Congestion pricing, infrastructure investment and redistribution

Russo, Antonio

Toulouse School of Economics - GREMAQ

15 February 2011

Congestion Pricing, Infrastructure Investment and Redistribution

-Preliminary version-

Antonio Russo*

February 15, 2011

Abstract

We study congestion pricing by a government that has redistributive concerns, in the presence of optimal income taxation. Individuals differ in (unobservable) earning ability and consumption technology for commodities using a congestible network (e.g. roads, Internet). We find, assuming separable preferences, that when efficiency of consumption technology is either invariant or postively correlated with earning ability, low ability individuals should face higher marginal congestion charges than high ability ones. Moreover, reducing congestion (by raising charges or expanding network capacity) enables government to increase redistribution. We also find that means tested congestion pricing may be necessary to implement the second-best allocation.

JEL Code: R41, H23, H21, H41, H54

Keywords: congestion pricing, income taxation, redistribution, infrastructure investment

1 Introduction

Network congestion is becoming an increasingly serious issue, in various domains, from transport networks (e.g. roads) to telecommunications networks (including the Internet). One of the most commonly suggested remedies to the problem is the introduction of congestion pricing, which has been the subject of a large number of studies

*Toulouse School of Economics (antonio.russo@tse-fr.eu). I thank Georges Casamatta, Jean-Marie Lozachmeur, Emmanuel Thibault and especially Helmuth Cremer for useful comments and suggestions. All errors and opinions expressed are mine alone.

in the transport and network economics literature.[1] Governments and firms are becoming increasingly receptive to this idea. In the case of road networks, for instance, we have the example of congestion charging schemes recently introduced by the cities of London, Stockholm and Milan. Telecom utilities, such as AT&T and Verizon in the US and BT in the UK, have recently put forward proposals to introduce congestion based pricing for Internet and data services on their networks, as well as to create a "two-tier" Internet and email system, where access to a non-congested portion of the network would be granted only to customers willing to pay a premium, with the others confined to a congested portion. Congestion pricing si also being considered by providers of newly installed 4G wireless networks in the face of looming congestion problems in major metropolitan areas in the US (such as New York City and San Francisco).[2]

While there is a large consensus among economists on the merits of congestion pricing from an efficiency standpoint, its redistributive implications still seem to be debated.[3]There certainly are strong concerns among the public, regulators and policy makers about its possible regressivity. A recurrent argument is that since goods like transportation and telecommunications represent a higher share of total expenditures for low income families than for the well off (even if their consumption is positively correlated with income), an increase in their price would hurt the former relatively more than the latter (an argument quite similar to one of the main objections against raising fuel taxes). Moreover, time gains from reduced congestion would be beneficial mostly to high income individuals, whose opportunity cost of time is supposedly higher. Such concerns were, for instance, one of the main arguments against the introduction of a congestion tolling scheme for Manhattan (recently blocked by the New York State Assembly)[4], but also in other cities such as Manchester or Edinburgh. As for telecommunications networks and the Internet, members of the FCC and the Obama Administration in the US have expressed concerns about the pricing proposals mentioned above, motivated by the will to protect low income customers, as well as preserving "net neutrality".[5]

[1]see, e.g, Small and Verhoef (2007) for the case of transport networks. MacKie-Mason and Varian (1995) look at the issue of pricing for access to the Internet, dedicating significant attention to congestion issues.

[2]See for instance "Is 2010 the year of wireless congestion?", MSNBC, http://www.msnbc.msn.com/id/34634571/ns/technology_and_science-tech_and_gadgets/

[3]The literature on equity issues related to road pricing is reviewed in Levinson (2009). He concludes that designing road pricing schemes such that they have a progressive impact should be possible. A key aspect in this sense is the redistribution of the revenues they generate. Anyway, he reports a number of contributions that suggest a regressive impact.

[4]In a recent interview to the New York Times, NY Assemblyman Richard L. Brodsky said he opposed the introduction of congestion charging in Manhattan "for the reason that these schemes put the burden for paying the fees on blueblood and blue collar alike" (see New York Times, "Congestion Pricing: Just Another Regressive Tax?" www.nytimes.com)

[5]In a recent interview, FCC member Michael J. Copps stated that "so-called 'managed service' and 'paid priority' cannot be allowed to supplant the quality of the public Internet service available to us all"

Governments' intervention to curb network congestion can take different forms: they can directly impose charges or they can regulate utilities providing access to networks (in many cases, they directly own at least a minority share of these utilities). It is also well known that redistributive concerns may be addressed using instruments such as income taxes (or transfers). Finally, governments can invest in network capacity to alleviate congestion. This suggests some interesting questions: how should a government design its optimal fiscal and regulatory policy when it wants to reduce the inefficiencies of excessive network congestion and, at the same time, protect its less fortunate citizens? Can the introduction of congestion charges be justified on redistributional grounds (on top of the standard Pigouvian motives), once all the other available instruments are optimally set? These questions seem relevant in themselves, but more so when we consider that they are closely related to the issue of political acceptability of pricing measures, which, as testified by the anecdotal evidence provided above, is often a crucial obstacle holding back their implementation.

The first part of this paper tries to shed light on the issues raised above using a framework quite close to the well known optimal taxation problem introduced by Mirrlees (1971). We look at congestion charges as part of a general tax system in which the government uses both income and commodity (nonlinear) taxes[6], where the key aspect is the presence of self selection constraints that limit the amount of redistribution which can be put in place. This is, to the best of our knowledge, a novel approach in the literature on congestion pricing. In addition, we also look at the distributional impact of investments in network capacity (such as expanding network infrastructures, spending on their maintenance or improving traffic managment technologies). A crucial component of our model is that consumption of commodities requires time (as in Becker (1965)), which is a scarce resource for individuals. Congestion increases the time costs of consuming the commodities (e.g. car trips, phone calls, data for applications using the Internet...) that, by making use of the network, contribute to its congestibility. Individuals may differ in both earning ability and consumption technology for such commodities (both are assumed to be private information). Higher ability in consuming commodities is assumed to reduce,

(www.washingtonpost.com/posttech/2010/12/fcccopps_net_neutrality_requ.html). See also, for instance, the Save the Internet campaign initiative (www.savetheinternet.com). Concerns about preserving "net neutrality" also revolve around freedom of provision and transmission of content over the Internet and possible stifling of innovation coming from small firms or individuals. While these issues are certainly very important, their study is outside of the scope of this work.

[6]The nonlinearity assumption is motivated by the information reasonably available to the government. We also take the simplifying assumption that nonlinearities in prices faced by consumers depend only on governmental levies. Alternatively, one could interpret the problem as that of nonlinear pricing by a public utility providing access to the congestible piece of network, in the presence of optimal income taxation, as in Cremer and Gahvari (2002). In both cases, the objective of the model is to define the optimal pricing policy that a welfare maximising government would choose if it could determine pricing schedules. See Section 2 below for further clarifications on both assumptions.

for a given level of congestion, the time cost of consumption for the individual (i.e. it means the individual is endowed with a more efficient consumption technology). The rationale for this assumption is that the amount of effort (and time) required to use some of the goods we have in mind may vary among individuals: for example, many Internet applications (even "consumer" applications such as shopping online or searching for information) are relatively complex and require "cognitive skills" that differ across individuals. According to sociological literature, such differences are a key contributor to "digital inequality" (i.e. inequality in ease, effectiveness and quality of use of digital technologies (DiMaggio *et al.* (2004)), which is feared to widen existing inequalities as greater amounts of services and information become digitalised.[7] Now, cognitive skills may be developed with education, but are also innate (indeed, they may determine the individual's educational proficiency). Moreover, they are likely to explain a relevant part of an individual's earning ability. It may thus be interesting to consider also the case in which ability in consuming network commodities is correlated (either positively or negatively) to earning ability. Such correlation turns out to play an interesting role for the optimal pricing policy, when the government cares for protecting the less fortunate, as will be shown below.

