	×	Optimal prices, bus lanes	×	×
Small (1983)	×	×	×	(b)	×	Bus lanes, congestion pricing	✓	✓
Viton (1983)	✓	×	✓	(b)	×	Optimal prices	×	×
De Borger et al. (1996)	✓	✓	×	(b)	×	Pricing (optimal + restrictions)	×	×
Huang (2000)	(a)	×	×	×	×	Pricing	✓	✓
Proost and Van Dender (2008)	✓	✓	×	(b)	✓	Transit subsidies, congestion pricing	✓	×
Kutzbach (2009)	×	✓	×	(b)	×	Congestion pricing versus bus lanes	✓	×
Parry and Small (2009)	✓	✓	×	(b)	×	One public transport fare at the time	×	×
Basso et al. (2011)	×	✓	✓	✓	×	Transit subsidies, congestion pricing, bus lanes	✓	×
Basso and Jara-Díaz (2012)	×	×	✓	×	×	Transit subsidies, congestion pricing	×	×

Notes: In many cases, some features that authors include in their theoretical models are then “simplified away” in the numerical part. This table refers to what is actually included in the numerical analysis of the papers. (a) This is a bottleneck/scheduling model and therefore the length of the congested period is endogenous. (b) These papers consider that cars and buses congest each other while in motion, but do not consider that bus stop operations might affect car speed.

by using an efficiency measure, and we also study the substitutability of policies which sheds light on what a reasonable order of implementation could be. The lack of welfare comparisons in the literature can be tracked back to the demand models used. Table 1 shows that many of them are rather sophisticated, by featuring not only mode substitution but also inter-temporal and total demand elasticities, but they might not allow for a theoretically sound calculation of consumer surplus; we come back to this later with more detail.

Our main results show that the benefit that each stand-alone measure induces is different between applications. In London, congestion pricing and bus lanes increase social welfare significantly and by similar amounts, while optimal subsidization (free buses) achieves much less. In Santiago, on the other hand, bus lanes yield a much higher benefit than congestion pricing and optimal subsidization. However, in both applications there is large efficiency substitutability among these three policies, that is, once one is implemented, adding another does not increase welfare as much. In particular, the marginal contribution of transit subsidies to welfare is large only when none of the other urban policies considered in this paper are in place. In addition, segregating traffic through bus lanes seems to be particularly appealing, as it achieves large welfare improvements without subsidies or cumbersome car congestion tolling, affecting generalized prices through quality of service (speeds) instead of monetary prices. Moreover, the bus lane policy induces

the largest increase in frequency for both cities, and it does so without the use of subsidies, which are often defended as a mean to, indeed, increase frequency. Finally, the conjecture that congestion pricing and subsidies are roughly equivalent does not hold in our model.

For the case of Santiago, where data enables a full distributional analysis, we further find that congestion pricing happens to be a progressive measure if the transit system improves to cover the new demand, even before using revenues from tolls. This is an important point as one of the usual critiques against congestion pricing is that it would be regressive, by pricing off the streets those who benefit least from the speed increase due to low values of time, and therefore, with lower income (see Hau 2005a, b). We further find that dedicated bus lanes are also progressive, and that optimal transit subsidization is a Pareto-improving measure. These results have an important implication for a regulator with distributional concerns. When assigning different welfare weights to different income groups, to account for benefits from vertical equity, we find that the desirability of subsidies and its complementarity with other policies may be rekindled. If the distributional concerns are sufficiently high, optimal subsidization can be the best-stand alone measure and a complement for bus lanes and congestion pricing.

Our results may have important policy implications: they show that there might be ways to reduce transit deficits without affecting welfare or consumer surplus, but this requires careful planning of the order in which policies are implemented, because there is a clear best stand-alone measure for each application and substitutability between policies is high. Certainly, our results have to be qualified according to our assumptions. We do this in what follows.

The paper is structured as follows. Section I describes the analytical model, showing how key aspects of the problem are captured. Section II describes the data and the parameter values we use. Section III contains our main results, including welfare and distributional analyses. Section IV presents the analysis with welfare weights, while Section V concludes and elaborates on the generality of our results.

I. The Model

We model a representative kilometer of the road network of a city where bus service is offered, and we look at one day of operation. Travelers choose whether to travel in one of the two possible periods, peak and off-peak, or not to travel at all; furthermore, if they do travel, they choose between the two modes available in both periods: car and bus.

The planner can optimize the design variables of the transit system: bus frequency in each period q, f^q [bus/hr], bus capacity k [pax/bus], and number of equidistant bus stops per kilometer p . The planner can also set prices: the bus fare for each period P_{qb} [\$/km] and the congestion toll for cars for each period P_{qc} [\$/km]. Finally she can also decide on a traffic management measure, namely, the fraction of the capacity that is exclusively dedicated to bus lanes, n . Note that transit subsidies are not a variable *per se*, but will be obtained as a result of the optimization problem and the constraints considered: an X percent of subsidy is obtained by imposing that the revenue from fares covers $(100 - X)$ percent of the transit cost.

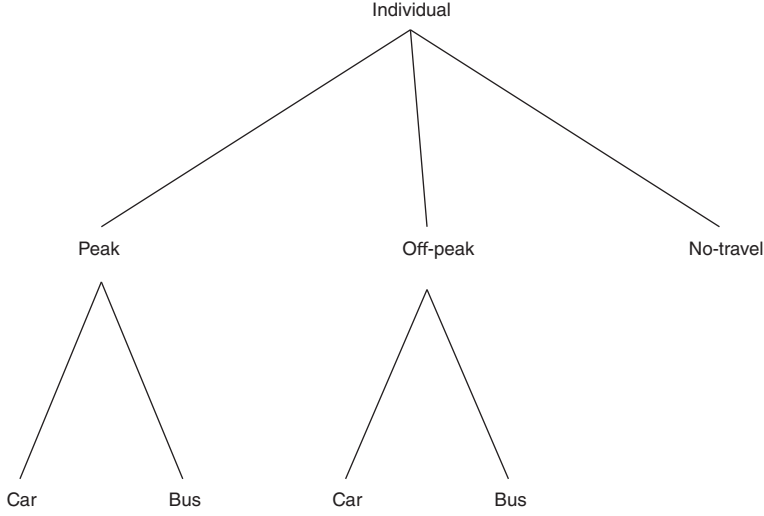


FIGURE 1. DECISION TREE

A. Demand

We deem important that the demand model allows, in addition to substitution between transport modes, for elasticities of substitution between peak and off-peak periods and for total travel demand elasticity. For this purpose, we use the nested Logit model introduced by Ben-Akiva (1973), well-rooted in the random utility theory framework; a detailed discussion on the assumptions underlying this model can be found in Anderson and De Palma (1992). In our model, people choose according to the two-stage sequential process illustrated in Figure 1: an individual from income group $i \in I$ first chooses between traveling during the peak, during the off-peak or not traveling at all, based on the expected benefit of each of the nests (choice of nest, upper level).

Then, conditional on the chosen period, she chooses between car and bus, based on the utility she gets from each alternative (mode choice, lower level). We start by describing this second decision. The utility that an individual from group i obtains if she uses mode m in period q , is given by:

$$(1) \quad U_{qm}^i = \theta_{qm}^i + \lambda^i \cdot \text{cost}_{qm} + \beta_{qm}^i \cdot \text{gt}_{qm},$$

$$q \in \{\text{Peak}, \text{Off-peak}\},$$

$$m \in M_q = \{\text{Car}, \text{Bus}\},$$

where θ_{qm}^i is the alternative-specific constant, cost_{qm} is the monetary cost of the trip, λ^i is the cost parameter, gt_{qm} is the generalized travel time, and β_{qm}^i is the marginal utility of time. As each income group has different marginal utilities of income and time, they most likely have different values of travel time savings.

The monetary cost of a bus trip is simply the bus fare, while for a car trip, it is the congestion toll plus an operational cost (c_{0c}) related to expenses on fuel, lubricants, tires, and so on. Assuming that auto commuters face a cost per kilometer that is evenly distributed among the a passengers of a car, we get, for each period q , $cost_{qb} = P_{qb} \cdot l$ and $cost_{qc} = (P_{qc} + c_{0c}) \cdot l/a$, where l is the average trip length. The generalized travel time is a weighted sum of in-vehicle time, waiting time, and walking time, where weights capture the fact that people perceive these times differently (for example, waiting is always more unpleasant). Because these times depend, through congestion, on all of the optimization variables, we discuss them in detail in the next subsection.

Since in a nested Logit the decision at each level is modeled as multinomial Logit, the proportion of commuters that choose mode m , conditional on the choice of period q , is:

$$(2) \quad \mathcal{P}_{m|q}^i = \frac{\exp(U_{qm}^i)}{\sum_{r \in M_q} \exp(U_{qr}^i)}, \quad q \in \{\text{Peak, Off-peak}\}, \quad m \in M_q = \{\text{Car, Bus}\}.$$

We now move to the upper level. The choice of nest is again modeled with a multinomial logit, but this time considering expected utilities. As shown by Ben-Akiva (1973), the expected utility of a nest is given by the logsum formula:

$$(3) \quad A_q^i = \ln \left(\sum_{r \in M_q} \exp(U_{qr}^i) \right) \quad q \in \{\text{Peak, Off-peak}\}.$$

The expected utility of no-travel ($A_{no-travel}^i$) is set to a constant, whose value affects the elasticity of total demand. Then, the proportion of commuters that choose each nest is therefore given by:

$$(4) \quad \mathcal{P}_n^i = \frac{\exp(\mu \cdot A_n^i)}{\sum_{u \in N} \exp(\mu \cdot A_u^i)}, \quad n \in N = \{\text{Peak, Off-peak, No-travel}\}.$$

In this model, the scale parameter of the upper-level, μ , represents the degree of substitutability between periods. When $\mu = 1$, nesting does not matter and nests (both periods and not traveling) become perfect substitutes. This would be equivalent to having one multinomial logit for the five alternatives. On the other hand, the smaller μ is, the lower the substitutability between nests is, making periods more independent. Thus, a larger μ and a larger value of utility for no-travel, implies a larger intertemporal and total demand elasticity, respectively.

Finally, the number of people choosing alternative nm (mode m in nest n), that results from the nested logit model is:

$$(5) \quad Y_{nm} = \sum_{i=1}^I Y^i \cdot \mathcal{P}_n^i \cdot \mathcal{P}_{m|n}^i, \quad n \in N, \quad m \in M_n,$$

where Y^i is the number of people per kilometer that belong to income group i , and in the no-travel nest there is only alternative that is chosen with probability 1. The resulting demand Y_{nm} is also per kilometer.⁴

B. Transport Times

In this section, we write the generalized travel time of each mode as a function of the optimization variables. In addition, we need to differentiate when cars and buses share the road capacity and when bus lanes are in place.

