TSE working paper series

12-353

November 2012

Research Group : Public Economics

“Pricing of Transport Networks, Redistribution and Optimal Taxation”

Antonio Russo

Revised version November 2012

Toulouse School of Economics (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31015 TOULOUSE Cedex 6
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.tse-fr.eu** - contact@tse-fr.eu

Pricing of Transport Networks, Redistribution and Optimal Taxation*

Antonio Russo†

This version: November 2012

Abstract

We study optimal pricing of roads and public transport in presence of nonlinear income taxation. Individuals are heterogeneous in unobservable earning ability. Optimal transport tariffs depend on time costs of travel and work schedule adjustments (days and hours worked per day) as a response to commuting costs. We find that discounts for low income individuals are optimal only if the time cost of a trip is small enough. Lower travel time costs facilitate screening: therefore, redistribution provides an additional motive for congestion pricing. Finally, we investigate the desirability of means-testing of transport tariffs.

JEL Code: R41, H23, H21

Keywords: road pricing; public transport pricing; income taxation; means-testing

*A previous version of this paper circulated with the title: "Pricing and Provision of Transport Infrastructure with Nonlinear Income Taxation". I thank Georges Casamatta, Helmuth Cremer, Bruno De Borger, Philippe De Donder, Jean-Marie Lozachmeur, Stef Proost and Emmanuel Thibault for useful comments and suggestions. I also thank audiences in Marseille (LAGV 2011), Montpellier (DMM 2011), Oslo (EEA 2011), Stockholm (Kuhmo-Nectar Conference 2011) and Toulouse. All errors are mine.

†Toulouse School of Economics and RSCAS, European University Institute. Email: antonio.russo@eui.eu

1 Introduction

It is often argued that prices on urban transport networks should reflect social costs of travel. For instance, as roads suffer from congestion externalities, economic theory suggests that road pricing can increase efficiency. Clearly, this may also have an impact on the distribution of welfare across society. Indeed, policymakers often care about redistribution when designing tariffs for publicly provided transport infrastructure. Concerns of a possible regressive effect recently impeded the introduction of road pricing in New York City and Paris.[1] Plans for a road pricing scheme in San Francisco include a tariff discount to low income drivers, while discounted public transport fares are often granted to people qualifying for certain criteria, including income. Moreover, governments often subsidize commuting expenditures (e.g. through tax exemptions) for reasons that include helping disadvantaged workers.

Economic literature has looked at redistributive issues in pricing of transportation infrastructure (see Small and Verhoef (2007) for a comprehensive review). However, it has done so (with an important exception discussed below) ignoring the presence of income taxation. This leaves open the question of whether such concerns are actually relevant, as they could possibly be addressed with appropriately designed income taxes. The main objective of this paper is to study such a question.

We consider the problem of a welfare-maximizing government that designs both income taxes and tariffs for roads and public transportation.[2] Individuals are heterogeneous in (exogenous) earning ability, which is assumed to be private information, as is their labor supply. Thus, the government faces self-selection constraints that may limit welfare redistribution. To keep the setup as simple as possible, we use a model with only two types of individuals (*à la* Stiglitz (1982)).

It is well-established that nonlinear tariffs are a crucial ingredient of efficient pricing policies in network industries (Wilson (1993)). They are drawing increasing interest also in transportation, although their potential redistributive role (recognized in other regulated industries, e.g. energy or telecommunications) has not been explored.[3] This is why they are studied in the current paper. Nonetheless, nonlinear pricing may not always be imple-

[1] In a recent interview, New York State Assemblyman Richard L. Brodsky said he opposed its introduction “for the reason that these schemes put the burden for paying the fees on blue blood and blue collar alike” (see New York Times, “Congestion Pricing: Just Another Regressive Tax?” www.nytimes.com)

[2] The term "tariff" should be given a broad interpretation here: since the government controls taxes and prices, tariffs we describe may result not only from fares or tolls, but also from commuting subsidies in the form of tax deductions.

[3] see Wang *et al.* (2011) for a study of nonlinear pricing of tolled roads and Batarce and Ivaldi (2011) for public transportation. Cremer and Gahvari (2002) study nonlinear pricing by a regulated firm in the presence of optimal income taxation. Their setup, however, neglects important features for transportation, such as time costs of consumption.

mentable (at least at reasonable costs).[4] This is why we also look at the case in which the government is constrained to use linear tariffs.

Previous public finance literature has studied how (if at all) a government that can use income taxes should deviate, due to distributional concerns, from correcting market failures (Cremer *et al.* (1998), Bovenberg and Goulder (2002), Kaplow (2006)). However, it has disregarded two relevant features for transportation, which are central in our analysis. The first is that consumption of transport goods requires travel time. Boadway and Gahvari (2006) and Gahvari (2007) consider time of consumption in an optimal redistributive taxation framework. They do not consider externalities.[5] Mayeres and Proost (1997) study optimal redistributive taxation in the presence of congestion externalities, but restrict attention to linear taxes. Our approach is complementary, since it does not assume restrictions on the design of income taxes (it is constrained only by the available information).

A second key feature of our setup is that we explicitly model the relation between travel and labor supply. Individuals can decide the number of days at the workplace (which require commuting) and the length of their working day (or their daily effort). This can be interpreted as the choice between jobs offering different time schedules. For instance, one may choose a job with a four-days-a-week schedule or a five-days-a-week one but requiring shorter daily shifts. Intuitively, increased commuting costs encourage, all else equal, to choose the former.[6] However, we assume substituting working days for more hours worked per day implies a penalty in terms of productivity. This is due to diminishing returns in daily hours caused, for example, by fatigue. [7] While labor supply plays a central role in models of income taxation, little attention has been dedicated to the impact of policies that affect commuting to work. Parry and Bento (2001) and Van Dender (2003) consider the issue, although in a setup with homogeneous individuals. Moreover, their model of labor supply is more rigid, allowing only the choice of working days (of fixed length). This matters because

[4]It is indeed quite demanding in informational terms, since observability of individual trip quantities is necessary. This information is not rarely available though: for example, most road pricing schemes involve the use of electronic tolling systems that keep track of individual accesses to the tolled road. Moreover, governments often have access to commuting data collected by employers. We discuss feasibility issues at the end of Section 2 below

[5]Cremer *et al.* (1998) and Kaplow (2006) studied environmental levies in the presence of nonlinear income taxation. They consider a model where commodities do not require any time for consumption. Moreover, they focus on externalities that do not affect the marginal cost of consuming goods, unlike traffic congestion. An optimal taxation model with time as input for activities and congestion externalities is also studied in De Borger (2011). He uses a representative agent framework.

[6]Commuters may also have other margins of flexibility in responding to changes in travel costs: they may change their residence or shift travel to off-peak hours (Arnott *et al.* (1993)). A discussion of their likely impact on our results is provided in the concluding remarks.

[7]This can also be interpreted as capturing the trade-off between workdays at the office and telework, as long as working outside the office (e.g., at home) is less productive than working on-the-job This is discussed in some detail in Section 3 below.

if individuals can adjust daily hours, increasing the cost of commuting does not necessarily result in lower labor supply. This has important implications for optimal transport tariffs.[8]

We show, to begin, that transport pricing has a redistributive role even in the presence of nonlinear income taxation. It can be used to improve screening of types, relaxing the self selection constraints.[9] In our setup, individuals face a trade-off when deciding on their work schedule. On the one hand, commuting less often saves time spent on travel. On the other, it requires (*at constant income*) to increase workday length and, hence, total hours worked. This is because when people work more hours per day, (average) hourly productivity is reduced. Our findings suggest, roughly speaking, that if this reduction is large (resp. small) compared to the time cost of a trip on a given mode, it is optimal to have low ability individuals pay a smaller (resp. larger) tariff to use that mode. The reason is that low ability types have less free time than high ability mimickers. Hence, encouraging low types to commute more often (via a lower tariff) improves their welfare more than that of mimickers only if the benefit in terms of reduced labor supply outweighs the additional time spent on travel. It is only in that case that lowering the tariff intended for low ability types relaxes self-selection constraints. To put it differently, our results suggest that, for a given transport mode, discounted tariffs for low income individuals are optimal if and only if the time cost of travel on that mode is small enough. If travel by public transport is more time-consuming than car travel, this suggests that "social" tariffs and discounts for low income households may be more effective, in redistributive terms, when concerning cars than public transportation.

The above implies that individuals of different earning ability should not pay the same tariff for a given transport mode. Hence, nonlinear tariffs are necessary to implement the second-best allocation (constrained, that is, only by self-selection). However, as mentioned above, the government may also face additional information constraints: it may only be able to observe aggregate trip quantities (anonymous transactions). In that case, only linear tariffs are implementable. Even so, the trade-off described above is a key determinant of optimal prices. In essence, the optimal (linear) tariff for a given mode tends to increase with travel time cost, but decreases with the extent of productivity losses when commuting is discouraged and more daily hours induced.

Furthermore, different value of time for individuals of different ability (at a given quan-

[8]It is quite intuitive that the costs of commuting have an impact on labor supply. However, there exists some empirical evidence (Gutierrez-i-Puigarnau and van Ommeren (2010)) suggesting this impact is small. Since further research is needed to corroborate or qualify these results, the assumptions on which our analysis is based look reasonable.

[9]This is true in spite of the fact that individuals' preferences are separable in goods and leisure. In the absence of time costs of travel and of diminishing returns in daily hours, separability would make distributive concerns irrelevant when designing pricing of transport infrastructure (Atkinson and Stiglitz (1976)).

tity of goods and income) implies that screening of types can be sharpened by reducing the time costs of journeys. Hence, curbing road congestion may also make redistribution more effective. This gives an additional motive to raise car tariffs for all types. Interestingly, a redistribution-minded government has, therefore, an additional reason to implement congestion pricing. This is in line with the results of Kreiner and Verdelin (2012), who focus on provision of public goods.

As the anecdotal evidence mentioned above shows, there are concerns that transport tariffs that closely follow marginal social costs of travel may be hurtful to the poor. The government may thus want to differentiate them based on income, introducing means-testing. The suitability of means-testing for urban transportation is part of the current policy debate (see, e.g. Estupinan *et al.* (2007)). The results we obtained suggest that, when nonlinear tariffs can be implemented, individuals of different income should not pay the same *per-trip* tariff.[10] However, this does not necessarily imply that individuals differing in income should be offered different tariff schedules (i.e., that transport tariffs should be means-tested). In the last part of the paper, we turn our attention to such a question. We show that when modal split, in the second-best allocation, is such that high income individuals commute more by car than low income ones and public transport trips have larger time costs than car trips, transport tariffs can be independent of income. That is, under a reasonable condition, individuals of different income can be offered the same tariff schedules and means-testing avoided. We conduct some numerical simulations in the final section of the paper. We find only very few counterexamples in which implementability with separable tax and tariff functions cannot be achieved.

The rest of the paper is organized as follows: Section 2 presents the model. We present optimal tariffs in Section 3. Section 4 considers implementation and means-testing. Section 5 presents some numerical illustrations of the results. Section 6 concludes. Proofs of all propositions are provided in an Appendix.

[10]The results described above are obtained under the assumption that the government uses a general "tax-and-pricing" function, based on income and trip quantities. This means that optimal transport tariffs should, a priori, be conditional on income. Income taxes may also have to be conditional on travel quantities. Note also that if the government is constrained to use linear tariffs, it is because individual trip quantities are unobservable and all transactions are anonymous. In that case, tariff differentiation is never incentive-compatible.

2 The model

2.1 Setup

We consider a population composed of two types of individuals $i = 1, 2$. They differ in earning ability (a measure of their productivity at work), identified by the parameter w_i, with $w_2 > w_1$. The size of group i is denoted π_i, with $\sum_{i=1,2} \pi_i = 1$.