The results we obtain are the following: first, in the presence of optimal income taxes, nonlinear congestion charges are optimal. Whether the marginal tax rate (or price) faced by low ability individuals should be higher or lower than that faced by high ability individuals depends on the correlation earning ability and ability in using the network good. When these are positively correlated or when the latter is invariant in the population, then low ability individuals should face higher marginal tax rates. When correlation is negative, the opposite may occur. Second, reducing congestion can help achieve a higher level of redistribution, by letting the government relax the self selection constraints it faces when trying to redistribute from the more to the less able. The intuition is as follows: as time spent consuming commodities competes with labour supply and pure leisure for the individuals' time, reducing it lowers the marginal value of leisure relatively more for those who, at a given consumption allocation, work more. This is precisely the case of the low ability individuals who would, out of equilibrium, be mimicked by high ability types. This effect is stronger (resp. weaker) if earning and consumption ability are positively (negatively) correlated. Therefore, all individuals should face a tax on congestion generating goods (or activities) not only for "Pigouvian" reasons, but also because this allows to improve the government's

[7] Freese *et al.* (2006), focusing on elderly people, provide evidence of a correlation between cognitive skills and propensity to use Internet applications. Other evidence underlines the relevance of such skills even among younger individuals (DiMaggio *et al.* (2004)).

redistributive capabilities. This conclusion contrasts with regressivity concerns on congestion charges presented above. Third, similar considerations apply to investment in network capacity: it is a public good that, by reducing the overall amount of congestion, affects the government's redistributive ability and should be thus provided following a "modified Samuleson's rule" accounting for this additional effect.[8] Again, provision of capacity has a positive redistributive effect when abilities are positively correlated or the latter is invariant in the population, while it may be negative when correlation is negative. Fourth, all the mentioned results are derived assuming separability of goods (including public goods or bads, such as congestion) and leisure in individuals' utility function and no taste variations across the population. This is important because, according to all previous literature, such assumptions were supposed to block all redistributive and efficiency properties of commodity taxes on goods responsible for externalities (except for Pigouvian taxes), as long as the government could tax income optimally. In our model, such properties survive and provide quite clear policy implications, even under these two assumptions.

Finally, congestion charges may be seen as fees governments ask their citizens to pay in order to finance the provision of certain services (for instance, access to a congestible road). Such fees are often subject to means testing. The second part of the paper is dedicated to the investigation of the question of implementability of the optimal fiscal policy (and allocation of goods and income) without recurring to means tested congestion charges. In particular, we ask whether the second best allocation derived in the presence of self selction constraints can be implemented using separable tax functions (one for income and one for the "congestion" commodity), instead of a general tax function conditional on both consumed commodities and income. Means testing would be necessary if and only if the answer to this question were negative. While, unfortunately, we are not able to provide analytical conditions answering the question (numerical analysis may be needed and will be developed in the future—-work in progress), we prove that means testing is necessary in the quite important case of quasilinear preferences.

[8]This recalls results first obtained by Boadway and Keen (1993) and Kreiner and Verdelin (2009).

Previous literature

The issue of optimal fiscal policy in the presence of externalities has been studied quite extensively in the literature.[9] Mayeres and Proost (1997) have studied the design of optimal fiscal policy (where the government can use income and commodity taxes, as well as invest in abatement technologies) in the presence of congestion externalities.[10] The results of their analysis suggest that taxes on congestion generating goods and investment in congestion abating technologies should not be significantly affected by redistributional concerns (Mayeres and Proost, 1997, Table 2a), at least as long all other instruments are optimally adjusted. However, in their model the government uses only linear tax schedules. This leaves open the question of whether the positive redistributive impact of taxing congestion would be found even when the government uses optimal tax schedules (constrained only by the available information). We provide an answer to this question, which confirms the finding of a positive redistributive impact of congestion charges and capacity investment, as part of an optimally designed tax and spending system. However, given that we allow for nonlinearities, while also recognizing the asymmetric information problems faced by the government, the redistributive effects of congestion pricing come from completely different sources.[11] Cremer *et al.* (1998, 2001) and Kaplow (2006), studied the role of environmental levies in the presence of asymmetric information and nonlinear income taxation: however, they consider a standard model of labour supply where commodities do not require any time of consumption. Their individuals also have preferences defined over some measure of environmental quality (the amount of pollution, for instance). In our model, on the contrary, individuals have no taste for congestion in itself, but they suffer from it due to the fact that it increases the time required to consume commodities that make use of the network. This is the key reason why, contrary to what they find, redistributive properties of charges on commodities generating congestion survive even under separability assumptions.

Recently, Kleven (2004), Boadway and Gahvari (2006) and Gahvari (2007) have looked at the issue of commodity taxation, in various setups, in light of household production theory first introduced by Becker (1965),

[9] See Bovenberg and Goulder (2002) for a survey.

[10] Optimal congestion charges in the presence of redistributive concerns were studied also in Leuthold (1976). She finds that, when the government cannot redistribute through income taxes and transfers, congestion charges have to be adjusted according to the welfare weight attached to consumers using the infrastructure: charges will be lower than the Pigouvian first best if the congestible facility is used relatively more by those consumers the government wants to protect.

[11] Mayeres and Proost's model is more sophisticated than ours on the production side, allowing for firms using and producing the congestion affected (and affecting) good. We abstract from these issues.

showing that standard optimal tax results may have to be modified.[12]Our paper is obvously very much related to these works. However, these papers do not consider the presence of externalities. They also do not allow for heterogeneous abilities in consuming goods (i.e. consumption technologies) by individuals. Our paper extends this literature in both directions.

Our results on the optimal provision of network capacity relate to those of Kreiner and Verdelin (2009), who, in turn, generalized the results of Boadway and Keen (1993). They find that, as long as the government can optimally adjust income taxes, the optimal provision of public goods should follow a modified Samuleson's rule suggesting overprovision (underprovision) is optimal if low ability individuals have, at a given allocation, a higher (lower) marginal valuation for the public good than the high ability ones. In our setup, the first case is verified since, at a given allocation, low ability individuals value the reduction in the time cost of consuming network goods (generated by additional network capacity) more than high ability ones, who need to provide less labour.[13] Both of these works, differently from the present paper, are in any case not concerned with externalities and optimal taxation (or congestion charging) of commodities.

Finally, the paper relates to the literature on nonlinear pricing in the presence of redistributive concerns. Cremer and Gahvari (2002) considered pricing of public sector goods (such as energy or telecommunications) and showed that, when the government optimally sets nonlinear income taxes, nonlinear pricing of goods is optimal. They also show that means testing in pricing of such goods is not necessary. Their results are obtained under separability assumptions, but in the presence of heterogenous preferences for the public sector good. They neglect the presence of time costs of consumption as well as congestion externalities. Our results differ from theirs due essentially to these assumptions.

The rest of the paper is organised as follows: Section 2 presents the model. We present the optimal congestion charging scheme in Section 3. Section 4 discusses optimal network capacity investments. Section 5 looks at implementation and means testing. Section 6 concludes.

[12]An important point of Gahvari's (2007) paper is that standard optimal tax formulae have to be modified under the assumption of separability in the utility function among commodities and leisure, but not under separability of commodities and labour supply. In this paper, we consider only the former case, as is done, for instance, in Boadway and Gahvari (2006) and Kleven (2004).

[13]Sandmo (1973) studied optimal provision of public goods as intermediate goods that need to be combined with private ones (such as the individual's time) to produce a given final good. He also considers the possibility of a congestible public good. His approach however abstracts from the need for distortionary taxation to finance provision and redistribute welfare. Engel *et al.* (2009) look at the optimal design of institutions for infrastructure provision and maintenance. We abstract from that issue.

2 The model

The setup of the model is similar to that of Stiglitz (1982). There are only two types of individuals $i = 1, 2$. They differ in (up to) two dimensions: the first is earning ability, identified by the parameter w with

$$w_2 > w_1$$

The second is ability in consuming commodities: this will be illustrated below. For simplicity, the size of both groups is the same and normalized to one.

There are only two commodities in the economy: composite consumption C (the numeraire) and a "network" commodity that makes use of a congestible network D (e.g. car trips along a road, phone calls or data travelling on a telecom network, such as those of Internet applications). The economy's production technology is linear in labour for both commodities, with constant marginal costs normalised to one. The production sector is perfectly competitive so producer prices are equal to unity and constant.[14]

Individuals are assumed to face the following budget constraint

$$C^i + q_D D^i \leq w^i L^i + \tau^i \quad i = 1, 2$$

where q_D the market price of good D (post tax), L is the individual's labour supply and τ is the lump-sum transfer\tax from the government.

Commodities are also assumed to require some time for consumption. Ability in consuming commodities (which is assumed to determine the time required to consume them) is the second dimension along which individuals may differ. It is also assumed to be not observable by the government. We work under two simplifying assumptions: the first is that consumption ability and earning ability are either perfectly correlated (either positively or negatively). Otherwise, consumption ability is simply homogenous across the population.[15] The

[14]This implies that all nonlinearities in prices faced by consumers for commodities come from nonlinear charges levied by the government. This is clearly a simplification as, for instance, industries such as telecomunications are clearly not perfectly competitive and firms, though regulated, do have some pricing power. An alternative interpretation of the model is to consider, as in Cremer and Gahvari (2002), access to the network as being under the control of a single monopolistic firm with constant marginal costs that is either owned by the state or whose prices are set entirely via regulation. Whatever the interpretation chosen, the reader should keep in mind that the objective of the model is to define the optimal pricing policy that a welfare maximising government would choose if it could fully determine pricing schedules through charges or regulation.