Dedicated Bus Lanes.—The generalized travel time for a transit user in period q is given by: $gt_{qb} = t_{bus}^q \cdot l + \phi_1 \cdot t_w^q + \phi_2 \cdot t_{acc}$, where t_{bus}^q is in-vehicle travel time per kilometer in period q , t_w^q is time waiting at the bus stop, and t_{acc} is access time, namely walking to and from bus stops; $\phi_1 > 1$ and $\phi_2 > 1$ are weights to be obtained empirically, which capture the fact that people dislike more waiting and walking than being on the bus. Describing the last two components of generalized travel time in terms of optimization variables is simple: waiting time (t_w^q) is a fraction ϑ of the interval between buses ($t_w^q = \vartheta/f^q$); since commuters are uniformly distributed along the corridor, the average walking distance to access the bus stop is one fourth of the distance between them ($1/4p$) and the average walking distance from the bus stop to the destination is the same. If the walking speed is V_w , then average walking time is $t_{acc} = 1/(2 \cdot p \cdot V_w)$.

The time that a bus takes to travel one kilometer, when dedicated bus lanes are in place, is given by:

$$(6) \quad t_{bus}^q = t_f \cdot \left(1 + \alpha \cdot \left(\frac{f^q \cdot b(k)}{n \cdot C} \right)^\beta \right) + p \cdot \left(\frac{Y_{qb}}{H^q \cdot f^q \cdot p} \cdot t_{sb} + t_d \right).$$

The first term on the right-hand side corresponds to the time that a bus spends while in motion. The free-flow travel time is t_f , α is a parameter related to speed reductions caused by congestion, f^q is bus frequency in period q , $b(k)$ is an equivalence factor between buses and cars which increases with bus size, C is the capacity of the road (in cars/hour), n is the fraction of capacity dedicated only to buses, and β is a parameter. This function, with parameter values $\alpha = 0.15$ and $\beta = 4$, is known as the BPR function (see Small and Verhoef 2007 for discussion) and is commonly used in transportation analyses to model congestion.

The second term on the right-hand side of (6) is the time spent at bus stops, and it is given by the number of stops that a bus makes in each kilometer (p) multiplied by the time spent at each stop. The time each passenger takes to board a bus is t_{sb} , and the number of passengers boarding a bus at each stop is $Y_{qb}/(H^q \cdot f^q \cdot p)$, i.e.,

⁴Some readers might find our decision tree too restrictive in that, perhaps, some people may choose first the transport mode and then the period. What we want to stress is that, in the end, what the model delivers are the proportions of people choosing a given mode in a given period (see equation (5)). McFadden (1978) shows that the resulting proportions in a nested logit model can also be obtained from a multinomial discrete choice model that is consistent with (nonsequential) individual utility maximization. Thus, our sequential model can also be interpreted as a simultaneous discrete choice model.

the period bus demand per km divided by the number of hours of the period, H^q , the bus frequency in the period, and the number of bus stops per km.⁵ Finally, t_d is a nonlinear function representing bus congestion at the bus stop, that is, buses queuing to get in and out of the bus stop. It depends on several variables including bus frequency, bus stop capacity, and number of passengers boarding, and therefore it is not the same in both periods. The specific functional forms behind bus stop congestion were obtained from microsimulation exercises by Fernandez, Valenzuela, and Gálvez (2002), but are not particularly informative so we omit them here; see the online Appendix for details.

Next, the generalized travel time for a car driver is given only by the in-vehicle travel time, since she does not need to wait or walk; thus $gt_{qc} = t_{car}^q \cdot l$. Further, when dedicated bus lanes are in place, cars do not interact with buses or bus stop operations, so car travel time (per kilometer) is:

$$(7) \quad t_{car}^q = t_f \cdot \left(1 + \alpha \cdot \left(\frac{l \cdot Y_{qc} / (H^q \cdot a)}{(1 - n) \cdot C} \right)^\beta \right),$$

where $l \cdot Y_{qc} / (H^q \cdot a)$ is the car flow, since Y_{qc} is car demand per period per kilometer, H^q is the period duration in hours, l is the trip length, a is the (constant) car occupancy and $(1 - n) \cdot C$ the capacity they have available.

Mixed Traffic Conditions.—Travel times in mixed traffic conditions are more difficult to describe because, in addition to the obvious interaction that happens when vehicles are in motion, bus stop operations cause delay on cars. As far as we are aware, the transportation engineering literature has not yet delivered a treatment that allows us to capture this interaction in terms of all the variables we would like (traffic mix, distance between bus stops, bus size, and so on). Yet, since this interaction does occur, what we do is to add to the car travel time a fraction (ϵ) of the time a bus needs for bus stop operations. We make this fraction a function of frequency, such that when f^q goes to zero ϵ is zero, and grows until reaching one when f^q is very large. This is, we believe, a sensible assumption as cars would not suffer bus stop effects when there are no buses, but will have to essentially behave like a bus when the mix of traffic is heavily tilted towards buses; ϵ will usually be between zero and one because overtaking is possible.⁶ Considering that buses and cars now share the capacity, the new travel times, for each period q , are given by:

$$(8) \quad t_{bus}^q = t_f \cdot \left(1 + \alpha \cdot \left(\frac{l \cdot Y_{qc} / (H^q \cdot a) + f^q \cdot b(k)}{C} \right)^\beta \right) + p \cdot \left(\frac{Y_{qb}}{H^q \cdot f^q \cdot p} \cdot t_{sb} + t_d \right)$$

$$(9) \quad t_{car}^q = t_f \cdot \left(1 + \alpha \cdot \left(\frac{l \cdot Y_{qc} / (H^q \cdot a) + f^q \cdot b(k)}{C} \right)^\beta \right) + \epsilon \cdot p \cdot \left(\frac{Y_{qb}}{H^q \cdot f^q \cdot p} \cdot t_{sb} + t_d \right).$$

⁵ Here we assume that per-passenger boarding takes more time than per-passenger alighting.

⁶ As explained in Table 1, this effect has usually not been incorporated in the modeling. In the model proposed by Basso et al. (2011) this fraction is constant and equal to 0.5. Here, however we (arbitrarily) use $\epsilon(f^q) = 1 - 1.01^{-f^q}$, which delivers the right intuition at least. For example, for a frequency of 60 buses per hour this fraction is 0.45 and for a frequency of 30 buses per hour the value of ϵ is 0.26.

C. Bus Operating Costs

We model the operating costs of the bus system (G , in \$/day) as a function of the bus fleet (B), the total number of vehicle-kilometers of each period (V_q), and the bus size (k) in the following way:

$$(10) \quad G = G_b(k) \cdot B + \sum_q G_v(k) \cdot V_q.$$

The first-term on the right hand side is mainly labor and vehicle-capital expenses while the second captures operational expenses. Both types of expenses are a function of the vehicle size through the (linear) functions G_b and G_v , that give the cost per bus per day and cost per vehicle-km respectively.

What we need next is to express both the fleet and the number of vehicle-kilometers in terms of optimization variables. The required fleet is $B = \max_q \{f^q \cdot t_{bus}^q\} \cdot L$, where L is the total distance that a bus covers before starting a new cycle; t_{bus}^q is travel time for one kilometer in period q , and is given either by (6) or (8), depending on whether dedicated bus lanes are in place or not. This reflects that there is one period (usually the peak) that defines the amount of buses required for operation, while there will be idle capacity (spare buses) in the other period. The daily vehicles-kilometers are the sum of the vehicles-kilometers of both periods: $V_{peak} = H^p \cdot f^p \cdot L$ and $V_{off-peak} = H^o \cdot f^o \cdot L$. Since all previous demand functions were expressed in per-day and per-kilometer terms, we also need to obtain transit costs for a day of operation and one kilometer of the network. This is simple, we only need to divide G in equation (10) by L , which makes L disappear as variable so we do not need to estimate it. To simplify notation, we denote the operating cost per day per kilometer as OC_b .

D. Optimization Problem

The objective function we consider is unweighted social welfare, for one kilometer of a day of operation. It includes consumer surplus (CS), the financial result of the bus system, congestion pricing revenues, and implementation costs of any policy in place. Consumer surplus in the nested-logit model is obtained through what is known as the logsum formula:⁷

$$(11) \quad CS = \sum_i \left(\frac{1}{\mu} \cdot \frac{Y^i}{(-\lambda^i)} \cdot \ln \left[\sum_n \exp(\mu \cdot A_n^i) \right] \right).$$

⁷This measure of consumer surplus deserves some discussion. Small and Rosen (1981) established that for discrete choice models, the change in the area to the left of the compensated demand as it shifts in response to changes in attributes (quality) or price is a valid measure of the compensating variation. If the marginal utility of income is approximately constant (here given by $-\lambda^i$), then compensated and Marshallian demands coincide and, therefore, one can use Marshallian demand to calculate an exact measure of consumer surplus. Anderson, De Palma, and Thissse (1992) extended the result to the nested logit formulation. Note that, in order to be able to correctly assume that marginal utilities of income are the same for a group of people, an adequate segmentation of the population has to be performed. For the case of Santiago, the population was segmented in five groups, each of which has its own demand model.

Note that, since consumer surplus is an unweighted sum of each individual commuter surplus, the Marshallian measure will value more the time savings of those with higher willingness to pay, which is related to higher income levels through smaller marginal utilities of income. Therefore, the measure could be considered as regressive. An alternative is to assign different weights on individual consumer surplus according to income, something that we explore in Section IV.

Social welfare is completed by adding the revenue from bus fare and car tolls, and by subtracting the cost of the bus system (OC_b) and the implementation costs of congestion pricing and bus lanes (if they are in place). We assume the costs of congestion pricing to be a fraction η of the revenues, that the costs of implementing dedicated bus lanes is fixed (per unit of distance), and that public funds are costly ($mcpf > 1$),⁸ so the final expression for social welfare is:

$$(12) \quad SW = CS + \left[\left(\sum_t Y_{tb} \cdot P_{tb} \cdot l \right) - OC_b + \left(\sum_t Y_{tc} \cdot P_{tc} \cdot l/a \right) \cdot (1 - \eta) - OC_{dl} \right] \cdot mcpf,$$

where OC_{dl} is the additional cost of implementing and operating dedicated bus lanes with respect to the case of mixed traffic (i.e., it does not include road construction), and it is included in the welfare function only when bus lanes are being considered. Note that we do not consider the possibility that subsidies induce cost inefficiencies on the transit system, because the extent of these inefficiencies, if they exist, depend on the contract between the transit operator and the regulator (see e.g., Gagnépain and Ivaldi 2002); the $mcpf$ larger than one, though, works as a proxy for cost inefficiencies induced by subsidies.

