DEPARTMENT OF ECONOMICS

The political economy of pricing and capacity decisions for congestible local public goods in a federal state

Bruno De Borger & Stef Proost

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

Stadscampus
Prinsstraat 13, B.226
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ECONOMICS

The political economy of pricing and capacity decisions for congestible local public goods in a federal state

Bruno De Borger & Stef Proost

RESEARCH PAPER 2013-020
SEPTEMBER 2013

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@uantwerpen.be

D/2013/1169/020

The political economy of pricing and capacity decisions for congestible local public goods in a federal state

Bruno De Borger and Stef Proost (*)

Abstract

This paper studies the political economy of pricing and investment for excludable and congestible public goods in a federal state. Currently, we observe a wide variety of practices, ranging from federal gasoline taxes and road investment to the local supply of -- and sometimes free access to -- libraries, parking spaces and public swimming pools. The two-region model we develop allows for spill-overs between regions, it takes into account congestion, and it captures both heterogeneity between and within regions. Regional decisions are taken by majority voting; decisions at the federal level are taken either according to the principle of a minimum winning coalition or through cooperative bargaining. We have the following results. First, when users form the majority in at least one region, decentralized decision making performs certainly better than centralized decision making if spill-overs are not too large. Centralized decisions may yield higher welfare than decentralization only if users have a large majority and the infrastructure in a given region is intensively used by both local and outside users. Second, if non-users form a majority in both regions, centralized and decentralized decision making yield the same socially undesirable outcome, with prices that are much too high. Third, both bargaining and imposing uniform price restrictions across regions improve the performance of centralized decisions. Fourth, the performance of decentralized supply is strongly enhanced by local self-financing rules; it prevents potential exploitation of users within regions. Self-financing rules at the central level are not necessarily welfare-improving. Finally, the results of this paper contribute to a better understanding of actual policy-making.

Keywords: congestible local public goods, pricing, capacity decisions, fiscal federalism
JEL-codes: D62, H23, R41, R48

(*) De Borger : University of Antwerp (bruno.deborger@ua.ac.be); Proost (stef.proost@kuleuven.be). The paper was presented at the IEB-conference The Political Economy of Decentralization (Barcelona, June 2013), the conference of the International Association of Transport Economists (ITEA, Northwestern University, July 2013), and at the universities of Antwerp and Leuven. For useful suggestions we are grateful to Jan Brueckner, Federico Boffa, Massimo Bordignon, Eckhard Janeba and Amedeo Piolatto.

1. Introduction

This paper studies the political economy of pricing and investment decisions for congestible local public goods in a federal state. The model allows for interregional spill-overs, it takes into account congestion, and it captures heterogeneity both between and within regions. Although it is quite generally applicable to many local congestible public facilities such as libraries, public transport, museums, public swimming facilities, etc., one of our main motivations for this paper is the problem of providing and pricing road infrastructure in federal states[1]. In most countries, the provision and financing of urban and regional roads share some remarkable characteristics. First, with the exception of parking charges, local pricing instruments (taxes, user fees, road pricing) are almost never used, neither to finance regional roads nor -- despite some recent successes in London and Stockholm -- to control local congestion and pollution problems. Second, many (but not all) countries rely on pricing instruments that are uniform across regions; such instruments are obviously not well suited to deal with spatially differentiated conditions. European countries, for example, heavily rely on fuel taxes. These are typically determined at the national level, even in explicitly federal states such as Belgium and Spain. In the US, however, fuel taxes do differ between states, although this differentiation does not necessarily match local conditions. Third, often there is heavy involvement of the federal government in capacity decisions and in financing such infrastructure. This is not only true in the case of cross-jurisdictional infrastructure such as the interstate highway system in the US (see, for example, Levinson (2002)) or the Trans European networks in the EU (see Proost et al. (2013)), it also applies to urban or regional projects with rather localized benefits (Knight (2004)).[2]

Although our focus will be on pricing and capacity of transport infrastructure decisions, it is clear that the problem of dividing responsabilities over different levels of government also exists for many other public services. A recent study by the OECD (see Blöchliger (2008)) analyzes how decision authority and financing responsability for a number of public services (including education, hospitals, public transport and nursing homes) is divided between different layers of government in different countries. As in transportation,

[1] The explicit reference to ‘federal’ states is made for convenience. In principle, the model applies to all political structures with multi-layered governments. For example, it applies equally well to decision making processes of a regional government versus local urban governments.

[2] There is also clear evidence that governance and financing responsability is not constant but evolves over time. The information provided Xie and Levinson (2009) suggests that the role of local versus higher level governments moved from strong decentralization in the early years of infrastructure provision towards more centralization when transport systems developed. In recent decades -- due to increasing interjurisdictional demands, aging infrastructure and rising congestion -- the impact of higher level governments (state and federal) has been steadily increasing.

substantial differences are seen, both between different public services and across countries. Uniform pricing is often observed, and the use of local user fees is typically quite limited; if they are used (as in the case of public transport in most countries), they cover only a small share of expenditures. The study further finds a widespread use of restrictions on users; for example, users are often limited to the services provided in the own jurisdiction.

The examples mentioned above raise a number of questions. For example, under what conditions are the outcomes of decentralized political decision-making socially preferable to those of centralization? Under what conditions does the use of uniform pricing instruments improve the welfare performance of political decisions? Are there other institutional restrictions (for example, earmarking of transport tax revenues) that are socially desirable, in the sense that they improve the outcomes of the political process? Why are decisions concerning local infrastructure projects (both user pricing and capacity) often taken at the federal level? Why do we observe so few cases of pricing of road infrastructure that spatially differentiate according to local conditions? Why is there widespread use of parking charges in many cities and municipalities, but are there almost no examples of some form of road pricing (cordon pricing, electronic road pricing)?

In this paper, we develop a political economy model of pricing congestible public goods in a two-region federation that may cast some light on these issues. In our model, each regional road is used both by local inhabitants and by users from the other region, leading to spill-overs between regions. Furthermore, in each region the voting population involved in decision-making consists of both users and non-users. The model captures congestion of the local public good (road infrastructure) and it introduces pricing to deal with the congestion externality. The user fees are returned to the population via reductions in local or centralized head taxes. The model introduces heterogeneity both within regions (users versus non-users) and between regions (both the level of spill-overs and the share of users may differ between regions). We assume that decentralized (urban or regional) decisions are taken by simple majority voting. At the central level, decisions are taken in a legislature of regionally elected representatives; both decision-making by a minimum winning coalition (a non-cooperative process) and through bargaining (a cooperative process) are studied. Moreover, we consider the role of several constitutional restrictions (uniform pricing, self-financing rules) on the performance of the decision-making process.

Within this setting, we first study the problem of pricing existing capacity, focusing on specific questions. How would, for the current road network, user prices be determined by regional and federal authorities? Under what conditions are decentralized regional decisions welfare-superior to centralized decisions? How can restrictions that lead to favorable welfare outcomes be embedded in federal constitutions; i.e., under what conditions will regions voluntarily transfer the decision-making power to the central level? Next, we extend the model to joint decisions on pricing and capacity provision, focusing on the implications of earmarking of revenues and regional or federal budgetary restrictions for the relative welfare performance of decentralized and centralized decisions.

The heterogeneity captured by the model introduces two types of potential conflict. On the one hand, depending on the characteristics of the regions and on how political decisions are made at the central level, elected representatives may be biased towards favoring their own region (for convincing empirical evidence see, for example, Knight (2004)). On the other hand, within each region, the policy preferences of the two groups of voters we consider (users and non-users) strongly differ, and this will show up both under decentralized and centralized decisions. It is the interaction of the two types of heterogeneity that drives many of our results.

A brief description of the main results follows. Consider pricing decisions for a given infrastructure. First, if there are no spill-overs, we show that a sufficient condition for decentralized decision making to perform better than centralized decisions taken by a minimum winning coalition is for users to have the majority in at least one of the two regions. However, even if substantial spill-overs exist, decentralization in many cases yields higher welfare than centralized decision-making. More precisely, centralized decisions are to be preferred only when two conditions simultaneously hold: users have large majorities, and spill-overs are such that the demand for road use in a given region is approximately equal for local and foreign users. Second, if non-users form a majority in both regions, centralized decision making and decentralized decision making yield the same outcomes. Importantly, in welfare terms these outcomes are equally undesirable, with prices that are much too high. Third, both bargaining between elected regional representatives and requiring user prices to be uniform across regions greatly improve the efficiency of centralized decision making. Whether they do better than decentralized decision-making depends: for symmetric regions, we find that both bargaining and uniform pricing yield higher welfare than decentralized outcomes if drivers have large majorities in the regions and there are large spill-overs. Decentralization is better otherwise.

Fourth, we show that if regions are symmetric and drivers have a majority, both regions will agree to transfer decision power to the central level if constitutional arrangements either impose price uniformity across regions or an explicit bargaining decision-making process. However, if non-users are a majority in a given region we find that they will never agree to transfer decision power to the central level. Furthermore, we also show that, to avoid exploitation of users by non-users, restrictions may have to be introduced that limit tolls to be equal to or below marginal external cost.

Fifth, we find that imposing self-financing rules on the individual regions strongly enhances the performance of both decentralized and centralized decision making. In fact, under a mild additional assumption, they allow to attain the first-best social optimum. One of the main advantages of a regional self-financing rule is that it protects users from being exploited by non-users. Interestingly, a federal budget restriction does not necessarily produce favorable welfare outcomes.

Although one should be careful with 'explaining' the real world on the basis of very stylized theoretical models, the analysis of this paper has some relevance to understand actual policy making. For example, as we will argue below, it may contribute to understanding why some decisions are taken at the central level whereas others are not, why we observe so many instances of uniform pricing, why the use of user fees for road infrastructure is not widespread whereas parking charges are, why decentralization is often accompanied by user restrictions that limit the consumer's choice to the regional market, etc.

The paper builds upon several strands of literature. First, it obviously relates to the literature on the provision of local public goods. For a long time, Oates' (1972) decentralization theorem was the dominant paradigm. It suggests that, unless spill-overs of the benefits of the public good to other regions are substantial, supply decisions should be left to local authorities. Intuitively, decentralized decisions allow variation in the supply of public goods in function of local preferences, and they have an accountability advantage. Moreover, when citizens can move easily between regions, mobility between regions makes regions more homogeneous; this enhances the match between local preferences and public good supply (Tiebout (1956)). It is only when benefits of the local public good spill over to other regions that the efficiency advantages of the decentralized solution are watered down. With identical regions, spill-overs imply that centralization in fact outperforms decentralization. With heterogeneous regions, the advantages of decentralization (accounting for taste differences) have to be traded off against the disadvantages (regions ignore spill-overs). Centralization is better given a sufficiently high level of spill-overs.

When applied to infrastructure decisions, however, the decentralization theorem did not receive much empirical support (see, for example, Hulten and Schwab (1997)). Moreover, the 'second generation' literature on fiscal federalism -- initiated by, among others, Persson and Tabellini (2000), Lockwood (2002) and Besley and Coate (2003) -- casts serious doubt on the implicit assumption underlying the theorem, viz., that centralized decisions necessarily imply uniform provision of the public good across regions. This assumption is unattractive. Why would the central government impose it, knowing that regions are not identical and may in fact prefer different levels of the public good? Moreover, it was found to be inconsistent with the available empirical evidence (see, for example, Knight (2004)). Equally important, the second generation literature has taken a political economy approach, focusing on both cooperative (for example, legislative bargaining) and non-cooperative (for example, decisions according to a minimum winning coalition) decision-making procedures. Recent contributions include, among many others, Redoano and Scharf (2004), Oates (2005), and Hatfield and Padro i Miquel (2012).

A number of papers have endogenized the division of authority over different layers of government, and emphasized the role of constitutional constraints. For example, Lörz and Willmann (2005) add a constitutional bargaining stage where regions negotiate the degree of centralization (in essence, what goods will be supplied centrally) as well as the associated regional cost shares (modeled by introducing side-payments between regions), showing that the level of centralization will be inefficiently low. Hickey (2013) analyzes under what federal institutions regions will agree to transfer the decisions on local public goods to a federal government. He shows that uniform taxation facilitates federation formation in a unicameral system. More generally, both uniform taxation and federal bicameralism are institutions leading to more federal competence. Most recently, Kessler (2013) studies the role of communication in federal political structures, showing that uniform provision of local public goods may be the result of the difficulties of credible transmission of information from the regional to the federal level.

Our paper is probably most closely in spirit to the seminal paper by Besley and Coate (2003)[3]. They let regional decisions be taken by majority voting. Decisions at the federal level are made by a legislature of locally elected representatives, allowing for different political mechanisms to reach conclusions. Within this framework they study the provision of an uncongested local public good that is financed by a head tax, and they compare decentralized and centralized solutions. Unlike the standard approach, with identical districts and spill-

[3] The papers by Lockwood (2002) and Besley and Coate (2003) are complementary. The former deals with decision-making on agendas of region-specific projects. The latter considers both cooperative and non-cooperative behavior in deciding on quantities of local public goods, assuming very specific central decision-making rules. Among others, they emphasize that even cooperative centralized decision-making leads to inefficiencies due to strategic delegation of elected representatives.

overs, they find that decentralization may yield higher welfare than centralization. With non-homogenous regions, centralized decisions perform better for sufficiently high levels of spill-overs. Compared to their analysis, our model mostly focuses on pricing rather than public good provision, it explicitly exploits the interaction of heterogeneity between and within regions, it studies under what conditions regions will voluntarily transfer decision power to the central level, it considers a more extended set of institutional constraints and, finally, it contributes to an understanding of the use of pricing instruments in the transport sector in federal states.

Second, our model is obviously related to the literature on pricing and investment of transport services[4]. One line of research has extensively studied the role of congestion and other externalities for pricing, investment and cost recovery in a single region (see, among many others, Kidokoro (2006), Small and Verhoef (2007), Calthrop, De Borger and Proost (2010)). It has been shown that charging users the marginal external cost of congestion improves overall efficiency and signals capacity needs; moreover, it allows to recover an important part of the capacity costs (also see De Palma and Lindsey (2007)). When there are spillovers -- in the sense of foreign users of domestic transport infrastructure -- decentralized decisions may imply large welfare losses, depending on the pricing instruments used and the nature of the network structure between regions (De Borger, Proost and Van Dender (2005), De Borger, Dunkerley and Proost (2007), Ubbels and Verhoef (2008)).

Finally, a growing recent literature analyzes the political economy of decision-making in the transport sector. However, these studies typically focus on pricing in a setting with a single government (Borck and Wrede (2005), Brueckner and Selod (2006), De Borger and Proost (2012)). A few available papers do model the political economy of infrastructure decisions with multiple regions. For example, Knight (2004, 2008) uses a legislative bargaining framework as one of the ingredients of the political mechanism to explain the allocation of highway funds in the US. He shows that political power can be used at the federal level to favor certain regions, and the empirical results support his prediction. He further finds that the geographic distribution of federal funds operates via two channels: one is the 'proposal power' of a representative in the legislature, the other the 'vote cost' channel that favors the participation of "easy to please" jurisdictions in minimum winning coalitions. Using a serial road network structure (as in De Borger, Dunkerley and Proost (2007)), Xie and

[4] Congestion has also been introduced into the literature on the optimal provision of local public goods (see, e.g., Brueckner (1981), Scotchmer (1988), and Craig (1989)).

Levinson (2009) study the political economy of governance choice in a federation in the provision of transport infrastructure. However, unlike the current paper, their model ignores congestion, it is restricted to symmetric regions and it exclusively focuses on infrastructure. Finally, recent papers by Brueckner (2013) and Ferguson (2013) study transport decision making in models of multi-jurisdictional monocentric cities that include both transport and land markets. The former shows that decentralized capacity choices (made by individual zones within the city) generate the social optimum, despite the presence of spillovers[5]. The analysis in Ferguson (2013) focuses on the implications of differences in the trade-offs between money and time for central city and suburban residents.

Structure of the paper is the following. In the next section, we describe the model used for the analysis. In Section 3 we study optimal pricing on a given infrastructure; this allows us to explain the political mechanisms used under both decentralized and centralized decision-making. Moreover, it provides insight into the main driving forces of different systems of political decision-making that will also be at play in the remainder of the paper. In section 4 we study the effect of two common restrictions on the central decision mechanism, a uniform pricing constraint and interregional bargaining, and we show under what conditions this improves welfare. In Section 5 we study the case of flexible capacity and pricing. We consider the question whether institutional constraints like self-financing rules improve the efficiency of centralized and decentralized decision making. We end with a brief conclusion and some policy implications.