[15]I am working on a version of the model where this assumption is relaxed.

second is that consumption ability matters only for the consumption of the network good D. In particular, consuming a unit of commodity D by type i requires $a_D^i \quad i = 1, 2$ units of time, while the unit time cost for consuming C is equal to $a_C = 0$ for both types.[16] Individuals thus face the following time constraint

$$a_D^i D^i + L^i + l^i \leq 1 \quad i = 1, 2$$

where l is leisure (assumed to be "pure", no market good is involved in its "consumption"). We normalize individuals' time endowment to one (assumed ot be the same for all types). To capture network congestion, we assume that the time cost of consuming D is in fact an increasing and convex function of the aggregate consumption of D itself. Therefore

$$a_D^i = \theta^i \varphi(D, K) \quad i = 1, 2$$

where $\frac{\partial \varphi}{\partial D}, \frac{\partial^2 \varphi}{\partial D^2} > 0$, with $D = D^1 + D^2$, and $\frac{\partial \varphi}{\partial K} < 0$. We have $\theta^1 \gtreqless \theta^2 \quad i = 1, 2$ depending on whether earning and consumption ability are respectively positively correlated, negatively correlated or the latter is simply homogeneous across the population. In the last case, $\theta^1 = \theta^2$ and and $a_D^1 = a_D^2 = a_D$ for any φ.

Increasing network capacity K reduces, for any given total level of consumtpion of D, the amount of congestion. We assume that K has a constant unit cost p. Individuals are assumed to take φ as given when deciding how much of D to consume, which generates the standard congestion externality problem.

Individuals derive utility from consuming the two commodities C and D and leisure. The utility function is

$$U = U(C, D, l)$$

increasing in all of its arguments and having standard properties. Since labour supply $L = \frac{y}{w}$ and earning ability w, as well as consumption ability a_D^i are not observable by the government, while total earned income y is, we rewrite, as customary in the optimal taxation literature, the individual's utility function in terms of observables,

[16] The assumption is taken in order to focus on the implications of the presence of time costs of consuming the network good D. However, it may also be justified by saying that, contrary to the time required for accessing the congestible network (which is here modelled as a substitute for labour time, thus a pure waste for the individual) generic consumption requires time that is a substitute for leisure, which is valuable in itself. Boadway and Gahvari (2006) elaborate on this point.

saturating the time constraint, as

$$U^i = U(C^i, D^i, l^i) \equiv \Omega^i(C^i, D^i, 1 - a_D^i D^i - \frac{y^i}{w^i}) \quad i = 1, 2$$

The government is able to set nonlinear charges on commodities and income taxes.[17]It also decides on the level of network capacity K. Given the cost of production normalisation discussed above, each individual faces a (possibly type specific) marginal price p_j^i for commodities given by

$$p_j^i = 1 + t_j^i \quad i = 1, 2 \quad j = C, D$$

where $t_j^i \quad i = 1, 2 \quad j = C, D$ is the (possibly type specific) marginal charge rate imposed by the government on commodity j.

The government's budget constraint is

$$\sum_{i=1,2} (I^i - C^i - D^i) \geq pK + R$$

We assume the objective of the government is to maximize the following utilitarian social welfare function[18]

$$W = U^1 + U^2$$

with respect to C^i, D^i and $y^i \quad i = 1, 2$, subject to the usual revenue constraint and the self selection constraints. The government directly chooses allocations of individuals' consumption and income, following the Taxation Principle (Stiglitz, 1982). We assume that the government wants to redistribute from high to low ability

[17]This rests on the assumption that personal consumption levels of D can be observed: for the case of road congestion charges, tolling systems involve, generally, the use of electronic systems allowing to track individual accesses (or at least those of a given household). In telecommunications, it is almost always the case that personal consumption levels are observable: indeed pricing schemes for telecom services are often nonlinear. As for C, with two commodities and observable income, if one commodity's consumption level is observable then the other's must be as well. A version of the model with many commodities (for which personal consumption levels are not observable) gives the same qualitative results (and is available from the author), though they would be more difficult to interpret. In the case D is produced by utilities (such as telecommunication services) we assume, as in Cremer and Gahvari (2002), that there are no information asymmetries between government and utilities about individuals' consumption levels.

[18]The assumption of a utilitarian SW function is taken only to minimize the amount of notation but is of no consequence for the results.

individuals: therefore, only one self selection constraint will be relevant:

$$U^2 \geq U^{21} \tag{1}$$

where $U^{21} = \Omega(C^1, D^1, 1 - a_D^2 D^1 - \frac{I^1}{w^2})$ is the utility of the only (potentially) mimicking type. The Lagrangean of the government's problem can thus be written as

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

The FOC of the government's problem are (we omit those with respect to individuals' income y as they are not the focus of the paper)

$$\frac{\partial \mathcal{L}}{\partial C^1} = U_C^1 - \mu - \lambda U_C^{21} = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial C^2} = U_C^2 - \mu + \lambda U_C^2 = 0 \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial D^1} = U_D^1 - \frac{\partial \varphi}{\partial D}\theta^1 D_1 \Omega_l^1 - \frac{\partial \varphi}{\partial D}\theta^2 D_2 \Omega_l^2 - \mu - \lambda U_D^{21} + \lambda \frac{\partial \varphi}{\partial D}\theta^2 D_1 \Omega_l^{21} - \lambda \frac{\partial \varphi}{\partial D}\theta^2 D_2 \Omega_l^2 = 0 \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial D^2} = U_D^2 - \frac{\partial \varphi}{\partial D}\theta^1 D_1 \Omega_l^1 - \frac{\partial \varphi}{\partial D}\theta^2 D_2 \Omega_l^2 - \mu + \lambda U_D^2 + \lambda \frac{\partial \varphi}{\partial D}\theta^2 D_1 \Omega_l^{21} - \lambda \frac{\partial \varphi}{\partial D}\theta^2 D_2 \Omega_l^2 = 0 \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial K} = -\Omega_l^1 \frac{\partial \varphi}{\partial K}\theta^1 D_1 - \Omega_l^2 \frac{\partial \varphi}{\partial K}\theta^2 D_2 - \mu p + \lambda \Omega_l^{21} \frac{\partial \varphi}{\partial K}\theta^2 D_1 - \lambda \Omega_l^2 \frac{\partial \varphi}{\partial K}\theta^2 D_2 = 0 \tag{6}$$

where subscripts denote derivatives with respect to either goods or leisure and where the derivatives

$$U_j^i \equiv \Omega_j^i - a_j^i \Omega_l^i \quad i = 1,2 \quad j = C, D$$

denote the marginal utility individual $i = 1,2$ derives from good $j = C, D$, *net of the opportunity cost of the time* required to consume it, taking as given the unit time costs a_j (we assume the net marginal utilities to be always positive for simplicity). Moreover

$$U_l^i \equiv \Omega_l^i \quad i = 1,2$$

denotes the marginal utility of pure leisure.

3 Optimal congestion charges in the presence of redistributive concerns

3.1 Benchmark

Let us start from a benchmark case in which the self selection constraint (1) is not binding, so that $\lambda = 0$. In that case, it is easy to show, starting from the FOC provided above, that optimal commodity marginal tax rates would be as follows:

$$t_C^i = 0 \quad i = 1,2 \qquad t_D^1 = t_D^2 = \tau = \sum_{i=1,2} \frac{\frac{\partial \varphi}{\partial D} \theta^i D_i U_l^i}{U_C^i}$$

that is, if the government does not face binding self selection constraints, as in the ideal case of fully observable individuals' ability levels, then the only commodity taxes that would be justified, as long as income taxes are optimally set, are only Pigouvian taxes. Congestion pricing has no redistributive role in this case.

3.2 Optimal marginal congestion charge rates with binding self selection constraints

Following the steps of Cremer *et al.* (1998) we can, after some manipulations, write the (type specific) optimal tax rates $t_j^i \quad i = 1,2 \quad j = C, D$ (full derivation of the solution to the problem is provided in the Appendix).