There are five goods in the economy: composite consumption C (the numeraire), (peak-hour) trips by car D and public transportation B, leisure x and labor supply L. The production technology is linear in labor, with constant marginal costs normalized to one, for C and D. The production sector is perfectly competitive. The marginal cost of a public transport trip, sustained by the government (assumed to be the provider of the service), is constant and equal to c_B.

A trip by car or public transport requires a_j $j = D, B$ units of time, for all individuals.[11] We ignore heterogeneity in location: hence, all trips cover the same distance. We assume the time spent consuming C to be a (perfect) substitute for leisure. Thus, contrary to time on travel and at work, it has no opportunity cost (see Boadway and Gahvari (2006)). Individuals face the time constraint

$$a_D D^i + a_B B^i + L^i + x^i \leq 1 \quad i = 1, 2$$

Suffix i stands for individual quantities, which may vary depending on the individual's type. We normalize the time endowment to one (same for all types). To capture road congestion, we assume that a_D is an increasing and convex function of the aggregate amount of car trips. Congestion on public transport is ignored for simplicity: it would make the optimal tariff formulae more complicated without adding much to the results. Therefore $a_D = \varphi^D(\bar{D})$ and a_B is fixed, with $\bar{D} = \sum_{i=1,2} \pi_i D^i$ denoting the total level of road traffic. We assume also that a_D is taken as given when deciding how many car trips to take, which generates a congestion externality.

Individuals choose the amount of labor supply deciding on two key parameters: N, the number of working days, and h, the amount of hours worked per day on-the-job. Labor supply is thus $L = N \cdot h$. This represents the choice between jobs offering different workdays-hours schedules. For example, the individual may have the choice between a job offering a four-days-a-week schedule but requiring longer daily shifts (e.g. start early or finish late) or one with a five-days-a-week schedule but with shorter daily shifts. Moreover, h may also be

[11] Transport trips can be seen as activities obtained combining goods and time. We assume a fixed-proportions household production technology, as in, e.g., Kleven (2004), so our formulation is consistent with that representation.

interpreted as a measure of effort provided (for a given number of hours) per day on the job. All individuals are assumed to be commuters and to use the transport network only for this purpose (which we consider a reasonable simplification given our focus on peak-hour travel, for which commuting is a dominant contributor). A day at the workplace requires a return commuting trip, on one of the two modes. Therefore: $N^i = D^i + B^i$. Finally, individuals' income is obtained as

$$I^i = N^i w^i f(h^i) = \left(D^i + B^i\right) w^i f(h^i) \quad i = 1, 2$$

where, importantly $f' > 0,\ f'' < 0$. The assumption of decreasing returns in hours per day captures diminishing productivity when working longer hours or when increasing daily effort (as in Gutierrez-i-Puigarnau and van Ommeren (2010)). This may be because of fatigue, but also because opportunities to interact and coordinate with colleagues or customers may be smaller when working at early or late hours. The more these effects constrain hourly productivity, the greater the concavity of function f. For a given individual of type w^i, total labor supply (which is unobservable) can be rewritten as

$$L^i = N^i \cdot h^i = (D^i + B^i) g\left(\frac{I^i}{w^i\left(D^i + B^i\right)}\right) \qquad \text{where} \qquad h^i = g\left(\frac{I^i}{w^i\left(D^i + B^i\right)}\right)$$

using the fact that $N^i = D^i + B^i$ and that $f(h^i) = {}^{I^i}/_{w^i\left(D^i+B^i\right)}$. Function $g(.)$ is the inverse of $f(.)$. Therefore $g' > 0,\ g'' > 0$. As may already be understood, diminishing returns generate an important trade-off when deciding on the work and travel schedule. If an individual travels one more day to work, she has to sustain the monetary and time costs of a commuting trip. On the other hand, doing so allows, for given income, to reduce total labor supply. This is because hours per day are reduced and average hourly productivity goes up. We will see below that such a trade-off has important implications for the optimal tariff schemes.

All individuals have the same utility function

$$U(C, D, B, x) = \Omega(C) + \gamma(D, B) + \phi(x)$$

Note the separability between leisure and goods. In the absence of time costs of travel and of diminishing returns on hours worked per day, this would yield the Atkinson-Stiglitz (1976) result of redundancy of marginal tariffs (except for pigouvian ones). However, in our model this result does not hold. We assume $\Omega(.)$ and $\phi(.)$ to be increasing and concave. As for $\gamma(.)$, it may be increasing or decreasing in D and B. Transport trips, though necessary for commuting, may provide some utility to the individual (which could be interpreted as an additional purpose of the trip, such as escorting kids to school, i.e. "trip chaining"), or disutility (e.g. stress). In this we follow Parry and Bento (2001) and Van Dender (2003).

The objective of the government is to maximize the social welfare function

$$W = \sum_{i=1,2} \delta^i U^i$$

where δ^i are positive weights, with the normalization $\sum_{i=1,2} \delta^i = 1$. We impose no a priori restriction on the instruments it may use, except from the information at its disposal. Assuming individual's income to be observable, the government has access to a nonlinear income tax schedule. As for transport trips, we are going to study two alternative scenarios. In the first, for each type $i = 1, 2$, individual trip quantities D^i and B^i can be observed by the government, This is crucial for implementability of nonlinear transport tariffs. While obstacles to the use of nonlinear tariffs exist, in reality road and public transport pricing schemes are not rarely nonlinear, at least to some extent. From a technological standpoint at least, observability of individual trip quantities seems feasible.[12] Moreover, in many countries information on commuting travel is collected directly by employers (and passed on to governments) to be used ex-post as the basis to compute commuting subsidies (in the form of discounts on transport tariffs or tax deductions and rebates). Indeed, the pricing schedules we discuss below can be interpreted as resulting also from those subsidies. Nonetheless, we also consider a second scenario, in which only aggregate trip quantities are observable and transactions are anonymous. In that case, only linear tariffs are feasible.

3 Optimal transport tariffs

3.1 Nonlinear transport tariffs

3.1.1 Government's maximization problem

When, on top of income I, individual consumption of D and B can be observed, the design of nonlinear transport tariffs is essentially akin to that of nonlinear commodity (as well as income) taxes.[13] We begin by rewriting the utility function of a given type in terms of

[12] Urban road pricing schemes usually involve the use of electronic systems allowing to track each car's access to the tolled road. As for public transportation, many cities have adopted the use of smart cards (e.g. the Oyster Card in London or the Passe Navigo in Paris) which require personal registration and allow to keep track of trips taken.

[13] As for C, with observable income, if transport trips are observable then individual's consumption level is observable as well.

observable quantities. We also saturate the time constraint and replace for x, so

$$U^i = \Omega(C^i) + \gamma\left(D^i, B^i\right) + \phi\left(1 - a_D D^i - a_B B^i - (D^i + B^i) g\left(\frac{I^i}{w^i(D^i + B^i)}\right)\right) \quad i = 1, 2$$

U^i is type-specific since, for a given allocation, it depends on w_i. We proceed as if the government directly chose allocations, for each type of individual, of C, D, B and I. This follows the Taxation Principle (Stiglitz (1982)). The government's problem is

$$\max_{\{C^i, D^i, B^i, I^i\}} \quad W$$

subject to the budget constraint

$$\sum_{i=1,2} \pi_i \left(I^i - C^i - D^i - c_B B^i\right) \geq R \tag{1}$$

(where R is an exogenous revenue requirement) and, assuming only one self selection constraint is relevant (this is a reasonable assumption in a two-type setup like ours)

$$U^2 \geq U^{21} \tag{2}$$

where

$$U^{21} = \Omega(C^1) + \gamma\left(D^1, B^1\right) + \phi\left(1 - a_D D^1 - a_B B^1 - (D^1 + B^1) g\left(\frac{I^1}{w^2(D^1 + B^1)}\right)\right)$$

is the utility of a high ability type mimicking a low ability one. Constraint (2) tells us that the optimal allocations designed by the government have to be such that individuals of high earning ability do not chose the "bundle" (of income, travel quantities and consumption) intended for low ability ones. Note that when mimicking, an high ability type will need to work less while earning the same income and consuming the same amount of C, D and B as the type she mimics. In this framework, mimickers commute to work the same number of days as the mimicked, but provide less hours of work per day (or daily effort).

To implement the optimal allocation, the government sets nonlinear tariffs for the transport network (i.e. road and public transport pricing schedules) as well as nonlinear income taxes. More precisely, the government designs a general general tax function $\Theta(C, D, B, I)$ based on all observable quantities. In what follows, (as is customary in the literature) we will focus on marginal (i.e. per-trip) tariffs t_D^i, t_B^i (the marginal tax on income t_I is presented in the Appendix).[14] We assume, without loss of generality, that good C is untaxed. The

[14] This is a slight abuse of notation, since they are part of nonlinear schedules, which, a priori, depend on

Lagrangian of the government's problem is

$$\mathcal{L} = W + \mu \left(\sum_{i=1,2} \pi_i (I^i - C^i - D^i - c_B B^i) - R \right) + \lambda \left(U^2 - U^{21} \right)$$

The first order conditions of this problem are provided in the Appendix.

It is useful to illustrate the adjustment in labor supply induced by a marginal change in the number of workdays (i.e. commuting trips). For a given type, the latter writes as

$$m_i = \frac{\partial L^i}{\partial N^i} = -g' \left(\frac{I^i}{w^i(D^i + B^i)} \right) \cdot \left(\frac{I^i}{w^i(D^i + B^i)} \right) + g \left(\frac{I^i}{w^i(D^i + B^i)} \right) \quad i = 1, 2$$

Note that $m_i < 0$ due to convexity of $g(.)$. Given that hours per day at the workplace have diminishing returns, marginally increasing the number of commuting days (i.e. trips) brings the individual, for a given income, to reduce total labor supply. This is an interesting feature of our model, that comes from the fact that we allow the choice not only of working days, but also of daily labor supply. We have also

$$m_{21} = -g' \left(\frac{I^1}{w^2(D^1 + B^1)} \right) \cdot \left(\frac{I^1}{w^2(D^1 + B^1)} \right) + g \left(\frac{I^1}{w^2(D^1 + B^1)} \right)$$

as the adjustment for the individual of type 2 mimicking an individual of type 1. It is easy to see that $m_1 < m_{21} < 0$, as $w^2 > w^1$. Given their smaller daily effort, mimickers can substitute hours worked for days at the workplace suffering smaller productivity losses than low skilled types. This has relevant implications for optimal tariffs.

3.1.2 Benchmark

As a benchmark, consider the ideal case in which the government observes w, L, or both. Then $\lambda = 0$. The optimal per-trip tariffs are

$$t_D^1 = t_D^2 = \tau_D \equiv \frac{\partial \varphi^D}{\partial \bar{D}} \sum_{i=1,2} \pi_i \left(D^i \frac{\phi_x^i}{\Omega_C^i} \right) \quad t_B^i = c_B \quad i = 1, 2$$

Let us begin from tariffs for car trips t_D: they should consist simply of a Pigouvian tax. Their only component is τ_D, the marginal external cost of a trip. This is given by the increase

all quantities observed. For instance

$$t_D^i \equiv \frac{\partial \Theta(C^i, D^i, B^i, I^i)}{\partial D^i}$$

The general tax function may also include lump-sum tax/transfers, as well as "fixed" components of transport tariffs (e.g. the fixed part of a two part tariff).

in time of journeys (on aggregate) due to additional congestion on the road, weighted by the individuals' the marginal rate of substitution between leisure and consumption $\frac{\phi_x^i}{\Omega_C^i}$, for $i = 1, 2$. Such ratio provides a measure of the individual's valuation of time.[15] Tariffs for public transportation t_B should be equal to the marginal cost of providing the trip, c_B. Thus, in the presence of optimal income taxation, and if self selection constraints are not relevant, optimal tariff schedules should not deviate from the marginal social cost of a trip. This is because the government can use differentiated lump sum taxes to redistribute welfare and cover the eventual fixed costs of service provision.