In order to compare benefits and service levels of the different transport policies, we build scenarios defined as the maximization of social welfare subject to different constraints, which model the application or absence of specific transport policies. For instance, one scenario may include dedicated bus lanes but no congestion pricing; then the travel times to be used are (6) and (7), while we need to impose that the car toll is zero. The constraints to be used in each scenario are made explicit below, but we first discuss the ones that are common to all scenarios. First, note that demand depends on travel times but travel times depend on demand through congestion effects (equations (6)–(9)), so there is a fixed-point problem. What we do is to include demands as optimization variables and impose the equilibrium (fixed-point equation) as a constraint, i.e., that the number of people that choose each alternative has to be consistent with the mode split equilibrium. These constraints are:

$$(13) \quad 0 \leq Y_{nm}^i \leq Y^i \quad \forall i \in I, \forall n \in N, m \in M_n,$$

$$Y_{nm} = \sum_{i=1}^I Y^i \cdot \mathcal{P}_n^i \cdot \mathcal{P}_{m|n}^i \quad \forall n \in N, m \in M_n.$$

⁸The marginal cost of public funds ($mcpf$) measures the cost of each unit of public funds, taking account of the deadweight loss from the additional taxes associated with those funds (Auerbach and Hines Jr. 2002).

Frequency is constrained to be positive and less than the capacity of bus stops for each period.⁹ In addition, bus fare and car toll are restricted to be non-negative in every period. With respect to dedicated bus lanes, since in our application road capacity is three lanes, the number of bus lanes can be one or two and therefore the fraction of capacity dedicated to buses can be one or two thirds. Finally bus size must be equal or larger than the passenger load but, since in this model having idle capacity only decreases the value of the objective function, this constraint will always bind.¹⁰ Hence, bus size is:

$$(14) \quad k = \max_q \frac{Y_{qb} \cdot l}{f^q \cdot H^q}.$$

Thus, we can now fully describe each of the scenarios we study, which are identified by a short name in capital letters.

Reference Scenario (REF).—In our reference scenario there is no congestion toll, the bus system is self-financed, and the road is shared by buses and cars. We further impose that the bus fare is the same for both periods so that we can later study the benefits of optimally differentiating peak and off-peak bus fares. In fact, the only *policy* we do consider is the optimization of frequency, bus size, and bus stop spacing which take place in all scenarios. We think this is a useful reference point to then obtain the incremental benefits of implementing transport policies. The problem that has to be solved, together with the constraints discussed above, is:

$$\begin{aligned} & \max_{f^P, f^O, P, P_{pb}, P_{ob}, Y_{pb}, Y_{pc}, Y_{ob}, Y_{oc}, Y_{nt}} SW \\ \text{s.t. } & \sum_q Y_{qb} \cdot P_{qb} \cdot l = OC_b; \quad P_{pb} = P_{ob}; \quad P_{qc} = 0 \forall q. \end{aligned}$$

Bus Fare Differentiation (CROSS).—This scenario allows for bus fares that are differentiated by time of the day (peak and off-peak), and, as a consequence, there is potential for cross-subsidization between periods. The optimization problem is:

$$\begin{aligned} & \max_{f^P, f^O, P, P_{pb}, P_{ob}, Y_{pb}, Y_{pc}, Y_{ob}, Y_{oc}, Y_{nt}} SW \\ \text{s.t. } & \sum_q Y_{qb} \cdot P_{qb} \cdot l = OC_b; \quad P_{qc} = 0 \forall q. \end{aligned}$$

Transit Subsidization (SUBX).—This scenario allows for an X percent of subsidization of the bus system, which translates in a change of the budget constraint.

⁹The capacity of a bus stop is a function of boarding and alighting times, which in turn depend on demand, frequencies, and many of the optimization variables. Details can be found in the online Appendix.

¹⁰Having idle capacity might be optimal if, for example, commuters dislike crowding. The demand models we have access to, however, do not incorporate that feature.

We maintain the constraint of charging the same bus fare during the day, therefore the problem is:

$$\begin{aligned} & \max_{f^p, f^o, p, P_{pb}, P_{ob}, Y_{pb}, Y_{pc}, Y_{ob}, Y_{oc}, Y_{nt}} SW \\ \text{s.t. } & \sum_q Y_{qb} \cdot P_{qb} \cdot l = OC_b \cdot (100 - X)/100; \quad P_{pb} = P_{ob}; \quad P_{qc} = 0 \forall q. \end{aligned}$$

Car Congestion Pricing (CON).—The implementation of a congestion pricing policy in isolation is represented by this scenario. It is as in the reference case (REF), but without imposing that the car toll is zero.

$$\begin{aligned} & \max_{f^p, f^o, p, P_{pb}, P_{ob}, P_{pc}, P_{oc}, Y_{pb}, Y_{pc}, Y_{ob}, Y_{oc}, Y_{nt}} SW \\ \text{s.t. } & \sum_q Y_{qb} \cdot P_{qb} \cdot l = OC_b; \quad P_{pb} = P_{ob}. \end{aligned}$$

Transit Subsidization and Car Congestion Pricing (SUBX + CON).—Now, we apply both pricing policies together by considering a X percent of subsidization as explained before, together with congestion pricing:

$$\begin{aligned} & \max_{f^p, f^o, p, P_{pb}, P_{ob}, P_{pc}, P_{oc}, Y_{pb}, Y_{pc}, Y_{ob}, Y_{oc}, Y_{nt}} SW \\ \text{s.t. } & \sum_q Y_{qb} \cdot P_{qb} \cdot l = OC_b \cdot (100 - X)/100; \quad P_{pb} = P_{ob}. \end{aligned}$$

Dedicated Bus Lanes (DL).—This scenario analyzes the dedicated bus lanes policy. It is very similar to the reference scenario: the bus system is self-financed and there is no congestion pricing. The difference, however, is that road capacity is no longer shared by cars and buses and, therefore, it is no longer the case that cars congest buses or that bus operations disturb cars flow. For this scenario we use the travel time functions for dedicated bus lanes (equations (6) and (7)). Additionally, a new optimization variable is incorporated: the fraction of capacity that is dedicated to buses (n).¹¹ This scenario sets the problem to:

$$\begin{aligned} & \max_{f^p, f^o, p, P_{pb}, P_{ob}, Y_{pb}, Y_{pc}, Y_{ob}, Y_{oc}, Y_{nt}, n} SW \\ \text{s.t. } & \sum_q Y_{qb} \cdot P_{qb} \cdot l = OC_b; \quad P_{pb} = P_{ob}; \quad P_{qc} = 0 \forall q; \quad n = 1/3 \vee n = 2/3. \end{aligned}$$

¹¹ Note that including one more optimization variable does not imply that the result will necessarily be better because the objective function has changed now that travel time functions are different.

The rest of the scenarios are almost identical to the ones in mixed traffic conditions, with the only difference that dedicated bus lanes are used in addition. Optimization problems are omitted but the name of these scenarios together with their abbreviation are:

- Transit subsidization and dedicated bus lanes (SUBX + DL)
- Car congestion pricing and dedicated bus lanes (CON + DL)
- Transit subsidization, car congestion pricing, and bus lanes (SUBX + CON + DL)

This set of scenarios enable a quite complete analysis and comparison of different transport policies. For example, the social benefit of implementing dedicated bus lanes can be obtained by subtracting the (optimal) objective function value from scenarios DL and REF. Or, the marginal benefit of increasing subsidies when other policies are already in place can be obtained by subtracting optimal objective function values in slightly different values of X .

E. Relation to Other Studies

At this point, we think it is useful to compare our modeling strategy with what has been done previously. The demand model we propose—a sequential discrete choice model—allows for all relevant substitutions and elasticities, and, most importantly, enables the calculation of consumer surplus. As argued in the introduction, this is highly relevant as it allows for studying the efficiency of each urban policy. Earlier papers, while considering demand models that are also very flexible, did not provide welfare measures for the policies they analyzed. The reason might lie, precisely, in the demand models they use. Take, for example, Glaister and Lewis (1978) or De Borger et al. (1996). In both cases the authors use a system of demands where the demand for a mode in a given period depends in log-linear fashion on all generalized prices (all modes and periods); in other words, all demand elasticities are constant. These demand models, while widely used in applied economics, present the problem that the Slutsky symmetry conditions, necessary for integrability of the demand functions into an expenditure function, might hold locally at a single reference point but will generally not hold globally, unless expenditures are proportional between commodities (travel in a specific mode and period), a very restrictive situation (e.g., LaFrance 1986). Hence, in general, the calculation of consumer surplus measures is not possible with these demand systems. In the case of Parry and Small (2009), they assume that travel demands have constant elasticities with respect to own generalized price, while they adjust to other prices according to constant modal diversion ratios. So, for example, suppose that the price of peak bus diminishes and that, as a consequence, the peak bus demand increases in X commuters, according to its constant price elasticity. Then peak-car and peak-rail demands will diminish by an amount that is a fixed percentage of X ; 50 percent and 30 percent respectively. These modeling assumptions have two consequences: first, the authors cannot (and actually do not) modify more than one price (mode and period specific) at the same time and, therefore, simultaneous pricing of rail and bus, or even of peak and off

TABLE 2—MAIN PARAMETER VALUES

	London	Santiago
<i>General parameters</i>		
Reference peak passenger load [pax/hr]	7,000	10,000
Trip length [km]	9.6	10.0
Passenger car equivalence factor for buses $b(k = 80)$	2.06	2.06
Car operating cost [\$ / km]	1.187	0.357
Car occupancy [pax / car]	1.41	1.50
Road capacity [veq / hr]	3,600	3,600
Peak duration [hours]	6	7
Off-peak duration [hours]	14	13
<i>Demand parameters</i>		
Nested logit scale parameter (μ)	0.25	0.25
Value of peak car travel time [\$ / hr]	11.56	0.74
Value of peak bus travel time [\$ / hr]	13.88	1.31
Value of off-peak car travel time [\$ / hr]	9.25	0.35
Value of off-peak bus travel time [\$ / hr]	11.10	0.86
Waiting time weight in generalized travel time (ϕ_1)	2.00	1.93
Walking time weight in generalized travel time (ϕ_2)	2.50	3.63
<i>Transport time parameters</i>		
Free flow speed [km / hr]	60	60
Boarding time [seconds / pax]	2.5	2.5
Constant time that a bus spends in every bus stop [seconds]	15	15
Walking speed [km / hr]	3.6	3.6
Fraction of bus interval for waiting time, ϑ	0.5	0.5
<i>Bus parameters</i>		
$G_h(k = 80)$ [\$ / vehicles-day]	1,155	859
$G_v(k = 80)$ [\$ / vehicles-km]	2.40	1.78
<i>Welfare function parameters</i>		
Marginal cost of public funds	1.15	1.15
Congestion pricing operational cost [share of revenues]	0.45	0.45
Dedicated bus lanes operational cost [\$ / km]	73.71	22.11

Notes: We show the passenger car equivalence factor for buses and the cost parameters for bus operations evaluated in a reference capacity instead of giving the parameters for the slope and intercept of both linear functions because it gives a better picture of costs in our model. The values of time presented for Santiago are a weighted average over each income group. Monetary values are presented in 2009 US dollars.

peak bus, is not possible. Second, the authors cannot integrate the demand functions to obtain a measure of consumer surplus.