PROPOSITION 1: *The optimal tax rates* $t_j^i \quad i = 1,2 \quad j = C, D$ *write as*

$$t_C^i = 0 \quad i = 1,2$$

$$t_D^1 = \frac{U_D^1}{U_C^1} = k^1 + \tau + \delta \qquad\qquad t_D^2 = \frac{U_D^2}{U_C^2} = \tau + \delta$$

where

$$k^1 \equiv \frac{\lambda}{\mu} U_C^{21} \left(\frac{U_D^{21}}{U_C^{21}} - \frac{U_D^1}{U_C^1} \right) \qquad \tau \equiv \sum_{i=1,2} \frac{\frac{\partial \varphi}{\partial D} \theta^i D_i U_l^i}{U_C^i} \qquad \delta \equiv \frac{\lambda}{\mu} \frac{\partial \varphi}{\partial D} \theta_1 D_1 \frac{\Omega_C^{21} \Omega_C^1}{U_C^1} \left(\frac{\Omega_l^1}{\Omega_C^1} - \frac{\Omega_l^{21}}{\Omega_C^{21}} \frac{\theta_2}{\theta_1} \right)$$

We will elaborate below on the interpretation of the three terms. For the moment, observe only that k^1 and δ are "incentive" terms that are nonzero only as long as the self selection constraint binds ($\lambda > 0$). Their role is to relax such constraint and improve the redistributive capabilities of the government (as long as income taxes are optimally adjusted, of course).

The term τ is the standard Pigouvian component of the tax formula for the "network" commodity D (the sum of the marginal benefits, in terms of reduced opportunity costs of time, of aggregate reduction in consumption of D). Finally, δ is a term that accounts for the role the reduction in congestion plays in relaxing the self selection constraint faced by the government (intuition for the role of this term is given below).

Let us now take the following assumption, which is standard in the optimal taxation literature:

ASSUMPTION: *Weak separability of preferences between goods and leisure*

$$U^i = U(C^i, D^i, l^i) \equiv U\left(\Omega^i(C^i, D^i), \phi(1 - a_D D^i - \frac{I^i}{w^i})\right) \quad i = 1, 2$$

where both $\Omega(.)$ *and* $\phi(.)$ *are increasing and concave functions of both of the respective arguments*

Under this assumption, we have the following

PROPOSITION 2: *Under the above assumption, the components of the optimal tax rates* $t_D^i \quad i = 1, 2$ *write as*

$$k^1 = \frac{\lambda}{\mu U_C^1} a_D^1 \Omega_C^{21} \Omega_C^1 \left(\frac{\phi_l^1}{\Omega_C^1} - \frac{\phi_l^{21}}{\Omega_C^{21}} \frac{\theta_2}{\theta_1}\right) \qquad \delta = \frac{\lambda}{\mu} \frac{\partial \varphi}{\partial D} \theta_1 D_1 \frac{\Omega_C^{21} \Omega_C^1}{U_C^1} \left(\frac{\phi_l^1}{\Omega_C^1} - \frac{\phi_l^{21}}{\Omega_C^{21}} \frac{\theta_2}{\theta_1}\right)$$

so that

$$k^1, \delta \gtreqless 0 \Longleftrightarrow \frac{\phi_l^1}{\Omega_C^1} \gtreqless \frac{\phi_l^{21}}{\Omega_C^{21}} \frac{\theta_2}{\theta_1}$$

The assumption of separability of preferences between goods and leisure is taken for two reasons: first of all, it allows to identify in a clear way the channels through which (nonlinear) charges on the good using the congestible network play a redistributive role, even with optimal income taxation in place. Moreover, previous literature has argued that separability between goods (including public goods and bads, such as congestion in our model) and leisure in the utility function blocks, at least with homogeneous preferences and in the presence of optimally adjusted income taxation, all redistributive properties of taxes on externality generating commodities.[19] Our results go in a different direction.

The above results suggest that, when considered as part of a general tax system, nonlinear congestion charges can help the government improve its ability to redistribute, by relaxing binding self selection constraints. This happens through two channels. Let us start by the first, identified by the term k^1: this is a type-specific "Atkinson-Stiglitz" term accounting for the distortion in the consumption of D, that is optimally introduced by modifying tax rates for the low ability individuals, in order to make the allocation of goods and income designed for them less interesting for mimicking high ability types (Stiglitz, 1982). Gahvari (2007), showed that, in the presence of nonlinear commodity taxes and when consumption of commodities is costly in time terms (with all individuals having the same consumption technology), such a distortion is called for even if utility is separable in leisure and commodities. In our model, the term differs from that of Gahvari's given that we allow for heterogeneous abilities in consuming the network good.[20] What we obtain is that this term is surely positive in the case of either positively correlated earning and consumption abilities or when the latter is the same for all the population:[21] in both cases, a downward distortion in the consumption of D for low ability types is warranted. Such distortion is likely to be stronger the larger the difference in consumption abilities (i.e. the smaller the ratio $\frac{\theta_2}{\theta_1}$). The sign of the term k_1 may be reversed if (and only if) we have a negative correlation between earning and consumption abilities.

Let us now consider the second channel, identified by the non-type specific term δ: congestion charges discourage (aggregate) consumption of the good D, reduce φ and, therefore, the unit time cost a_D. This affects the self selection constraint faced by the government. In particular, the constraint is relaxed in the case of positive correlation between earning and consumption abilities, or int he simplified case in which consumption

[19] See Cremer *et al.* (1998, 2001) and Kaplow (2006).

[20] Indeed, if all individuals had the same θ, this term would look exactly like the one presented in Gahvari (2007, Proposition 3). However, we restrict attention to the case in which consumption of the numeraire C is not costly in time terms, so $a_C = 0$. Gahvari does not impose such restriction and finds that the term has a generally ambiguous sign.

[21] Observe that, under separability, and in both cases mentioned, we surely have $\frac{\phi_l^1}{\Omega_C^1} > \frac{\phi_l^{21}}{\Omega_C^{21}}$ at a given allocation of commodities and income, since the high ability type supplies a lower amount of labour, so that $\Omega_C^1 = \Omega_C^{21}$ while $\phi_l^1 > \phi_l^{21}$. If high (earning) ability types also require less time to consume D, then this also contributes to making leisure marginally more valuable to low than to high ability types (mimickers). Moreover $\frac{\theta_2}{\theta_1} < 1$ in that case.

ability is homogenous in the population (in both cases $\delta > 0$).[22] The intuition is simple: a high earning ability individual mimicking an individual of lower ability earns the same income, consumes the same bundle of goods but (having to work less and, in the case of positively correlated abilities, incurring less time costs in using the "network" good) enjoys more leisure than the mimicked; therefore a reduction in the time cost of consuming the network good (e.g. driving home during rush hours or searching for information over the Internet) is less valuable to him than it is for the individual whom he mimicks (at a given allocation). A relaxed self selection constraint allows the government to redistribute additional amounts through income taxes and transfers: redistribution is more effectively done in a world where the consumption of goods is less costly in time terms. As a consequence, the non type specific component $\tau + \delta$ of the tax rate on D should be set strictly above a standard "Pigouvian" tax, for redistributive reasons on top of efficiency ones. This contrasts with seemingly widespread beliefs that congestion charges have a regressive impact on the economy, presented in the introduction.

Notice, however, that the sign of δ may be reversed in the case of negative correlation between earning and consumption abilities (i.e. if $\frac{\theta_2}{\theta_1} > 1$): in that case, a reduction in the amount of congestion would have an ambiguous effect: while high earning ability individuals would be working less, they would spend more time consuming good D (at a given allocation), and this could make their marginal utility of leisure possibly higher than that of the low ability individuals they want to mimick. Reducing congestion might thus tighten the self selection constraint and would be undesirable from a redistributive perspective.

Cremer *et al.* (2001) obtained that reducing environmental damage (here identified by congestion φ) may have a positive redistributive effect only the latter is a substitute for labour supply. However, they do not provide any theory explaining why it should be the case. Our paper provides such a theory for the case of congestion externalities. We think this is a contribution for the following reasons: first of all, it is not immediately clear why the amount of network congestion and labour supply should be complements (or substitutes). Take the case of roads: congestion, by increasing the time costs of commuting, may indeed be a substitute for labour supply.[23]However, while commuting is certainly a very important purpose for car trips and an important contributor to road congestion, evidence suggests that an important portion of trips, even at rush hour, are not

[22] The explanation follows that of footnote 18 above. In addition, the reduction in congestion is more valuable, at the margin, to low ability individuals due to the fact that θ_i enters a_D^i multiplicatively. If we used other functional forms (for instance, one in which $a_D^i = \theta_i + \varphi$) this particular effect may disappear, but the qualitative result would not change. What is required for the argument to be valid is that $a_D^2 \leq a_D^1$, for any level of φ.