3.1.3 Optimal marginal tariffs with binding self-selection constraints

Consider now the case in which w and L are unobservable and the self selection constraint binds, so $\lambda > 0$. The following holds.

PROPOSITION 1: *When nonlinear transport tariffs are feasible, the optimal per-trip tariffs for cars and public transport $t_j^i \quad i = 1, 2 \quad j = D, B$ are*

$$t_D^1 = \tau_D + \eta_D + z_D \quad t_D^2 = \tau_D + \eta_D \quad t_B^1 = c_B + z_B \quad t_B^2 = c_B$$

where

$$\eta_D = \frac{\lambda}{\mu} \frac{\partial \varphi^D}{\partial \bar{D}} \left(D^1 \Omega_C^{21} \left(\frac{\phi_x^1}{\Omega_C^1} - \frac{\phi_x^{21}}{\Omega_C^{21}} \right) \right)$$

$$z_j = \frac{\lambda}{\mu} \frac{\Omega_C^{21}}{\pi_1} \left(a_j \left(\frac{\phi_x^1}{\Omega_C^1} - \frac{\phi_x^{21}}{\Omega_C^{21}} \right) + \left(m_1 \frac{\phi_x^1}{\Omega_C^1} - m_{21} \frac{\phi_x^{21}}{\Omega_C^{21}} \right) \right) \qquad j = D, B$$

For both cars and public transport, marginal tariff formulae are different than in the benchmark case. They contain additional "incentive" terms, whose role is to improve screening of types (in spite of leisure-goods separability in preferences). Their presence depends on travel being time consuming and (partially) complementary to labor supply, due to diminishing returns in daily hours.

Let us focus first on t_D. While the marginal tariff (e.g. a road toll) t_D^2 intended for high skilled individuals contains only two terms (on which we comment below), the tariff intended for individuals of low ability t_D^1 carries the additional incentive component z_D. As is quite customary in these models, there is an additional distortion to low types' use of cars that the government optimally introduces in order to improve screening of types. this is obtained

[15] There is a large literature on the value of time in transportation (Jara-Diaz (2008)). Generally, it corresponds to the wage rate corrected for the additional utility (or disutility) of time spent on travel, in monetary terms. In our model, a unit of time at work and on travel have the same opportunity cost in terms of foregone leisure.

by raising (or lowering, depending on the sign of z_D) the marginal tariff they face. It is interesting to look at what determines its direction. This depends on the trade-off between daily productivity (at given income) and commuting time that we introduced above. To illustrate, assume $z_D = 0$, so all individuals pay the same (marginal) road toll. Suppose now the government decided to raise the toll intended for low ability types, thereby increasing their commuting costs. By commuting less, low types (and high types who wanted to mimic them) would save time otherwise spent on travel. This is captured by the first term in parenthesis in z_D: the time cost of a trip by car a_D, multiplied by difference of marginal valuation of time for low ability type and high ability mimicker. For a mimicker time is less valuable, at the margin, than for a low ability individual (all else given, the mimicker needs to work less and has more free time). Indeed, $\frac{\phi_x^1}{\Omega_C^1} - \frac{\phi_x^{21}}{\Omega_C^{21}} > 0$. As a consequence, the toll increase would hurt the low type less than the mimicker, if there was no change in labor supply. However, commuting less often (i.e. adopting a work schedule with less workdays, but of greater length, at given income) increases total labor supply. This is particularly true for low skilled types: their daily effort is larger than the mimicker's. Hence, they stand to lose more, in productivity terms, by having to further increase daily hours (recall that $m_1 < m_{21} < 0$). The sign of z_D depends on which of these two effects has the greater magnitude. If the increase in labor supply is higher (resp. lower) than the time cost of the trip itself, the tariff raise hurts low ability types more (less) than high ability mimickers. As a consequence, the self-selection constraint would be tightened (resp. relaxed). Loosely speaking, suppose (all else given) an additional workday lets the individual reduce total labor supply and the commuting trip by car is not too lengthy. Then, it is optimal to have low skill individuals pay a smaller marginal tariff for car trips than high skill types (and vice-versa).[16]

The formulae for t_D also contain two non-type specific terms. The first is the τ_D "pigouvian" term described above. The second is η_D: this is strictly positive and accounts for how a reduction in road congestion can foster screening of types. The reason is that a marginal reduction in a_D is always going to benefit low ability individuals more than high ability mimickers, whose time is less valuable for them at the margin. Unlike in the benchmark case, the marginal external cost of a trip is not only the "classic" pigouvian one but has to include the extra cost of congestion in making redistribution less effective. The incentive effect of public goods (or bads, as in this case), in the presence of nonlinear income taxation, has been previously analyzed by Boadway and Keen (1993), Pirttilä and Tuomala (1997) and,

[16] It is not easy to say, a priori, which of the two effects above is of greater relevance. To fix ideas, consider two extreme cases. Suppose, first, that the time cost of a car trip were negligible, so $a_D \to 0$. Then, we would have $z_D < 0$. Suppose, instead, that daily hours had constant returns, so that the length of the working day does not affect productivity ($g'' = 0$). As long as the time cost of a car trip is non-negligible, so $a_D > 0$, we would have $z_D > 0$.

more recently, Kreiner and Verdelin (2012). However, in the setup of the first two papers, this effect does not survive if individuals have separable preferences for goods and leisure, which we assume. Kreiner and Verdelin pointed out that such an effect exists as long as there is positive correlation between an individual's ability and her willingness to pay for a public good, at a given income and consumption bundle. This is indeed the case here since individuals have to allocate time to labor, leisure *and* travel. By relaxing the time constraint at the individual level, reductions in road congestion benefit more mimicked than mimickers. This is interesting from a policy perspective because it means that redistribution provides an additional motive to raise tariffs (for all types), in order to curb network congestion.[17]

It is also interesting to note that the optimal marginal tariff for high ability types is strictly higher than a standard pigouvian tax. Note, also, that η_D is larger the more low ability types commute by car, D^1.[18]

Finally, we can look at public transport tariffs t_B: marginal tariffs for low ability types t_B^1 also carry the component z_B, whose nature is similar to z_D discussed above. Except, of course, that the relevant time cost of a trip is a_B. It is however interesting to note that the formulae for z_j only differ in the per trip time cost a_j. Hence, the extent to which the government wants, for redistributive reasons, to encourage low income individuals to use of a given mode (by lowering the marginal tariff they face) is generally larger the smaller the time cost of travel on that mode. Hence, if travel by car is less costly in time terms than by public transport (assuming the distance to be traveled is invariant), it is more desirable to discount tariffs for cars than for public transport.

3.2 Optimal linear transport tariffs

Let us now consider the case in which individual trip quantities are not observable and all transactions are anonymous. Then, the government has to design a "mixed" tax system with nonlinear income taxes and linear tariffs for transportation.

[17] A similar effect would be observed if we allowed the government also to control investment in infrastructure: as long as greater network capacity allows, all else given, to reduce travel times, its provision can also produce a positive redistributive effect.

[18] Previous analyses of transport pricing with redistributive concerns (see, e.g., Mayeres and Proost (1997)) have suggested that (linear) tariffs on a given mode should be higher when the mode is used to a large extent by high income individuals. In our model, this does not apply. High income (and ability) types pay, with respect to the pigouvian tax, a "premium" which is higher the more low income types use cars. Moreover, it is not necessarily the case that high ability/income types should be charged more, for a trip, than low ability/income ones.

3.2.1 Government's maximization problem

We proceed, following Cremer *et al.* (1998), under the assumption that the government designs an optimal revelation mechanism consisting of a set of type-specific before-tax incomes I^i, disposable incomes y^i (expenditures on consumption and travel) and a vector of transport tariffs $\mathbf{t} = (t_D, t_B)$, which are akin to commodity taxes. Equivalently, the mechanism designs trip prices $\mathbf{q} = (q_D, q_B)$ where $q_D = 1 + t_D$ and $q_B = t_B$. Again, without loss of generality, we assume C is untaxed. The mechanism assigns the bundle $(\mathbf{q}, y^i, I^i)$ to an individual that reports type $i = 1, 2$. The couple (y^i, I^i) is such that $I^i - T(I^i) = y^i$, where $T(I)$ is the income tax schedule. Given prices and disposable income, the individual decides consumption and travel quantities. That is, given $(\mathbf{q}, y^i, I^i)$, a type-i individual solves

$$\max_{C,D,B} \quad U^i(C, D, B, y, I) \quad i = 1, 2$$

(note that the utility function $U^i(C, D, B, y, I)$ is type specific because, at a given allocation, it depends on w_i) subject to the budget constraint

$$C + q_D D + q_B B = y$$

We denote the resulting conditional demand functions as

$$D^i = D^i\left(\mathbf{q}, y^i, I^i\right) \quad B^i = B^i\left(\mathbf{q}, y^i, I^i\right) \quad C^i = C^i\left(\mathbf{q}, y^i, I^i\right)$$

again, demands are type specific (given $\mathbf{q}, y^i, I^i$) since utility depend on w_i. We denote the (type-specific) indirect utility function as $V^i(\mathbf{q}, y^i, I^i) = U^i(D^i, B^i, C^i, y^i, I^i)$. Finally, we define

$$D^{21} = D^2\left(\mathbf{q}, y^1, I^1\right) \quad B^{21} = B^2\left(\mathbf{q}, y^1, I^1\right)$$
$$C^{21} = C\left(\mathbf{q}, y^1, I^1\right) \quad V^{21}\left(\mathbf{q}, y^1, I^1\right) = U^2\left(D^{21}, B^{21}, C^{21}, y^1, I^1\right)$$

as demands and indirect utility function for a mimicker. Once again, given the presence of only two types in our setup, we can safely focus only on cases in which high ability types want to mimick low ability ones. The government's problem is[19]

$$\max_{\mathbf{q}, y^i, I^i, \bar{D}} \quad \sum_{i=1,2} \delta^i V^i$$

[19] It is convenient to solve this problem assuming that the government also decides on the amount of road congestion $\bar{D}$.

subject to the budget constraint

$$\sum_{i=1,2} \pi_i \left(I^i - y^i + t_D D^i + (t_B - c_B) B^i \right) \geq R \tag{3}$$

and the self-selection constraint

$$V^2 \geq V^{21} \tag{4}$$

we still denote by μ and λ the Lagrange multipliers for these constraints. The solution to this problem is presented in the Appendix.

3.2.2 Benchmark

With no self-selection constraints binding, so $\lambda = 0$, optimal tariffs are $t_D = \frac{\partial \varphi^D}{\partial \bar{D}} \sum_{i=1,2} \pi_i \frac{\phi_x^i D^i}{\partial V^i / \partial y^i}$ and $t_B = c_B$. As in the previous section, they have no redistributive role. [20]

3.2.3 Optimal transport tariffs with binding self-selection constraints

Consider now the case in which w and L are unobservable and the self selection constraint binds, so $\lambda > 0$. The following holds.