II. Parameter Values

We solve each optimization problem numerically, using data from two different metropolitan areas: London and Santiago. Both have highly congested peak-hours but, not surprisingly, Santiago has a larger market share for transit than London. Tables in the online Appendix depict observed situations. The main parameters we use are presented in Table 2 and a brief explanation of how they were obtained follows; a detailed discussion is available in the online Appendix. The main data sources for the London application are publications from Transport for London (TfL) and parameters used by Parry and Small (2009); for the case of Santiago we use data from studies carried out by Secretaría Interministerial de Planificación de Transporte (SECTRA) (dependent of the Planning Ministry of Chile).

A. General Parameters

We choose a road capacity that corresponds to a three-lanes setting, and set the peak and off-peak period duration based on the daily motorized trip patterns. The total demand is set in order to produce a congested condition in the peak period and in absence of policies (i.e., in the REF scenario). The resulting peak passenger load is approximately 7,000 pax/hr in London and 10,000 pax/hr in Santiago, which were chosen by looking at average traffic conditions in peak periods together with the average trip length.

The equivalence factor between buses and cars, which enters travel time equations (6), (8), and (9), is a function of bus size estimated with a linear regression using values that are common in project appraisal in the United Kingdom and Chile. We obtain figures such as 1.6 and 2.5 cars for buses ranging from 40 to 120 passengers respectively.¹²

Car operating costs are quite different across cities, as expected. For London, the parameter is set at almost a dollar per kilometer, while for Santiago it is about one third of a dollar.

B. Demand

The parameters required for the logit models are marginal utilities of income (the cost parameter) and time, and the modal constants. Marginal utilities can be obtained directly from estimated modal choice logit models or derived from observed elasticities and values of time; modal constants, however, must be calibrated for this particular case.

In the case of Santiago, data enables us to use a demand model that includes heterogeneity. SECTRA (2005) provides an estimation of several logit models, with five income groups, for trips in Santiago based on revealed preferences. Details on these groups are available in Section IV and in the online Appendix, but two main features are that they do not encompass the same number of people (they are not quintiles), and they differ significantly in values of time: the ratio between the highest and the lowest is 4 for peak travel and 1.6 for off-peak. The car is heavily used only by the two groups with higher income while the two groups with lower income have a large use of transit and negligible car trips (see online Appendix Table A.1). We use the main demand parameters from SECTRA (2005), including marginal utilities of income for each income group. Having income groups for Santiago allows us to perform a detailed distributional analysis, by assessing the consumer surplus variation of each group.

Data and models for London are not as detailed, so we are only able to consider one nested logit model. We calculate the marginal utility of car-peak time from the elasticity of car-peak demand with respect to car-peak travel time. The data needed is the observed car share of trips (Transport for London 2007), the travel

¹²The equivalence factors used in cost benefit analysis in the United Kingdom go from 2.0 to 2.5 (Department for Transport 2004); in Santiago they range from 1.6 to 3 (Ministerio de Transportes y Telecomunicaciones 2004).

time elasticity for London (Litman 2012)¹³ and the travel times from average traffic speeds (Transport for London 2007). The marginal utility of income is calculated with the marginal utility of travel time and the values of travel time savings, which were obtained from the transport analysis guidance published by the UK Department for Transport (UK DfT 2009).

The scale parameter (μ) and the expected utility of no-travel ($A_{no-travel}$) for London are set such that:

- our implied elasticity of total travel demand with respect to the peak bus fare in the reference scenario is similar to the one implied by Parry and Small (2009) in their reference scenario (-0.002);
- our implied elasticity of bus peak demand with respect to bus peak fare, -0.25 , is similar to the one reported by Litman (2012) and the one used by Parry and Small (2009), -0.24 and -0.4 respectively. For the case of Santiago, we transfer the values directly, i.e., we use the same scale parameter and set $A_{no-travel}^i$ for each income group such that the percentage of people not traveling implied by the observed data is the same as in London. Finally, the modal constants of the nested logit models are calibrated simply by imposing that the observed modal share (adjusted for a two mode system in each period) is equal to the one that the model predicts with the observed values of each attribute.

The values of time presented in Table 2 show the difference between each city: travelers in London are willing to pay significantly more money in order to save one minute of travel time than the ones in Santiago.¹⁴ This is a feature that allows us to study how the implications of each policy (e.g., social benefit or consumer surplus) change under different conditions, specifically different income, values of time, and taste.

C. Transport Times

For the BPR-type congestion function (equations (6)–(9)), we assume that the speed at capacity is reduced to one third, that the free flow speed is 60 km/hr, and that $\beta = 4$.¹⁵ The parameters for the functions describing time at bus stops are obtained from a microsimulation study (Fernández, Valenzuela, and Gálvez 2002) and from empirical surveys (e.g., Transportation Research Board 1985). The time a passenger takes to board t_{sb} is set to 2.5 seconds, which is consistent with a system of contactless cards.

¹³Litman summarizes several studies about travel elasticity with respect to travel time. We use -0.41 based on a study made specifically for European cities (Table 8), but similar values are found for other cities.

¹⁴The difference, almost by a factor of ten, is explained by the difference in GDP per capita but also by factors specific to developing countries. Shires and de Jong (2009) perform a meta-analysis of values of time for different countries (including the United Kingdom and Chile) finding a large degree of variation and that, for commuting, only 30 percent is explained by variation in GDP per capita.

¹⁵Verhoef and Small (2004) and Kutzbach (2009) use $\beta = 4$; Parry and Small (2009) use 3.7.

D. Bus Operating Costs

The functions $G_b(k)$ and $G_v(k)$ which multiply bus fleet and vehicle-kilometers in equation (10) are assumed to be linear. In the case of Santiago, the functions are estimated with a linear regression over some cost studies made for four different firms with different bus size varying from 40 passengers to 160 passengers. In the case of London, available data (e.g., Parry and Small 2009) does not allow for using a cost function that depends on the fleet, the vehicles-kilometers, and the bus size at the same time. Because of this, we use the cost parameters estimated for Santiago multiplied by a factor that makes the bus fare in the reference scenario similar to the observed fare.

E. Welfare Function Parameters

Regarding the marginal cost of public funds, Parry and Small (2009) state that a typical estimate is 1.15, which is what we use. This value is consistent with actual estimates: Harrison, Rutherford, and Tarr (2002) find a *mcpf* for Chile that is between 1.08 and 1.18 depending on the tax considered. Ballard, Shoven, and Whalley (1985) estimate a range of 1.17 to 1.33 for the United States, while Auriol and Warlters (2012) find an average *mcpf* for 38 African countries of 1.2.

The share of congestion pricing revenues that is spent operating the system is set to the average of the reported values by Transport for London for the period 2004–2008. We believe this is a sensible assumption because, despite the changes during the mentioned period, the share has been fairly constant (between 0.42 and 0.49). The cost of operating dedicated bus lanes is estimated by Tirachini, Hensher, and Jara-Díaz (2010) for Australia, and it includes the operation and maintenance of track, right-of-way, signaling, communications, and so on. We use that value for London and a 30 percent of that cost for Santiago, to account for differences in cost in developing countries (e.g., labor).

III. Main Results

Before moving on to the simulation results we think it is useful to describe first the economic forces that are at play, and then how the policies tackle them. Consider as a starting point the REF scenario—where no price measures are being used—and consider for a moment only car travel. Then it would be welfare improving to move people from the peak to the off-peak: this would reduce heavy peak congestion while only marginally increasing off-peak times. If one considers only bus travel, the same holds true: moving commuters from the peak to the off-peak decreases a negative externality (congestion, particularly at bus stops) while increasing a positive one (the so called Mohring positive effect of increased ridership on frequencies). If, on the other hand, we consider both modes but only in the peak, then moving people from cars to buses reduces congestion externalities—which are stronger for a person in a car than in a bus—and therefore increases welfare. Looking at the two modes but in the off-peak only, it is also welfare improving to move people from car to buses: it decreases a negative (although probably not

large) congestion externality while inducing exploitation of the positive externality of ridership on frequencies.

Now, because in mixed traffic conditions bus travel time will be always larger than car travel time, the main objective of the two pricing policies we consider (congestion pricing and transit subsidies) is to achieve a substantial difference between the amount that a car user pays for making the trip and the bus fare. As explained before though, these two will actually not be equivalent in welfare terms. First, congestion pricing increases the generalized cost of traveling in the peak and thus it pushes commuters off the peak (or the road altogether), while a subsidization policy makes the peak, *ceteris paribus*, more attractive. Second, public funds have a shadow price and, therefore, subsidization is costly while congestion pricing creates revenue, diminishing dead-weight losses from taxation elsewhere in the economy. A final point is that there is no limit on how much the congestion fare might increase; yet the bus fare cannot (realistically) decrease below zero.

The way dedicated bus lanes work is different. It eliminates the negative congestion externality that cars exert on buses, thus generating differences in full-prices by improving bus speed (and perhaps diminishing car speed). This change in speeds will attract commuters to the transit system—particularly during the peak—a change that, as discussed, should be welfare improving.

A. General Results

We use the model and data described to solve each optimization problem and with the optimal results of each scenario, we compute variables describing service levels, financial results, and the surplus of each agent; main results are summarized in Table 3 and the following figures.

A first glance at Table 3 shows that all variables and results end up taking reasonable values. For example, bus stops are (optimally) spaced at around 330 to 430 meters and bus size reaches a maximum value consistent with an 18m long bus (i.e., articulated). We think this speaks well of the simulation exercises and the model, because we did not impose variables to be constrained to reasonable values. The distance between bus stops varies less than 8 percent across scenarios, showing that the actual value is more associated with the weight of walking time in the generalized time function than the policy implemented. We also find that the optimal number of lanes exclusively dedicated to buses is always one.