[23] Several optimal taxation models in the literature are built on that assumption: Parry and Bento (1999) and Calthrop (2001) assume commuting to be complementary to labour supply

for commuting, but are related to shopping, visiting relatives or friends, taking kids to school, etc...[24] It is not immediately clear, then, that less road congestion should increase, rather than reduce, the attractiveness of labour supply compared to other activities. Similar questions arise when one considers congestion on telecom and data networks: many consumer applications have little to do with labour supply, but look rather complementary to leisure, such as watching videos, shopping online, playing online games, chatting with friends. The theory provided here abstracts from any exogenously assumed complementarity (or substitutabilty) between goods using the network and labour supply or leisure (neither technological nor in preferences): the interaction we are identifying comes from a different source and the substitutability of labour supply and congestion is endogenous. It comes from the fact that labour and consuming goods are activities that compete with leisure for the individual's limited time. This also explains why our results predict a redistributive role for reducing congestion even in the case of separability in preferences between goods and leisure. In fact, in our setup, the amount of congestion φ does not even enter individuals' utility: they have no taste for it. By contrast, in Cremer et *al.*, if preferences were weakly separable in leisure and goods (including public goods such as environmental quality), such role would not survive, $\delta = 0$ and Pigouvian taxes would suffice (Cremer *et al.* 2001, Proposition 5). [25]

4 Optimal investment in network capacity

In this section, we consider investment in network capacity as an abatement technology enabling the government to reduce the amount of congestion. The network is modelled as a public good whose provision is measured in capacity terms. As examples, one may think of investments in new network infrastructures but also in the maintenance of existing ones, as well as the improvement of traffic management technologies. The question is what criteria should a government caring about redistribution (on top of economic efficiency) follow when

[24] The 1995 NTPS (Figure 12, p.14) for the US reports that only around 37% of all peak hour trips are commuting trips by car. The 2001 NHTS for the US reports that only around 27% of vehicle miles travelled by car are for commuting purposes (Hu and Reuscher, 2004, Table 12), while if we restrict attention to rush hour trips the portion is around a half (Small and Verhoef, 2007). As for average daily person trips, work trips account for around 20% of them (Hu and Reuscher, 2004, Table 11). The National Travel Survey conducted in 2009 for the UK shows that among total car trips, only 31% are either commuting or business trips, while 51% are leisure or shopping trips (UK DfT, 2009, Table NTS0401). A survey conducted in London shows that around 60% of morning peak hour car trips are commuting trips, while for the evening peak hour the figure is 50% (LRC, 1994).

[25] Kreiner and Verdelin (2009), in the conclusion to their paper, suggest a redistributive effect of marginal changes in externalities when the public damage is correlated, for a given income level, with individuals' ability. This is in line with what our results.

deciding how much to invest in network capacity.[26]

In order to shed light on the issue, we may start from the FOC (6) of the government's problem. This expression, after some rearrangements, can be rewritten in the form of a modified Samuelson's rule (derivations are provided the Appendix):

PROPOSITION 3: *The optimal rule for provision network capacity K is the modified Samuelson's rule:*

$$\sum_{i=1,2} -\frac{D_i \Omega_l^i}{U_C^i}\frac{\partial \varphi}{\partial K} + \gamma = p$$

where

$$\gamma = -\frac{\lambda}{\mu}\frac{\partial \varphi}{\partial K}\frac{D^1 \Omega_C^{21} \Omega_C^1}{U_C^1}\left(\frac{\Omega_l^1}{\Omega_C^1} - \frac{\Omega_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\right)$$

so

$$\gamma \gtreqless 0 \Longleftrightarrow \frac{\Omega_l^1}{\Omega_C^1} \gtreqless \frac{\theta_2}{\theta_1}\frac{\Omega_l^{21}}{\Omega_C^{21}}$$

Now, the first term on the left hand side is the sum of the direct marginal benefits of the addition of capacity in terms of reduction of congestion. In addition, we have the term γ: this accounts for the effect that raising capacity (when taxes are optimally set) has on relaxing the self selection constraint faced by the government (notice that $\gamma = 0$ if $\lambda = 0$ so the standard Samuelson's rule applies if self selection constraints are not binding). Since the effect of additional capacity is that of reducing the aggregate amount of congestion φ, the discussion for the sign of this term follows the one for the term δ in the above section. When individuals' earning and ability in making use of the "network" commodity D are positively correlated or the latter does not vary across individuals, the sign of the term is surely positive, implying that raising capacity entails an additional extra social benefit due to the opening up of additional redistributive opportunities. The sign of γ may instead be negative in the case of a negative correlation among abilities, for the same reasons that explain why the term δ in Section 3 could be negative in such a case.

These results follow those of Kreiner and Verdelin (2009) who, generalizing those of Boadway and Keen (1993), obtained a modified Samuleson's Rule for the provision of a public good (as is network capacity in

[26] Our formulation considers the network as a public good that has to be combined with individuals' time and money to produce a given final consumption good (D in our case). This is close to Sandmo's (1973) modelling of public goods as intermediate goods for individual production.

our setup), when the marginal willingness to pay for it, at a given allocation, varies with individuals' ability (and in the presence of optimally adjusted income taxes). In our setup, such variation is due to the fact that consumption of commodities making use of the network is a costly activity in terms of time (a scarce resource for the individual) and, as such, competes for it with labour supply and leisure. Higher capacity reduces such time costs, at least as long as optimal pricing for access is in place, reducing the opportunity cost of additional labour time, which, at a given allocation, is more valuable to low ability individuals.[27]

5 Means testing and implementation of optimal congestion pricing

Congestion charges may be seen as "user fees" to be paid to be granted access to a given piece of infrastructure. The question we ask in this section is whether congestion charges should be means-tested, like other charges levied by governments to finance directly the provision of certain services, and explicitly depend on individuals' income or not. In order to answer this question, we look to implement the second best allocation using two separable functions $T(I)$ and $P(D)$ for income and car driving taxation respectively. If this is possible, then means testing is unnecessary. Otherwise, the government should use a more general tax function $\tau(D, I)$ that depends on both observable income and consumption of D: in that case, congestion charges should be conditional on income and means testing would be called for.

The analysis for this part is conducted assuming no correlation between earning and consumption abilities, so $a_D^1 = a_D^2 = a_D$. We retain the separability assumption, so the utility function takes the following form

$$U(C, D, l) = \Omega^i(C^i, D^i, 1 - a_D D^i - \frac{I^i}{w^i}) = \Omega(C^i, D^i) + \phi(1 - a_D D^i - \frac{I^i}{w^i}) \quad i = 1, 2$$

h, g and ϕ are all increasing and concave functions of the respective arguments. We assume, for simplicity, that $a_C = 0$.

[27] Indeed, in their conclusions, Kreiner and Verdelin argue that public transportation is likely to benefit, for a given income level, low ability individuals more than high ability ones, by reducing travel time to and from the workplace, leaving more time for other activities. The results we obtain develop from essentially the same intuition: however, our results suggest that the purpose for which the public good is used (whether related to the act of supplying labour or not) is not relevant: what is relevant is that additional provision reduces the time required to make use of it, for whatever purpose.

The implementation problem

The government wants to implement the second best allocation

$$A^{SB} = \left((I^1 - (T^1 + P^1), D^1, I^1); (I^2 - \left(T^2 + P^2\right), D^2, I^2)\right)$$

derived in the previous sections, using the separable tax functions $T(I)$ and $P(D)$. T^i and P^i $\quad i = 1, 2$ denote respectively the type specific payments of income and congestion taxes. Therefore $C^i = I^i - (T^i + P^i) \quad i = 1, 2$. Incentive compatibility calls for types 1 and 2 to choose payments and allocations $\left((T^1, I^1); (P^1, D^1)\right)$ and $\left((T^2, I^2); (P^2, D^2)\right)$ respectively. However, when individuals face separable tax functions, they have additional possibilities to deviate from the allocation intended for them: for instance, they may choose to consume a quantity of commodity D intended for the other type, while choosing the amount of income intended for their type. Or they could choose to mimick the other's type income, while consuming the "right" amount of D. This implies that, in order to be implementable in separable tax functions, the second best allocation A^{SB} has to respect seven incentive constraints (i.e. the standard "global" IC constraint (1) and six additional incentive constraints ensuring domination of "partial" mimicking strategies). We present the constraints in the Appendix.