PROPOSITION 2: *When the government is constrained to use linear transport tariffs, the optimal tariffs $t_j \quad j = D, B$ satisfy the following*

$$\begin{pmatrix} t_D - \varepsilon \\ t_B - c_B \end{pmatrix} = A^{-1} \cdot \begin{pmatrix} \frac{\lambda}{\mu} \frac{\partial V^{21}}{\partial y_1} \left(D^1 - D^{21} \right) \\ \frac{\lambda}{\mu} \frac{\partial V^{21}}{\partial y_1} \left(B^1 - B^{21} \right) \end{pmatrix}$$

where

$$A = \begin{pmatrix} \sum_{i=1,2} \pi_i \frac{\partial \tilde{D}^i}{\partial q_D} \chi & \sum_{i=1,2} \pi_i \left(\frac{\partial \tilde{B}^i}{\partial q_D} + \frac{\partial \tilde{B}^i}{\partial a_D} \cdot \frac{\partial \varphi^D}{\partial D} \cdot \frac{\partial \tilde{D}^i}{\partial q_D} \chi \right) \\ \sum_{i=1,2} \pi_i \frac{\partial \tilde{D}^i}{\partial q_B} \chi & \sum_{i=1,2} \pi_i \left(\frac{\partial \tilde{B}^i}{\partial q_B} + \frac{\partial \tilde{B}^i}{\partial a_D} \cdot \frac{\partial \varphi^D}{\partial D} \cdot \frac{\partial \tilde{D}^i}{\partial q_B} \chi \right) \end{pmatrix}$$

$$\varepsilon = \frac{\partial \varphi^D}{\partial \bar{D}} \left(\sum_{i=1,2} \pi_i \frac{\phi_x^i D^i}{\partial V^i / \partial y^i} + \frac{\lambda}{\mu} \frac{\partial V^{21}}{\partial y_1} \left(\frac{\phi_x^1 D^1}{\partial V^1 / \partial y^1} - \frac{\phi_x^{21} D^{21}}{\partial V^{21} / \partial y^1} \right) \right) \qquad \chi = \frac{1}{1 - \frac{\partial \varphi^D}{\partial \bar{D}} \sum_{i=1,2} \pi_i \frac{\partial \tilde{D}^i}{\partial a_D}} < 1$$

where $\tilde{D}^i$ and $\tilde{B}^i$ denote hicksian demands for, respectively, car and public transport travel. χ is a feedback term that stands for the net effect of a change in prices on the demand for car trips, after accounting for the change in road congestion.

[20] We here write the value of time as $\frac{\phi_x^i}{\partial V^i / \partial y^i} \quad i = 1, 2$ since this form is more convenient for solving the problem below. With no binding self-selection constraints and nonlinear income taxation, one has $\partial V^i / \partial y^i = \mu = \Omega_C^i \; i = 1, 2$. So the benchmark value of t_D is the same as that of Section 3.1.

The structure of optimal tariffs is affected by two "incentive" terms. Their presence is not novel. First, a reduction in road congestion affects the self-selection constraint in a similar way as in the case of nonlinear tariffs (compare the second component of ε above and term η_D in Proposition 1). It is not possible to precisely determine what direction the effect takes here. This is because, when tariffs are linear, mimicker and mimicked do not necessarily commute the same number of times to the workplace (even if they face the same budget constraint). Nonetheless, except if the mimicker drives much more than the mimicked, it is reasonable to expect the sign of this term to be positive. The second incentive term (right hand side of the equalities in the proposition) is positive if and only if mimickers use more the given mode (car or public transport) than mimicked. Again, this cannot be immediately determined. To get some more insight, we will now present a simplified example with a single travel mode. It will show that the signs of the incentive terms just described may depend crucially on the trade-off between time cost of commuting trips and labor supply changes when days at the workplace are substituted with more hours per day.

A single mode example Consider the case in which cars are the only travel mode (we could focus on public transport with similar outcomes). Then the optimal tariff is simply

$$t_D = \varepsilon + \frac{1}{\sum_{i=1,2} \pi_i \frac{\partial \tilde{D}^i}{\partial q_D} \chi} \left(\frac{\lambda}{\mu} \frac{\partial V^{21}}{\partial y_1} \left(D^1 - D^{21} \right) \right) \Longrightarrow D^{21} \gtreqless D^1 \Leftrightarrow t_D \gtreqless \varepsilon$$

In such a simplified setup, the budget constraint of mimicker and mimicked is the same. The only difference between them, at a given (D_0, C_0) couple, is the I/w ratio (and, given this, the number of daily work hours). Thus, whether a mimicker drives more than a mimicked depends simply on whether her indifference curves in the (D, C) plane are flatter than those of a mimicked. That is

$$D^{21} \gtreqless D^1 \Leftrightarrow \sigma \left(D, C; I^1/w^2 \right) \gtreqless \sigma \left(D, C; I^1/w^1 \right)$$

at a given allocation. The slope of an indifference curve, computed at a given allocation (D_0, C_0), is

$$\sigma \left(D_0, C_0; I/w \right) = -\frac{\partial U/\partial D}{\partial U/\partial C} = -\frac{\gamma_D \left(D_0 \right) + \phi_x \cdot \left(-a_D - m \right)}{\Omega_C \left(D_0 \right)}$$

where

$$m = g \left(\frac{I}{w D_0} \right) - g' \left(\frac{I}{w D_0} \right) \cdot \left(\frac{I}{w D_0} \right) < 0$$

taking the derivative of σ with respect to I/w, one obtains

$$\frac{\partial\sigma}{\partial\left(I/w\right)}=-\frac{\phi_{xx}\cdot g'\left(\frac{I}{wD_0}\right)\cdot\left(-a_D-m\right)-\phi_x\cdot g''\left(\frac{I}{wD_0}\right)\cdot\left(\frac{I}{wD_0^2}\right)}{\Omega_C}$$

The sign depends on the trade-off between days at the workplace and commuting trips that drives the difference between marginal tariffs when nonlinear pricing is feasible (see term z_j in Proposition 1). Indeed, when the time cost of a commuting trip is larger (resp. smaller) than the reduction in labor supply with more day on-the-job, then $\frac{\partial\sigma}{\partial(I/w)}<0$ (resp. >0). Therefore, $D^{21}<D^1$ (resp. $>D^1$).

Since the sign of $\frac{\partial\sigma}{\partial(I/w)}$ is not immediately determined, it is once again useful to look at two extreme cases. Suppose that the time cost of a trip were negligible, so $a_D\to 0$ while $f''<0$ (and $g''>0$) . In that case, $\frac{\partial\sigma}{\partial(I/w)}>0$ so $\sigma\left(D,C;I^1/w^2\right)<\sigma\left(D,C;I^1/w^1\right)$. Then, $t_D<\varepsilon$. Moreover, if a low ability type drives more than a mimicker, the incentive term in ε is will certainly be positive. Suppose, instead, that hours worked per day had constant returns, so $f''=g''=0$, and the time cost of a car trip were non-negligible, $a_D>0$. In that case, m would be equal to zero, so $\sigma\left(D,C;I^1/w^2\right)>\sigma\left(D,C;I^1/w^1\right)$ and $t_D>\varepsilon$.

3.3 Telework

The model presented above can easily be adapted to consider telework (i.e. work done outside the standard workplace, e.g. at home), with little impact on the results. It is generally recognized that telework has the potential to ease the pressure on transport networks in peak hours, by reducing travel demand. However, it may also lead to lower productivity than work done while physically on the job, as coordination with colleagues and supervisors (or supervisees) is more difficult. Also, monitoring of work safety and data protection is more complicated.

If indeed there is a productivity penality for telework, one more day at the workplace, for given income, reduces total labor supply. However, it requires a time-consuming commuting trip. Whether commuting should thus be more or less encouraged than for high ability ones depends on a trade-off that is essentially the same as in the case of choice of workday lenght.[21] This is why, in terms of optimal pricing schedules (in the presence of optimal income taxes), the results would not differ from those derived above.

Let us sketch how the model could be adapted to include telework. Denote by s the number of days worked outside the workplace. In order to neatly identify the trade-off

[21] De Borger and Wuyts (2011) study a model with telework, though without looking at distributional concerns.

between working off and at the workplace, we assume a fixed lenght for working days and normalize such lenght to unity. An amount N of days at the workplace provides $w_i N$ $i = 1, 2$ units of income. Each requires a commuting trip. An amount s of days off the workplace brings instead $w_i f(s)$ $i = 1, 2$ units of income, where f is an increasing and concave function, with $f(0) = 0$, $f'(0) \leq 1$. Therefore, $I^i = w^i\left(N^i + f(s^i)\right)$. Concavity of f captures increasing losses in productivity as more days of telework replace days at the workplace. With such a setup, we have $s_i = g\left(\frac{I}{w^i} - N_i\right)$ with g being the inverse of f, thus increasing and convex. Then $m_i = 1 - g'\left(\frac{I_i}{w^i} - N_i\right)$, so $m_1 < m_{21} < 0$.

4 Implementation of optimal tax and tariff schedules

The question we investigate now is whether means-testing is a useful tool for a redistribution-minded government designing both transport tariffs and income taxes. This responds to some questions raised in the policy debate on reforming transport pricing (see the Introduction). If only linear tariffs are feasible and all transactions anonymous, the question is moot. This is why we focus on the case in which the government can use nonlinear tariff schedules.

In Section 3.1, we have assumed that the government implements the second-best allocation (defined as A^{SB}) using a generalized tax-and-tariff function $\Theta(C, D, B, I)$. This means that, a priori, it may have to design tariff schedules for transportation that are differentiated according to income. Moreover, the income tax schedule may have to depend on commuting trips. Following Cremer and Gahvari (2002), we are now going to study whether using an income tax function $T(I)$ and a separate transport tariff schedule $P(D, B)$ is enough to implement A^{SB}.[22] If such a thing is feasible, then transport tariffs do not need to be means-tested.

The government looks to implement the second-best allocation

$$A^{SB} = \left((C^1, D^1, B^1, I^1); (C^2, D^2, B^2, I^2)\right)$$

that solves the problem presented in Section 3.1, using the functions $T(I)$ and $P(D, B)$. T^i and P^i $i = 1, 2$ denote respectively the payments of income taxes and transport tariffs for individuals of type 1 and 2. Therefore $C^i = I^i - (T^i + P^i)$ $i = 1, 2$. Incentive compatibility of the tax and tariff schedules calls for types 1 and 2 to choose quantities and payments

[22] The setup of our problem is similar to that of Cremer and Gahvari. Our results are different. The reason is that, even with separable preferences, consumption of transport trips affects the marginal utility of leisure. This makes implementation with separable functions more difficult to achieve, as labor supply (and income) and consumption decisions cannot be separated. In addition, we do not assume any difference in tastes between individuals of different ability. Finally, our problem is of greater complexity due to the presence of two goods that the government has to price.

$((T(I^1));(P(D^1+B^1)))$ and $((T(I^2));(P(D^2+B^2))$ respectively. The increased complexity stems from the fact that individuals have additional possibilities to deviate from the "bundle" designed for them. For instance, they may choose to consume a quantity of trips $D+B$ intended for the other type, while choosing the amount of I intended for them. Or they could choose to mimick the other's type income, while consuming the "right" amount of $D+B$. Therefore, in order to be implementable through separable payment functions, A^{SB} has to respect the "standard" incentive compatibility constraint (2), the government's budget constraint (**??**), plus *four additional incentive constraints* ensuring domination of "partial" mimicking strategies (each of them for $i=1,2 \quad \tilde{i}\neq i$):

$$\Omega(I^i-T^i-P^i)+\gamma(D^i,B^i)+\phi(1-a_D D^i-a_B B^i-\left(D^i+B^i\right)g\left(\frac{I^i}{w^i\left(D^i+B^i\right)}\right))\geq$$
$$\Omega(I^{\tilde{i}}-T^{\tilde{i}}-P^i)+\gamma(D^i,B^i)+\phi(1-a_D D^i-a_B B^i-\left(D^i+B^i\right)g\left(\frac{I^{\tilde{i}}}{w^i\left(D^i+B^i\right)}\right))\quad (5)$$

$$\Omega(I^i-T^i-P^i)+\gamma(D^i,B^i)+\phi(1-a_D D^i-a_B B^i-\left(D^i+B^i\right)g\left(\frac{I^i}{w^i\left(D^i+B^i\right)}\right))\geq$$
$$\Omega(I^i-T^i-P^{\tilde{i}})+\gamma(D^{\tilde{i}},B^{\tilde{i}})+\phi(1-a_D D^{\tilde{i}}-a_B B^{\tilde{i}}-\left(D^j+B^j\right)g\left(\frac{I^i}{w^i\left(D^{\tilde{i}}+B^{\tilde{i}}\right)}\right))\quad (6)$$

The first two ensure that an individual of type i will not, while choosing the number of transport trips intended for his type, choose income level intended for the other type ("partial mimicking" on income). The second set of constraints ensures an individual of type i , while choosing the income intended for his type, will not mimick the other on transport trips. The solution of this problem is provided in the Appendix.