Frequencies and bus size: Optimal peak frequency in the REF scenario is 24 buses/hour in London and 32 buses/hr in Santiago. It halves in the off-peak in Santiago but decreases by a 25 percent in London. After a stand-alone policy is implemented, peak frequencies increase roughly by 70 percent to 130 percent in London and by 15 percent to 30 percent in Santiago; the increase is smaller in the off-peak for both cities, showing that the policies will increase the frequency gap between times of day. In all cases (cities and periods), the bus lane policy is the one that induces the largest increase in frequency and it does this without the use of subsidies, which are often defended as a mean to, indeed, increase frequency. Overall, implementing any

TABLE 3—MAIN RESULTS

	London					
	REF	CROSS	SUB100	DL	CON	CON+DL
Social benefit [\$ /day-km]	0	634	3,641	7,907	9,146	9,170
CS change [\$ /day-km]	0	634	7,404	7,977	2,998	4,905
Bus fare peak [\$ /km]	0.134	0.060	0.000	0.098	0.097	0.098
Bus fare off-peak [\$ /km]	0.134	0.263	0.000	0.098	0.097	0.098
Car toll peak [\$ /km]	0.000	0.000	0.000	0.000	0.670	0.509
Car toll off-peak [\$ /km]	0.000	0.000	0.000	0.000	0.000	0.000
Bus frequency peak [bus/hr]	24.2	25.0	42.0	57.5	47.3	59.4
Bus frequency off-peak [bus/hr]	18.2	17.5	22.3	28.4	19.1	29.7
Bus size [pax]	58	66	49	62	75	67
Car speed peak [km/hr]	15.1	15.8	17.6	15.1	31.3	23.8
Bus speed peak [km/hr]	12.7	13.1	14.8	27.7	22.7	26.8
Car speed off-peak [km/hr]	46.2	45.7	46.9	39.8	46.7	39.1
Bus speed off-peak [km/hr]	30.2	30.5	30.7	33.1	30.7	33.0
Peak share [percent]	42.77	43.64	44.29	45.46	43.53	44.51
Off-peak share [percent]	48.39	47.56	47.26	46.11	47.80	46.91
No-travel share [percent]	8.84	8.80	8.45	8.42	8.68	8.58
Car modal share peak [percent]	79.47	76.53	70.90	51.34	48.97	44.41
Bus modal share peak [percent]	20.53	23.47	29.10	48.66	51.03	55.59
Car modal share off-peak [percent]	80.50	84.70	74.54	75.93	79.03	75.69
Bus modal share off-peak [percent]	19.50	15.30	25.46	24.07	20.97	24.31
Bus stops per km	2.35	2.36	2.39	2.45	2.29	2.48
Dedicated bus lanes	0	0	0	1	0	1
	Santiago					
	REF	CROSS	SUB55	DL	CON	CON+DL
Social benefit [\$ /day-km]	0	4	359	1,110	482	1,110
CS change [\$ /day-km]	0	4	2,834	1,135	−182	1,099
Bus fare peak [\$ /km]	0.053	0.051	0.022	0.047	0.050	0.047
Bus fare off-peak [\$ /km]	0.053	0.057	0.022	0.047	0.050	0.047
Car toll peak [\$ /km]	0.000	0.000	0.000	0.000	0.063	0.003
Car toll off-peak [\$ /km]	0.000	0.000	0.000	0.000	0.000	0.000
Bus frequency peak [bus/hr]	32.3	32.7	37.1	42.1	38.5	42.2
Bus frequency off-peak [bus/hr]	13.7	13.3	15.4	15.8	14.2	15.7
Bus size [pax]	173	173	171	153	170	153
Car speed peak [km/hr]	22.5	22.7	26.5	17.1	28.8	17.5
Bus speed peak [km/hr]	14.8	14.9	16.5	20.7	17.5	20.7
Car speed off-peak [km/hr]	47.5	47.6	47.3	51.0	47.3	51.0
Bus speed off-peak [km/hr]	21.9	21.8	21.9	23.0	21.9	22.9
Peak share [percent]	50.65	50.87	51.28	51.22	50.41	51.19
Off-peak share [percent]	41.02	40.79	40.81	40.63	41.25	40.65
No-travel share [percent]	8.34	8.34	7.91	8.15	8.34	8.16
Car modal share peak [percent]	44.92	44.37	38.17	37.17	35.25	36.94
Bus modal share peak [percent]	55.08	55.63	61.83	62.83	64.75	63.06
Car modal share off-peak [percent]	46.48	47.35	40.20	44.90	45.84	44.91
Bus modal share off-peak [percent]	53.52	52.65	59.80	55.10	54.16	55.09
Bus stops per km	3.01	3.01	3.05	3.04	3.04	3.05
Dedicated bus lanes	0	0	0	1	0	1

Note: Monetary values for London are in 2002 US dollars; for Santiago they are in 2009 US dollars.

of the policies—particularly bus lanes in London—would have a sizable effect on decreasing waiting times. Regarding optimal bus sizes, they do not change much across scenarios. Bus size in London is rather small (about 60 passengers), while in Santiago it is large, consistent with an articulated bus.

Bus fare: Congestion pricing and dedicated bus lanes reduce the cost recovery bus fare by 6 percent and 12 percent respectively in Santiago and 27 percent (both) in London. This effect is caused by larger bus ridership (which decreases the average cost), and, in the case of dedicated bus lanes, also by increased bus speeds that enable larger frequencies with the same fleet. As expected, an X percent of subsidization for bus operating costs yields approximately an X percent reduction of bus fare. When subsidies are used in isolation, our results show that the optimal subsidy for Santiago reaches 55 percent of operational cost. The (restricted) optimal subsidy in London is 100 percent, implying that buses should be free. Note that this is a corner solution, because we restricted fares to be non-negative; in other words, the actual optimal subsidy in London is beyond a 100 percent. Our result is in line with Parry and Small (2009) who find that optimal subsidies in London should be above 90 percent in both periods.¹⁶

Time-of-day fares: If in the reference scenario, where the budget constraint is active, we allow for differentiated peak and off-peak fares, we obtain that the peak fare diminishes and the off-peak fare increases. This shows that it is more efficient to try to move people out of cars and into buses in the peak, than to try to entice people to move to the off-peak. This might be explained because the two negative effects of increasing the off-peak fare are mild. The first effect is that a larger off-peak fare might move people to the peak but the intertemporal elasticity is small. The second effect is that people might move, in the off-peak, from transit to cars, but congestion will not increase much. As a result, the off-peak is used to cross-subsidize the peak, a result that was also previously found by Glaister and Lewis (1978). This cross-subsidy is small in Santiago, but large in London, because in London peak congestion costs are higher—since the initial car share is large—and values of time are larger.

Market shares and speeds in pricing policies: Implementing any pricing measure (congestion pricing or subsidization) produces an increase in the bus share. In the off-peak the increase is small in both cities, ranging from 1 percentage point to 6 percentage points. In the peak, in Santiago, the increase in bus ridership is 7 percentage points for the optimal subsidization and 10 percentage points in the case of congestion pricing. London achieves more: 9 percentage points in the case of optimal subsidization while there is a change of 30 percentage points for congestion pricing. The explanation for the better performance in London of congestion

¹⁶Parry and Small (2009) do not report the actual optimal value and therefore, we cannot say whether they would also find that negative bus fares are optimal, as we do. Negative optimal bus fares have been found before in the literature; see, e.g., Mohring (1979).

pricing as compared to subsidization is simple: as explained before, bus fares reach zero there, implying that the actual optimal fare—which would be negative—is not achieved. Also, policies in London can achieve a larger modal split change than Santiago because the initial car market share is larger.

If instead of optimal subsidies authorities decide on something smaller, reductions of car ridership will obviously be smaller. In our view, this may explain why transit subsidies (if used in isolation) may not always be seen as an efficient tool to decrease congestion. Probably, in many real cases actual subsidies are well below optimal levels (subsidies in Santiago reach about 35 percent while in London about 60 percent). One could further argue that if subsidies are not delivered in a fashion that leads to cost minimization, then a fraction of the subsidy that is received by the transit authorities is lost in inefficiencies, decreasing the power subsidies have to affect the modal split.

As expected, optimal subsidization attracts people from the off-peak to the peak, and from the nontraveling group to both periods. Congestion pricing does the opposite in Santiago: it moves people from car peak to bus peak, to the off-peak, and off the road. In London, however, it brings people to the peak: the reason is that pricing achieves more of a change of mode than in Santiago. The new less congested situation is attractive.

Speed changes are consistent with market share changes. In the off-peak they do not change much; in the peak, if (only) optimal subsidies are in place, speeds increase in about 9 percent for both cities and modes. Congestion pricing, again, achieves quite more. Car speed increases in London from 15 to 31 km/hr and in Santiago from 22 to 29 km/hr. Bus speed increases in London from 12 to 22 km/hr and in Santiago from 14 to 17 km/hr.

Market shares and speeds with bus lanes: Dedicated bus lanes induce differences in full-prices by improving bus speed while keeping self-financing bus fares, as opposed to the transit subsidization policy that achieves differences in monetary prices, but the bus remains slower than the car because of mixed traffic conditions. Therefore, in London, where values of time are large, dedicated bus lanes achieve a higher full price difference in favor of the bus in the peak period than transit subsidization, particularly because the bus fare hits zero. This explains why bus lanes increase bus ridership by 26 percentage points as opposed to the 9 percentage points of optimal subsidization. In Santiago, on the other hand, where values of time are lower and the optimal bus fare is interior (above zero), the change in modal split is similar: 7 percentage points for optimal subsidization, 8 percentage points for bus lanes.

Changes in peak speeds are consistent with both the new market shares and the fact that traffic is segregated: bus speed in London more than doubles (from 13 to 27 km/hr) while car speed remains almost unchanged at 15 km/hr. In Santiago, bus speed increases from 15 to 21 km/hr while car speed decreases from 22 to 17 km/hr. With bus lanes, off-peak speeds improve for all cases but for cars in London, where there is a drop from 44 to 40 km/hr. These results for off-peak speeds are relevant: it implies that there seem to be no need to use somewhat cumbersome time-of-day operation rules for bus lanes.

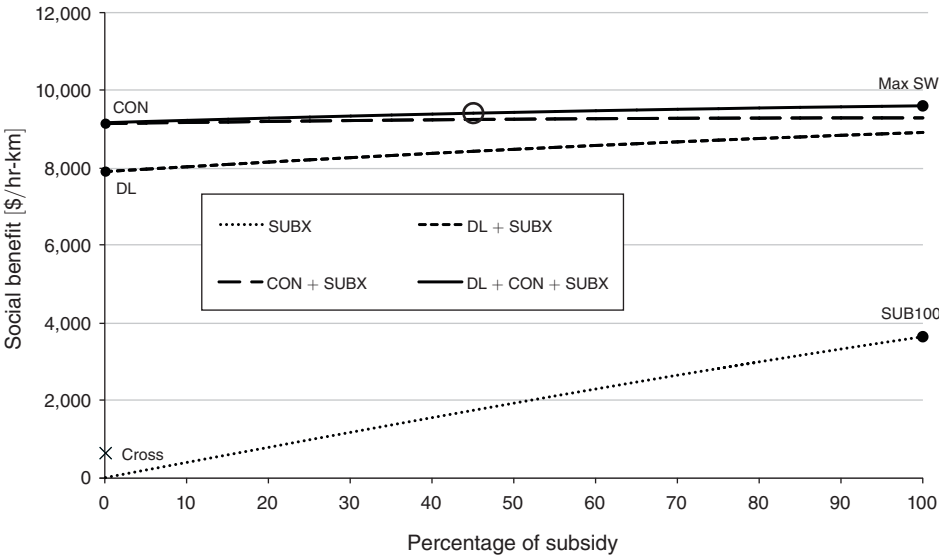


FIGURE 2. SOCIAL BENEFIT WITH RESPECT TO REFERENCE SCENARIO. LONDON

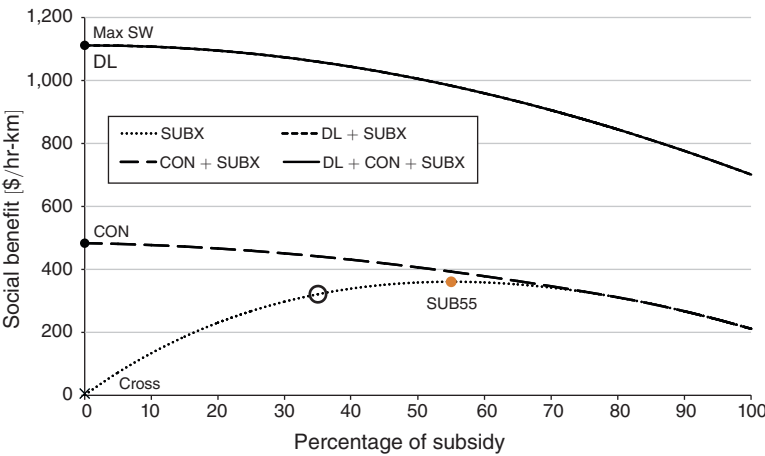


FIGURE 3. SOCIAL BENEFIT WITH RESPECT TO REFERENCE SCENARIO. SANTIAGO

B. Welfare Analysis

The results discussed in the previous subsection might be difficult to digest without a simple way to measure the goodness of a policy. In this section, therefore, we explore the welfare implications of the different scenarios. Welfare results for the London application are presented in Figure 2 and for Santiago in Figure 3. Both figures contain all the information concerning social benefit. A line represents the benefit of a policy together with an X percent of subsidization, thus, the welfare gain of congestion pricing and dedicated bus lanes as stand-alone measures are given by the intercept of their respective curves, which are marked with solid dots for visibility. Additionally we include a × in each figure, which

represent the social benefit of allowing for differentiated time-of-day bus fares (the CROSS scenario).