Results

In the Appendix, we prove that all of the above constraints are surely satisfied at the second best allocation A^{SB}, except the constraint ensuring no partial mimicking on income by type 2 individuals. So we can say that if and only if that constraint is verified, then A^{SB} is implementable with separable tax functions, and means testing is not necessary. Unfortunately, it is difficult to identify analytical conditions ensuring that the constraint whether it will be the case or not. The issue can be analysed numerically (work in progress). However, we are able to establish that the constraint will surely fail in a particular but quite interesting case, that of quasilinearity of $U(.)$. We summarize the results in the following Proposition:

PROPOSITION 4: *Assume that the government wanted to implement the second best allocation A^{SB} with separable tax functions $T(I)$ and $P(D)$. Then:*

- *A necessary and sufficient conditon for A^{SB} to be implementable in separable functions is* *that it satisfies the PMI2 constraint.*

- *The constraint is surely not satisfied if $U(.) = C + g(D) + \phi(l)$, i.e. preferences are separable and quasilinear. In that case, then, means testing is necessary*

6 Conclusion

This paper has considered the problem of optimal congestion pricing for access to a public network in the presence of optimal income taxation, when the government has redistributive concerns and faces self selection constraints. We consider this a novel approach in the study of congestion pricing, which may contribute to the still open debate on the redistributive implications of such measures. The model also extends previous literature on optimal taxation by modelling congestion as deteriorating individuals' consumption technology (i.e. increasing the time required for consumption) of commodities making use of the network and allowing for heterogeneities in individuals' consumption technologies (which may capture, for instance, relevant issues such as "digital inequalities" that may be relevant in determining individuals' ease and effectiveness in using digitalised services and information provided through the Internet).

The results we obtain are the following: first, in the presence of optimally set income taxes, nonlinear congestion charges are optimal. Whether the marginal tax rate (or price) faced by low ability individuals should be higher or lower than that faced by high ability individuals depends on the correlation between earning ability and ability in using the network good. When these are positively correlated or when the latter is undifferentiated across the population, then low ability individuals should face higher marginal tax rates. When correlation is negative, the opposite may occur. Second, reducing congestion can help achieve a higher level of redistribution, by letting the government relax the self selection constraints it faces. This effect is stronger (resp. weaker) if earning and consumption ability are positively (negatively) correlated. Therefore, all individuals should face a tax on congestion generating goods (or activities) not only for "Pigouvian" reasons, but also to improve the government's redistributive capabilities. This conclusion contrasts with regressivity concerns on congestion charges that seem to be quite persistent among policy makers and the public. This indication may be even more cogent in the presence of a positive correlation between earning ability and ability in making use of network goods, such as Internet services: this is interesting because it suggests that if, as seems likely, more "skilled" individuals have both higher earning ability and are better suited to exploit network goods and services, then consumption of such goods for low ability individuals should, perhaps counterintuitively, be distorted even more in order to increase the amount of redistribution that can be put in place. Moreover, in that case, reducing network congestion, either by raising congestion charges or by investing in additional network capacity, has an even stronger effect in increasing the government's redistributive capabilities. Both such measures would be even more desirable from a social welfare perspective.

Finally, we have considered the problem of implementing the optimal (second best) allocation with separable tax functions forcommodities using the network (and contributing to congestion) and income. If such feat is feasible, then it means that the optimal allocation can be achieved without making use of means tested congestion charges. Our results suggest that means tested congestion charges may be necessary, for instance in the case of quasilinear preferences. Numerical analysis is however needed to investigate the issue in more depth.

Our model is very simple: we have considered a discrete, two-type model of the economy (though our results would generalize to n-types setups), with only two goods. We believe, however, that the results would generalize to environments with multiple types and goods. We have also considered individuals differing only in earning and consumption ability, disregarding differences in tastes. Moreover, our results rest on two important assumptions: the first is that the government is able to fully determine the pricing schedule for the goods using the network, which may not always be the case in reality. However, one should interpret the results as indicative of what a welfare maximising government should do when being interested in both curbing excessive congestion and protecting the less fortunate, if it faced no constraint (except self selection) in implementing the optimal allocation. The second important assumption is that the government is able to adjust, when introducing congestion pricing, the income tax in an optimal way. This may not always be the case, for instance, with multiple levels of government involved in setting charges and income taxes. Relaxing these assumptions looks like an interesting and challenging path for future research.

References

[1] Atkinson A. and Stiglitz J. (1976), "The Design of Tax Structure: Direct versus Indirect Taxation". *Journal of Public Economics* 6 (1976): 55–75.

[2] Becker G. (1965) A theory of the allocation of time, *Economic Journal* 75, 493–517

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

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

[5] Bovenberg L. and Goulder L. (2002), Environmental taxation and regulation. in A. J. Auerbach & M. Feldstein (ed.) *Handbook of Public Economics*, chapter 23, 1471-1545.

[6] Calthrop E. (2001), On subsidizing auto commuting. CESIfo Working paper N.566

[7] Cremer H., Gahvari F. and Ladoux N. (2001), Second best pollution taxes and the structure of preferences, *Southern Economic Journal*, 68-2, 258-280

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

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

[10] Department for Transport UK (2009), National Travel Survey.

[11] DiMaggio P., Hargittai E., Celeste C. and Shafer S. (2004), Digital Inequality: from unequal access to differentiated use. In K. M. Neckerman (ed.) *Social Inequality*. Russel Sage. New York.

[12] Engel E., Fischer R. and Galetovic A. (2009), On the efficient provision of roads. Working Paper, Yale University

[13] Freese J., Rivas S. and Hargittai E. (2006), Cognitive ability and Internet use among older adults. *Poetics*, Volume 34, Issues 4-5, Pages 219-318 (August-October 2006)

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

[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. Harvard Law School Discussion Paper n. 547

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

[18] Kreiner C. T. and Verdelin N. (2009), Optimal provision of public goods: a synthesis, forthcoming in *The Scandinavian Journal of Economics*

[19] Leuthold J. H. (1796), The optimal congestion charge when equity matters, *Economica*, 43, 77-82

[20] Levinson D.(2009), Equity Effects of Road Pricing: A Review. *Transport Reviews.* 30(1) 33-57

[21] London Research Center (1994), Travel in London: London Area Traffic Survey 1991. HMSO, London.

[22] MacKie-Mason J.K. and Varian H.R. (1995), "Pricing the Internet" in B.Kahin & J.Keller, eds, *Public Access to the Internet*, Prentice Hall, Englewood Cliffs.

[23] 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 (Jun., 1997), pp. 261-279

[24] Sandmo A. (1973), Public goods and the technology of consumption, *Review of Economic Studies*, Vol. 40, N. 4, pp. 517-528

[25] Parry I. (2008), Pricing Urban Congestion. *Resources for the Future Discussion Paper N. 08-35.*

[26] Parry I. and Bento A. (1999), Revenue Recycling and the Welfare Effects of Road Pricing. *Scandinavian Journal of Economics*

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

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

Appendix

Proof of Proposition 1 and 2

Start from the FOC (2)-(5) and rearrange to get to

$$\frac{U_D^1}{U_C^1} = \frac{1 + \frac{\lambda}{\mu} U_D^{21} + \frac{\sum_{i=1,2} \frac{\partial \varphi}{\partial D} \theta_i D_i U_l^i}{\mu} - \frac{\lambda}{\mu} \frac{\partial \varphi}{\partial D} \theta_2 \left(D_1 U_l^{21} - D_2 U_l^2 \right)}{1 + \frac{\lambda}{\mu} U_C^{21}}$$

multiplying both sides by $1 + \frac{\lambda}{\mu} U_C^{21}$ and rearranging we get

$$\frac{U_D^1}{U_C^1} = 1 + \frac{\lambda}{\mu} U_C^{21} \left(\frac{U_D^{21}}{U_C^{21}} - \frac{U_D^1}{U_C^1} \right) + \frac{\sum_{i=1,2} \frac{\partial \varphi}{\partial D} \theta_i D_i U_l^i}{\mu} - \frac{\lambda}{\mu} \frac{\partial \varphi}{\partial D} \theta_2 \left(D_1 U_l^{21} - D_2 U_l^2 \right) \tag{7}$$

Similarly, we get

$$\frac{U_D^2}{U_C^2} = 1 + \frac{\sum_{i=1,2} \frac{\partial \varphi}{\partial D} \theta_i D_i U_l^i}{\mu} - \frac{\lambda}{\mu} \frac{\partial \varphi}{\partial D} \theta_2 \left(D_1 U_l^{21} - D_2 U_l^2 \right)$$