We now provide a sufficient condition under which using separable functions $T(I)$ and $P(D,B)$ is enough to implement A^{SB}. As long as the condition holds, means-testing is not required.

PROPOSITION 3: *Assume that the government wanted to implement the second best allocation A^{SB} using a separate payment schedule for income $T(I)$ and transportation $P(D,B)$. Then a sufficient condition for A^{SB} to be implementable is that it satisfies $D^2+B^2\geq D^1+B^1$ and $a_D D^1+a_B B^1\geq a_D D^2+a_B B^2$*

The condition requires that high ability/income households travel more, but their total travel time is smaller than for the others. This can be the case if high income households commute more by cars than low income ones, while public transport trips have larger time costs than trips by car.[23]

In the numerical examples below, the condition given in Proposition 5 generally holds. In fact, even when it fails, we find no counterexample in which implementation with separable functions is unfeasible. We also go one step further. Instead of using functions $T(I)$ and $P(D,B)$, we study whether implementation of A^{SB} can be achieved by complementing the income tax schedule $T(I)$ with *two separate tariff schedules*, $P(D)$ for cars and $Q(B)$ for public transportation. The theoretical problem is similar to the one presented above, but considerably more complex to solve. The volume of conditions to be checked would make treating the problem in an analytical way simply too tedious. This is why we only investigate the issue numerically. The results obtained seem to support the conclusion that implementation is feasible even using fully separable transport tariff functions.

5 Numerical illustration

We present here a numerical example to illustrate the features of the optimal tariff schemes derived above. We are also interested in verifying that conditions for implementability in separable functions, as discussed in Section 5, reasonably hold. In order to focus on these two aspects, we only look at the case of nonlinear tariff schemes and consider fixed network capacities. The examples are based on the following utility and daily productivity functions

$$U(C,D,B,x) = C^{\frac{1}{3}} + 0.05\left(D^{\frac{1}{2}} + B^{\frac{1}{2}}\right) + 3x^{\frac{1}{2}} \qquad f(h) = h^{\frac{5}{6}}$$

and we assume that $\pi_i = \delta^i = \frac{1}{2} \quad i = 1,2$. We also use the following function for time of car trips: $a_D = a + 0.00015\bar{D}$. We are going to describe three scenarios, each characterized by different relative qualities (measured in terms of trip time costs) of cars and public transport. They are obtained by varying the intercept a for car trips, as well as the time cost of public transport trips a_B. In Scenario 1, we have $a = 0.005$ and $a_B = 0.01$, in Scenario 2 $a = 0.003$ and $a_B = 0.015$. Finally, $a = 0.001$ and $a_B = 0.02$ in Scenario 3. Recall that individuals' endowment of time is normalized to one. We fix the monetary cost of a car trip to one and set $c_B = 0.1$ for a public transport trip. In each scenario, w_2 is set at 100 and we vary w_1 from 50 to 90. This produces differences in earned income, at the second-best allocation,

[23]Empirical evidence suggests that travel (and commuting) tend to be increasing in income. This is particularly true for car travel (Hu and Ruscher (2004), Table 32). A modal split such that high income households travel more by car than low income households is, thus, not unlikely. Moreover, the UK Department for Transport reports a value of time for a commuting trip by car, on average, which is about one third of that of a commuting trip by public transport (DfT (2011), Table 9).

that go from the low ability type earning about 25% to about 75% of the (pre-tax) income of the other type. For each scenario, we report individuals' earned income I^i, their amount of travel on each mode and the optimal per-trip tariffs (all computed at the second-best allocation).

Concerning implementability, we refer to Condition I as implementability of the second-best allocation using separable transport tariffs and income taxes, making use, possibly, of a joint payment schedule for cars and public transport. Condition II identifies instead implementability using fully separable transport tariffs (i.e. separate payment schemes for cars and public transport), on top of a separate income tax schedule. For each scenario, we verify whether such conditions hold. Results suggest that implementability can be achieved in many circumstances, even when using three separate payment schedules for cars, public transport and income.

Scenario 1. In the first scenario, public transportation is a good alternative to cars. Good enough, in fact, to have both high and low income individuals make it their main commuting mode. This scenario may represent cities in which public transportation is very effective and the primary commuting mode for most of the population. Fitting examples might be European cities like Zurich and Stockholm. We can see that trip quantities are increasing with income, as individuals supply more labor and need to commute increasingly often. Note, in particular, that as her productivity increases, the low ability type works and commutes more, though always less than the high income type. Due to low road congestion, the pigouvian tax τ_D on car trips is quite small (about 5% of the monetary cost of a car trip). The per-trip tariff t_D^2 is strictly higher than that (though by a small amount), while t_D^1 is smaller. Low ability types pay the smaller per-trip tariff also on public transport. This is because, at the margin, the cost (in terms of lower daily productivity) of reducing the amount of commuting is larger than the time cost of a journey, on both modes (see the expression for the term z in optimal tariffs of Proposition 1). However, the difference between the marginal tariff intended for high and low types is larger for cars than for public transport.[24] This is due to the fact that public transport has higher time costs. Finally, considering implementability of the second-best allocation, the sufficient condition of Proposition 5 fails. Nonetheless, implementability is achievable using fully separable payment functions (i.e. Condition I and II hold), in all cases considered.

[24] The situation in which $w_1 = 50$ is an exception only because t_D^1 is constrained to be nonnegative.

SCENARIO 1			a_D=0.005+0.0015D			a_B	0.01			*Implementability*	
W_1=50				a_D	0.005	τ_D	0.048				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
23.38	93.03	0.493	3.650	1.844	11.886	0.048	0.000	0.100	0.031	OK	OK
W_1=66				a_D	0.005	τ_D	0.053				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
45.24	89.76	0.656	9.046	2.112	9.978	0.053	0.002	0.100	0.058	OK	OK
W_1=75				a_D	0.005	τ_D	0.057				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
56.97	88.42	0.908	9.580	2.072	9.877	0.058	0.025	0.100	0.074	OK	OK
W_1=90				a_D	0.005	τ_D	0.063				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
66.99	87.55	1.179	9.824	2.021	9.834	0.063	0.046	0.100	0.087	OK	OK

Scenario 2. Compared to Scenario 1, we consider here a situation in which the car, though more expensive, is significantly more attractive than public transportation. As a consequence, modal split is such that public transport is popular only among low income individuals, while the others mostly travel by car (except in the case in which $w_1 = 90$ and earning abilities are very similar). The reason is that low income types work less than the others *in equilibrium* (this is optimal given their lower productivity), are less time constrained and can better cope with a more time-consuming (but cheaper) travel mode. The higher volume of car trips implies the pigouvian tax τ_D is at about 4 times higher than in Scenario 1. Once again, optimal per-trip tariffs are smaller when intended for low than for high ability types, with the difference being larger for cars than for public transport. Implementability in separable functions is achievable in all the cases presented. The sufficient condition of Proposition 5 holds, except in case $w_1 = 50$. In that case, however, it is impossible to implement the second-best allocation with separate tariff schedules for cars and public transport (as long as they do not depend on income). Implementation is feasible, instead, if a joint transport tariff scheme (independent of income) is used.

SCENARIO 2			a_D=0.003+0.0015D			a_B	0.015			*Implementability*	
W_1=50				a_D	0.004	τ_D	0.184				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
19.99	96.56	0.443	4.076	8.936	1.681	0.184	0.067	0.100	0.001	OK	NO
W_1=66				a_D	0.004	τ_D	0.188				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
41.35	92.67	1.444	4.984	8.432	1.780	0.188	0.107	0.100	0.039	OK	OK
W_1=75				a_D	0.005	τ_D	0.198				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
53.61	90.60	2.607	4.610	7.980	1.923	0.198	0.146	0.100	0.063	OK	OK
W_1=90				a_D	0.005	τ_D	0.231				
I_1	I_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	*I*	*II*
73.93	87.62	5.069	3.333	7.025	2.324	0.231	0.221	0.100	0.093	OK	OK

Scenario 3. In this scenario, public transport travel is significantly more time consuming than car travel (time cost being more than five times that of a car trip). Cars are thus the

preferred mode by both high and low income households, except in the case in which low income ones earn (and work) much less than the others. This scenario seems consistent with the situation of many car-dependent cities. Fitting examples may be American ones such as Atlanta or Los Angeles. Note, however, that low income types commute to a much smaller extent than their high income counterparts. Optimal tariffs follow similar patterns as in Scenario 2, except that the pigouvian tax for cars is larger, given stronger road congestion. As in Scenario 2, the sufficient condition of Proposition 5 holds in all cases presented, except case $w_1 = 50$. Implementability of the second-best allocation is feasible using separate tariffs and income taxes. This is true except when $w_1 = 50$. In that case, a joint tariff schedule for both transport modes (separate from the income tax schedule) is necessary.

SCENARIO 3				a_D=0.001+0.0015D		a_B	0.020			*Implementability*	
W_1=50				a_D	0.003	τ_D	0.244				
l_1	l_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	I	II
17.94	102.22	0.671	1.071	12.681	0.443	0.244	0.106	0.100	0.013	OK	NO
W_1=66				a_D	0.003	τ_D	0.255				
l_1	l_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	I	II
41.18	97.25	3.979	2.496	10.782	0.597	0.256	0.159	0.100	0.040	OK	OK
W_1=75				a_D	0.003	τ_D	0.276				
l_1	l_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	I	II
55.29	94.34	5.084	1.994	10.094	0.673	0.278	0.216	0.100	0.065	OK	OK
W_1=90				a_D	0.004	τ_D	0.321				
l_1	l_2	D_1	B_1	D_2	B_2	t_D^2	t_D^1	t_B^2	t_B^1	I	II
77.02	91.01	7.632	1.139	9.186	0.805	0.321	0.310	0.100	0.099	OK	OK

6 Concluding remarks

Our findings suggest that transport tariffs can, if properly designed, be used to improve the redistributive capabilities of the tax system. In a nutshell, this is because low ability types and high ability mimickers may have, at the same allocation, different values of time and changes in commuting costs affect their labor supply in different ways. This has led us to results which are perhaps counterintuitive, such as the fact that low income individuals may optimally have to pay higher (marginal) tariffs for using a given mode than high income individuals. Moreover, redistributive concerns may actually provide an additional justification for congestion pricing.

Our results rest, anyway, on some important assumptions. First, we have assumed that the income tax is optimally designed, which may not always be the case in reality. Yet, we have no reason to believe that the results would not stand even if the income tax schedule is suboptimal, as long as it can be flexibly adjusted to account for changes in transportation policy (as in, e.g., Kaplow (2006)).