In the case of London, results show that the best stand-alone policy is car congestion pricing, followed by bus dedicated lanes and then by a subsidy equal to 100 percent of operating costs. For Santiago, results place dedicated bus lanes as the measure that gives the highest benefit, followed by congestion pricing and then optimal subsidization (55 percent of operational costs). It is interesting to note that the benefit that each stand-alone policy induces is quite different: in Santiago, the benefit of bus lanes is more than double of the benefit of congestion pricing or optimal subsidization. In London, congestion pricing and bus lanes reach similar levels of welfare but this is much larger than what optimal subsidization (free buses) achieves. Three important comments follow; first, the implementation path of policies matter, in that there is a clear first policy that should be pursued. Second, that Mohring (1979) conjecture—also sustained by Basso et al. (2011)—that congestion pricing and subsidies are roughly equivalent may not hold, once one adds total and intertemporal demand elasticities, costly public funds, non-negative bus fares, and congestion pricing operation costs. And third, that dedicated bus lanes are indeed an attractive policy and, since they are in many cases considered a “traffic” measure, as opposed to a pricing measure, it might be simpler to implement as it probably does not need to go through congress or city parliament.

We now answer a relevant question: why does congestion pricing perform so well in London but not as well in Santiago? In London, the optimal car toll achieves a very large mode change in favor of bus in the peak (30 percentage points) while, in Santiago, the mode shift in the peak is modest (10 percentage points). Thus, congestion pricing in London is able to bring larger benefits because it can increase car speeds more—as discussed above—but, also, because it increases bus speeds and leads to much larger frequencies, which benefit transit users. The question that follows is why the mode shift is more pronounced in London than in Santiago. The first drivers are the values of time which are much larger in London than in Santiago and, therefore, people are more responsive there to time savings. Two other suspects are (i) the initial bus share of trips (in the reference scenario), which is relatively low in London as compared to Santiago and (ii) the aggregated model used for London, as opposed to the five-income groups used in Santiago. To disentangle this, and show that in fact it is issue (i)—the potential population that can actually be moved from car to buses through car tolls—that dominates, we performed two exercises. First, we reassessed the benefits of each policy using one representative income group in Santiago (homogeneous population), maintaining the average values of time and aggregate initial mode and period shares, and second, we tilted the initial mode share in London towards buses, so that it looks similar to Santiago. The results from these experiments show that, when forcing homogeneity in the Santiago application, results do not change much: congestion pricing still yields roughly half of the benefit that bus lanes generates, and very similar to optimal subsidization. On the other hand, when the car share in the reference scenario in London is similar to the bus share, in both periods, the results change significantly: the benefit of implementing congestion pricing is lower than the benefits from dedicated bus lanes, but not as pronounced as in Santiago. The outcome of these experiments suggest that the

cost, income, and preferences differences between Santiago and London—that are reflected in our model in the values of time and initial modal shares, among others—play a much more important role in the efficiency assessment of policies than the aggregation used for London due to limited data availability.¹⁷

An implication of congestion pricing inducing a larger change in London than in Santiago is that, there, the benefits are associated much more with improved service than revenue collection. It follows that, in a sense, in London there will be benefits even if revenues are “burnt,” because congestion pricing increases consumer surplus (see Table 3). To complement this we conduct two experiments. First, we impose that 90 percent of revenues are lost in the operation of the system; the results show that in London the benefit of congestion tolls halves but remains positive, while in Santiago there are welfare losses. We then calculate the operational cost of congestion pricing—as a percentage of revenues—that would make it equivalent, in welfare terms, to bus lanes. We find 56 percent for London and 6 percent for Santiago.

Results also show that there is little to none efficiency complementarity of policies; that is, once one is implemented, the incremental contribution of additional policies is very small. Consider first congestion pricing and dedicated bus lanes. In the two cities, implementing both policies together lead to almost the same result as implementing only the most efficient stand-alone one (congestion tolls in London, bus lanes in Santiago). Moreover, in London, the maximum benefit that can be achieved is obtained by implementing the three policies together—including free buses—but this is not even 5 percent larger than the benefit given by congestion pricing itself.

Regarding subsidization—a widely discussed subject in the literature—Figures 2 and 3 show that the marginal benefit of increasing transit subsidies, in mixed traffic conditions and without congestion pricing (the slope of the SUBX dashed line), is rather large in London and large only for low levels of subsidies in Santiago. Equivalently, reducing transit subsidies below optimal levels will imply a considerable welfare reduction, something that is particularly true for London, a result that—as mentioned—is in agreement with the recent findings of Parry and Small (2009). Importantly though, when dedicated lanes or optimal congestion pricing are in place the situation is very different. In London, although it is nearly always welfare improving to raise subsidies, their marginal contribution is strongly diminished. In Santiago, adding any subsidy causes no extra welfare and, may actually, harm efficiency. Finally, the fact that with costly public funds the optimal subsidy is 100 percent in London but much lower in Santiago is mainly due to the difference in values of time. With low values of time, in order to have time benefits that are higher than the monetary costs, the congestion reduction has to be very large, something that is obviously more difficult in a situation that is already tilted towards buses (Santiago), and it becomes more difficult as the subsidy is higher. With high

¹⁷ Note that during the first few months after the implementation of congestion pricing in London, automobile traffic declined about 20 percent and average traffic speed during charging days increased 37 percent (Litman 2011). These real changes are large but somewhat smaller than what we find; mind, though, that we consider first-best congestion pricing and the implemented scheme in London has time-of-day and spatial limitations, among others.

values of time, even a moderate decrease in congestion may still bring benefits that are higher than the monetary cost of giving subsidies, which also explains why the marginal benefit of subsidization is almost constant in London.

It seems then that bus lanes and congestion pricing are policies that are powerful enough to create the necessary generalized price difference between modes. Adding subsidies to this, something that one would think should help to fine tune the difference, is a somewhat expensive instrument given the marginal cost of public funds. In fact, additional simulations—not reported here—show this very clearly: if the cost of public funds is imposed to be closer to 1, then the marginal benefit of increased subsidies increases importantly. But if, conversely, the $mcpf$ is larger, the benefits of subsidization decrease importantly, particularly for Santiago. For example, if $mcpf = 1.4$, the optimal subsidization level is only 15 percent in Santiago and while in London it would still be 100 percent, welfare decreases by 17 percent.

We can also use Figures 2 and 3 to make a rough assessment of the current situation in each city. London features cordon congestion pricing, dedicated lanes on several main roads, and about 50 percent of subsidy of bus operating cost, hence its situation would be represented by the circle in Figure 2 placed on 50 percent of subsidy and close to the top lines. This may suggest that London is in a situation with limited possibilities of welfare improvement at least from any of the studied policies. Conversely, Santiago has no congestion pricing, a subsidy of approximately 35 percent of operating costs and very few dedicated bus lanes, so it would be represented by the circle in Figure 3 placed on 35 percent of subsidy and over the dashed curve. It is apparent that there seems to be plenty of room for welfare improvement through the implementation of dedicated lanes and/or optimal congestion pricing. This, in principle, should also help to reduce the amount of subsidy.

As discussed in the previous section, optimally designing the transit system for each urban policy leads to different frequencies and bus sizes. A relevant question naturally follows: if the authority acquires the optimal rolling stock for one scenario but then decides to implement another, how costly is it, in welfare terms, to be stuck with the wrong bus fleet? In order to analyze this question, we first solve the optimization problem with the constraint of having the bus-stop spacing and the bus-size set to those of the reference scenario, while allowing for the possibility of increasing the number of buses. The benefits of implementing stand-alone measures in London and Santiago remain almost unchanged (the benefit loss is below 5 percent), showing that the inefficiencies caused by having the wrong fleet in terms of bus size is small as long as more buses of the same type can be purchased. On the other hand, if starting from the REF scenario a policy is implemented but both the number and size of buses remain the same, then welfare losses—as compared to what is obtained with the optimal fleet—increases importantly, reaching 10 percent in Santiago and 70 percent in London.

C. Distributional Impacts

To forecast support or resistance towards the policies, we look into distributional impacts. We look at the distribution of surplus between consumers and government

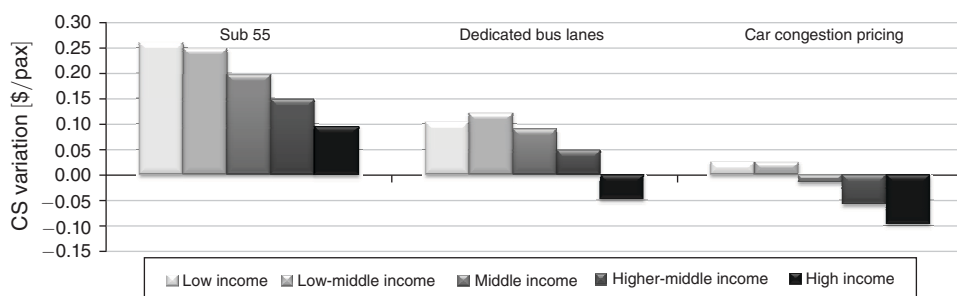


FIGURE 4. CS VARIATION BY INCOME GROUP WITH RESPECT TO REFERENCE SCENARIO. SANTIAGO.

and the distribution of surplus between different income groups for the case of Santiago.¹⁸

Optimal Transit Subsidization.—With optimal bus subsidization (55 percent of operational costs in Santiago and 100 percent in London), government spending is large and costly because of the dead-weight losses from taxation elsewhere in the economy. On the other hand, bus users benefit directly from reduced fares and all individuals are better off before considering general equilibrium effects, because demand changes (from car to buses in both periods and from the outside option to peak or off-peak traveling) reduce travel times in the peak while keeping off-peak speeds almost unchanged. Figure 4 shows the variation of consumer surplus in Santiago, with respect to the reference scenario, for a representative consumer of each income group: the lower the income, the higher the benefit from optimal subsidization. This is due to the fact that most of the benefit is perceived by bus users and transit usage is higher the lower the average income of a group is.