We now set out to derive the terms τ and δ that we identify in the tax formulas above: rewrite $\frac{\sum_{i=1,2} \frac{\partial \varphi}{\partial D} \theta_i D_i U_l^i}{\mu}$ as

$$\frac{\sum_{i=1,2} \frac{\partial \varphi}{\partial D} \theta_i D_i U_l^i}{\mu} = \frac{\partial \varphi}{\partial D} \left(\frac{D_1 \theta_1 U_l^1 U_C^1}{\mu U_C^1} + \frac{D_2 \theta_2 U_l^2 U_C^2}{\mu U_C^2} + \frac{\lambda \theta_2 D_2 U_l^2}{\mu} - \frac{\lambda \theta_2 D_1 U_l^{21}}{\mu} - \frac{\lambda}{\mu} \theta_2 \left(D_2 U_l^2 - D_1 U_l^{21} \right) \right)$$

now, using the FOC (2) we have

$$\frac{\frac{\partial \varphi}{\partial D} \theta_1 D_1 U_l^1 U_C^1}{\mu U_C^1} = \frac{\frac{\partial \varphi}{\partial D} \theta_1 D_1 U_l^1 \mu}{\mu U_C^1} + \frac{\frac{\partial \varphi}{\partial D} \theta_1 D_1 U_l^1 \lambda U_C^{21}}{\mu U_C^1}$$

and using FOC (3) we have

$$\frac{\frac{\partial \varphi}{\partial D} \theta_2 D_2 U_l^2 U_C^2}{\mu U_C^2} = \frac{\frac{\partial \varphi}{\partial D} \theta_2 D_2 U_l^2 \mu}{\mu U_C^2} - \frac{\frac{\partial \varphi}{\partial D} \theta_2 D_2 U_l^2 \lambda U_C^2}{\mu U_C^2}$$

so that we can rewrite

$$\frac{\sum_{i=1,2}\frac{\partial\varphi}{\partial D}\theta_i D_i U_l^i}{\mu}=\frac{\partial\varphi}{\partial D}\left(\frac{\theta_1 D_1 U_l^1}{U_C^1}+\frac{\theta_2 D_2 U_l^2}{U_C^2}+\frac{\lambda\theta_1 D_1 U_l^1 U_C^{21}}{\mu U_C^1}-\frac{\lambda\theta_2 D_1 U_l^{21}}{\mu}+\frac{\lambda}{\mu}\theta_2\left(D_1 U_l^{21}-D_2 U_l^2\right)\right)$$

finally, replacing the above expression in (7) we have

$$\frac{U_D^1}{U_C^1}=1-\frac{\lambda}{\mu}U_C^{21}\left(\frac{U_D^1}{U_C^1}-\frac{U_D^{21}}{U_C^{21}}\right)+\tau+\frac{\lambda\frac{\partial\varphi}{\partial D}\theta_1 D_1}{\mu}U_C^{21}\left(\frac{U_l^1}{U_C^1}-\frac{U_l^{21}}{U_C^{21}}\frac{\theta_2}{\theta_1}\right)$$

and

$$\frac{U_D^2}{U_C^2}=1+\tau+\frac{\lambda\frac{\partial\varphi}{\partial D}\theta_1 D_1}{\mu}U_C^{21}\left(\frac{U_l^1}{U_C^1}-\frac{U_l^{21}}{U_C^{21}}\frac{\theta_2}{\theta_1}\right)$$

where

$$\tau\equiv\frac{\frac{\partial\varphi}{\partial D}\theta_1 D_1 U_l^1}{U_C^1}+\frac{\frac{\partial\varphi}{\partial D}\theta_2 D_2 U_l^2}{U_C^2}$$

is the Pigouvian term as described in the text. Up to now, we have essentially followed the derivations in Cremer *et al.* (2001). It now only remains to use the fact that

$$U_j^i\equiv\Omega_j^i-a_j\Omega_l^i\quad i=1,2\quad j=C,D$$

and

$$U_l^i=\Omega_l^i\quad i=1,2$$

to be able to write, after some rearrangements,

$$\frac{\lambda\frac{\partial\varphi}{\partial D}\theta_1 D_1}{\mu}U_C^{21}\left(\frac{U_l^1}{U_C^1}-\frac{U_l^{21}}{U_C^{21}}\frac{\theta_2}{\theta_1}\right)=\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial D}\theta_1 D_1\frac{\Omega_C^{21}\Omega_C^1}{U_C^1}\left(\frac{\Omega_l^1}{\Omega_C^1}\left(1-\frac{\Omega_l^{21}}{\Omega_C^{21}}a_c\right)-\frac{\Omega_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\left(a_C-1\right)\right)$$

the sign of this term is not immediately determined. Assuming that $a_C=0$ yields however

$$\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial D}D_1\frac{\Omega_C^{21}\Omega_C^1}{U_C^1}\left(\frac{\Omega_l^1}{\Omega_C^1}\left(1-\frac{\Omega_l^{21}}{\Omega_C^{21}}a_c\right)-\frac{\Omega_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\left(a_C-1\right)\right)=\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial D}\theta_1 D_1\frac{\Omega_C^{21}\Omega_C^1}{U_C^1}\left(\frac{\Omega_l^1}{\Omega_C^1}-\frac{\Omega_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\right)$$

which is the definition of the term δ given in the text.

We now focus on rewriting the term k_1 in the tax formula for the individual of type 1. We have

$$\frac{\lambda}{\mu}U_C^{21}\left(\frac{U_D^{21}}{U_C^{21}}-\frac{U_D^1}{U_C^1}\right)=\frac{\lambda}{\mu U_C^1}\left(\left(\Omega_D^{21}-a_D^2\Omega_l^{21}\right)\left(\Omega_C^1-a_C\Omega_l^1\right)-\left(\Omega_D^1-a_D^1\Omega_l^1\right)\left(\Omega_C^{21}-a_C\Omega_l^{21}\right)\right)$$

which, after some rearrangements, can be written as

$$\frac{\lambda}{\mu U_C^1}\left[\Omega_C^1\Omega_C^{21}\left(\frac{\Omega_D^{21}}{\Omega_C^{21}}-\frac{\Omega_D^1}{\Omega_C^1}\right)+\Omega_l^1\Omega_D^{21}\left(a_D^2\frac{\Omega_C^{21}}{\Omega_D^{21}}-a_C\right)-\Omega_l^{21}\Omega_D^1\left(a_D^2\frac{\Omega_C^1}{\Omega_D^1}-a_C\right)+a_C\left(a_D^2-a_D^1\right)\Omega_l^{21}\Omega_l^1\right]$$

Up to now, we have not used the separability assumption. Let us assume indeed that preferences are separable as in the ASSUMPTION in the text. Then

$$\frac{\Omega_C^1}{\Omega_D^1}=\frac{\Omega_C^{21}}{\Omega_D^{21}}$$

so that the expression above gives

$$\frac{\lambda}{\mu U_C^1}\left[a_D^1\Omega_C^1\Omega_C^{21}\left(\frac{\Omega_l^1}{\Omega_C^1}-\frac{\Omega_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\right)+a_C\Omega_D^1\Omega_D^{21}\left(\frac{\Omega_l^{21}}{\Omega_D^{21}}-\frac{\Omega_l^1}{\Omega_D^1}\right)+a_C\left(a_D^2-a_D^1\right)\Omega_l^{21}\Omega_l^1\right]$$

now, using $a_C=0$, we obtain that the expression above simplifies to

$$\frac{\lambda}{\mu U_C^1}\left[a_D^1\Omega_C^1\Omega_C^{21}\left(\frac{\Omega_l^1}{\Omega_C^1}-\frac{\Omega_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\right)\right]$$

which is the expression for the term k_1 in the text.