Second, we have assumed that commuters can respond to increased travel costs by raising daily work hours and ignored other margins of flexibility, such as changing residence or shifting travel to off-peak hours (Arnott *et al.* (1993)). Including the first feature in the model would require modelling also the urban land market, which is out of our scope. Moreover, fixed residence is often assumed in labor economics models studying commuting costs (Gutierrez-i-Puigarnau and van Ommeren (2010)). As for changes in travel times, while they would certainly add depth to the model, we can speculate that they would not significantly affect our results. Indeed, a likely response by commuters to increased peak-hour travel costs (e.g. the introduction of a road toll) would be to leave home earlier and/or stay longer at work. Hence, an increase in commuting costs would increase daily and total labor supply (at given income, at least), as it is already the case in our model.

Finally, we have neglected the presence of multiple government levels (e.g. local and national ones), which may have different powers as well as divergent objectives. [25]

We plan, in future work, to extend our research study to incorporate these features.

[25] We could consider the presence of an additional part of the population living outside the urban area. Assuming these people do not use its transport network (so that they do not care for D and B), fixed residential location, and that tax schedules are flexible enough to be differentiated between people belonging to a given urban agglomeration and those who do not, our results would not change. They would also not change with multiple urban areas and, again, income tax schedules may be differentiated across them.

Appendix

Proof of Proposition 1

The first order conditions of this problem are

$$\frac{\partial \mathcal{L}}{\partial C^1} = \delta^1 U_C^1 - \pi_1 \mu - \lambda U_C^{21} = 0 \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial C^2} = \delta^2 U_C^2 - \pi_2 \mu + \lambda U_C^2 = 0 \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial D^1} = \left(\delta^1 U_D^1 - \lambda U_D^{21}\right) - \pi_1 \mu - \lambda \pi_1 \frac{\partial \varphi^D}{\partial \bar{D}} \left(D_1 \phi_x^{21} - D_2 \phi_x^2\right) - \pi_1 \sum_{i=1,2} \delta^i D^i \frac{\partial \varphi^D}{\partial D} \phi_x^i = 0 \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial D^2} = U_D^2 \left(\delta^2 + \lambda\right) - \pi_2 \mu - \lambda \pi_2 \frac{\partial \varphi^D}{\partial \bar{D}} \left(D_1 \phi_x^{21} - D_2 \phi_x^2\right) - \pi_2 \sum_{i=1,2} \delta^i D^i \frac{\partial \varphi^D}{\partial D} \phi_x^i = 0 \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial B^1} = \delta^1 U_B^1 - \lambda U_B^{21} - c_B \pi_1 \mu = 0 \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial B^2} = U_B^2 \left(\delta^2 + \lambda\right) - c_B \pi_2 \mu = 0 \tag{12}$$

$$\frac{\partial \mathcal{L}}{\partial I_1} = \delta^1 U_I^1 - \lambda U_I^{21} + \pi_1 \mu = 0 \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial I_2} = U_I^2 \left(\delta^2 + \lambda\right) + \pi_2 \mu = 0 \tag{14}$$

where subscripts denote partial derivatives, $U_x^i \equiv \phi_x^i$ is the marginal utility of pure leisure and $U_j^i \equiv \gamma_j^i - (a_j + m^i)\, \phi_x^i$ $j = D, B$ denotes the marginal utility individual $i = 1, 2$ derives from a commuting trip $j = D, B$. This is net of the opportunity cost of trip time, as well as the induced adjustment in labor supply m, at a given income and goods bundle. Take (7), (9) and (11) and rearrange to get to

$$\frac{U_j^1}{U_C^1} = \frac{v_j + \frac{\lambda U_j^{21}}{\mu \pi_1} + \frac{\frac{\partial \varphi^{\mathcal{D}}}{\partial \bar{D}} \frac{\partial \bar{\mathcal{D}}}{\partial Q_j^1} \sum_{i=1,2} \delta^i D_i \phi_x^i}{\mu} - \frac{\lambda}{\mu} \frac{\partial \varphi^{\mathcal{D}}}{\partial \bar{D}} \frac{\partial \bar{\mathcal{D}}}{\partial Q_j^1} \left(D^1 \phi_x^{21} - D^2 \phi_x^2\right)}{1 + \frac{\lambda U_C^{21}}{\mu \pi_1}} \qquad j = D, B$$

where $\frac{\partial \bar{\mathcal{D}}}{\partial Q_j^i} = 1$ if $j = D$ and 0 otherwise, as public transport trips do not contribute to road congestion. Note that $v_j = 1$ if $j = D$ and $v_j = c_B$ if $j = B$. Multiplying both sides by

$1+\frac{\lambda U_C^{21}}{\mu\pi_1}$ and rearranging we get

$$\frac{U_j^1}{U_C^1}=v_j+\frac{\lambda U_C^{21}}{\mu\pi_1}\left(\frac{U_j^{21}}{U_C^{21}}-\frac{U_j^1}{U_C^1}\right)+\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\frac{\partial\bar{\mathcal{D}}}{\partial Q_j^1}\sum_{i=1,2}\delta^i D_i\phi_x^i}{\mu}-\frac{\lambda}{\mu}\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\frac{\partial\bar{\mathcal{D}}}{\partial Q_j^1}\left(D^1\phi_x^{21}-D^2\phi_x^2\right)\quad j=D,B \tag{15}$$

Similarly, using (8), (10) and (12) we get

$$\frac{U_j^2}{U_C^2}=v_j+\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\frac{\partial\bar{\mathcal{D}}}{\partial Q_j^2}\sum_{i=1,2}\delta^i D_i\phi_x^i}{\mu}-\frac{\lambda}{\mu}\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\frac{\partial\bar{\mathcal{D}}}{\partial Q_j^2}\left(D^1\phi_x^{21}-D^2\phi_x^2\right)\quad j=D,B$$

In the optimal allocation, we must have

$$\frac{U_D^i}{U_C^i}=1+t_D^i\ \text{ and }\ \frac{U_B^i}{U_C^i}=t_B^i\quad i=1,2$$

Using these relations, we can obtain the marginal tariff rates t_j^i provided in the Proposition. We now focus on $j=D$ and derive τ_D and η_D. Rewrite

$$\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\sum_{i=1,2}\delta^i D_i\phi_x^i}{\mu}=\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\left(\frac{\delta^1 D_1U_x^1U_C^1}{\mu U_C^1}+\frac{\delta^2 D_2U_x^2U_C^2}{\mu U_C^2}\pm\left(\frac{\lambda D_2U_x^2}{\mu}-\frac{\lambda D_1U_x^{21}}{\mu}\right)\right)$$

now using (7) we have

$$\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}D_1U_x^1\delta^1U_C^1}{\mu U_C^1}=\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}D_1U_x^1\pi_1\mu}{\mu U_C^1}+\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}D_1U_x^1\lambda U_C^{21}}{\mu U_C^1}$$

and using (8) we have

$$\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}D_2U_x^2\delta^2U_C^2}{\mu U_C^2}=\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}D_2U_x^2\pi_2\mu}{\mu U_C^2}-\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}D_2U_x^2\lambda U_C^2}{\mu U_C^2}$$

so that we can rewrite

$$\frac{\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\sum_{i=1,2}\delta^i D_iU_x^i}{\mu}=\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\left(\sum_{i=1,2}\pi_iD_i\frac{U_x^i}{U_C^i}+\frac{\lambda D_1U_x^1U_C^{21}}{\mu U_C^1}-\frac{\lambda D_1U_x^{21}}{\mu}+\frac{\lambda}{\mu}\left(D_1U_x^{21}-D_2U_x^2\right)\right)$$

finally, replacing the above expression in (15) for $j=D$ and rearranging we have

$$\frac{U_D^1}{U_C^1}=1-\frac{\lambda U_C^{21}}{\mu\pi_1}\left(\frac{U_D^1}{U_C^1}-\frac{U_D^{21}}{U_C^{21}}\right)+\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\sum_{i=1,2}\pi_iD_i\frac{U_x^i}{U_C^i}+\frac{\partial\varphi^{\mathcal{D}}}{\partial\bar{D}}\frac{\lambda}{\mu}U_C^{21}D_1\left(\frac{U_x^1}{U_C^1}-\frac{U_x^{21}}{U_C^{21}}\right)$$

and
$$\frac{U_D^2}{U_C^2} = 1 + \frac{\partial \varphi^{\mathcal{D}}}{\partial \bar{D}} \sum_{i=1,2} \pi_i D_i \frac{U_x^i}{U_C^i} + \frac{\partial \varphi^{\mathcal{D}}}{\partial \bar{D}} \frac{\lambda}{\mu} U_C^{21} D_1 \left(\frac{U_x^1}{U_C^1} - \frac{U_x^{21}}{U_C^{21}} \right)$$

where the terms τ_D and η_D as described in the text can be recognized (note that $U_x^i = \phi_x^i \quad U_C^i = \Omega_C^i$). We now focus on z_j $j = D, B$. We can write

$$\frac{\lambda U_C^{21}}{\mu \pi_1} \left(\frac{U_j^{21}}{U_C^{21}} - \frac{U_j^1}{U_C^1} \right) = \frac{\lambda}{\mu \pi_1 \Omega_C^1} \left(\left(\gamma_j^{21} - a_j \phi_x^{21} - m^{21} \phi_x^{21} \right) \Omega_C^1 - \left(\gamma_j^1 - a_j^1 \phi_x^1 - m^1 \phi_x^1 \right) \Omega_C^{21} \right) \quad j = D, B$$

the right hand side can also be written as

$$\frac{\lambda}{\mu \pi_1} \left(\Omega_C^{21} \left(\frac{\gamma_j^{21}}{\Omega_C^{21}} - \frac{\gamma_j^1}{\Omega_C^1} \right) + a_j \left(\phi_x^1 \frac{\Omega_C^{21}}{\Omega_C^1} - \phi_x^{21} \right) + \left(\phi_x^1 \frac{\Omega_C^{21}}{\Omega_C^1} m_1 - \phi_x^{21} m_{21} \right) \right) \quad j = D, B$$

Since $U(.)$ is such that $\frac{\gamma_j^1}{\Omega_C^1} = \frac{\gamma_j^{21}}{\Omega_C^{21}}$ (by separability), the expression above becomes

$$\frac{\lambda \Omega_C^{21}}{\mu \pi_1} \left(a_j \left(\frac{\phi_x^1}{\Omega_C^1} - \frac{\phi_x^{21}}{\Omega_C^{21}} \right) + \left(\frac{\phi_x^1}{\Omega_C^1} m_1 - \frac{\phi_x^{21}}{\Omega_C^{21}} m_{21} \right) \right)$$

which is z_j $j = D, B$ in the text.

Optimal income tax rates

Using (13) and (7) we obtain

$$\frac{U_I^1}{U_C^1} = -1 + \frac{\lambda}{\mu \pi_1} U_C^{21} \left(\frac{U_I^{21}}{U_C^{21}} - \frac{U_I^1}{U_C^1} \right)$$

now, using the fact that

$$U_I^1 = -g' \left(\frac{I_1}{w_1 (D_1 + B_1)} \right) \cdot \frac{\phi_x^1}{w_1} \quad U_I^{21} = -g' \left(\frac{I_1}{w_2 (D_1 + B_1)} \right) \cdot \frac{\phi_x^{21}}{w_2}$$

we have

$$t_I^1 = 1 + \frac{U_I^1}{U_C^1} = \frac{\lambda}{\mu \pi_1} \Omega_C^{21} \left(\frac{g' \left(\frac{I_1}{w_1(D_1+B_1)} \right) \cdot \frac{\phi_x^1}{w_1}}{\Omega_C^1} - \frac{g' \left(\frac{I_1}{w_2(D_1+B_1)} \right) \cdot \frac{\phi_x^{21}}{w_2}}{\Omega_C^{21}} \right) > 0$$

while, using (14) and (8), we have $t_I^2 = 1 + \frac{U_I^2}{U_C^2} = 0$.