Dedicated Bus Lanes.—A very important point of the dedicated bus lanes policy is that, when applied in isolation, it does not impose losses for the government (other than the implementation costs): the benefits are reached by charging a bus fare that covers the system costs while car travel remains untolled. Therefore, the benefit of dedicated lanes comes from an increase in consumer surplus. In the case of London, consumer surplus increases with this measure because the large reduction in bus travel times and a smaller bus fare overcome the small reduction in car speeds (less than 1 percent in the peak and 14 percent in the off-peak). For Santiago, however, while total surplus increases, car speed actually diminishes in the peak by 24 percent. Fortunately, here we are able to take a closer look at winners and losers by analyzing distributional impacts for different income groups. Figure 4 shows that implementing dedicated bus lanes benefits most income groups, with the exception of the highest income group, which is actually worse off. What happens is that car users face a moderate increase of travel time in the peak and a small reduction in the off-peak, while bus users face a much improved system (bus speed and frequency

¹⁸Note that transit operator will always cover its costs through bus fare plus subsidies when relevant; one could consider that nonmonetary benefits for the transit operator comes from size, that is, its market share but we do not do this here.

almost increase by 50 percent while the fare is reduced by more than 10 percent). Since the bus share was already large for the two lowest income groups the benefit is direct; for the next two income groups the modal shift from car to bus is large, which explains the benefits; finally, for the highest income group, the negative effect of reduced car speed in the peak dominates because car use remains large.

Congestion Pricing.—Car congestion pricing is probably one of the most debated ideas in transportation. Economists usually advocate its use, while authorities and commuters dislike the idea *ex ante*. The usual explanation for this resistance, that comes from models without explicit mode substitution, is that consumer surplus would decrease: the group of people that was not driving initially are indifferent, another group of people is priced off the streets and therefore dislike the policy, while those that remain driving face less congestion but pay a higher generalized price after the tax (which is necessary to achieve a smaller traffic flow).¹⁹ Moreover, it is likely that those priced off the road are those who benefit least from the speed increase due to low values of time (Verhoef and Small 2004), so that the measure will be regressive if lower values of time are due to lower incomes. This is what gave rise to the literature on Pareto-improving schemes for recycling the congestion tolls (see, e.g., Nie and Liu 2010 and references therein).

In our model—where mode substitution is allowed and the transit system is optimally modified to accommodate new demand—we obtain very different insights: when applying congestion pricing, total consumer surplus may increase despite the car tolling. The difference is clear: here, those commuters that were riding the bus in the first place are now not indifferent but better off, because they face an improved system generated by the larger bus ridership. And some of those that changed mode will also benefit if their previous decision was only marginally in favor of the car. In addition, the government collects tolls, that could later be used to induce other desirable outcomes. In fact, in terms of distribution of the efficiency gains, results are quite different across applications: in London 33 percent of the social benefit comes from an increase in consumer surplus while in Santiago consumer surplus decreases.

Importantly, the Santiago case enables a more detailed analysis of distributional effects, showing that the change in consumer surplus is actually not regressive. As shown in Figure 4, the high-income users (with high values of time) face a large decrease in consumer surplus while lower-income people actually benefit. The middle-income group is slightly worse off. This is because higher-income users keep driving more than lower-income individuals after implementation of the policy; they face decreased congestion but have to pay the tax. Travelers with low incomes, on the other hand, benefit because larger bus ridership—which came about mostly because of mode change of middle income users—improves the bus system. Middle income groups have commuters that keep driving, commuters that change mode because it is beneficial, and commuters that are already using the bus, who benefit.

¹⁹ See Hau (2005a,b).

TABLE 4—INCOME GROUPS AND WELFARE WEIGHTS

Income group	Size (share of population)	Average income per household [\$]	Income ratio	Welfare weight, $j = 0.25$	Welfare weight, $j = 0.50$
Low income	0.14	228	5.52	1.53	2.35
Middle-low income	0.28	505	2.49	1.26	1.58
Middle income	0.32	963	1.31	1.07	1.14
Middle-high income	0.17	1,913	0.66	0.9	0.81
High income	0.09	5,003	0.25	0.71	0.5

Note: Monetary values are in 2009 US dollars.

IV. Social Consumer Surplus Measure

Our analysis has used Marshallian consumer surplus, a measure that may be objected because it may assign a larger weight to richer people since they have larger willingness to pay mostly because of income. We explore here an alternative by simply assigning different welfare weights to the different income groups in Santiago and reassessing the social benefit of each of the policies studied in Section III. We consider that every dollar raised has a cost equal to the marginal cost of public funds and by combining the *mcpf* with the use of welfare weights we have a simplified way to include distributional concerns of a government in the assessment of transport policies. A more detailed analysis would explicitly consider a specific revenue source for the subsidies, the particular (potentially distortionary) effect on the economy, or the distribution of the net revenues from congestion pricing, but this is outside the scope of this paper, so we assume that the government’s deficit that results from implementing any policy is being paid from general taxation.

With the purpose of illustrating the effect of different weights on the results, we (arbitrarily) use as the welfare weight for the income group i the ratio between national average income and group-specific average income ($\bar{I}/I_i$) raised to the power of j . We consider welfare weights that are relatively high for low-income groups ($\bar{I}/I_i > 1$) and relatively low for high-income groups ($\bar{I}/I_i < 1$), thus incorporating benefits of vertical equity. With this measure, the lower the exponent j is, the closer the weights are to unity, and the weighted welfare function approaches the welfare function considered in Section III as the exponent goes to zero. A very large exponent will imply virtually no consideration for the income groups whose income is too high with respect to the average, and almost exclusive sensitivity towards low-income groups. Table 4 shows, for each income group, the size, the average income per household, the income ratio, and the welfare weights considered.

Figure 5, to be compared with Figure 3, shows the reassessment of the social benefit of implementing transit subsidies, car congestion pricing, dedicated bus lanes, and the combination of them. The main result of this exercise is that bus subsidization becomes a much more appealing alternative when welfare weights are considered. Both its benefit as a stand-alone measure as well as the complementarities with other policies increase significantly. For the two sets of weights considered, the optimal amount of subsidy is 100 percent of operational costs, and, for the case of $j = 0.5$, it becomes the stand-alone measure that yields the highest social benefit, largely outperforming bus lanes and congestion pricing. Naturally, the larger j is,

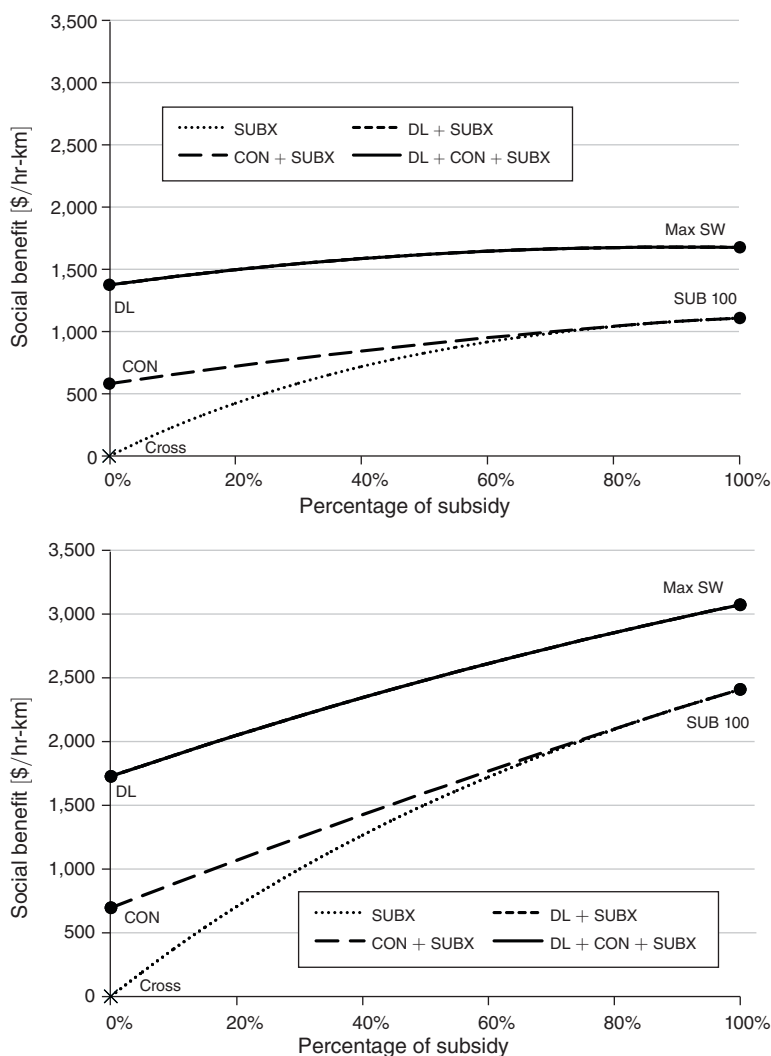


FIGURE 5. SOCIAL BENEFIT WITH WELFARE WEIGHTS: $j = 0.25$ (top), $j = 0.50$ (bottom). SANTIAGO

the more beneficial subsidization becomes. These results follow directly from the fact that subsidization benefits bus users and that lower-income groups are the ones that use more intensively the bus, therefore subsidization becomes a (imperfect) means to achieve equity objectives. Importantly, congestion pricing and bus lanes also yield a higher benefit the larger the value of the exponent is, but the increase is much more moderate. Recall that, in absence of weights, these policies are not regressive (see Figure 4), and, as a consequence, none of them should be hurt by using welfare weights that are concerned with vertical equity.

V. Conclusion

The main results of our analysis suggest that there is large efficiency substitutability between policies in both applications. This implies that—in the absence of

other policies—transit subsidies may indeed be a very good measure for reducing transport negative externalities and increase social welfare, at least if the marginal cost of public funds and induced cost inefficiencies are not very large. Nevertheless, their contribution might be severely diminished if other policies, such as congestion pricing and dedicated bus lanes, are implemented first. This speaks of the importance of the implementation path of urban transport policies: a well-planned road map might save the authorities both money and conflicts. Importantly, bus lanes seem to be an effective way to improve service levels and decrease fare, without injecting public funds. Our results also show that the best stand-alone policy is not the same in the two cities, and that the policies—including congestion pricing—are progressive if the transit system responds optimally to changes in demand.