2. Model setting

In this section, we describe the model. We first specify the composition of the regions, the demand for the use of the public facility in each region, and the average cost function for the user. Next we discuss the budget restriction of the regional and federal governments, and we describe the political mechanisms at the regional and federal level.

We use a setting with two regions, indexed $i=1,2$. The population of each of the regions consists of two groups: a group of users D_i, and a group of inhabitants N_i that does not use any road infrastructure (for example, they may not own a car). Users make two types of trips: trips in the home region and trips in the other region. To simplify the exposition without

[5] We return to this finding in Section 5 below.

affecting the qualitative insights to be derived from the model, we assume that the demand for both types of trips is independent.[6]

In order to focus on the role of spill-overs and the share of users in a given region, we assume regions have the same population *R*, and that demand and cost functions are the same in both regions. However, regions differ in two dimensions. First, the composition of the population between users and non-users can differ. For example, car users might form the majority in one region but not in the other. Our assumption implies

$$D_1 + N_1 = R = D_2 + N_2$$

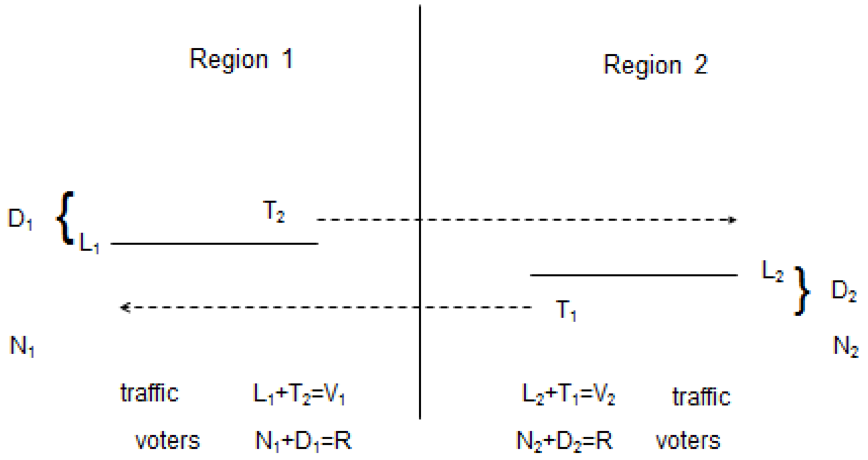
Second, for the group of users, the proportion of trips made at home and in the other region can be different. Specifically, total demand for trips in each region is given by, respectively:

$$L_1 + T_1; \quad L_2 + T_2$$

Note that T_1 is the number of trips in region 1 made by inhabitants of region 2. Similarly, the number of trips people from region 1 make in region 2 is denoted as T_2. In other words, the index refers to where the transport takes place. Figure 1 represents the different groups schematically.

[6] Admittedly, this strong assumption is more appropriate for some congested public goods than for others. For example, it is quite realistic in the case of libraries, public swimming facilities, museums, etc. However, in the case of transport trips it is more plausible that there are two types of trips: local trips in one region and border-crossing trips that use the infrastructure of the two regions. This latter case can easily be modeled as well, but it raises two additional complications. First, it would imply nonzero cross-price elasticities between the demand for trips in the two regions. Second, it introduces horizontal tax competition, as local governments share part of the tax base. The extra cost in terms of additional complexity is substantial (see, for example, De Borger, Dunkerley and Proost (2007)) and, given the focus of the current paper, the benefit in terms of extra insights is small. We therefore stick to the assumption of independent demands throughout the paper and return to this issue in our concluding section.

Figure 1 Model setting



We now specify aggregate transport demand and travel costs for an arbitrary region; we leave out the regional index, as parameters in aggregate transport demand and cost functions are assumed to be the same in both regions. First, total demand V for miles on the local road system is described by the linear inverse demand function

$$P(V) = a - bV \tag{1}$$

Therefore, demand can be written as

$$V = \frac{a - P}{b}$$

Total demand consists of local traffic by inhabitants of the region plus traffic in the region by inhabitants of the other region:

$$V = V_L + V_T$$

We finally assume that, conditional on a given generalized price, demands of local users and users that live in the other region are proportional. This allows us to define a 'spill-over' parameter, in analogy with the literature on local public goods referred to in the Introduction. Demands for local and transit demand are specified as

$$V_L = \theta\left(\frac{a - P}{b}\right)$$

$$V_T = (1-\theta)\left(\frac{a-P}{b}\right)$$

The parameter θ $(0 \le \theta \le 1)$ will play an important role in the analysis; it is the share of trips or kilometers users from a given region demand in their own region. The fraction $(1-\theta)$ can be interpreted as an indicator of 'spill-overs'.

The generalized user cost function for road users[7] is assumed to be linearly rising in the volume-capacity ratio

$$C(V) = \alpha + \beta \frac{V}{K} \tag{2}$$

Inclusive of a potential user charge (for example, a road toll) τ on road use, we have the gross user cost *g(V)*

$$g(V) = C(V) + \tau = \alpha + \beta \frac{V}{K} + \tau \tag{3}$$

Either the regional governments or the federal government are responsible for the costs of road capacity. They balance their budget via head taxesand/or via user prices on road use.[8] The rental cost of current capacity K^0 amounts to ρK^0.[9]

Our objective is to compare decentralized and centralized decision making. The results will of course depend on the precise political mechanism in place. For decision making at the regional level, the obvious candidate is simple majority voting. When preferences satisfy certain conditions, the median voter is decisive. Our basic model setting follows Besley and Coate (2003)[10]. They suggest, for decision making at the federal level, that a legislature of locally elected representatives makes the decisions by forming a minimum winning coalition[11]. In a two-region model, this assumes that one of the two regional representatives is

[7] We focus on road use, as this is the most common congested public good. We could also consider public transport, as this also suffers from congestion. We do not consider the interaction between private and public transport; this would again introduce much additional complexity without new insights. A sufficient condition to only consider private (or public) transport is that the other mode is priced at marginal social cost.

[8] We use head taxes (and not income taxes) because we want to limit the model to two sources of heterogeneity. In the concluding section we briefly discuss differences in income and speculate on the implications of introducing proportional income taxes as alternative tax base.

[9] One could include an operation cost per user of the local public good. The toll charged to the user can then be higher or lower than the variable operation cost; if it is below the variable cost, users are subsidized.

[10] Other systems of political decision-making at the central level, including cooperative legislative bargaining, are studied in Section 4.

[11] As the regions are not homogenous, centralized decisions may imply strategic delegation in which each region elects a representative whose preferences guarantee the best regional outcome at the federal decision level. See Besley and Coate (2003) and Lörz and Willmann (2005) for details.

made agenda setter and is ultimately decisive in taking the federal decision; the representatives from both regions are assumed to have an equal probability of being decisive.

3. Pricing the use of existing capacity

In this section, we study user pricing of existing capacity under decentralized as well as centralized decision making. Before moving to political economy issues, however, let us briefly consider the social optimum in an arbitrary region. It is assumed that the planner cares about the net benefits of the congested public good to users and the revenues of user pricing; moreover, the cost of infrastructure matters. Specifically, we assume the socially optimal user price is given by the solution of the following optimization problem

$$\underset{\tau}{\text{Max}} \ \{\int_0^V P(V)dV - V.g(V)\} + \tau V - \rho K^0$$

The first term between brackets in the objective function is net consumer surplus (gross consumer surplus minus total generalized costs), the second term captures government revenues on user prices, and the final term reflects the capital cost associated with given infrastructure (this will become relevant when we consider capacity choices in Section 5).

Differentiating the objective function, using the equality of the generalized price and the generalized cost in equilibrium (*P(V)=g(V)*), straightforward algebra produces the optimal user price rule:

$$\tau = \frac{\beta V}{K^0} \qquad (4)$$

Obviously, optimal pricing implies equality of the user price and marginal external congestion cost; this is given by the right-hand side of (4). To see this, note that a marginal increase in the number of users raises the user cost by $\frac{\beta}{K^0}$, and all users *V* are confronted with this higher cost.

3.1. Decentralized decision making

We focus on one arbitrary region and, for notational convenience, we leave out the regional index. We assume simple majority voting at the regional level. To study the outcome of this process, note that there are two groups of voters in the region that have clearly different preferences regarding transport decisions: the group of users *D*, and the group of people *N* that live in the region but do not use the regional road infrastructure.

First, suppose users have a majority in the region. A member of the group of users will choose the user price level that maximizes his welfare:

$$\underset{\tau}{Max}\ \frac{\theta}{D}\{\int_0^V P(V)dV - V.g(V)\} + \frac{\tau V - \rho K^0}{R} \tag{5}$$

Total transport demand in the region is the sum of demands by local users and non-local users, $V = V_L + V_T = \theta V + (1-\theta)V$. To interpret (5), note that welfare of an individual member of the group of users D consists of two components: (i) her consumer surplus as a user; expressed per person, this is a fraction $\frac{\theta}{D}$ of total surplus; (ii) the net revenues from user pricing that she will receive under the form of a head subsidy or reduced head tax – this is shared by the whole population R.

To derive the optimal user price rule, we take the first-order condition, use the definition of the generalized cost (3), and note that the total effect of a user price increase on travel demand can be written as

$$\frac{dV}{d\tau} = \frac{\frac{\partial V}{\partial \tau}}{1 - \frac{\beta}{K}\frac{\partial V}{\partial \tau}}$$

Straightforward algebra then produces the following user price rule

$$\tau^d = \frac{\beta V}{K^0} + \{\frac{\theta}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau}} \tag{6}$$

where the superscript 'd' stands for the 'decentralized' case, and where we have defined

$$\eta = \frac{D}{R}.$$

This parameter captures the fraction of voters that are users; under our assumption that users have a majority, $0.5 < \eta \leq 1$. Remember that θ is between zero (no spill-overs at all: all traffic in a region is by locals) and one (extreme spill-overs: there is no local traffic at all in a given region).

With this information in mind, turn to the interpretation. To get started, assume there is no transport demand by inhabitants from outside the region ($\theta = 1$). As $\frac{\partial V}{\partial \tau} < 0$, the preferred user price will then be smaller than the marginal external congestion cost $\frac{\beta V}{K^0}$. The reason is that the net revenue of user pricing is redistributed over all voters, road users and non-road users alike. So although the group of road users as a whole gains from efficient user pricing, the redistribution of revenues makes them select an inefficiently low user price. It also follows from (6) that, with zero spill-overs, the larger the majority that users in a region have (the larger η), the higher the user price. Intuitively, this is because the larger the majority of users, the smaller is his relative share in net consumer surplus and the larger the relative weight of net revenues.

Now introduce spill-overs, so that $\theta < 1$. More demand from outside the region (a reduction in θ) raises the preferred user price level of members of group *D*, reflecting tax exporting behavior. In the extreme case that all transport demand comes from people living outside the region ($\theta = 0$), user price rule (6) becomes the revenue maximizing user price (see (7) below).

Previous discussion implies that in a decentralized political system the user price a driver wants is determined by two forces. On the one hand, for a given number of users in his own region, a driver wants a higher user charge when more drivers from other regions use the local infrastructure. This is just tax exporting behavior. On the other hand, for a given level of spill-overs, a user wants a higher toll the higher the driver majority in his own region. This reflects an unwillingness to share net revenues with non-users. If net revenues are positive, a larger driver majority means he has to share with fewer non-users; if net revenues are negative (not fully covering capacity costs), there are fewer non-users to help covering the deficit. In both cases, the elected user prefers a higher toll. The implication is that the user charge in a region can range from far below marginal external cost or even no pricing at all (few spill-overs and a small majority of users) to substantially above marginal external cost (large spill-overs and a large user majority.

Note two other implications of (6). First, it is clear that with very low congestibility (small β), large spill-overs and a large driver majority in the region, the optimal user charge from the viewpoint of a user can be negative; we then have user subsidies instead of tolls. Our modeling framework remains perfectly valid to analyze such cases, but they are not the focus of this paper. We assume throughout the paper that all user fees are non-negative. Second, (6) boils down to the first-best outcome if the share of local demand in total traffic in the region (θ) equals the share of users in the number of local voters (η). In this case, the incentives for tax exporting compensate exactly the incentives to limit redistribution to non-users.

What happens if voters in the region that do not use the regional infrastructure have a majority, so that $0 \le \eta < 0.5$? They will then opt for the revenue maximizing user price: they do not pay but do share in the excess revenues[12]. Indeed, an individual of the group of non-users prefers a user price that

$$\underset{\tau}{\text{Max}} \ \frac{\tau V - \rho K^0}{R}$$

The resulting user price satisfies:

$$\tau^d = \frac{\beta V}{K^0} - \frac{V}{\frac{\partial V}{\partial \tau}} \tag{7}$$

The user price (7) equals the marginal external cost plus the monopoly margin.

3.2. Centralized decision making

Now consider centralized decision-making by a minimum winning coalition. This is implemented by assuming that in each region a member of each of the two groups (users and non-users) can be elected – by majority voting -- as representative and, once elected, has a 50% probability of being decisive at the central level. We implicitly assume that the central decision maker has perfect information on the preferences and costs in both regions. This is a strong assumption; however, it can be justified in our model, because the use of the local public good we are dealing with is observable (for example, road use) and it can be priced[13].

What user pricing decisions will be the result of the described decision-making process? To analyze this issue, we first have to understand the user pricing decisions a typical member of each of the different groups of voters in, say, region 1 would take if he would become decisive at the central level; that is, if he would be allowed to decide on user prices in both regions. The results when the elected representative from region 2 is decisive follow by analogy.

First, assume that the representative from region 1 is a user and that he has to decide on user prices on the existing capacity in both regions. His problem is to solve

$$\underset{\tau_1,\tau_2}{\text{Max}} \ \frac{\theta_1}{D_1}\{\int_0^{V_1} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\tau_1 V_1 - \rho K_1^0 + \tau_2 V_2 - \rho K_2^0}{2R}$$

In this expression, V_i is transport demand in region *i*. The first term in the objective function is the net consumer surplus he enjoys from driving in his own region, the second term is his net

[12] Of course, a less extreme result might be obtained under slightly different assumptions. For example, one could assume that non-users do suffer from car externalities (such as pollution).

[13] Recently, Kessler (2013) studies the role of information in a model where the central decision level has to rely on information provided by the regions to obtain estimates of the costs and benefits of the local public goods, showing that uniform provision may be the result of communication failures. Uniform pricing is studied in Section 4. Here we allow user prices to be different in both regions.

surplus when driving in region 2 (note that $(1-\theta_2)$ is the fraction of drivers in region 2 that are resident of region 1). The third component in the objective function is his share in total federal revenues generated by user pricing the existing capacity in both regions.

Straightforward algebra, using the same steps as in the case of decentralization above, leads to the following desired user price levels for a user from region 1:

$$\tau_1^c(1) = \frac{\beta V_1}{K_1^0} + \{\frac{2\theta_1}{\eta_1} - 1\} \frac{V_1}{\frac{\partial V_1}{\partial \tau_1}} \tag{8}$$

$$\tau_2^c(1) = \frac{\beta V_2}{K_2^0} + \{\frac{2(1-\theta_2)}{\eta_1} - 1\} \frac{V_2}{\frac{\partial V_2}{\partial \tau_2}} \tag{9}$$

The notation $\tau_i^c(j)$ stands for the toll in region *i* that is preferred by a representative from region *j* under 'centralized' decisions. To interpret these expressions, it is useful to start from a situation with zero spill-overs ($\theta_1 = \theta_2 = 1$), so that users only use the infrastructure in the own region. The representative of region 1 will in that case opt for a very low user price in his own region (see (8)); this is even lower than under decentralization, because now he has to share the excess user pricing revenue with the inhabitants of both regions. In the other region (where he does not drive), (9) implies that he will set the user price at the revenue maximizing level. Now introduce spill-overs. Expression (8) then implies that tax exporting will lead to higher user prices in the own region; at the same time, the preferred user price in the other region (where he does drive now) will decline as the decisive user will also have to pay it (see (9)). Finally, note the role of the majority of users in region 1. As in the decentralized case, a larger user majority in the own region leads to a higher desired toll; at the same time, it induces a lower desired user price in the other region.