Proof of Proposition 3

To derive the modified Samuleson's rule presented in the text, start from the FOC (6) and add and substract to the left hand side the following

$$\lambda\frac{\partial\varphi}{\partial K}D_1U_C^{21}\frac{U_l^1}{U_C^1}$$

so that (6) can be rewritten as

$$-\left(U_C^1-\lambda U_C^{21}\right)\left(\frac{\partial\varphi}{\partial K}D_1\frac{U_l^1}{U_C^1}\theta_1\right)+\left(1+\lambda\right)U_C^2\left(-\frac{\partial\varphi}{\partial K}\theta_2D_2\frac{U_l^2}{U_C^2}\right)+\lambda U_C^{21}\left(\frac{\partial\varphi}{\partial K}D_1\frac{U_l^{21}}{U_C^{21}}\theta_2-\frac{\partial\varphi}{\partial K}D_1\frac{U_l^1}{U_C^1}\theta_1\right)-\mu p=0$$

now, using expressions (2)-(3)we can rewrite the above as

$$-\mu\left(\frac{\partial\varphi}{\partial K}D_1\frac{U_l^1}{U_C^1}\theta_1\right)+\mu\left(-\frac{\partial\varphi}{\partial K}D_2\frac{U_l^2}{U_C^2}\theta_2\right)+\lambda U_C^{21}\left(\frac{\partial\varphi}{\partial K}D_1\frac{U_l^{21}}{U_C^{21}}\theta_2-\frac{\partial\varphi}{\partial K}D_1\frac{U_l^1}{U_C^1}\theta_1\right)-\mu p=0$$

which can finally be rewritten as

$$\sum_{i=1,2}-\frac{D_i\Omega_l^i\theta_i}{U_C^i}\frac{\partial\varphi}{\partial K}-\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial K}D^1U_C^{21}\left(\frac{U_l^1}{U_C^1}\theta_1-\frac{U_l^{21}}{U_C^{21}}\theta_2\right)=p$$

now, we can rewrite the second term on the right hand side as

$$-\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial K}D^1U_C^{21}\left(\frac{U_l^1}{U_C^1}\theta_1-\frac{U_l^{21}}{U_C^{21}}\theta_2\right)=-\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial K}\frac{D^1}{U_C^1}\left(\theta_1U_l^1\Omega_C^{21}-a_C\theta_1\Omega_l^{21}\Omega_l^1+\theta_2\Omega_l^{21}a_C\Omega_l^1-\theta_2\Omega_l^{21}a_C\Omega_C^1\right)=$$

$$=-\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial K}\frac{D^1}{U_C^1}\Omega_C^{21}\Omega_C^1\left(\left(\frac{\phi_l^1}{\Omega_C^1}-\frac{\phi_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\right)+a_c\frac{\Omega_l^{21}\Omega_l^1}{\Omega_C^{21}\Omega_C^1}\left(1-\frac{\theta_2}{\theta_1}\right)\right)$$

finally using $a_C=0$, we obtain that the expression above simplifies to

$$-\frac{\lambda}{\mu}\frac{\partial\varphi}{\partial K}\frac{D^1}{U_C^1}\Omega_C^{21}\Omega_C^1\left(\frac{\phi_l^1}{\Omega_C^1}-\frac{\phi_l^{21}}{\Omega_C^{21}}\frac{\theta_2}{\theta_1}\right)$$

which is the term γ presented in the text. The modified Samuelson's rule obtains as presented in the text then.

Proof of Proposition 4

We define $P^1=r^1$ and $P^2=P^1+r^2$ the payments the consumer has to make to consume the quantities D^1 and D^2 respectively, such that he is indifferent (given he has choosen the "right" post-tax income level) between choosing the quantity of D intended for him and the one intended for the lower ability individual (or none at all if he is type 1). In particular, r_1 is such that

$$\Omega(I^1-T^1-P^1,D^1)+\phi(1-a_DD^1-\frac{I^1}{w^1})=\Omega(I^1-T^1,0)+\phi(1-\frac{I^1}{w^1}) \quad (8)$$

and r_2 is such that

$$\Omega(I^2 - T^2 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^2}{w^2}) = \Omega(I^2 - T^2 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^2}{w^2}) \tag{9}$$

We then take the following assumptions:

Assumptions:

- $D^2 > D^1$
- $r^2 > 0$
- $I^2 > I^1$
- $I^2 - T^2 > I^1 - T^1$

The constraints

The seven constraint the SB allocation A^{SB} will have to satisfy in order for implementability in separable tax functions to be feasible are the following

- Global IC

  $$\Omega(I^2 - T^2 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^2}{w^2}) \geq \Omega(I^1 - T^1 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^1}{w^2}) \tag{10}$$

  which holds, by assumption, at equality in the second best allocation A^{SB}, given that it is binding.

- "Participation" type 2:

  $$\Omega(I^2 - T^2 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^2}{w^2}) \geq \Omega(I^2 - T^2, 0) + \phi(1 - \frac{I^2}{w^2}) \tag{11}$$

  this constraint ensures that individuals of type 2 will consume the intended quantity D^2 rather than not consume any D at all.

- "Participation" type 1:

$$\Omega(I^1 - T^1 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^1}{w^1}) \geq \Omega(I^1 - T^1, 0) + \phi(1 - \frac{I^1}{w^1}) \tag{12}$$

  same as the above constraint, for individuals of type 1.

- Partial mimicking on I, type 2:

$$\Omega(I^2 - T^2 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^2}{w^2}) \geq \Omega(I^1 - T^1 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^1}{w^2}) \tag{13}$$

- Partial mimicking on I, type 1:

$$\Omega(I^1 - T^1 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^1}{w^1}) \geq \Omega(I^2 - T^2 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^2}{w^1}) \tag{14}$$

- Partial mimicking on D, type 2:

$$\Omega(I^2 - T^2 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^2}{w^2}) \geq \Omega(I^2 - T^2 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^2}{w^2}) \tag{15}$$

- Partial mimicking on D, type 1:

$$\Omega(I^1 - T^1 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^1}{w^1}) \geq \Omega(I^1 - T^1 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^1}{w^1}) \tag{16}$$

Simplification of the problem: By definition of r^1 and r^2 the constraints (15) and (12) are satisfied at equality at the second best allocation A^{SB}. We can then prove that also (11) is surely satisfied, at least as long as functions Ω and ϕ satisfy standard Inada conditions. We can also prove that (15) implies (14). Finally, we can prove that (13) implies (16). (Proof of all these results is available upon request). Therefore, the only relevant constraint to be verified is (13): that the constraint is satisfied at A^{SB} is therefore a necessary and sufficient condition for implementability of A^{SB} with separable tax functions.

Validity of (13) **at** A^{SB} We know that

$$\Omega(I^2 - T^2 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^2}{w^2}) = \Omega(I^1 - T^1 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^1}{w^2}) \tag{17}$$

at A^{SB}. Therefore, (13) can be rewritten as

$$\Omega(I^1 - T^1 - P^1, D^1) + \phi(1 - a_D D^1 - \frac{I^1}{w^2}) \geq \Omega(I^1 - T^1 - P^2, D^2) + \phi(1 - a_D D^2 - \frac{I^1}{w^2}) \tag{18}$$

assume, to simplify the problem further, that the utility function is of the form

$$f(C) + g(D) + \phi(l)$$

with all subfunctions being increasing and (nonstrictly) concave in their respective arguments. Constraint (18) rewrites as

$$f(I^1 - T^1 - P^1) - f(I^1 - T^1 - P^2) \geq g(D^2) - g(D^1) + \phi(1 - a_D D^2 - \frac{I^1}{w^2}) - \phi(1 - a_D D^1 - \frac{I^1}{w^2})$$

now, using the partial mimicking on D for type 2 (15) we have that

$$g(D^2) - g(D^1) = f(I^2 - T^2 - P^1) - f(I^2 - T^2 - P^2) + \phi(1 - a_D D^1 - \frac{I^2}{w^2}) - \phi(1 - a_D D^2 - \frac{I^2}{w^2})$$

which replaced in the constraint above, allows us to rewrite (18) as

$$\begin{aligned} \big(f(I^2 - T^2 - P^2) - f(I^1 - T^1 - P^2)\big) - \big(f(I^2 - T^2 - P^1) - f(I^1 - T^1 - P^1)\big) \quad &\geq \\ \left(\phi(1 - a_D D^1 - \frac{I^2}{w^2}) - \phi(1 - a_D D^2 - \frac{I^2}{w^2})\right) - & \\ + \left(\phi(1 - a_D D^2 - \frac{I^1}{w^2})\phi(1 - a_D D^1 - \frac{I^1}{w^2}) - \phi(1 - a_D D^2 - \frac{I^1}{w^2})\right) & \end{aligned} \tag{19}$$

now, by concavity of both $f(.)$ and $\phi(.)$, both sides of the equality are positive. One can, however, make the observation that the left hand side is likely to be larger than the right hand side if $f(.)$ is strongly concave (i.e.

there are strong income effects) compared to $\phi(.)$, and viceversa. Indeed, we prove below that in the case income effects are absent, i.e. if $f(.)$ is linear and preferences are quasilinear, the constraint (19) is surely not verified.

Proof that (13) ia not valid at A^{SB} if preferences are quasilinear

Start from the constraint rewritten in the form (19) as done above. Assume $f(C) = \alpha C \quad \alpha > 0$. We obtain that the left hand side is equal to zero, evidently. Given that, by concavity of $\phi(.)$, the right hand side of (19) is strictly positive, then the inequality cannot be verified. (13) cannot hold.