Proof of Proposition 2

We solve this problem assuming that the government can directly determine the level of congestion (public bad), denoted $\bar{D}$. When solving the problem, we have thus an additional equality constraint given by $\bar{D} = \sum_{i=1,2} \pi_i D^i$. We denote by β the Lagrange multiplier for this constraint. Thus, the Lagrangian is

$$\begin{aligned}\mathcal{L} =& W + \mu \left(\sum_{i=1,2} \pi_i \left(I^i - y^i + t_D D^i + (t_B - c_B) B^i \right) - \sum_{j=D,B} c_{K_j} K_j - R \right) + \\ & +\lambda \left(V^2 - V^{21} \right) + \beta \left(\bar{D} - \sum_{i=1,2} \pi_i D^i \right)\end{aligned}$$

The first order condition of this problem are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial q_j} &= \delta^1 \frac{\partial V^1}{\partial q_j} + \left(\delta^2 + \lambda\right) \frac{\partial V^2}{\partial q_j} - \lambda \frac{\partial V^{21}}{\partial q_j} + \mu \left[\sum_{i=1,2} \pi_i \left(Q_j^i + (q_D - 1) \frac{\partial D^i}{\partial q_j} + (q_B - c_B) \frac{\partial B^i}{\partial q_j} \right) \right] + \\ & \quad -\beta \sum_{i=1,2} \pi_i \frac{\partial D^i}{\partial q_j} = 0 \quad j = D, B\end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial y^i} = \delta^1 \frac{\partial V^1}{\partial y^i} + \left(\delta^2 + \lambda\right) \frac{\partial V^2}{\partial y^i} - \lambda \frac{\partial V^{21}}{\partial y^i} + \mu \pi_i \left[-1 + (q_D - 1) \frac{\partial D^i}{\partial y^i} + (q_B - c_B) \frac{\partial B^i}{\partial y^i} \right] - \beta \pi_i \frac{\partial D^i}{\partial y^i} = 0 \quad i = 1, 2$$

$$\frac{\partial \mathcal{L}}{\partial I^i} = \delta^1 \frac{\partial V^1}{\partial I^i} + \left(\delta^2 + \lambda\right) \frac{\partial V^2}{\partial I^i} - \lambda \frac{\partial V^{21}}{\partial I^i} + \mu \pi_i \left[1 + (q_D - 1) \frac{\partial D^i}{\partial I^i} + (q_B - c_B) \frac{\partial B^i}{\partial I^i} \right] - \beta \pi_i \frac{\partial D^i}{\partial I^i} = 0 \quad i = 1, 2$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \bar{D}} &= \frac{\partial \varphi^D}{\partial \bar{D}} \left(\delta^1 \frac{\partial V^1}{\partial a_D} + \left(\delta^2 + \lambda\right) \frac{\partial V^2}{\partial a_D} - \lambda \frac{\partial V^{21}}{\partial a_D} + \mu \sum_{i=1,2} \pi_i \left[(q_D - 1) \frac{\partial D^i}{\partial a_D} + (q_B - c_B) \frac{\partial B^i}{\partial a_D} \right] \right) + \\ & \quad +\beta \left(1 - \sum_{i=1,2} \pi_i \frac{\partial D^i}{\partial a_D} \frac{\partial \varphi^D}{\partial \bar{D}} \right) = 0\end{aligned}$$

note that

$$\frac{\partial V^i}{\partial a_D} = -\phi_x^i D^i \quad i = 1, 2 \quad \frac{\partial V^{21}}{\partial a_D} = -\phi_x^{21} D^{21}$$

To start, we are going to focus on $\frac{\partial \mathcal{L}}{\partial \bar{D}}$. Add $\lambda \frac{\partial V^{21}}{\partial y^1}\left(\frac{\partial V^1}{\partial a_D}/\frac{\partial V^1}{\partial y^1}\right)\frac{\partial \varphi^D}{\partial \bar{D}}$ to both sides and rearrange to get

$$\frac{\partial \varphi^D}{\partial \bar{D}}\left(\left(\delta^1 \frac{\partial V^1}{\partial y^1}-\lambda \frac{\partial V^{21}}{\partial y^1}\right)\left(\frac{\partial V^1}{\partial a_D}/\frac{\partial V^1}{\partial y^1}\right)+\lambda \frac{\partial V^{21}}{\partial y^1}\left(\frac{\partial V^1}{\partial a_D}/\frac{\partial V^1}{\partial y^1}-\frac{\partial V^{21}}{\partial a_D}/\frac{\partial V^{21}}{\partial y^1}\right)\right)+$$
$$+\frac{\partial \varphi^D}{\partial \bar{D}}\left(\left(\delta^2+\lambda\right)\frac{\partial V^2}{\partial y^2}\left(\frac{\partial V^2}{\partial a_D}/\frac{\partial V^2}{\partial y^2}\right)+\left(\mu \sum_{i=1,2}\pi_i\left[\left(q_D-1\right)\frac{\partial D^i}{\partial a_D}+\left(q_B-c_B\right)\frac{\partial B^i}{\partial a_D}\right]\right)\right)+$$
$$+\beta\left(1-\sum_{i=1,2}\pi_i \frac{\partial D^i}{\partial a_D}\frac{\partial \varphi^D}{\partial \bar{D}}\right)=0$$

now substituting $\delta^1 \frac{\partial V^1}{\partial y^1}-\lambda \frac{\partial V^{21}}{\partial y^1}$ and $\left(\delta^2+\lambda\right)\frac{\partial V^2}{\partial y^2}$ from the first order conditions for $\frac{\partial \mathcal{L}}{\partial y}$ above, we obtain, after some rearrangements

$$\frac{\partial \varphi^D}{\partial \bar{D}}\left(\mu\left(\sum_{i=1,2}\pi_i \frac{\partial V^i}{\partial a_D}/\frac{\partial V^i}{\partial y^i}\right)+\lambda \frac{\partial V^{21}}{\partial y^1}\left(\frac{\partial V^1}{\partial a_D}/\frac{\partial V^1}{\partial y^1}-\frac{\partial V^{21}}{\partial a_D}/\frac{\partial V^{21}}{\partial y^1}\right)\right)+$$
$$+\frac{\partial \varphi^D}{\partial \bar{D}}\left[\mu \sum_{i=1,2}\pi_i\left(q_D-1\right)\left(\frac{\partial D^i}{\partial a_D}-\frac{\partial D^i}{\partial y^i}\left(\frac{\partial V^i}{\partial a_D}/\frac{\partial V^i}{\partial y^i}\right)\right)\right]+$$
$$+\frac{\partial \varphi^D}{\partial \bar{D}}\left[\mu \sum_{i=1,2}\pi_i\left(q_B-c_B\right)\left(\frac{\partial B^i}{\partial a_D}-\frac{\partial B^i}{\partial y^i}\left(\frac{\partial V^i}{\partial a_D}/\frac{\partial V^i}{\partial y^i}\right)\right)\right]+$$
$$+\beta \frac{\partial \varphi^D}{\partial \bar{D}}\left(\sum_{i=1,2}\pi_i \frac{\partial D^i}{\partial y^i}\left(\frac{\partial V^i}{\partial a_D}/\frac{\partial V^i}{\partial y^i}\right)\right)+\beta\left(1-\sum_{i=1,2}\pi_i \frac{\partial D^i}{\partial a_D}\frac{\partial \varphi^D}{\partial \bar{D}}\right)=0$$

To simplify further, we need to use the following Slutsky-type property obtained by Pirttilä and Tuomala (1997)

$$-\pi_i \frac{\partial D^i}{\partial y^i}\left(\frac{\partial V^i}{\partial a_D}/\frac{\partial V^i}{\partial y^i}\right)\frac{\partial \varphi^D}{\partial \bar{D}}=\pi_i\left(\frac{\partial \tilde{D}^i}{\partial a_D}-\frac{\partial D^i}{\partial a_D}\right)\frac{\partial \varphi^D}{\partial \bar{D}} \quad i=1,2$$

where a tilde denotes hicksian demands. Using these properties, the condition above rewrites as

$$\begin{aligned}\beta &= -\chi\left(\mu\left(\sum_{i=1,2}\pi_i \frac{\partial V^i}{\partial a_D}/\frac{\partial V^i}{\partial y^i}\right)+\lambda \frac{\partial V^{21}}{\partial y^1}\left(\frac{\partial V^1}{\partial a_D}/\frac{\partial V^1}{\partial y^1}-\frac{\partial V^{21}}{\partial a_D}/\frac{\partial V^{21}}{\partial y^1}\right)\right)\frac{\partial \varphi^D}{\partial \bar{D}}+\\ &\quad -\chi\left(\mu \sum_{i=1,2}\pi_i\left[\left(q_D-1\right)\frac{\partial \tilde{D}^i}{\partial a_D}+\left(q_B-c_B\right)\frac{\partial \tilde{B}^i}{\partial a_D}\right]\right)\frac{\partial \varphi^D}{\partial \bar{D}}\end{aligned}$$

where $\chi = \frac{1}{1-\frac{\partial\varphi^D}{\partial D}\sum_{i=1,2}\pi_i\frac{\partial D^i}{\partial a_D}}$. Let us now proceed by multiplying $\frac{\partial\mathcal{L}}{\partial y^i}$ by D^i for $i = 1, 2$ and adding the resulting expressions to $\frac{\partial\mathcal{L}}{\partial q_D}$. Then multiply $\frac{\partial\mathcal{L}}{\partial y^i}$ by B^i for $i = 1, 2$ and add the resulting expressions to $\frac{\partial\mathcal{L}}{\partial q_B}$. The equations obtained as a result can be simplified making use of Roy's identity and using the Slutsky equations $\frac{\partial Q_j^i}{\partial q_j} = \frac{\partial\tilde{Q}_j^i}{\partial q_j} - Q_j^i\frac{\partial V^i}{\partial y_i}$, where $Q_D^i = D^i \quad Q_B^i = B^i$ and where a tilde denotes hicksian demands. As a results, we obtain

$$\mu\sum_{i=1,2}\pi_i\left(\left(q_D - \frac{\beta}{\mu} - 1\right)\frac{\partial\tilde{D}^i}{\partial q_j} + (q_B - c_B)\frac{\partial\tilde{B}^i}{\partial q_j}\right) = \lambda\frac{\partial V^{21}}{\partial y^1}\left(Q_j^1 - Q_j^{21}\right) \quad j = D, B$$

Finally, one needs to replace for β as obtained above and rearrange to obtain, from the last two expressions above, the optimal tariffs as expressed in the proposition.

Proof of Proposition 3

We proceed assuming the following conditions hold at A^{SB}: $\frac{I^2}{w^w} > \frac{I^1}{w^1}$, $I^2 - T^2 > I^1 - T^1, I^2 > I^1$.[26]

Payments P^1 and P^2 are defined as the payments such that

$$\Omega\left(I^2 - T^2 - P^2\right) + \gamma\left(D^2, B^2\right) + \phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right)g\left(\frac{I^2}{w^2\left(D^2 + B^2\right)}\right)\right) =$$
$$\Omega\left(I^2 - T^2 - P^1\right) + \gamma\left(D^1, B^1\right) + \phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right)g\left(\frac{I^2}{w^2\left(D^1 + B^1\right)}\right)\right) \quad (16)$$

that is, $P^2 - P^1$ is the extra payment that needs to be asked to a type 2 individual in order to ensure that she (when choosing the level of income I^2) will consume trip quantities $D^2 + B^2$ rather than $D^1 + B^1$.[27]

Proof of validity of (5) **for** $i = 1$ **at** A^{SB} Rewrite the left hand side of (16) for *i=2*, using (2) (we know this constraint to be satisfied at equality since, by assumption, it constraint

[26] Their meaning is the following: with $\frac{I^2}{w^w} > \frac{I^1}{w^1}$, we assume that the amount of labor supplied by the high ability type is larger than that of the low ability type. We also assume that both the pre-tax and the post-tax income of individuals of high ability is higher than that of low ability types.