Note by analogy that user prices preferred by a user from region 2 that becomes decisive at the federal level are:

$$\tau_1^c(2) = \frac{\beta V_1}{K_1^0} + \{\frac{2(1-\theta_1)}{\eta_2} - 1\} \frac{V_1}{\frac{\partial V_1}{\partial \tau_1}} \tag{10}$$

$$\tau_2^c(2) = \frac{\beta V_2}{K_2^0} + \{\frac{2\theta_2}{\eta_2} - 1\} \frac{V_2}{\frac{\partial V_2}{\partial \tau_2}} \tag{11}$$

The pricing rules (8)-(11) clearly point at potential exploitation of one region by the other. For example, if spill-overs are limited users from a region that become decisive at the central level have incentives to impose low charges in the own region (where they drive) and high charges in the other region (because they share in the revenues). Although this type of behavior whereby policy makers favor the own region at the expense of others may seem extreme, it has been empirically documented in other contexts (see, for example, Knight (2004)).

A final observation related to expressions (8)-(11) is that centralized decisions are first best when $\theta_1 = \theta_2 = 0.5; \eta_1 = \eta_2 = 1$. This is not surprising: if all voters are drivers and they use the infrastructure of both regions equally intensively, the two incentives to deviate from socially optimal pricing (sharing with non-users and exploitation of other regions) disappear.

Second, assume that the representative of region 1 who is chosen as agenda setter belongs to group N_1: he is not a car user at all. Obviously, he will select revenue maximizing user prices for both regions:

$$\tau_1^c(1) = \frac{\beta V_1}{K_1^0} - \frac{V_1}{\frac{\partial V_1}{\partial \tau_1}} \tag{12}$$

$$\tau_2^c(1) = \frac{\beta V_2}{K_2^0} - \frac{V_2}{\frac{\partial V_2}{\partial \tau_2}} \tag{13}$$

Similarly, a non-user from region 2 that becomes decisive at the federal level will opt for the same user prices.

3.3. Comparing centralized and decentralized decision making

We are interested in comparing user charges and welfare of policy outcomes under decentralized and centralized decision making. However, note that one important observation can be made right at the outset: depending on parameter values, both decentralization and centralization may lead to the highest welfare. We pointed out before that when there are no spill-overs and all inhabitants of both regions are users (so $\theta_1 = \theta_2 = \eta_1 = \eta_2 = 1$) decentralization is first-best. More generally, if users have a majority in both regions, all combinations such that $\theta_i = \eta_i$ in both regions give first-best under decentralization (see (6)). However, centralization was found to be first-best when all voters are drivers in both regions and people drive as much in the other as in their own region ($\theta_1 = \theta_2 = 0.5; \eta_1 = \eta_2 = 1$);

decentralized decisions would imply user prices exceeding first-best (see (6)). Hence, centralization performs better[14].

In the remainder of this section we perform a more general comparison of decentralized and centralized decision making. We start with a comparison of toll rules; this will facilitate the interpretation of the analysis of the relative welfare performance of the two political systems that follows.

3.3.1. Tolling outcomes: centralized versus decentralized decisions

To allow a simple graphical interpretation, we focus on the case of symmetric regions: road users have a majority in both regions, and all relevant parameters are the same for the two regions. In the graphical illustrations below, we show three types of equilibrium in 'tolling' space: the first-best (denoted *FB*), the decentralized equilibrium (denoted *D*), and the two equilibria under centralized decisions (denoted *C(1)* and *C(2)*, respectively, depending on whether a driver from region 1 or from region 2 is decisive at the central level). For convenience, we here reproduce the relevant expressions – obtained by imposing symmetry on (6) and (8)-(11) -- describing the various equilibria:

$$D \qquad \tau^d = \frac{\beta V}{K^0} + \{\frac{\theta}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau}} \tag{14a}$$

$$C(1) \qquad \tau_1^c(1) = \frac{\beta V}{K^0} + \{\frac{2\theta}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau_1}}; \qquad \tau_2^c(1) = \frac{\beta V}{K^0} + \{\frac{2(1-\theta)}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau_2}}$$

$$C(2) \qquad \tau_1^c(2) = \frac{\beta V}{K^0} + \{\frac{2(1-\theta)}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau_1}}; \quad \tau_2^c(2) = \frac{\beta V}{K^0} + \{\frac{2\theta}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau_2}} \tag{14b}$$

In Figures 2, 3 and 4 we illustrate the role of the parameters for the position of the different equilibria. Due to the symmetry assumption, the first-best and decentralized pricing equilibria are both situated on the 45° line; the equilibria under centralized decisions are symmetrically off the 45° line. Their position is affected by the parameters in a predictable way. For example, if $\theta = \eta$ decentralization is first-best (hence, D coincides with *FB*). For $\eta > \theta$, *D* involves higher tolls than the first-best (*D* is to the right of *FB*), and vice versa when $\eta < \theta$ (*D* to the left of *FB*). Expressions (14b) imply that the position of the centralized

[14] If users do not have a majority in both regions, neither decentralization nor centralization can achieve the first-best, because at least one party will always want revenue maximizing user prices.

equilibria relative to the 45° line depends on the level of spill-overs, whereas their position relative to the first-best largely depends on the voting shares.

On Figure 2 we illustrate the various toll equilibria for the case $\theta > \eta$. In this case, the tax exporting motive in the decentralized equilibrium is dominated by the revenue sharing motive, so that the tolls are below first-best (see (14a)). The two equilibria under centralization will involve a very low toll in the own region combined with a very high toll in the other region (see (14b)). In graphical terms this implies that the centralized equilibria are not only off the 45° line, they are "further away" from the first best than point *D*. Since centralized decisions imply uncertainty as to the decisive representative and welfare is a concave function of toll levels, Figure 2 immediately suggests that $\theta > \eta$ is a sufficient condition for decentralized welfare to exceed expected welfare under centralization (at least for the symmetric case where drivers have a majority). We formally show this statement below.

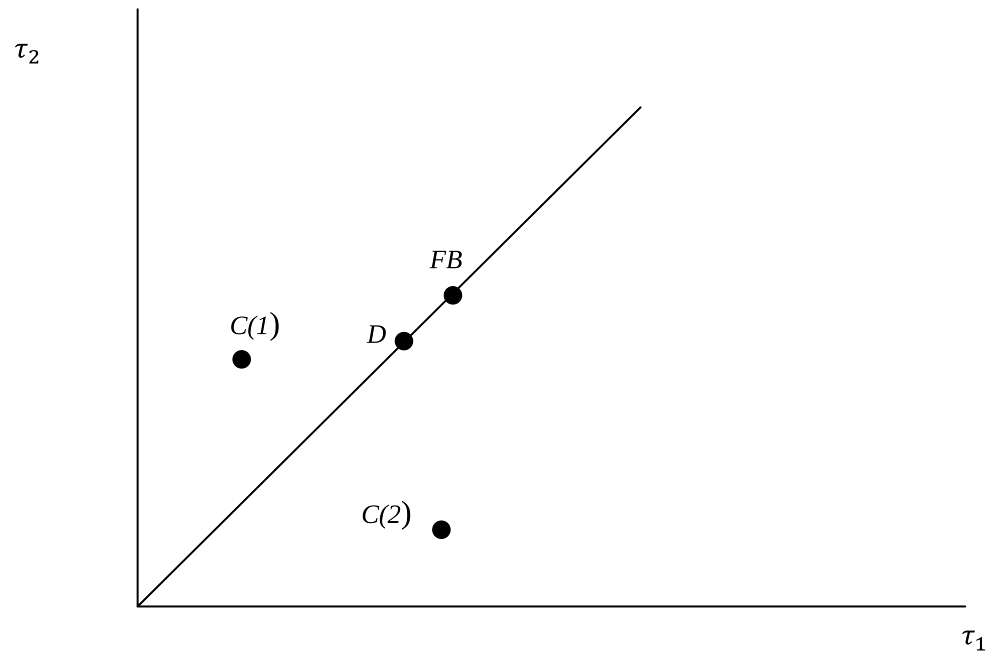


Figure 2: Comparing tolling equilibria (case $1 > \theta > \eta > 0.5$)

It was previously pointed out that for the specific parameter configuration $\theta_1 = \theta_2 = 0.5; \eta_1 = \eta_2 = 1$ centralized decisions were first-best. Figure 3 indeed suggests that for a range of parameter values (large majorities of drivers, and substantial spill-overs so that θ is close to 0.5) centralization will outperform decentralized decisions. In the decentralized equilibrium tax exporting incentives will now be more important than the unwillingness to share revenues with the minority of non-drivers, and tolls will be higher than first best (see (14a)). The centralized solutions wanted by the representatives from the two regions will now be quite close to one another; moreover, they will be not very far from the first-best (see (14b)). Given the properties of welfare functions, the tolling outcomes on Figure 3 suggest that centralization will yield higher welfare than decentralized decisions.

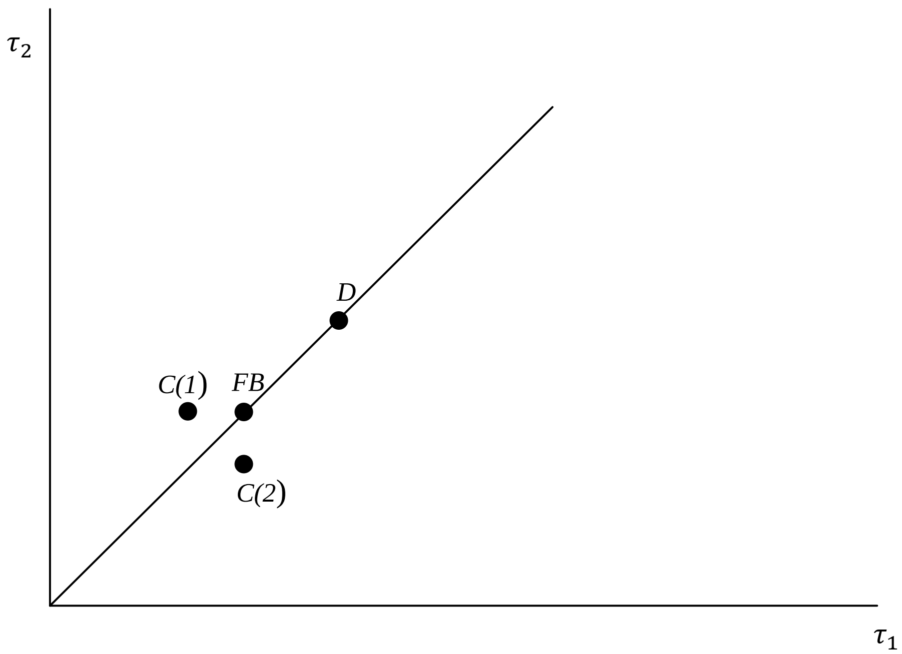


Figure 3. Comparing tolling equilibria (η large, θ close to 0.5)

Finally, on Figure 4 we depict the situation where drivers have a majority in both regions, but spill-overs are very large (θ much below 0.5). The decentralized equilibrium involves user tolls well above marginal external cost, so that D is now situated to the right of *FB*. Under centralized decision making, two things happen. First, as there is now more traffic in a region by drivers from the other region than by the region's own drivers, expressions (14b) imply that *C(1)* and *C(2)* switch positions relative to the 45° line (in the sense that the former will now be below the 45° line, whereas the latter will be above it). Each representative wants a very high toll in the own region (tax exporting) but a very low toll in the other region (where he does most of the driving). Second, the larger spill-overs are, the further both centralized equilibria move away from the first-best. This suggests that for very large spill-overs, decentralization will again welfare dominate centralized decision making.

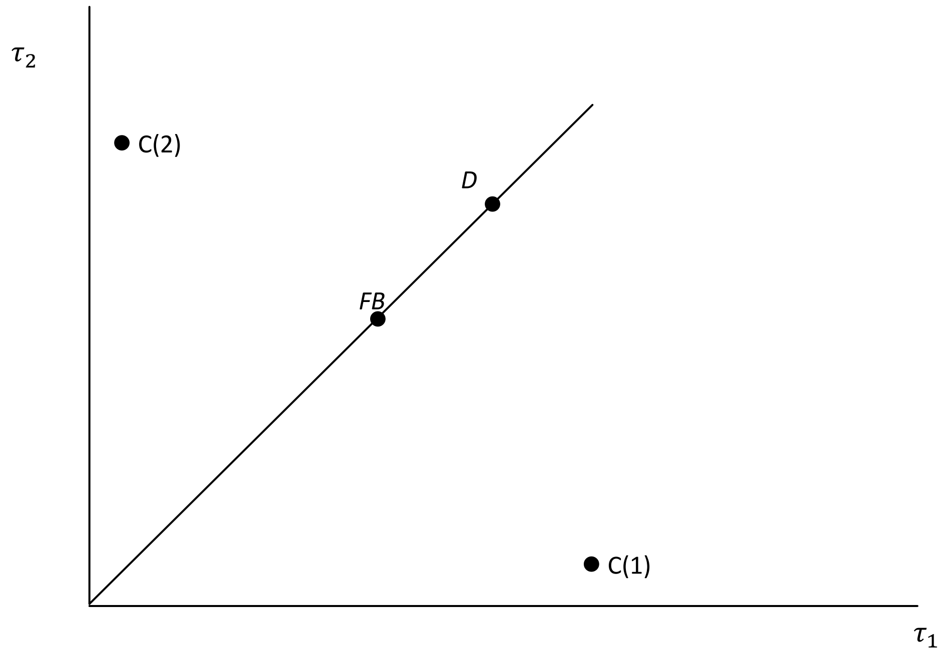


Figure 4: Comparing toll equilibria (θ very small)

3.3.2.Welfare comparison: centralized versus decentralized decisions

Next we turn to a formal welfare performance of decentralized and centralized decision making. Assuming a risk-neutral definition of welfare under centralization, we specifically want to know whether

$$(W_1^d + W_2^d) > or < \left\{0.5\left[W_1^c(1) + W_1^c(2)\right] + 0.5\left[W_2^c(1) + W_2^c(2)\right]\right\}$$

In this expression W_i^d is welfare in region *i* under decentralization, and $W_i^c(j)$ is welfare in region *i* under centralized decisions when the representative from region *j* is decisive at the central level. The right-hand side is expected welfare under centralized decisions; it reflects the fact that each representative has equal probability of being decisive.

The comparison depends in a complex way on the parameters. We therefore first zoom in on the case of zero spill-overs. Next we discuss findings for the general case and point out the role of the parameters. For a detailed derivation of the results reported in the remainder of this section, we refer to Appendix 1.

The case of zero spill-overs

We can distinguish four cases: one is where users have a majority in both regions, there are two cases where users have a majority in only one region and, finally, non-users can have a majority in both regions. The results derived in Appendix 1 when there are zero spill-overs are summarized for the various cases in Table 1. The pricing regimes are denoted as “LOW”, “LOWLOW” and “HIGH”. For example, with zero spill-overs and a majority of users, decentralized decision making leads to a low user price for the own region; the relevant rule is given by (6), but imposing $\theta=1$. We call this regime “LOW”. If non-users have a majority then we found a revenue maximizing user price, given by (7); we call this pricing regime “HIGH”. Under centralized decision making, we have very low user prices in the own region if the decision maker is a car user; it is given by (8) and (11), but imposing $\theta_1 = \theta_2 = 1$. Denote this regime as “LOWLOW”. In the other region, zero spill-overs give again regime “HIGH”. Finally, if the decision maker is not a car user, centralized decision making leads to regime “HIGH” in both regions (“HIGH”).

In the final column of Table 1, we report the result of the formal welfare comparison performed in Appendix 1. This yields two important insights. First, if there are no spill-overs, decentralization is better than centralized decision making provided users have a majority in at least one region. The intuition is that, first, whenever non-users decide on user prices, these will be inefficiently high and, second, when users decide on their own prices, they will choose too low prices to avoid too much redistribution to non-users. Second, if non-users have a majority in both regions, the political system does not matter but, importantly, both systems yield the same poor result: user prices will be too high and large welfare losses occur compared to the social optimum.

Cases	Decentralized decision making		Centralized decision making		Conclusion
1	User majority LOW	User majority LOW	Region 1 decisive LOWLOW and HIGH	Region 2 Decisive HIGH and LOWLOW	Decentralized decisions certainly better
2	User majority LOW	Non-user majority HIGH	Region 1 Decisive LOWLOW and HIGH	Region 2 Decisive HIGH and HIGH	Decentralized decisions certainly better
3	Non-user majority HIGH	Non-user majority HIGH	Region 1 Decisive HIGH and HIGH	Region 2 Decisive HIGH and HIGH	No difference between centralized and decentralized welfare
4	Non-user majority HIGH	User majority LOW	Region 1 Decisive HIGH and HIGH	Region 2 Decisive HIGH and LOWLOW	See Case 2

Table 1. Comparing decentralized and centralized decision making: pricing a given capacity in the case of zero spill-overs and symmetric regions. The meaning of the price regimes *HIGH, LOW* and *LOWLOW* is explained in the text.