Evidently, several caveats or qualifications apply to our analysis. First, we have used the demand models far from the cloud of estimation points; this is, of course, necessary if one wants to forecast the probable result of new and large policies. Second, we do not explore an actual network, but a homogeneous yet representative part of it. In actual transport networks, some parts will be congested while others will not, because demand is not homogeneous, as it depends on the distribution of firms, dwellings, and production centers. A transit authority will need to take this into account when deciding operation rules, routes, and fleet acquisition, all of which impacts on subsidies of course. Also, charging first-best congestion tolls in a real network is nothing short of impossible; if congestion pricing can only be applied in second best fashion and in limited parts of the network, subsidies will certainly look better and complementarities between policies may appear. Furthermore, as we set the road capacity of the congested representative kilometer of network to an equivalent of three lanes, we allow the fraction of capacity dedicated to buses to be $1/3$ or $2/3$. We believe that this is not a very strong limitation, as in a real networks the available capacity for trips is not limited to one highway with a fixed number of lanes, but there will be parallel roads that can be grouped. A good example of this, already applied in some cities, is having a complete road dedicated to buses with (substitute) parallel roads available for cars. The wide set of possibilities enhances the generality of the results on bus lanes with respect to the fraction of capacity, and suggests that what is important is to generate differences between car and bus speed in peak periods, while not damaging too much the off-peak operations, when congestion is not a big issue.

Finally, due to a lack of detailed demand models for London we are only able to use a single model and, therefore, the demand substitution patterns are homogeneous in the population. As explained, this does not seem to affect the overall conclusions, but it does preclude a detailed distributional analysis.

REFERENCES

- Allport, Roger J. 1981. "The Costing of Bus, Light Rail Transit and Metro Public Transport Systems." *Traffic Engineering and Control* 22 (12): 633–39.
- Anderson, Simon P., and André de Palma. 1992. "Multiproduct Firms: A Nested Logit Approach." *Journal of Industrial Economics* 40 (3): 261–76.
- Anderson, Simon P., André de Palma, and Jacques François Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge: MIT Press.

- Auerbach, Alan J, and James R. Hines, Jr. 2002. "Taxation and economic efficiency." In *Handbook of Public Economics*, Vol. 3, edited by Alan J. Auerbach and Martin Feldstein, 1347–1421. Amsterdam: Elsevier.
- Auriol, Emmanuelle, and Michael Warlters. 2012. "The marginal cost of public funds and tax reform in Africa." *Journal of Development Economics* 97 (1): 58–72.
- Ballard, Charles L., John B. Shoven, and John Whalley. 1985. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *American Economic Review* 75 (1): 128–38.
- Basso, Leonardo J., Cristián Angelo Guevara, Antonio Gschwender, and Marcelo Fuster. 2011. "Congestion pricing, transit subsidies and dedicated bus lanes: Efficient and practical solutions to congestion." *Transport Policy* 18 (5): 676–84.
- Basso, Leonardo J., and Sergio R. Jara-Díaz. 2006. "Distinguishing Multiproduct Economies of Scale from Economies of Density on a Fixed-Size Transport Network." *Networks and Spatial Economics* 6 (2): 149–62.
- Basso, Leonardo J., and Sergio R. Jara-Díaz. 2012. "Integrating congestion pricing, transit subsidies and mode choice." *Transportation Research Part A: Policy and Practice* 46 (6): 890–900.
- Basso, Leonardo J., and Hugo E. Silva. 2014. "Efficiency and Substitutability of Transit Subsidies and Other Urban Transport Policies: Dataset." *American Economic Journal: Economic Policy*. <http://dx.doi.org/10.1257/pol.6.4.1>.
- Ben-Akiva, Moshe Emanuel. 1973. *Structure of Passenger Travel Demand Models*. PhD diss. Massachusetts Institute of Technology.
- Boyd, J. Hayden, Norman J. Asher, and Elliot S. Wetzler. 1978. "Nontechnological innovation in urban transit: A comparison of some alternatives." *Journal of Urban Economics* 5 (1): 1–20.
- Corporación Andina de Fomento (CAF). 2010. *Observatorio de Movilidad Urbana para América Latina*. <http://publicaciones.caf.com/media/1130/0.pdf>.
- De Borger, Bruno, Inge Mayeres, Stef Proost, and Sandra Wouters. 1996. "Optimal Pricing of Urban Passenger Transport: A Simulation Exercise for Belgium." *Journal of Transport Economics and Policy* 30 (1): 31–54.
- De Borger, Bruno, Kristiaan Kerstens, and Álvaro Costa. 2002. "Public transit performance: What does one learn from frontier studies?" *Transport Reviews* 22 (1): 1–38.
- Department for Transport. 2004. *MSA: Cost Benefit Analysis. Transport Analysis Guidance (TAG) Unit 3.9.2*. Department for Transport. London, April.
- Department for Transport. 2009. *Values of Time and Vehicle Operating Costs. Transport Analysis Guidance (TAG) Unit 3.5.6*. Department for Transport. London, October.
- Elgar, Ilan, and Christopher Kennedy. 2005. "Review of Optimal Transit Subsidies: Comparison between Models." *Journal of Urban Planning and Development* 131 (Special Issue): 71–78.
- Fernandez, R., E. Valenzuela, and T. Gálvez. 2002. "Incorporación de la capacidad y rendimiento de paraderos en el programa TRANSYT." In *Actas del XI Congreso Panamericano de Ingeniería Tránsito y Transporte*. Gramado, Brasil.
- Gagnepain, Philippe, and Marc Ivaldi. 2002. "Incentive regulatory policies: The case of public transit systems in France." *RAND Journal of Economics* 33 (4): 605–29.
- Glaister, Stephen, and Davis Lewis. 1978. "An integrated fares policy for transport in London." *Journal of Public Economics* 9 (3): 341–55.
- Harrison, Glenn W., Thomas F. Rutherford, and David G. Tarr. 2002. "Trade Policy Options for Chile: The Importance of Market Access." *World Bank Economic Review* 16 (1): 49–79.
- Hau, Timothy D. 2005a. "Economic Fundamentals of Road Pricing: A Diagrammatic Analysis, Part I—Fundamentals." *Transportmetrica* 1 (2): 81–117.
- Hau, Timothy D. 2005b. "Economic Fundamentals of Road Pricing: A Diagrammatic Analysis, Part II—Relaxation of Assumptions." *Transportmetrica* 1 (2): 119–49.
- Hensher, David A. 1998. "Establishing a Fare Elasticity Regime for Urban Passenger Transport." *Journal of Transport Economics and Policy* 32 (2): 221–46.
- Huang, Hai-Jun. 2000. "Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters." *Transportation Research Part E: Logistics and Transportation Review* 36 (4): 267–84.
- Jara-Díaz, Sergio. 2007. *Transport Economic Theory*. Bingley: Elsevier.
- Jara-Díaz, Sergio, and Antonio Gschwender. 2009. "The effect of financial constraints on the optimal design of public transport services." *Transportation* 36 (1): 65–75.
- Kenworthy, Jeffrey R., Felix B. Laube, and Jean Vivier. 2001. *Millennium cities database for sustainable transport*. Brussels: International Union of Public Transport.
- Kutzbach, Mark J. 2009. "Motorization in developing countries: Causes, consequences, and effectiveness of policy options." *Journal of Urban Economics* 65 (2): 154–66.

- LaFrance, Jeffrey T.** 1986. "The Structure of Constant Elasticity Demand Models." *American Journal of Agricultural Economics* 68 (3): 543–52.
- Litman, Todd.** 2011. *London Congestion Pricing: Implications for Other Cities*. Victoria Transport Policy Institute. Victoria, November.
- Litman, Todd.** 2012. *Transportation Elasticities: How Prices and Other Factors Affect Travel Behavior*. Victoria Transport Policy Institute. Victoria, March. <http://www.vtpi.org/elasticities.pdf> (accessed on March 20, 2013).
- McFadden, Daniel.** 1978. *Modelling the Choice of Residential Location*. In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, 75–96. Amsterdam: North Holland.
- Ministerio de Transportes y Telecomunicaciones (MTT).** 2004. *Actualización de las Programaciones de Semáforos como Parte del PTUS en el Marco de Transantiago, II Etapa*. MTT. Chile.
- Mohring, Herbert.** 1972. "Optimization and Scale Economies in Urban Bus Transportation." *American Economic Review* 62 (4): 591–604.
- Mohring, Herbert.** 1979. "The Benefits of Reserved Bus Lanes, Mass Transit Subsidies, and Marginal Cost Pricing in Alleviating Traffic Congestion." In *Current Issues in Urban Economics*, edited by Peter M. Mieszkowski and Mahlon R. Straszheim. Baltimore: Johns Hopkins University Press.
- Nie, Yu Marco, and Yang Liu.** 2010. "Existence of self-financing and Pareto-improving congestion pricing: Impact of value of time distribution." *Transportation Research Part A: Policy and Practice* 44 (1): 39–51.
- Parry, Ian W. H., and Kenneth A. Small.** 2009. "Should Urban Transit Subsidies Be Reduced?" *American Economic Review* 99 (3): 700–724.
- Proost, Stef, and Kurt Van Dender.** 2008. "Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds." *Transportation Research Part A: Policy and Practice* 42 (9): 1220–30.
- Savage, Ian.** 2004. "Management objectives and the causes of mass transit deficits." *Transportation Research Part A: Policy and Practice* 38 (3): 181–99.
- Secretaría Interministerial de Planificación de Transporte (SECTRA).** 2005. *Análisis y Actualización del Modelo Estraus*. SECTRA. Chile.
- Shires, J. D., and G. C. de Jong.** 2009. "An international meta-analysis of values of travel time savings." *Evaluation and Program Planning* 32 (4): 315–25.
- Small, Kenneth A.** 1983. "Bus priority and congestion pricing on urban expressways." *Research in Transportation Economics* 1: 27–74.
- Small, Kenneth A., and Harvey S. Rosen.** 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49 (1): 105–30.
- Small, Kenneth A., and Erik T. Verhoef.** 2007. *The Economics of Urban Transportation*. New York: Routledge.
- Tirachini, Alejandro, David A. Hensher, and Sergio R. Jara-Díaz.** 2010. "Restating modal investment priority with an improved model for public transport analysis." *Transportation Research Part E: Logistics and Transportation Review* 46 (6): 1148–68.
- Transportation Research Board.** 1985. *Highway Capacity Manual*. Transportation Research Board Special Report 209. National Research Council. Washington, DC.
- Transport for London.** 2007. *London Travel Report*. <http://www.tfl.gov.uk/cdn/static/cms/documents/London-Travel-Report-2007-final.pdf>.
- Verhoef, Erik T., and Kenneth A. Small.** 2004. "Product Differentiation on Roads." *Journal of Transport Economics and Policy* 38 (1): 127–56.
- Viton, Philip A.** 1983. "Pareto-optimal Urban Transportation Equilibria." *Research in Transportation Economics* 1: 75–101.
- Winston, Clifford, and Chad Shirley.** 1998. *Alternate Route: Toward Efficient Urban Transportation*. Washington, DC: Brookings Institution Press.