[27] Similarly, P^1 should be designed as the payment such that trip quantity $D^1 + B^1$ gives the same utility, to an individual of type 1 choosing to earn income I^1, as making no trips at all. However choosing no travel at all would always be a dominated alternative, given their commuting purpose (with no commuting, labor supply would be infinite) even if P^1 took away all of the individual's net income. We thus set P^1 arbitrarily. This also means that we can be sure that neither individuals of type 1 nor those of type 2 will prefer zero trips to, respectively, $D^1 + B^1$ and $D^2 + B^2$, as long as P^1 and P^2 are not unreasonably high.

binds at A^{SB}). We have the following

$$\Omega\left(I^1-T^1-P^1\right)+\gamma\left(D^1,B^1\right)+\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^1}{w^2\left(D^1+B^1\right)}\right)\right)= \\ \Omega\left(I^2-T^2-P^1\right)+\gamma\left(D^1,B^1\right)+\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^2}{w^2\left(D^1+B^1\right)}\right)\right)$$

therefore

$$\Omega\left(I^1-T^1-P^1\right)=\Omega\left(I^2-T^2-P^1\right)+\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^2}{w^2\left(D^1+B^1\right)}\right)\right)+ \\ -\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^1}{w^2\left(D^1+B^1\right)}\right)\right)$$

So constraint (5) for $i=1$ is verified if (replacing $\Omega\left(I^1-T^1-P^1\right)$ from the above and rearranging)

$$\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^1}{w^1\left(D^1+B^1\right)}\right)\right)+ \\ -\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^2}{w^1\left(D^1+B^1\right)}\right)\right)\geq \\ \phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^1}{w^2\left(D^1+B^1\right)}\right)\right)+ \\ -\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^2}{w^2\left(D^1+B^1\right)}\right)\right)$$

This is verified by convexity of $g(.)$ and concavity of $\phi(.)$.

Proof of validity of (6) **for** $i=1$ **at** A^{SB} Start from (6) for $i=2$. Using (16), we have

$$\Omega\left(I^2-T^2-P^1\right)-\Omega\left(I^2-T^2-P^2\right)-\gamma\left(D^2,B^2\right)+\gamma\left(D^1,B^1\right)= \\ \phi\left(1-a_D D^2-a_B B^2-\left(D^2+B^2\right)g\left(\frac{I^2}{w^2\left(D^2+B^2\right)}\right)\right)+ \\ -\phi\left(1-a_D D^1-a_B B^1-\left(D^1+B^1\right)g\left(\frac{I^2}{w^2\left(D^1+B^1\right)}\right)\right)$$

Now assume that $D^2+B^2 \geq D^1+B^1$ and $a_D D^1 + a_B B^1 \geq a_D D^2 + a_B B^2$. Then, by $\frac{I^2}{w^2} > \frac{I^1}{w^1}$ and concavity of $\phi(.)$, we have

$$\begin{aligned}
\phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^2}{w^2\left(D^2+B^2\right)}\right)\right) + \\
-\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^2}{w^2\left(D^1+B^1\right)}\right)\right) > \\
\phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^1}{w^1\left(D^2+B^2\right)}\right)\right) + \\
-\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^1}{w^1\left(D^1+B^1\right)}\right)\right)
\end{aligned}$$

therefore

$$\begin{aligned}
\Omega\left(I^2 - T^2 - P^1\right) + \gamma\left(D^1, B^1\right) - \Omega\left(I^2 - T^2 - P^2\right) - \gamma\left(D^2, B^2\right) > \\
\phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^1}{w^1\left(D^2+B^2\right)}\right)\right) + \\
-\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^1}{w^1\left(D^1+B^1\right)}\right)\right)
\end{aligned}$$

Now, by concavity of $\Omega(.)$ and since $I^2 - T^2 > I^1 - T^1$, we can write

$$\Omega\left(I^1 - T^1 - P^1\right) - \Omega\left(I^1 - T^1 - P^2\right) > \Omega\left(I^2 - T^2 - P^1\right) - \Omega\left(I^2 - T^2 - P^2\right)$$

therefore

$$\begin{aligned}
\Omega\left(I^1 - T^1 - P^1\right) + \gamma\left(D^1, B^1\right) - \Omega\left(I^1 - T^1 - P^2\right) - \gamma\left(D^2, B^2\right) > \\
\phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^1}{w^1\left(D^2+B^2\right)}\right)\right) + \\
-\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^1}{w^1\left(D^1+B^1\right)}\right)\right)
\end{aligned}$$

which, rearranged, gives us (6) for $i = 1$.

Validity of (5) **for** i =**2 at** A^{SB} Use (2) to rewrite the left hand side of (5) for $i = 2$. We can rearrange to get

$$\begin{aligned}\Omega\left(I^1 - T^1 - P^1\right) + \gamma\left(D^1, B^1\right) - \Omega\left(I^1 - T^1 - P^2\right) - \gamma\left(D^2, B^2\right) \geq \\ \phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^1}{w^2\left(D^2 + B^2\right)}\right)\right) + \\ -\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^1}{w^2\left(D^1 + B^1\right)}\right)\right)\end{aligned}$$

Now, by concavity of $\Omega(.)$ and since $I^2 - T^2 > I^1 - T^1$, we have

$$\Omega(I^1 - T^1 - P^1) - \Omega(I^1 - T^1 - P^2) > \Omega(I^2 - T^2 - P^1) - \Omega(I^2 - T^2 - P^2)$$

(5) is thus certainly satisfied if

$$\begin{aligned}\Omega(I^2 - T^2 - P^1) - \Omega(I^2 - T^2 - P^2) - \gamma(D^2, B^2) + \gamma(D^1, B^1) \geq \\ \phi\left(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^1}{w^2\left(D^2 + B^2\right)}\right)\right) + \\ -\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^1}{w^2\left(D^1 + B^1\right)}\right)\right)\end{aligned}$$

holds. Using (16) we can replace for the left hand side of the above and rearranging we have

$$\begin{aligned}\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^1}{w^2\left(D^1 + B^1\right)}\right)\right) + \\ -\phi\left(1 - a_D D^1 - a_B B^1 - \left(D^1 + B^1\right) g\left(\frac{I^2}{w^2\left(D^1 + B^1\right)}\right)\right) \geq \\ \phi(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^1}{w^2\left(D^2 + B^2\right)}\right)) + \\ -\phi(1 - a_D D^2 - a_B B^2 - \left(D^2 + B^2\right) g\left(\frac{I^2}{w^2\left(D^2 + B^2\right)}\right))\end{aligned}$$

On condition that $D^2 + B^2 \geq D^1 + B^1$ and $a_D D^1 + a_B B^1 \geq a_D D^2 + a_B B^2$, this is verified. Therefore, (5) holds as well.

References

[1] Atkinson, A.B., Stiglitz, J.E., 1976. The design of tax structure: direct versus indirect taxation. *Journal of Public Economics* 6, 55–75

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

[3] Batarce, M., and Ivaldi, M. (2011), Travel Demand Model with Heterogeneous Users and Endogenous Congestion: An Application to Optimal Pricing of Public Transport Services. CEPR Discussion Paper No. DP8416

[4] Boadway, R., and Gahvari, F. (2006), Optimal taxation with consumption time as a leisure or labor substitute, *Journal of Public Economics* 90, 1851–1878

[5] Boadway, R., and Keen, M. (1993), Public goods, self selection and optimal income taxation, *International Economic Review* 34, 463-478

[6] Bovenberg, A. L., and Goulder, L. H. (2002). "Environmental taxation and regulation," *Handbook of Public Economics*, in: A. J. Auerbach & M. Feldstein (eds.). Edition 1, volume 3, chapter 23, pages 1471-1545.

[7] Cremer, H., Gahvari, F., and Ladoux, N. (1998), Externalities and optimal taxation, *Journal of Public Economics*, 70, 343-364

[8] Cremer, H., and Gahvari, F. (2002), Nonlinear pricing, redistribution and optimal tax policy, *Journal of Public Economic Theory*, 4 (2), pp. 139-161

[9] De Borger, B., (2011), Optimal congestion taxes in a time allocation model. *Transportation Research Part B*, Volume 45, Issue 1, Pages 79-95

[10] De Borger, B., and Wuyts, B., (2011), The structure of the labor market, telecommuting and optimal peak period congestion tolls. A numerical optimization model. *Regional Science and Urban Economics* 41, 426–438.

[11] Department for Transport, UK government (2011), Values of Time and Operating Costs - Transport Analysis Guidance (TAG) Unit 3.5.6.

[12] Estupinan, N., Gomez-Lobo A., Munoz-Raskin, R. and Serebrisky, T. (2007). Affordability and Subsidies in Public Urban Transport: What Do We Mean, What Can Be Done? *World Bank Policy Research Working Paper No. 4440*

[13] Gahvari, F. (2007), On optimal commodity taxes when consumption is time consuming, *Journal of Public Economic Theory*, 9, 1-27

[14] Gutiérrez-i-Puigarnau, E. and van Ommeren, J. (2010), Labor supply and Commuting. *Journal of Urban Economics* 68 82–89

[15] Hu, P.S., and Reuscher, T. (2004), Summary of travel trends, the 2001 National Household Travel Survey, US Department of Transportation

[16] Kaplow, L. (2006). Optimal Control of Externalities in the Presence of Income Taxation. NBER Working Paper No. 12339

[17] Kleven, H.J., (2004). Optimal taxation and the allocation of time. *Journal of Public Economics* 88 (3-4), 545-557.

[18] Kreiner, C. T., and Verdelin, N. (2012), Optimal provision of public goods: a synthesis. Forthcoming in the *Scandinavian Journal of Economics*

[19] Jara-Diaz, S.R., 2008. Allocation and valuation of travel time savings. In: Hensher, D.A., Button, K.J. (Eds.), *Handbook of Transport Modelling*, second edition. Elsevier, 363-379.

[20] Mayeres, I., and Proost, S. (1997), Optimal Tax and Public Investment Rules for Congestion Type of Externalities, *Scandinavian Journal of Economics*, Vol. 99, No. 2, pp. 261-279

[21] Parry, I., and Bento, A. (2001), Revenue Recycling and the Welfare Effects of Road Pricing. *Scandinavian Journal of Economics* 103, 4, pp. 645-671

[22] Pirttilä, J., and Tuomala, M. (1997). Income tax, commodity tax and environmental policy. *International Tax and Public Finance* 4:379-93.

[23] Small, K., and Verhoef, E., (2007), *The economics of urban transportation.* Routledge.

[24] Stiglitz, J.E., (1982), Self-selection and Pareto efficient taxation, *Journal of Public Economics*, 17, 213-40

[25] Van Dender, K., (2003), Transport taxes with multiple trip purposes. *Scandinavian Journal of Economics*, Vol. 105, No. 2, pp. 295-310

[26] Wang, J., Lindsey, R., and Yang, H. (2011), Nonlinear Pricing on Private Roads with Congestion and Toll Collection Costs, *Transportation Research Part B* Vol 45, pp. 9-40

[27] Wilson R. (1993), *Nonlinear Pricing*