<u>*Introducing spill-overs*</u>

In the general case with spill-overs, few general theoretical statements can be made about the relative welfare performance of the two political systems. Only a few findings are worth reporting. Apart from that, we will resort to numerical analysis to get further insight.

Let <u>drivers have a majority in both regions.</u> In Appendix 1, we first formally show, for the case of linear demand, that a set of sufficient conditions for decentralization to be welfare superior to centralization is:

$$\theta_i \geq \eta_i \quad i = 1,2$$

Note that this confirms the discussion related to Figure 2 above. In general, decentralization certainly outperforms centralization if either spill-overs are very limited (θ_i close to 1) or, when spill-overs are substantial, if users only have a small majority so that they have to share revenues with a nontrivial group of non-users (η_i only slightly above 0.5). Second, if previous condition does not hold so that $\theta_i < \eta_i \quad (i = 1, 2)$, centralized decisions may outperform decentralized decisions. Unfortunately, a set of sufficient conditions for this to happen is not very insightful as it turned out to involve all demand and cost parameters in a highly nonlinear way. In view of the discussion surrounding Figure 3, however, we show in Appendix 1 that there will be a range of parameter values 'in the neighborhood' of ($\theta_1 = \theta_2 = 0.5; \eta_1 = \eta_2 = 1$) such that centralization dominates decentralized decisions. The charges wanted by the two representatives then do not differ much from one another, and they are close to first-best. The interpretation is that centralization yields higher welfare than decentralized decisions if driver majorities are large and, in any given region, use of the infrastructure comes about equally from local users and from people from the other region. Intuitively, the former condition means that revenue sharing is not an issue, the latter condition implies that the incentives for decisive representatives at the central level to favor their own region disappear. Outside this 'neighborhood', decentralization will do better[15]. This implies that decentralization will often yield higher welfare, even when spill-overs are very large. We illustrate this with a numerical example below.

Consider the case were <u>users have a majority in one region only</u>. We then again find that a sufficient condition for decentralization to be better is that $\theta_i \geq \eta_i$ (Appendix 1). Finally, if <u>non-users have a majority in both regions</u>, the two political regimes give the same outcome, and parameters do not make a difference. Importantly, note that the equal performance of centralized and decentralized decisions in this case yields outcomes that are equally undesirable, leading to much too high user prices everywhere.

The theoretical results yield useful but rather partial insights. Therefore, we illustrate the relative performance of the two political systems by a numerical example. The example assumes that demand, cost and capacity parameters are the same in both regions, and the demand function is assumed to be linear. All relevant expressions for tolls, transport volumes and welfare levels under the various political systems are given in Appendix 1. The numerical

---

[15] We also investigated in detail how the relative performance of centralized versus decentralized decisions depends on the parameters. We found that, if $0.5 \leq \frac{\theta_i}{\eta_i} \leq 1$, both an increase in spill-overs and in the driver majority raise the relative performance of centralized decisions. To get some intuition, start from the situation without spill-overs and with only users in both regions. Introduce spill-overs in region 1, so that θ_1 decreases below 1. This raises the decentralized user price for region 1 due to tax exporting, see (6). In the centralized case, the user price the decisive representative from region 1 wants for his own region increases for the same reason, but it increases more strongly. The user price the representative from region 2 wants for region 1 decreases. Together this suggests that a decrease in θ_1 improves welfare under both political systems, but it makes the centralized case relatively more attractive compared to decentralization.

exercise reported here is based on the following inverse demand function in each region *i* *(i=1,2)*:

$$P_i = 1.2 - 0.0001 * V_i .$$

The cost function parameters are $\alpha_i = 0.5, \beta_i = 0.75$. Capacity is assumed to be $K_i^0 = 3000$; capacity unit cost is $\rho = 0.1$.

Using this simple example, we calculated the relative welfare performance of the two systems. Although spill-overs and driver majorities can differ between regions, we first focus on the symmetric ($\theta_1 = \theta_2, \eta_1 = \eta_2$) case. This has the advantage that the relative welfare performance can be illustrated graphically in two-dimensional space[16]. We summarize the results in Figure 5. On the horizontal axis, we show the share of users in the region, on the vertical axis the degree of spill-overs. The figure illustrates for which parameter combinations decentralization outperforms centralization. If voters do not have a majority ($\eta_i < 0.5$), centralized and decentralized decisions yield the same (equally poor, because far from first-best) outcome. Provided voters have a majority ($\eta_i > 0.5$), decentralized decisions are first-best for all parameter combinations $\theta_i = \eta_i$; centralization gives first-best outcomes when all voters are drivers and drivers of a given region travel as much in the other as in the own region ($\eta_i = 1, \theta_i = 0.5$). With a majority of drivers, decentralization is better than centralization for a very wide range of parameter values. As argued before, only when drivers have a large majority and spillovers are close to $\theta_i = 0.5$ does centralization yield the highest welfare.

[16] The general asymmetric ($\theta_1 \neq \theta_2, \eta_1 \neq \eta_2$) case is analyzed in Section 4 below; there we compare a more extended set of political decision making mechanisms.

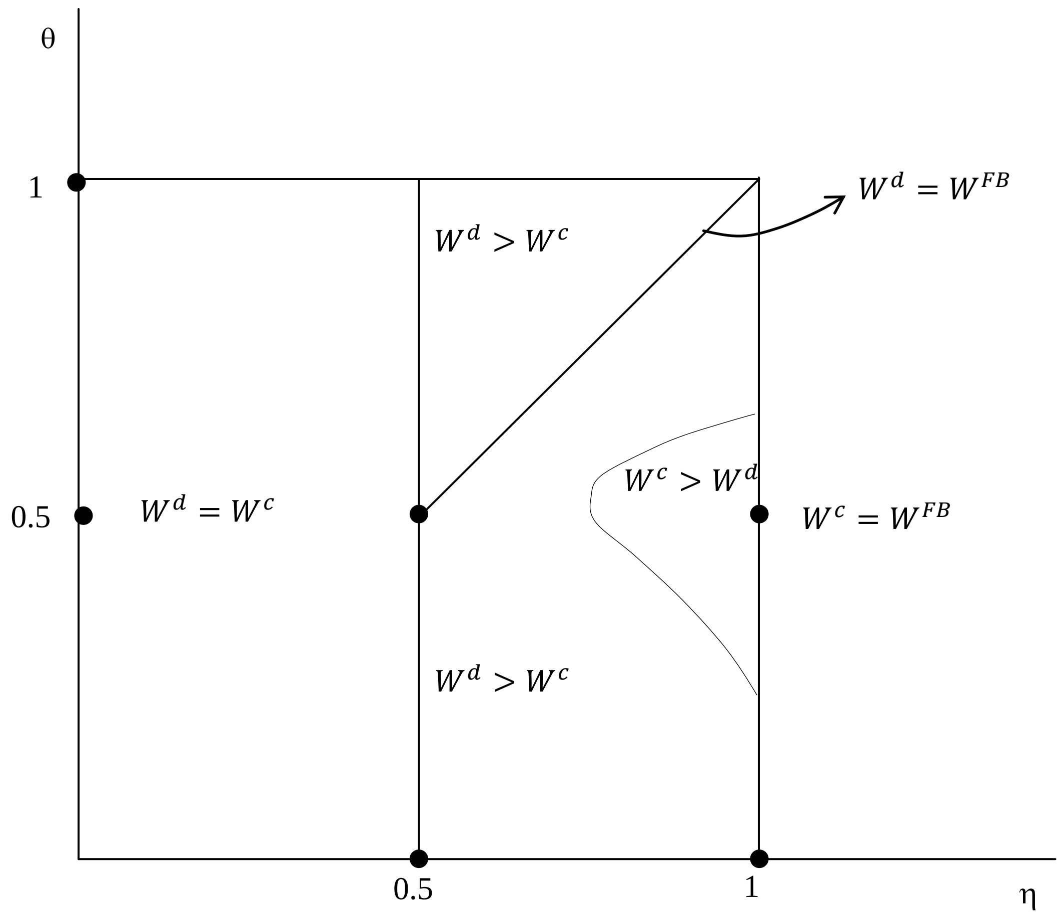


Figure 5: Welfare comparison: decentralization versus centralization (symmetric case)

We summarize our results so far in the following proposition.

PROPOSITION 1: THE CHOICE OF USER PRICES FOR EXISTING INFRASTRUCTURE

a. Whenever non-users form a majority, centralized decision making and decentralized decision making are equally bad in terms of welfare. User prices are too high.

b. If there are no spill-overs and users have a majority in at least one region, decentralized decisions yield higher welfare than centralization.

c. When users form the majority in at least one of the two regions, a sufficient condition for decentralized decision to perform better than centralized decision making is that either

spill-overs are very limited or, when spill-overs are substantial, if users have only a small majority.

d. Centralization only gives higher welfare than decentralized decisions if users have a large majority in both regions and the infrastructure in any given region is used approximately equally intensively by drivers from both regions (θ close to 0.5).

e. Zero spill-overs and symmetry are neither necessary nor sufficient for decentralized decisions to yield first-best outcomes.

Note that the final part of the proposition follows from the analysis above, although it was not explicitly mentioned before. It differs from the seminal paper by Besley and Coate (2003). With symmetric regions and zero spill-overs, they find that decentralization produces first-best outcomes. In our setting, this is not the case because of heterogeneity within regions: decentralization implies a welfare loss even without spill-overs, unless all voters in both regions are users. Moreover, unlike their paper, we find that decentralization can be first-best even in the presence of substantial spill-overs: heterogeneity within regions can compensate for the existence of spill-overs.

4. What institutional restrictions improve pricing decisions?

One conclusion of the analysis so far is that, when users form the majority in at least one of the two regions and unless very specific conditions hold there are good reasons to prefer decentralized decision making. But, of course, federal institutions may develop that avoid the potential exploitation of regions under centralized decision-making, improving its performance. As mentioned in the introduction, the development of such institutions has been studied in the literature before, although in a different setting (see Hickey (2013), Kessler (2013)). In what follows, we analyze two specific constraints and study to what extent they improve the performance of centralized decisions: requiring to charge uniform prices in both regions, and forcing regions to reach a solution on prices through bargaining. Both constraints can be either the result of a constitutional agreement when a federation is formed, or they can be the result of a game in trigger strategies where deviations by a regional representative are severely punished[17].

Of course, one could also imagine institutional constraints that improve decentralized solutions, because this may lead to exploitation of one group by another (for example, users by non-users in case the latter have a majority). In our model, one obvious constraint has already been built in: non-local users are charged the same price as local ones. The use of the non-discrimination principle in pricing policies is widespread in practice, and we will not discuss its efficiency effects here (see De Borger, Proost and Van Dender (2005) for such an analysis). Other possibilities would be to add caps on user fees, or to impose constraints on toll revenues. We briefly discuss price caps in Section 4.4. Budgetary restrictions will be discussed in an extended model with joint pricing and capacity choices (see Section 5 below).

4.1. Uniform user prices

The first constraint we consider is that regional representatives in the central legislature are restricted to select user price levels that are equal in both regions.

Consider the case where drivers have a majority in both regions. In Appendix 2 we show that the tolls wanted by the two regional representatives at the central level are given by, respectively

$$\tau^u(1) = \frac{\beta V^u(1)}{K} + \left[\frac{(1+\theta_1-\theta_2)}{\eta_1} - 1 \right] \frac{V^u(1)}{\frac{\partial V^u(1)}{\partial \tau}} \tag{15}$$

$$\tau^u(2) = \frac{\beta V^u(2)}{K} + \left[\frac{(1-\theta_1+\theta_2)}{\eta_2} - 1 \right] \frac{V^u(2)}{\frac{\partial V^u(2)}{\partial \tau}} \tag{16}$$

In these expressions, $\tau^u(i), V^u(i)$ are the tolls and transport volumes under the uniformity constraint when the representative from region *i* is decisive at the central level. For a given representative that is in charge, the transport volumes are equal in both regions, because the tolls are uniform and regional demand functions are the same.

Comparison of (15)-(16) with the pricing rules in the absence of the uniformity restriction (see (8)-(9) above) clearly shows how the constraint 'averages out' the price difference between the regions. Under symmetry, two observations also immediately follow. Imposing

---

[17] Proost and Zaporozhets (2012) also consider centralized public good provision, but foresee that a regional representative that becomes agenda setter at the central level will observe certain regional shares for the allocation. The cost shares can then be an equilibrium in trigger strategies, in the sense that the other regional representatives can punish deviations in the future.

uniform prices in both regions gives first-best if all voters are drivers; and if there are zero spill-overs, tolls and hence welfare levels are equal to what they are in the decentralized case (compare with (6)).

In Appendix 2 we use the toll expressions (15)-16) to show three policy-relevant results. First, imposing a uniform price restriction necessarily improves the welfare of centralized decision-making. The intuition is most easily explained when there are no spill-overs. The uniform price a representative wants is then a weighted average of the high user price (above marginal external cost) he wanted for the other region and the low user price (below marginal external cost) he wanted for his own region. The uniform user price increases the user price for his own region and, as it brings the user price closer to marginal external cost, it raises welfare in this region. At the same time, the uniform user price reduces the user price in the other region; as this user price was too high from a welfare perspective, welfare rises in the other region as well. Hence, uniform user prices outperform differentiated user prices. Second, uniform pricing is more likely to yield higher welfare than decentralized decision making when drivers have a large majority and there are substantial spill-overs. Third, one again easily shows that a sufficient condition for decentralization to yield higher welfare is that $\theta_i \geq \eta_i$ (see Appendix 2).

Finally, if drivers have a majority in just one region, similar findings are obtained as those just reported (see Appendix 2); and, as before, if non-drivers are dominating both regions, the two political systems yield equally undesirable effects.

We summarize findings in the following proposition.

PROPOSITION 2. IMPOSING A UNIFORM PRICING RESTRICTION ON CENTRALIZED DECISIONS

a. ***In the case of centralized decision-making, if the users form a majority in at least one region, imposing a uniform user price restriction across regions is welfare improving.***
b. ***A sufficient condition for decentralization to yield higher welfare than centralization under a uniformity constraint is that*** $\theta_i \geq \eta_i$.
c. ***Centralization with a uniformity restriction yields higher welfare than decentralization if there are large spill-overs and drivers have a large majority in the regions.***

d. ***Imposing uniform tolls on decision makers does not improve the outcomes when non-drivers have a majority in both regions.***

4.2. Centralized decisions through bargaining

One obvious alternative for the minimum winning coalition setup we used in section 3 is bargaining between the elected representatives from the different regions. In this section, we focus on a bargaining game without threat points[18]. It can be justified in different ways. First, it can be based on Weingast's (1979, 2009) idea of 'universalism'. There are no explicit threat points because there is a consensus ex ante (based on trigger strategies that punish the deviant) to maximize the joint surplus of the two regions. Second, it can also be justified by observing that many federal countries have bicameral federal decision structures. Hickey (2013) shows how, in a country with only two regions, alternating proposals lead to a Rubinstein (1982) bargaining game, where the solution corresponds to the outcome of a Nash bargaining solution. Finally, it could be argued that uncertainty with respect to who will be in charge at the central level leads regions to move towards a bargained solution; in expected value terms this may yield higher welfare (also see Section 4.4 below).

In Appendix 3, we formally study bargaining. If <u>drivers have a majority in both regions</u>, we show that the tax rules can be written as:

$$\tau_1^b = \frac{\beta V_1^b}{K_1^0} + \{\frac{\theta_1}{\eta_1} + \frac{(1-\theta_1)}{\eta_2} - 1\} \frac{V_1^b}{\frac{\partial V_1}{\partial \tau_1}} \tag{17}$$

$$\tau_2^b = \frac{\beta V_2^b}{K_2^0} + \{\frac{\theta_2}{\eta_2} + \frac{(1-\theta_2)}{\eta_1} - 1\} \frac{V_2^b}{\frac{\partial V_2}{\partial \tau_2}} \tag{18}$$

To interpret these rules, remember what regional representatives wanted under a standard minimum winning coalition. For example, if spill-overs are limited, they both would like low tolls in the own region and high tolls in the other. The outcome of bargaining given in (17)-(18) is that these extreme wishes are smoothed out.

Note from (17)-(18) that, if the two regions only consist of drivers (hence, $\eta_1 = \eta_2 = 1$), then bargaining yields the first-best outcome. Moreover, when there are zero spill-overs,

[18] An alternative would be to consider a bargaining game with explicit threat points, allowing for side payments between regions (by differentiating the federal head tax over regions). The threat point is the allocation preferred by the regional representative that is federal agenda setter. In this bargaining game, the agenda setter will only reduce the user price for the other region if he is compensated with lower federal head taxes.

centralized decisions by a legislative bargaining process and decentralized decisions yield the same welfare; this follows from comparing (17)-(18) with (6). In both cases, user charges will be below marginal external congestion cost (unless users make up 100% of the voters in both regions). When limited spill-overs do exist, user charges under decentralization increase, but bargaining leads to the same tolls as long as drivers' voting shares are unaffected. In general, bargaining will perform better than decentralization when drivers have large majorities and spill overs are substantial. We will further check this statement using numerical analysis below.

In Appendix 3 we further compare bargaining and uniform price restrictions as ways to smooth out potential exploitation of regions. We show that, when regions are symmetric, uniform price restrictions and bargaining over prices yield the same outcomes, hence welfare is equal in both cases: $W^b = W^u$. The extreme pricing decisions potentially taken by representatives from different regions can be smoothed out in two equivalent ways: impose a uniform pricing constraint on the centralized decisions taken by the representative in power, or decide over pricing by a bargaining process between representatives from different regions. When regions are asymmetric, it is shown that decisions by legislative bargaining will in most plausible cases lead to higher welfare than the type of uniform pricing studied in Section 4.1 above. Loosely speaking, bargaining generally outperforms uniform pricing whenever driver majorities are not too different between the regions. The intuition is that bargaining allows for differentiated pricing; this matters when regions are not symmetric. See Appendix 3 for details.

For completeness sake, consider bargaining when <u>drivers have a majority in one region</u> (say, region 1) but not in the other (say, region 2). We find the tax rules

$$\tau_1^b = \frac{\beta V_1}{K_1^0} + \{\frac{\theta_1}{\eta_1} - 1\} \frac{V_1}{\frac{\partial V_1}{\partial \tau_1}} \tag{19}$$

$$\tau_2^b = \frac{\beta V_2}{K_2^0} + \{\frac{1-\theta_2}{\eta_2} - 1\} \frac{V_2}{\frac{\partial V_2}{\partial \tau_2}} \tag{20}$$

The driver from region 1 wants the very low toll (if spill overs are limited) in his own region and a high toll in the other region, but the non-driver of region 2 wants the revenue maximizing toll everywhere. Bargaining leads to a mixture of these wishes, as expressed by (19)-(20). If spill-overs are absent, then the bargained outcome is again the same as under decentralization, see (6) and (7). Small spill-overs imply that decentralization outperforms centralized decisions, but for high spill-overs the opposite may hold.

Finally, bargaining between two regions where non-drivers have a majority gives revenue maximizing charges everywhere, as it does under decentralization.

We summarize our results on bargaining in the following proposition.

PROPOSITION 3. CENTRALIZED DECISIONS THROUGH BARGAINING

a. ***Bargaining between regions leads to the first-best outcome when all voters in both regions are drivers.***
b. ***Bargaining may perform better than decentralization if drivers have large majorities and spill-overs are substantial.***
c. ***Bargaining does not improve the outcomes if non-drivers have a majority in both regions.***
d. ***If regions are symmetric, bargaining leads to the same outcomes as centralized decisions under a uniformity constraint.***
e. ***If regions are asymmetric, either bargaining or uniform price restrictions may give the highest welfare. Bargaining outperforms uniformity unless driver majorities are quite different.***

Of course, we could combine the two constraints and study bargaining under a uniformity constraint for the user prices. That is, representatives would have to agree on one uniform toll level that applies universally in all regions. One easily shows that imposing uniformity (which was a good idea with a non-cooperative legislature) on the bargaining process is not necessarily a good idea[19]. The intuition is easy: imposing a uniform toll level destroys opportunities for differentiating tolls according to different local conditions. If spill-overs and voting shares differ, transport volumes and congestion will differ across regions, requiring non-uniform toll levels. As under bargaining extreme toll differences have been smoothed out, forcing tolls to be uniform may be welfare-reducing.

[19] A formal proof is available from the authors.

4.3. Comparing alternative institutions for pricing existing capacity

We have now studied decentralized tolling decisions on a given infrastructure and compared it with three alternatives for centralized decision making: a minimum winning coalition, a minimum winning coalition with a uniformity constraint, and a negotiated solution between regions. It will be instructive to illustrate comparative findings using the numerical example referred to before. To do so, we first consider a full comparison under symmetry, then we discuss the case of asymmetric regions. Formulas for toll and welfare levels have been provided in Appendices 1, 2 and 3.

Assuming symmetry, we can illustrate the results we obtained for the numerical example graphically, see Figure 6. Note that under symmetry uniform pricing and bargaining yield the same welfare outcome. We can then distinguish three parameter zones. In zone 1, decentralization outperforms uniform and bargained solutions, which in turn are better than pure centralization. This holds true when some voters are non-drivers and, conditional on the fraction of non-drivers, when spill-overs are not too important. Of the four political systems considered, decentralization is therefore more likely to be the best option when spill-overs are very small or, when they are not, when many voters in the region are non-users. In zones 2 and 3, uniform and bargained solutions welfare-dominate the other systems. These centralized decision-making systems are more likely to be optimal when there are high spill-overs and users have large majorities. Note that in zone 2 pure centralized decisions are the worst possible political system. In zone 3, decentralization performs worst.

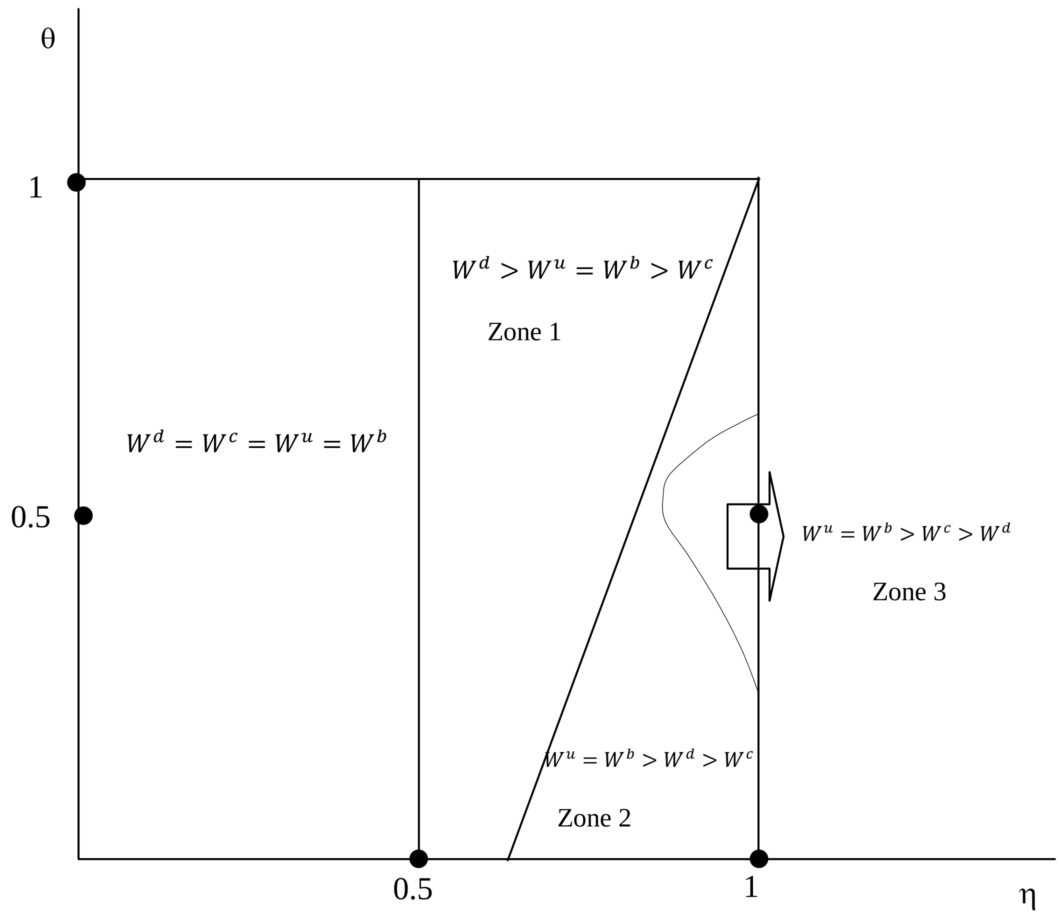


Figure 6: Welfare comparison between the four political systems (symmetric case)

If we allow for asymmetries between regions in the sense that both spill-overs and voting majorities can differ between regions, presenting the results graphically is not obvious. We therefore illustrate what happens in Table 2, where we numerically analyze asymmetries step by step. The table consists of three parts. In the upper part we make the degree of spill-overs different between the regions, for different levels of voting majorities. The middle part looks at the role of asymmetric voting majorities for different levels of spill-overs. The bottom part illustrates the possibility for uniform pricing to be optimal under rather extreme conditions. In each row, the alternative that yields the highest welfare is indicated in bold and underlined. For purposes of comparison, note that the first best welfare in the example is 517.

Several observations stand out from Table 2. First, the results confirm that, under symmetry, decentralization yields the highest welfare as long as spill-overs are not too important; bargaining and uniform pricing perform better if the opposite holds. Second, without further restrictions centralized decisions under a minimum winning coalition never yields higher welfare than all alternatives considered. But, as shown theoretically, requiring uniform prices across regions as well as bargained solutions improve the performance of centralized decisions. Third, when driver majorities are large, asymmetries in spill-overs improve the performance of bargaining relative to decentralization. However, bargaining becomes less attractive compared to decentralization when voting majorities are asymmetric. Finally, bargaining yields substantially higher welfare than uniform pricing when spill-overs are highly asymmetric between regions. But the lowest part of the table indicates that, when both voting majorities and spill-overs are very asymmetric, uniform pricing may be optimal[20].

[20] The conditions for this to happen are spelled out in more detail in Appendix 3. These conditions require very asymmetric voting majorities and (i) either asymmetric and large spill-overs or (ii) asymmetric and very small spill-overs.

θ_1	θ_2	η_1	η_2	W^d	W^c	W^u	W^b
0.9	0.9	0.9	0.9	**517**	485	516	516
0.5	0.9	0.9	0.9	515	506	511	**516**
0.1	0.9	0.9	0.9	510	495	495	**516**
0.9	0.9	0.7	0.7	**515**	461	512	512
0.5	0.9	0.7	0.7	**515**	486	500	512
0.1	0.9	0.7	0.7	509	461	461	**512**
0.9	0.9	0.5	0.5	**497**	274	484	484
0.5	0.9	0.5	0.5	**507**	379	441	484
0.1	0.9	0.5	0.5	**501**	274	274	484
0.9	0.9	0.9	0.9	**517**	485	516	516
0.9	0.9	0.9	0.7	**516**	478	514	514
0.9	0.9	0.9	0.5	**507**	384	500	503
0.5	0.5	0.9	0.9	513	**516**	**516**	**516**
0.5	0.5	0.9	0.7	514	514	514	**516**
0.5	0.5	0.9	0.5	**515**	500	500	508
0.1	0.1	0.9	0.9	503	494	**516**	**516**
0.1	0.1	0.9	0.7	504	478	514	**515**
0.1	0.1	0.9	0.5	**504**	385	500	503
0.1	0.5	0.5	1	509	502	**514**	513
0.2	0.5	0.5	1	511	506	**513**	512
0.3	0.5	0.5	1	**513**	507	511	512
0.1	0.5	0.5	1	509	502	**514**	513
0.1	0.5	0.6	0.9	509	503	512	**515**
0.1	0.5	0.7	0.8	509	498	508	**514**

Table 2: numerical example for regions asymmetric in parameters

4.4 Federal constitutional rules

Earlier in this section, we considered the ranking of different institutional settings in welfare terms. A logical follow-up question is whether and how restrictions that lead to favorable welfare outcomes can be embedded in federal constitutions? Constitutions can be seen as setting the stage for the long term game between regions or between groups of citizens within a region (for example, users versus non-users, different language groups, different ethnicities, etc.). Note that both dimensions are highly relevant in our setting: under some political systems, regions could be exploited by other regions; moreover, we found that within regions one group of voters might be subject to exploitation by another group (for example, users by non-users).

Compared to the previous sections, we need to add a constitutional stage to the analysis. Will the regions, or groups of citizens within regions, be willing to transfer decision making power to the central level? If so, what decision rules do they accept at the central level? Given the structure of our model, alternative central decision rules can be simple minimum winning coalition, minimum winning coalition with uniform pricing, or bargaining. Importantly, for regions to allow decisions to be taken at the federal level, it is typically not sufficient that central decisions always generate a higher total expected welfare than decentralized decision making. What is required is that a very large fraction of the regional population is better off with a federal allocation. In fact, many constitutions in federal countries require a two-third majority for changes in the decision power of the central and decentral levels. In our setting with only 2 regions and equal populations, what we need is the requirement that a large majority in both regions gain in the centralized allocation compared to the decentralized solution.

Although a complete analysis of constitutional design is outside the scope of this paper, in the remainder of this subsection we analyze, for a few specific cases, under what conditions regions will be in favor of transferring decision-making power to the central level. As our Propositions 2 and 3 suggest that the likelihood of this happening is much larger under uniform pricing and bargaining than under a simple minimum winning coalition, we do not discuss the latter in what follows.

First, suppose drivers have a large majority in both regions and assume, for simplicity, that regions are symmetric. In Appendix 4 we formally study under what conditions the drivers (who have, by assumption, a sufficiently large majority so as to be able to decide on transferring power to the central level) will agree to delegate decision making power to the central level. The proposal is that central decisions are taken by a minimum winning coalition subject to a uniformity restriction. If there are spill-overs, we show that drivers in both regions will be better off with centralized decisions and uniform tolls than with decentralized decisions. If there are no spill-overs, of course, drivers are indifferent between centralized and decentralized decisions. The same result also holds when considering a transition from decentralization to centralized bargaining because, under symmetry, uniformity and bargaining yield the same outcomes. Therefore, under the stated conditions drivers in the two regions will agree on a centralized decision-making process subject to uniform pricing constraints. These findings are in line with Hickey (2013), who found in a totally different context that federal bicameralism (which comes close to our bargaining framework) or price uniformity are important for a federation to be acceptable.

Second, assume there is one region with drivers in the majority and one region with non-drivers in the majority. We easily show (see Appendix 4) that the region where non-drivers have a majority will never agree to transfer decision power to the central level; the conflict of interest between drivers and non-drivers is too large. As a consequence, one expects that regions will not voluntarily agree on transferring power to the central level.

Third, consider the least realistic case where non-drivers have a majority in both regions. In that case the choice of constitution does not matter. Prices will always be too high.

Implicit in the previous discussion was the concern of potential exploitation *between* regions. Importantly, note that one also may have to build in additional federal guarantees for minority groups to prevent exploitation of different groups *within* regions[21]. In principle, therefore, constitutional guarantees may have to be included to protect the non-driving or driving minority, depending on the regional situation. For example, when the only public policy instrument available is the price to charge for access to a local public good and to redistribute via head taxes, to guarantee a Pareto improvement for all groups (drivers and non-drivers in both regions) one has to charge a road user fee that is smaller than the marginal external congestion cost[22]. This protects the drivers from being overcharged. The congestion fee has to be smaller than the marginal external congestion cost, because revenues are shared

[21] This is done in some federal states; for example, certain federal restrictions in Belgium serve to protect language groups.

[22] This type of constraint exists in the EU, but it was issued mainly to prevent exploitation of transit traffic in some member states. The constraint takes the form of a maximum limit to the fee in function of road capacity costs. These capacity costs are a very crude approximation of the marginal congestion costs. In theory, the mechanism can produce optimal pricing and capacity, even if the federal government does not know congestion costs (Van der Loo and Proost (2013)).

by non-drivers. Only a maximum needs to be imposed because regions or federations with a majority of non-drivers will charge the maximum fee that is allowed. When drivers have a small majority they will charge a congestion fee that is below the maximum, as fee revenues are shared with all other members of the region (under decentralization) or the federation (under centralized decisions).

We summarize our findings in the following proposition.

PROPOSITION 4: TRANSFERRING DECISION POWER TO THE CENTRAL LEVEL

a. ***If regions are symmetric and drivers have a majority in both regions, both regions will agree to centralize decisions under a uniform pricing constraint.***
b. ***If regions are symmetric and drivers have a majority in both regions, both regions will agree to centralize decisions if the constitution guarantees decision making by bargaining.***
c. ***If drivers have a majority in one region only, the region where non-users have a majority will never agree to transfer decision power to the central level.***
d. ***To protect drivers, price restrictions may have to be imposed that restricts tolls to be equal or less than marginal external cost.***

4.5. Policy implications

What do the findings of this and the preceding section imply for policy? Of course, drawing too strong conclusions from our very stylized models is risky, but a number of insights do seem worth reporting.

First, both centralized and decentralized decisions can be enhanced by constitutional constraints that protect regions and minorities within regions. For example, except under very specific conditions, centralized decisions only make sense if restrictions are imposed to prevent exploitation of regions. As another example, under decentralization, maximum prices may be needed to protect users against non-users.

Second, our results may partly explain the choice for centralization in some, but not all, federal states. If regions are reasonably symmetric, drivers have large majorities and there are regional spill-overs, we found that regions will be willing to transfer decision-making to the central level, provided pricing is uniform across regions. This is what may have happened in federal states such as Belgium, Spain and Switzerland, where the main pricing instrument is still a federal gasoline tax (uniform across different regions). Moreover, our model suggests that under the stated conditions uniform pricing may well be the welfare-optimal system.

However, in the US transport pricing is partially decentralized; apart from a federal gasoline tax which is uniform across states, there is also an additional fuel tax levied by the states (see, for example, Xie and Levinson (2010)). This makes sense. At the state level spill-overs – in terms of the percentage of users of state infrastructure that comes from other states -- are much more limited than in some of the European examples given before, but drivers have very large majorities in almost all states. Our model then suggests that decentralized decisions will be better than centralization. Note that in the US we also see some regional attempts at road pricing (for example, in California).

Third, in cases where pricing decisions are decentralized, the results may help us understand why some pricing instruments are used much more often than others. For example, use of regionally differentiated pricing instruments to control for congestion (such as road pricing) is very limited, while at the same time the use of parking fees is widespread in the center of cities (and even in many small municipalities)[23]. Within the framework of our model, the difference lies in the user majorities, not so much in the level of spill-overs. Although the share of non-users in many regions and urban areas is not small, drivers typically do have a majority. Provided spill-overs are not exceptionally large, drivers then prefer low charges on road use; in fact, the tolls they want may be negative. As a consequence, they will not be in favor of road pricing (certainly not if charges are close to marginal external cost). However, in city centers, local inhabitants typically form a minority of the overall demand for parking space, preferring to impose high charges on parking spots.

Fourth, when regions face quite asymmetric spill-overs, we found that uniform pricing is often dominated by bargained solutions. This will be the case when one region attracts a lot of commuter traffic from other regions (for example, in Belgium there is much more commuting from the Flemish region into the Brussels region than vice versa). Interestingly, negotiations about the introduction of some form of road pricing in Brussels are currently taking place.

Fifth, applying the results of the model in a broader perspective to pricing of public rather than private road transport, our findings seem consistent with the organization of rail and bus transport in some explicitly federal states. In European federal states, trains are typically used more for longer trips than buses; typically, therefore, rail has much larger spill-overs between regions. It follows that federal organization and pricing of rail services is appropriate. In regional bus transport (used more for short trips), however, there are much less

[23] For some other recent political economy explanations for the widely observed opposition to road pricing see, among others, De Borger and Proost (2012) and Russo (2013).

spill-overs. They are organized and priced at the regional level by separate public bus companies.

Finally, more in general, the results are consistent with decentralized decision making and with the limited use of user charges for many local public goods (see Blöchliger (2008)). If spill-overs are small and not all voters are users, decentralization was found to be the system that yields the highest welfare. These conditions may well hold for services such as child care, parks, libraries, hospitals, etc., services that typically are highly decentralized. Interestingly, it is often found that decentralized systems impose restrictions on the use of services in order to explicitly reduce spill-overs: although jurisdictions are typically not allowed to differentiate user fees between residents and non-residents, users do not always have the freedom to consume services in other jurisdictions (see Table 3 in Blöchliger (2008)). This formal restriction on spill-overs may make sense: as observed before, decentralization performs better when spill-overs are small. Moreover, given small spill-overs and a fair number of non-users (as the case may be for libraries, parks, etc.), we also showed that the user charges wanted by the user majority will typically be small.

5. Choosing capacity financed by user prices

In this section, we extend the model by assuming that decisions have to be taken on capacity levels and their financing. We assume user prices can be used to pay for all, or part of, the cost of capacity. In case only part of the cost is covered, the remaining funds come from other (head) taxes.

There is at least one reason why our model is better suited to discuss pricing solutions than to understand capacity decisions. Capacity decisions for most congestible public goods (consider roads) have a long lead time and are durable, lasting for several political terms. Our model should therefore ideally extend over several terms and take into account the possibility of regime switching at the federal level. However, this induces the elected representatives to try to guarantee the supply of their preferred public goods in the future by committing resources now (see, for example, Glazer (1989)). These issues will not be dealt with in the current paper; we continue to work within a static framework.

We start by analyzing the first best, then we look at decentralized and centralized decision making and provide a brief comparison of the various systems. Finally, we turn to the most important contribution of this section, viz., the role of budgetary restrictions on the outcomes of the decision making process, both under decentralized and centralized decisions.

5.1. The First Best solution

Looking at an arbitrary region and ignoring the regional index for convenience, the region solves

$$\underset{\tau,\,K}{\text{Max}}\ \{\int_0^V P(V)dV - V.g(V)\} + \tau V - \rho K \tag{21}$$

We show in Appendix 5 that, unsurprisingly, the optimal user price equals marginal external cost; furthermore, optimal capacity equates marginal benefit and marginal cost. We find

$$\tau = \frac{\beta V}{K}; \quad \beta\left(\frac{V}{K}\right)^2 = \rho \tag{22}$$

It easily follows from (22) that, at the optimum we have $\tau V = \rho K$: the first-best gives exact cost recovery. Given the assumptions underlying our model -- our specifications imply homogeneity of degree 0 of the user cost function in volume and capacity, and constant returns to scale in capacity costs -- this could be expected (see Mohring and Harwitz (1962), De Palma and Lindsey (2007)).

5.2. Decentralized and centralized solutions

In this subsection, we look at the various political systems studied before. However, we can be very brief, as the analysis yields results in line with previous sections. As we will see, for all cases studied, the pricing rules are the same as in Sections 3 and 4, and capacity rules are always the first-best rules (although, of course, the volumes are not first-best).

Decentralized decisions

First let drivers have a majority in an arbitrary region. What capacity and user price choices would a decisive road user want? She solves

$$\underset{\tau,\,K}{\text{Max}}\ \frac{\theta}{D}\{\int_0^V P(V)dV - V.g(V)\} + \frac{\tau V - \rho K}{R}$$
$$s.t. K \geq 0$$

The price and capacity rules obtained are the following (see Appendix 5):

$$\tau = \frac{\beta V}{K} + \{\frac{\theta}{\eta} - 1\} \frac{V}{\frac{\partial V}{\partial \tau}} \tag{23}$$

$$\beta \left(\frac{V}{K} \right)^2 = \rho \tag{24}$$

The user price rule boils down to what we had in section 3, see expression (6). Capacity will be optimally adjusted as in the first-best, but the adjustment is based on a different volume of use. If there are large spillovers, user prices will be higher and capacity lower than first-best.

Second, if non-users have a majority in the region, it is easily shown that they will prefer the toll that maximizes the net revenues of tolling (see (7)); the capacity rule will be the same as (24).

The centralized solution

Initially we consider decisions by a minimum winning coalition with no further constitutional restrictions.

First, if a car user in region 1 is decisive at the central level, he chooses user prices and capacity levels in both regions so as to solve the following optimization problem (the analysis is entirely analogous if the user from region 2 is decisive):

$$\begin{aligned} \underset{\tau_1, \tau_2, K_1, K_2}{\text{Max}} \quad & \frac{\theta_1}{D_1} \{ \int_0^{V_1} P(V_1) dV_1 - V_1 . g(V_1) \} + \frac{1-\theta_2}{D_1} \{ \int_0^{V_2} P(V_2) dV_2 - V_2 . g(V_2) \} + \frac{\tau_1 V_1 + \tau_2 V_2 - \rho K_1 - \rho K_2}{R} \\ s.t. \quad & K_1 \geq 0 \\ & K_2 \geq 0 \end{aligned} \tag{25}$$

The user price and capacity rules can be written as (using the same steps as described in Appendix 5):

$$\tau_1 = \frac{\beta V_1}{K_1} + \{\frac{2\theta_1}{\eta_1} - 1\} \frac{V_1}{\frac{\partial V_1}{\partial \tau_1}}; \qquad \beta \left(\frac{V_1}{K_1} \right)^2 = \rho \tag{26}$$

$$\tau_2 = \frac{\beta V_2}{K_2} + \{\frac{2(1-\theta_2)}{\eta_1} - 1\} \frac{V_2}{\frac{\partial V_2}{\partial \tau_2}}; \qquad \beta \left(\frac{V_2}{K_2} \right)^2 = \rho \tag{27}$$

Again, the toll rules are identical to what we found in Section 3, and the capacity rule is the first-best one. Together, these rules just strengthen possible exploitation of one region by the other as in the fixed capacity case of Section 3.3. To give an extreme example, start from a situation with no spill-overs. The user of region 1 then wants a very low user price in his own region and a high user price in the other region. The low user price in the own region leads to high demand, and the capacity rule then implies high investment in infrastructure. The high user price in the other region yields low demand and low investment. One interpretation is that the decisive policy maker uses the revenues gained in the other region to finance high capacity investments in his own region. Of course, if spill-overs are very large, exploitation goes the other way around. The user from region 1 (who drives more kilometers in the other region than in his own region) may now want a low user price and high investment in the other region and, at the same time, high user prices and limited investment in the own region (where he does not drive much).

Second, assume that the group that doesn't drive at all is decisive. This group will prefer a solution where the net revenue from user fees (on essence, revenues minus capacity costs) is maximized in both regions.

Finally, imposing uniformity restrictions or considering bargained solutions yield equally predictable rules. In all cases, we find that toll rules are the same as in Section 4, and in all cases the capacity rule is first best[24].

5.3. Comparing decentralized and centralized decisions

Noting that in all cases pricing rules are as before and capacity rules are the same as in first-best, it is clear that having capacity as an extra decision variable on top of user prices will not change the nature of the relative welfare results derived in Sections 3 and 4. Intuitively, as long as the same party decides on prices and capacity and we give it more degrees of freedom, this just reinforces the deviations from the first-best solution. The welfare ranking of the various decision making systems will depend on the same parameters as before and will move in the same direction when these parameters change[25]. In sum, as long as there are no other constraints imposed, the relative welfare results we found for the case of fixed

[24] We did not consider uniform provision of capacity. In that case, of course, the capacity rule will deviate from the first best rule.

[25] For example, using (23)-(24) it again immediately follows that, when there are no spill-overs and all voters are users ($\theta_1 = \theta_2 = \eta_1 = \eta_2 = 1$), decentralization yields the first-best outcome. Similarly, using (26)-(27) plus their equivalent when the representative from region 2 is decisive we find that, when $\theta_1 = \theta_2 = 0.5; \eta_1 = \eta_2 = 1$, centralization is first best. It will also still be the case that centralization will only perform better than decentralized decisions when spill overs are substantial (θ_i close to 0.5) and drivers have a large majority in both regions.

capacity (Section 3 and 4) carry over to the case with variable capacity. We summarize this observation in Proposition 4.

PROPOSITION 4. CHOICE OF CAPACITY AND USER PRICES.

When governments make decisions on both user prices and capacity, the ranking of political institutions is the same as in the case where capacity is fixed.

5.4. Institutions constraints: the role of budgetary constraints

As argued in the preceding subsections, as the objectives of the decision makers are unaffected and they have an extra policy instrument, no changes in the relative performance of different political systems will occur. It is only when we somehow constrain pricing and capacity choices that we can expect fundamentally new results. One of the promising additional constraints to be considered is a self-financing constraint for capacity. Exact cost recovery is indeed not guaranteed under the political decision making systems studied in this paper. The reason is that decision makers can always rely on other taxes to finance capacity, or they can use the revenues of user pricing to decrease other taxes.

In this section, we therefore discuss the effect of a self-financing or earmarking constraint on decentralized and centralized decisions. Many variants of this constraint exist but we will, for expositional reasons, focus on a simple and strict specification, requiring full cost recovery. This is particularly convenient because, as observed in section 5.1, under our assumptions the first-best user fee exactly covers capacity costs. We consider imposing the constraint at the regional and at the federal level.

5.4.1. Cost recovery at the regional level

To demonstrate the impact of the cost recovery constraint, suppose <u>drivers have a majority in the region</u>, and assume the decision makers are restricted by exact cost recovery. Consider the following maximization problem, where a multiplier λ is used for the budget constraint (and assuming an internal solution for capacity):

$$\underset{\tau,\,K}{\text{Max}} \;\; \frac{\theta}{D}\{\int_0^V P(V)dV - V.g(V)\} + \frac{\tau V - \rho K}{R}$$

$$s.t. \quad \frac{\tau V - \rho K}{R} = 0$$

$$K \geq 0$$

Using the same steps as described in Appendix 5, we find:

$$\tau = \frac{\beta V}{K} + \{\frac{\theta}{\eta}\frac{1}{1+\lambda} - 1\}\frac{V}{\frac{\partial V}{\partial \tau}} \tag{28}$$

$$\beta\left(\frac{V}{K}\right)^2 = \rho \tag{29}$$

Note that, in the absence of the budgetary restriction ($\lambda = 0$), the user price rule obviously boils down to what we had in the previous section, see expression (23). More importantly, in the presence of the constraint, substituting (28)-(29) in the budget constraint and rearranging we immediately find:

$$\tau = \frac{\beta V}{K}; \quad \beta\left(\frac{V}{K}\right)^2 = \rho \tag{30}$$

This the first-best outcome.

Ex ante, the fact that a regional cost recovery restriction produces the first-best was not obvious; ex post, of course, it is not surprising. We know that the first-best yields exact cost recovery, so imposing this as a constraint yields the first best[26]. The intuition is most easily understood when there are no spill-overs. Then we know the driving voter would like to impose a user price that is too low compared to the social optimum (see (28)), but the budget restriction prevents him from doing so. As he does suffer from congestion externalities, the best he can do is set a user price at marginal external cost. Introduce now spillovers. As long as the local decision maker cannot price discriminate between local and non-local users, both groups have the same preferences, and users pay all the costs, the solution chosen by the local user coincides with the best solution for the non-local user. We therefore achieve the first best.

Finally, what are the implications of the regional cost recovery constraint if non-users have a majority in the region. If there were no such constraint, they would prefer the (net) revenue maximizing toll and, hence, a low capacity level. However, when the budget constraint is strict, non-users are in theory indifferent between all solutions that balance the budget, including no capacity at all. One possibility is that they select, among all feasible solutions, the best solution for the users. All one needs is that the minority of users has a very

[26] As mentioned above, Brueckner (2013) showed a similar result in a different context. Furthermore, Ogawa and Wildasin (2009) use a tax competition framework in which externalities depend on the stock of capital in a jurisdiction; capital is mobile and can be taxed. Under some conditions they also find that decentralization yields efficiency, even in the presence of spill-overs.

small weight in the objective function of the majority of non-users[27]. Seen from this perspective, the major advantage of the regional budget constraint is then to protect users from exploitation by non-users.

5.4.2 Cost recovery at the federal level

For centralized decision-making, we focus on the solution by a minimum winning coalition (similar insights are obtained for uniform pricing and bargaining). We first consider the case with a budget restriction at the level of the federation, then we look at regional constraints.

Let there be a budget restriction at the federal level. If a local car user in region 1 is decisive, he chooses user prices and capacity levels in both regions but he must make sure *total revenues in the federation cover total capacity costs*. He specifically solves the following problem (the analysis is entirely analogous if the user from region 2 is decisive):

$$\underset{\tau_1,\tau_2,K_1,K_2}{\text{Max}} \ \frac{\theta_1}{D_1}\{\int_0^{V_1} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\tau_1 V_1 + \tau_2 V_2 - \rho K_1 - \rho K_2}{R}$$

$$\begin{aligned} s.t. \quad & \frac{\tau_1 V_1 + \tau_2 V_2 - \rho K_1 - \rho K_2}{R} = 0 \\ & K_1 \geq 0 \\ & K_2 \geq 0 \end{aligned}$$

Assuming an internal solution for both capacities, the user price and capacity rules can be written as (λ_1 is the Lagrange multiplier associated with the federal budget constraint)

$$\tau_1 = \frac{\beta V_1}{K_1} + \{\frac{2\theta_1}{\eta_1}\frac{1}{1+\lambda_1} - 1\}\frac{V_1}{\frac{\partial V_1}{\partial \tau_1}}; \qquad \beta\left(\frac{V_1}{K_1}\right)^2 = \rho \tag{31}$$

$$\tau_2 = \frac{\beta V_2}{K_2} + \{\frac{2(1-\theta_2)}{\eta_1}\frac{1}{1+\lambda_1} - 1\}\frac{V_2}{\frac{\partial V_2}{\partial \tau_2}}; \qquad \beta\left(\frac{V_2}{K_2}\right)^2 = \rho \tag{32}$$

It is easy to understand that the federal cost recovery restriction does not produce first-best results. It implies, using (31)-(32) and simple manipulations, that

$$V_1\left(\tau_1 - \frac{\beta V_1}{K_1}\right) = -V_2\left(\tau_2 - \frac{\beta V_2}{K_2}\right)$$

[27] One way to formalize this result would be to assume lexicographic preferences for non-users, in the sense that the welfare of users counts whenever non-users are indifferent.

The federal constraint therefore restricts the deviations between charges and marginal external cost in the different regions, but does not imply the first-best. It does not even necessarily improve federal decision making. Why not? One easily shows the intuitive result that, if the unconstrained revenues are insufficient to cover overall capacity costs, imposing the budget constraint raises tolls in both regions (the opposite holds if unconstrained revenues are excessive). If spill overs are small, then the higher user price in the own region brings the toll closer to marginal external congestion cost; it further reduces capacity investment in region 1 and hence 'requires' less extra funding from the other region. Welfare increases. However, the higher toll in the other region (where the toll already exceeded marginal external cost in the absence of the constraint) reduces welfare. The overall expected welfare effect is hence unclear. If in the current setting spill-overs are substantial, a similar story applies.

If the group that doesn't drive at all is decisive, we have a similar situation as under decentralization: this group will be indifferent among the many solutions that satisfy the federal budget constraint. This may yield a first best outcome if the minority of users in both regions has at least some weight in the objective function of non-users.

Finally, suppose the *federal level operated under regional budget restrictions*. In other words, decisions are taken centrally but have to satisfy regional cost recovery constraints. Again, one then easily shows that this does produce the first-best.

5.4.3 Decentralized and centralized equilibria under self-financing constraints

We just derived three useful results. First, the most important result is probably that decentralized decision making with a regional budget constraint is always first-best. The two main drivers behind this result are that, first, every decisive user is forced to treat the non-local users in the same way as he treats himself and, second, non-users can no longer benefit from excessively taxing users. The main improvement obtained by self-financing constraints at the regional level is indeed that non-users can no longer exploit users. Because of this, under a mild condition, a first-best solution has become possible. The required condition is that the decisive non-users are generous enough to allow this Pareto improvement in which they themselves are not affected. The second important result is that a budget constraint at the aggregate federal level may not be very helpful when users are decisive, but it may generate first-best results when non-users are decisive. A third finding is that centralized decisions also produce the first-best if the central level agrees to satisfy regional budgetary constraints.

We summarize in Proposition 5.

PROPOSITION 5. CHOICE OF CAPACITY AND USER PRICES WITH BUDGET CONSTRAINTS

a) ***When there is a regional cost recovery constraint, the decentralized equilibrium generates a first-best solution – this also holds for the case where users are a minority in as far as the non-users do not block a Pareto improvement for the users.***

b) ***A federal cost recovery constraint does not necessarily produce a first-best solution. It may not even improve the federal outcome.***

c) ***Central decision-making under regional cost recovery constraints also yield first-best outcomes.***

d) ***One advantage of regional cost recovery restrictions is that they protect users from exploitation by non-users.***

6. Conclusions and caveats

We considered a political economy model of pricing and investment decisions for congestible local public goods in a federal state. The model assumed majority voting at the regional level, and federal decisions were assumed to be taken by elected regional representatives. Different cooperative and non-cooperative political decision-making systems were considered. Moreover, we looked at the role of various institutional constraints, including uniform pricing and regional cost recovery restrictions. The model allowed for heterogeneity within regions by considering the preferences of both users and non-users of the regional infrastructure. Heterogeneity between regions was captured by allowing differences in spill-overs and in user majorities in different regions. The two types of heterogeneity captured by the model implied possible exploitation of regions and, within regions, exploitation of particular groups of voters (for example, infrastructure users).

The analysis produced a number of interesting results. For example, assuming centralized decisions are obtained according to the principle of a minimum winning coalition, it was shown that decentralization may well yield higher welfare than centralized decision-making, even with very large spill-overs. In fact, centralized decisions are to be preferred only when two conditions simultaneously hold: users have large majorities, and spill-overs are such that the demand for the use of the infrastructure in a given region is approximately equal for local users and users from outside the region. If non-users form a majority in both regions then we found that centralized and decentralized decision making yield the same outcomes; however, the outcome is very undesirable, with prices that are much higher than marginal social cost. We further showed that both bargaining between elected regional representatives and requiring user prices to be uniform across regions greatly improve the efficiency of centralized decision making. If regions are symmetric then we found both bargaining and uniform pricing to yield higher welfare than decentralized outcomes if drivers have large majorities in the regions and there are large spill-overs. Interestingly, to avoid exploitation of users by non-users, we argued that restrictions may have to be introduced that reduce tolls below marginal external cost.

Importantly, although fully endogenous governance structures were outside the scope of the paper, we showed that – in symmetric regions with a user majority -- both regions will agree to transfer decision power to the central level if constitutional arrangements either impose price uniformity across regions, or explicitly imply a decision-making process through bargaining. However, if non-users are a majority in a given region we show that they will never agree to transfer decision power to the central level.

Finally, when joint decisions have to be made on capacity provision and pricing of infrastructure use, we showed that imposing cost recovery constraints on the individual regions strongly enhances the performance of decentralized decision making. In fact, under a mild additional assumption, they allow to attain the first-best social optimum. One of the main advantages of a regional self-financing rule is that it protects users from being exploited by non-users. A federal budget restriction, however, does not necessarily produce favorable welfare outcomes.

The results of this paper have relevance for understanding actual policy making in countries with a multi-layered government structure, emphasizing the interaction between the conflicting objectives of users and non-users of the infrastructure and the biases introduced by the political process. First, they contribute to understanding why some decisions are taken at the central level whereas others are not and, relatedly, why we observe so many instances of uniform pricing. For example, the results are consistent with decision making in federal states such as Germany, Belgium, Spain and Switzerland, where the main pricing instrument for car use is a federal gasoline tax, uniform across different regions. At the same time, they are also consistent with partial decentralization of fuel taxes in the US. Second, the model provides an explanation why the use of user fees for road infrastructure (road pricing) is not widespread, whereas parking charges are. Finally, the results are consistent with the observation that decentralized systems often impose restrictions on the use of services in order to explicitly reduce spill-overs: although jurisdictions are typically not allowed to differentiate user fees between residents and non-residents, users do not always have the freedom to consume services in other jurisdictions (see Blöchliger (2008)). This formal restriction on spill-overs makes sense, because decentralization performs better when spill-overs are small.

Of course, it is clear that we made a series of restrictive assumptions that may have to be reconsidered in future work. Our model was restricted to two regions and, although they could be asymmetric in terms of spill-over and user share parameters, they were symmetric in all other dimensions, such as population size and income. One expects that size differences would strengthen the exploitation of the smaller region under centralized decisions by a minimum winning coalition, and that therefore the role of uniform taxes would become even more relevant. Introducing income differences would suggest extending the set of available tax instruments to include a (nonlinear) income tax. This would allow studying the political economy of centralization within a nonlinear optimal taxation framework. This was outside the scope of the current paper. Finally, generalization to an arbitrary number of regions is not a priori obvious. As long as regions are symmetric and they have equal probability of being decisive at the central level much of the analysis goes through. However, matters become much more complicated with multiple asymmetric regions, because the identification of minimum winning coalitions is no longer straightforward.

References

Besley T., Coate S. (2003), Centralized versus decentralized provision of local public goods: a political economy approach, Journal of Public Economics, 87, 2611-37.

Besley, T.,(2005) Principled Agents? The Political Economy of Good Government, Lindahl lectures, Oxford University Press.

Blöchliger, H. (2008), Market mechanisms in public service provision, OECD network on foscal relations across governments, working paper n° 6, OECD.

Borck, R. and M. Wrede (2005), Political economy of commuting subsidies, Journal of Urban Economics 57, 478-99

Brueckner, J.K. (1981), Congested Public Goods: The Case of Fire Protection, Journal of Public Economics 15, 45-58.

Brueckner, J.K. and Selod (2006), The political economy of urban transport-system choice, Journal of Public Economics 90, 983, 1005.

Brueckner, J.K. (2013), Efficiency of Decentralized Road Investment and Pricing in a Multi-Jurisdictional City with Spillovers, University of California-Irvine, August.

Calthrop, E., De Borger, B. and S. Proost (2010), Cost-benefit analysis in distorted economies, Transportation Research B 44, 850-869.

Craig, S.G. (1989), The Impact of Congestion on Local Public Good Production, Journal of Public Economics 32, 331-353.

De Borger, B. and S. Proost (2012), A political economy model of road pricing, Journal of Urban Economics 71, 79-92.

De Borger, B., F. Dunkerley and S. Proost (2007), Strategic Investment and Pricing Decisions in a Congested Transport Corridor, Journal of Urban Economics 62, 294-316.

De Borger, B., S. Proost, and K. Van Dender (2005), Congestion and Tax Competition on a Parallel Network, European Economic Review 49, 2013-2040.

De Palma, A. and R. Lindsey (2007), Transport user charges and cost recovery, in De Palma, Lindsey and Proost (eds.), Investment and the use of user price and tax revenues in the transport sector, Elsevier, 29-57.

Ferguson, G. (2013), Transportation system-choice in a fragmented metropolis, Department of Planning, Policy and Design, University of California-Irvine

Glazer, A. (1989), Politics and the choice of durability. American Economic Review 79 (5), 1207-1213.

Hatfield, J. and Padro i Miquel (2012), A Political Economy Theory of Partial Decentralization, Journal of the European Economic Association 10 (3), 605-633.

Hickey, R., (2013), Bicameral bargaining and federation formation, Public Choice 154, 217–241.

Hulten, C. and R. Schwab (1997), A fiscal federalism approach to infrastructure policy, Regional Science and Urban Economics 27, 139-159.

Kessler, A.(2013), Communication in Federal Politics: Universalism, Policy Uniformity, and the Optimal Allocation of Fiscal Authority, forthcoming Journal of Political Economy

Kidokoro, Y., 2006. Benefit estimation of transport projects – a representative consumer approach. Transportation Research Part B 40 (7), 521–542.

King, D., Manville, M. and D. Shoup (2007), The political calculus of congestion pricing, Transport Policy 14, 111–123.

Knight, B. (2004). Parochial interests and the centralized provision of local public goods: evidence from congressional voting on transportation projects. Journal of Public Economics 88, 845-866.

Knight, B. (2008), Legislative representation, bargaining power and the distribution of federal funds: evidence from the US congress, The Economic Journal, 118 (October), 1785–1803

Levinson, D. (2002), Financing transportation networks, Edward Elgar Publishing.

Lockwood, B. (2002), Distributive politics and the costs of centralization, Review of Economic Studies 69, 313-337.

Lockwood, B. (2006), The political economy of decentralization, in Ahmad, E. and G. Brosio, Handbook on fiscal federalism, Edgar Elgar.

Marcucci, E., Marini, M. and D. Ticchi (2005), Road pricing as a citizen-candidate game, European Transport 31, 28-45.

Oates, W.E. (1972), Fiscal federalism, Harcourt, Brace Jovanovich, New York.

Oates, W.E. (2005), Toward a Second-Generation Theory of Fiscal Federalism, International Tax and Public Finance, 12(4), 349–373.

Ogawa, H. and D. Wildasin (2009), Think locally, act locally: spill-overs, spillbacks, and efficient decentralized policymaking, American Economic Review 99, 1206-2017.

Persson, T. and G. Tabellini (2000), Political Economics: Explaining Economic Policy. MIT Press, Cambridge.

Proost, S., Dunkerley,F., Van der Loo,S., Adler, N., Bröcker,J., and A. Korzhenevych (2013), Do the selected Transeuropean investment projects pass the Cost Benefit Test, forthcoming in Transportation

Redoano, M. and K. Scharf (2004), The political economy of policy centralization: direct versus representative democracy, Journal of Public Economics 88, 799–817.

Rubinstein, A. (1982), Perfect equilibrium in a bargaining model, Econometrica 50, 97–109.

Russo, A. (forthcoming 2013), Voting on road congestion policy, Regional Science and Urban Economics.

Tiebout, C (1956), A pure theory of local expenditures, Journal of Political Economy 64, 416-24.

Xie, F. and D. Levinson (2009), Governance choice on a serial network, Public Choice 141, 189–212.

Ubbels, B. and E. Verhoef (2008), Governmental competition in road charging and capacity choice, Regional Science and Urban Economics 38, 174–190.

Van der Loo, S. and S. Proost (2013), The European road pricing game: how to enforce optimal pricing in high-transit countries under asymmetric information, forthcoming Journal of Transport Economics and Policy.

Weingast, B. (1979), A rational choice perspective on congressional norms. American Journal of Political Science 23, 245–262.

Weingast, B. (2009), Second generation fiscal federalism: The implications of fiscal incentives, Journal of Urban Economics 65, 279–293.

Appendix 1: Comparing welfare under centralized and decentralized decision making: user pricing on given capacity

In this appendix, we provide a more formal analysis of the welfare comparison between centralized and decentralized decisions. Welfare for region *i* *(i=1,2)* was in all cases studied defined as follows

$$W_i = \{\int_0^{V_i} P(V_i)dV_i - V_i.g(V_i)\} + \tau_i V_i - \rho K$$

We assume linear demand throughout the analysis. Welfare can then be rewritten as

$$W_i = \frac{b}{2}(V_i)^2 + (\tau_i)(V_i) - \rho K \quad \text{(A1.1)}$$

It is easy to show that welfare is a concave function of the toll in the region. This observation will be crucial in showing the relative welfare performance of the different political systems.

Of course, welfare, the toll level and the volume depends on the specific political system studied and, in the case of centralized decisions, on who is decisive at the central level. Let us denote welfare in region *i (i=1,2)* under decentralization by W_i^d. Similarly, we denote welfare in region *i (i=1,2)* under centralization when the representative from region *j (j=1,2)* is decisive at the central level by $W_i^c(j)$. Volumes and tolls in the different regimes are distinguished using analogous notation. As capacity is assumed to be given and does not play much of a role in this section, we assumed in (A1.1) for simplicity $K_1^0 = K_2^0 = K$.

We want to compare welfare under decentralized decisions with (expected) welfare under centralization. In other words, we are interested in whether

$$(W_1^d + W_2^d) > \left\{0.5\left[W_1^c(1) + W_1^c(2)\right] + 0.5\left[W_2^c(1) + W_2^c(2)\right]\right\}$$

We start by looking at the case drivers have a majority in both regions, then we briefly study the other cases.

<u>*Drivers have a majority in both regions*</u>

We first derive the tolls, volumes and welfare levels as a function of parameters only under the various regimes. Given linear demand (expression (1) in the main body of the paper), the user price rule (6) under decentralization can be written as[28]

[28] Positive tolls require $\frac{\beta}{K} + b - \frac{\theta_i}{\eta_i} b > 0$. We assume this condition to hold.

$$\tau_i^d = \frac{\beta V_i^d}{K} - \{\frac{\theta_i}{\eta_i} - 1\}\left(bV_i^d\right) \tag{A1.2}$$

Note further that the equality of generalized price and cost implies

$$a - bV_i^d = \alpha + \frac{\beta V_i^d}{K} + \tau_i^d$$

This leads, after rearrangement, to the following traffic volumes[29]:

$$V_i^d = \frac{a-\alpha}{2A - X_i^d} \tag{A1.3}$$

where

$$A = b + \frac{\beta}{K}; \quad X_i^d = b\frac{\theta_i}{\eta_i} \tag{A1.4}$$

Using (A1.3-A1.4) in (A1.2) gives the tolls in function of the parameters only; we find after simple algebra

$$\tau_i^d = \left[\frac{a-\alpha}{2A - X_i^d}\right]\left[A - X_i^d\right] \tag{A1.5}$$

Finally, substituting (A1.2)-(A1.3) in the welfare expression (A1.1) gives welfare for region *i* under decentralization as

$$W_i^d = \left\{\frac{a-\alpha}{2A - X_i^d}\right\}^2\left[A + \frac{1}{2}b - X_i^d\right] \tag{A1.6}$$

In this expression we have ignored the capacity cost ρK (see (A1.1)). In this section it plays no role whatsoever.

Next turn to centralized decisions. Starting from expressions (8) to (11) and using the same procedure, the respective traffic volumes are easily derived as

$$\begin{aligned} V_1^c(1) &= \frac{a-\alpha}{2A - X_i^c(1)}; \qquad V_1^c(2) = \frac{a-\alpha}{2A - X_1^c(2)} \\ V_2^c(1) &= \frac{a-\alpha}{2A - X_2^c(1)}; \qquad V_2^c(2) = \frac{a-\alpha}{2A - X_2^c(2)} \end{aligned} \tag{A1.7}$$

In these expressions,

$$\begin{aligned} X_1^c(1) &= 2b\frac{\theta_1}{\eta_1}; \qquad X_1^c(2) = 2b\frac{(1-\theta_1)}{\eta_2} \\ X_2^c(1) &= 2b\frac{(1-\theta_2)}{\eta_1}; \quad X_2^c(2) = 2b\frac{\theta_2}{\eta_2} \end{aligned} \tag{A1.8}$$

Tolls as functions of the parameters only are given by

[29] Note that the positive toll restriction mentioned in the previous footnote guarantees positive volumes.

$$
\begin{aligned}
\tau_1^c(1) &= \left[\frac{a-\alpha}{2A - X_1^c(1)} \right] \left[A - X_1^c(1) \right] \\
\tau_1^c(2) &= \left[\frac{a-\alpha}{2A - X_1^c(2)} \right] \left[A - X_1^c(2) \right] \\
\tau_2^c(1) &= \left[\frac{a-\alpha}{2A - X_2^c(1)} \right] \left[A - X_2^c(1) \right] \\
\tau_2^c(2) &= \left[\frac{a-\alpha}{2A - X_2^c(2)} \right] \left[A - X_2^c(2) \right]
\end{aligned} \tag{A1.9}
$$

Finally, for completeness sake, we can write welfare levels under the various regimes under centralization as:

$$
\begin{aligned}
W_1^c(1) &= \left\{ \frac{a-\alpha}{2A - X_1^c(1)} \right\}^2 \left[A + \frac{1}{2}b - X_1^c(1) \right] \\
W_1^c(2) &= \left\{ \frac{a-\alpha}{2A - X_1^c(2)} \right\}^2 \left[A + \frac{1}{2}b - X_1^c(2) \right] \\
W_2^c(1) &= \left\{ \frac{a-\alpha}{2A - X_2^c(1)} \right\}^2 \left[A + \frac{1}{2}b - X_2^c(1) \right] \\
W_2^c(2) &= \left\{ \frac{a-\alpha}{2A - X_2^c(2)} \right\}^2 \left[A + \frac{1}{2}b - X_2^c(2) \right]
\end{aligned} \tag{A1.10}
$$

It is important to note that all toll expressions have the same general form (see (A1.5) and (A1.9))

$$
\tau = \left[\frac{a-\alpha}{2A - X} \right] [A - X]
$$

where the definition of X depends on the regime considered. The characteristics of this relation between tolls and the X's will be crucial in what follows. More specifically, the toll is declining in X at an increasing rate; we have

$$
\begin{aligned}
\frac{\partial \tau}{\partial X} &= -A \left[\frac{a-\alpha}{(2A - X)^2} \right] < 0 \\
\frac{\partial^2 \tau}{\partial X^2} &= -2A \left[\frac{a-\alpha}{(2A - X)^3} \right] < 0
\end{aligned} \tag{A1.11}
$$

Finally, observe that at the first-best outcome we have $X^{FB} = b$.

Armed with previous results, we can now easily study the welfare performance of the different political systems. We first assume zero spill-overs, then we take the general case.

Zero spill-overs

If there are zero spill-overs, we have $\theta_1 = \theta_2 = 1$. It then follows from the definitions in (A1.4) and (A1.8) that

$$X_i^d = b\frac{1}{\eta_i}; \quad X_i^c(i) = 2b\frac{1}{\eta_i}; \quad X_i^c(j) = 0 \quad (i = 1,2;\, j = 1,2; i \neq j) \tag{A1.12}$$

Defining

$$\bar{X}_i^c = 0.5\left[X_i^c(i) + X_i^c(j)\right]; \quad X_i^c(j) = 0 \quad (i = 1,2;\, j = 1,2; i \neq j)$$

it immediately follows from (A1.12) that

$$\bar{X}_i^c = b\frac{1}{\eta_i} = X_i^d$$

Given the shape of the toll as function of *X* (see (A1.11)), this immediately implies $\bar{\tau}_i^c < \tau_i^d$, where $\bar{\tau}_i^c = 0.5\left[\tau_i^c(i) + \tau_i^c(j)\right]$ is the expected toll under centralization in region *i*. Moreover, note from before that under the assumptions made the decentralized toll is less than (provided $\eta_i < 1$) or equal to (if $\eta_i = 1$) the first-best toll. Therefore we have

$$\bar{\tau}_i^c < \tau_i^d \leq \tau_i^{FB} \tag{A1.13}$$

Finally, this in turn implies, given the concavity of welfare in tolls, that

$$E(W_i^c) < W_i^d \leq W_i^{FB}$$

where

$$E(W_i^c) = 0.5\left[W_i^c(i) + W_i^c(j)\right]; \quad i = 1,2;\, j = 1,2; i \neq j$$

is expected welfare in region *i*. Note that the strict inequality holds if $\eta_i < 1$.

It is instructive to illustrate the proof graphically for an arbitrary region, as similar reasoning will be used several times further in the paper. Consider Figure A1. On the lower panel we measure the toll as a function of the relevant *X*, as defined before. The shape of this relation was proven above. On the upper panel we present welfare as a concave function of tolls. The *X*'s on the lower panel are such that they satisfy $\bar{X}^c = X^d < X^{FB} = b$. This generates $\bar{\tau}^c < \tau^d < \tau^{FB}$ on the upper panel. It then immediately follows on the upper panel that $E(W^c) < W^d$.

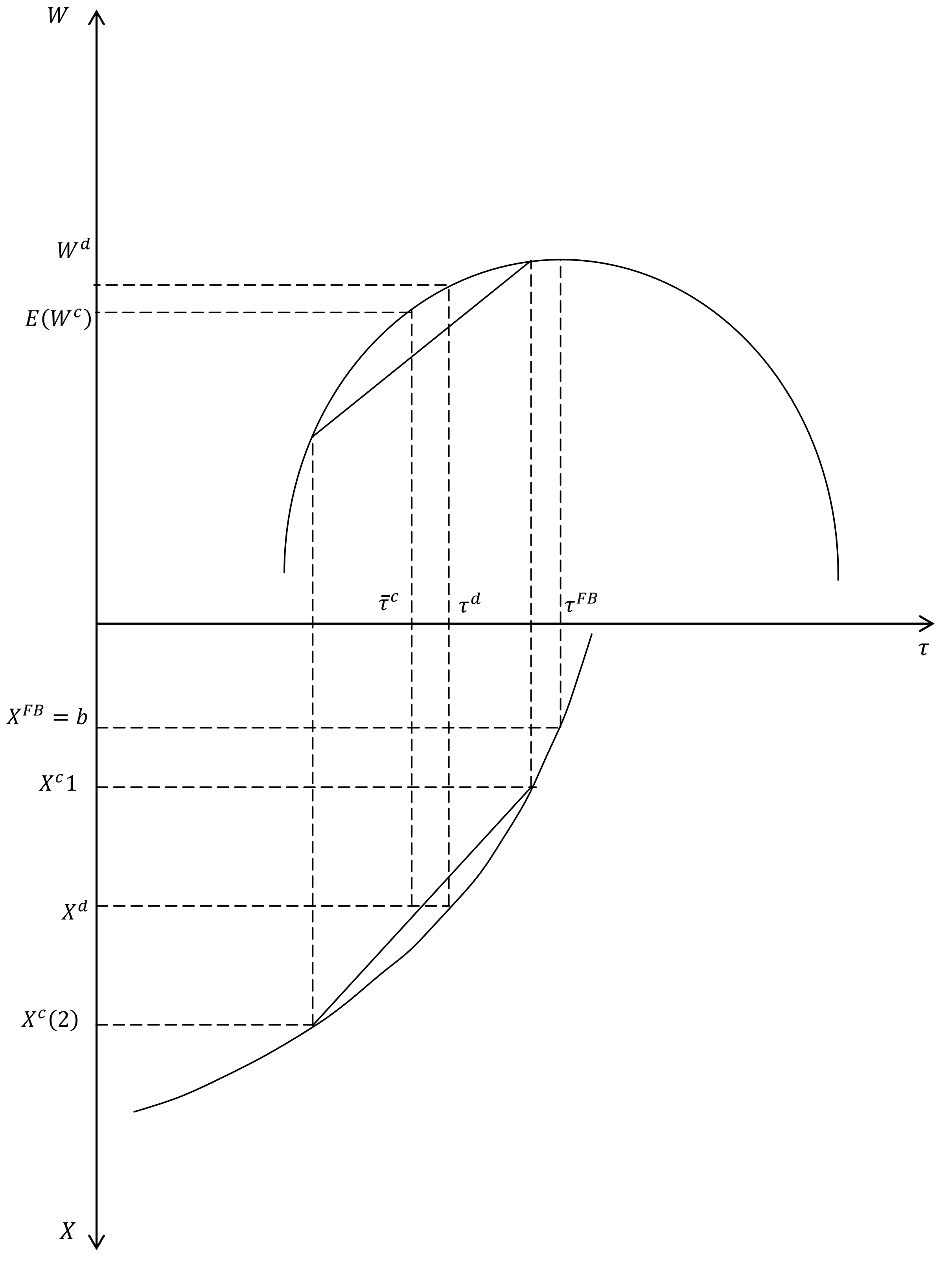


Figure A1. Centralized versus decentralized welfare: the zero spill-over case

The general case

If there are spill-overs, decentralization is not necessarily better than centralization, and few general results can be shown.

First, we can show that $\theta_i \geq \eta_i \quad (i=1,2)$ is a sufficient condition for decentralization to yield higher welfare. To see this, note from (A1.4) and (A1.8) that we now have

$$\bar{X}_1^c = 0.5\left[X_1^c(1) + X_1^c(2)\right] = 0.5\left[2b\frac{\theta_1}{\eta_1} + 2b\frac{(1-\theta_1)}{\eta_2}\right] > b\frac{\theta_1}{\eta_1} = X_1^d$$

Moreover, as long as $\theta_i \geq \eta_i \quad (i=1,2)$, the first-best toll exceeds the toll under decentralization. Given the properties of the toll as a function of X it then again unambiguously follows

$$\bar{\tau}_i^c < \tau_i^d \leq \tau_i^{FB} \quad (i=1,2)$$

The concavity of the welfare function in tolls then immediately yields that decentralized decisions outperform (in expected terms) centralized decisions:

$$W_i^c < W_i^d \leq W_i^{FB} \; (i=1,2)$$

The weak inequality holds with equality if $\theta_i = \eta_i \quad (i=1,2)$.

Second, we know that centralized decisions are first-best when, in both regions, all voters are users $\eta_i = 1$ and spill-over parameters are given by $\theta_i = 0.5$. Not surprisingly, we will see in the numerical illustration that centralization performs better than decentralized decisions if voter majorities are large and spill-overs are 'in the neighborhood' of 0.5.

<u>*Case 2: users have a majority in one region only*</u>

Suppose users have a majority in region 1, but not in region 2. Then we have

$$0.5 < \eta_1 \leq 1; \; \eta_2 < 0.5$$

Welfare under decentralization in region 1 is, as before, given by (A1.6). In region 2, however, the user price is revenue maximizing:

$$\tau_2^d = \frac{\beta V_2}{K} - \frac{V_2}{\frac{\partial V_2}{\partial \tau}}$$

Or, given linear demand

$$\tau_2^d = \frac{\beta V_2}{K} + bV_2$$

The traffic volume is

$$V_2^d = \frac{a-\alpha}{2A} \tag{A1.14}$$

where A is defined as before. Welfare is found to be

$$W_2^d = \left\{\frac{a-\alpha}{2A}\right\}^2 \left[A + \frac{1}{2}b\right] \tag{A1.15}$$

Finally, the toll as function of parameters only is

$$\tau_2^d = \left[\frac{a-\alpha}{2A}\right](A) \tag{A1.16}$$

Under centralized decisions we find the same results as before when the representative from region 1 is decisive; hence, (A1.13) holds:

$$\overline{\tau}_1^c < \tau_1^d \leq \tau_1^{FB}$$

However, when the non-driver from region 2 is decisive, she will charge revenue maximizing tolls everywhere. This implies

$$\tau_2^d = \tau_1^c(2) = \tau_2^c(2)$$

Using these observations, straightforward algebra shows that the two major results derived before remain valid when drivers have a majority in one region only. First, when there are no spill-overs, decentralization yields higher welfare than centralized decisions. Second,

$$\theta_1 > \eta_1$$

is sufficient for decentralization to perform better than centralization.

Case 3: non-drivers have a majority on both regions

Finally, if non-users have the majority in both regions, decentralized and centralized decisions make no difference. In all cases, we have revenue maximizing tolls in both regions.

Appendix 2. Centralized decisions with a uniform user price restriction

Suppose in region 1 users have a majority. If in power at the central level, they determine the uniform user price so as to

$$\underset{\tau}{\text{Max}} \ \frac{\theta_1}{D_1}\{\int_0^{V_1} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\tau(V_1+V_2) - \rho K_1^0 - \rho K_2^0}{2R}$$

Using the equality of generalized price and generalized cost, the first-order condition can be written as

$$-\frac{\theta_1}{D_1}V_1\frac{dg_1}{d\tau}-\frac{(1-\theta_2)}{D_1}V_2\frac{dg_2}{d\tau}+\frac{1}{2R}\left[\tau\left(\frac{dV_1}{d\tau}+\frac{dV_2}{d\tau}\right)+(V_1+V_2)\right]=0 \qquad \text{(A2.1)}$$

It is straightforward to show that

$$\frac{dg_i}{d\tau}=1+\frac{\beta}{K}\frac{dV_i}{d\tau}$$

$$\frac{dV_i}{d\tau}=\frac{\frac{\partial V_i}{\partial\tau}}{1-\frac{\beta}{K}\frac{\partial V_i}{\partial\tau}}$$

Substituting these expressions in the first-order condition (A2.1), multiplying by (*2R*) and rearranging, we solve for the uniform user price. We find

$$\tau=\frac{V_1\left[\frac{2\theta_1}{\eta_1}-1+\frac{\beta}{K}\frac{\partial V_1}{\partial\tau}\right]\left(1-\frac{\beta}{K}\frac{\partial V_2}{\partial\tau}\right)+V_2\left[\frac{2(1-\theta_2)}{\eta_1}-1+\frac{\beta}{K}\frac{\partial V_2}{\partial\tau}\right]\left(1-\frac{\beta}{K}\frac{\partial V_1}{\partial\tau}\right)}{\frac{\partial V_1}{\partial\tau}\left(1-\frac{\beta}{K}\frac{\partial V_2}{\partial\tau}\right)+\frac{\partial V_2}{\partial\tau}\left(1-\frac{\beta}{K}\frac{\partial V_1}{\partial\tau}\right)} \qquad \text{(A2.2)}$$

Now note that, for a given representative being decisive, the volumes will be equal, as the tolls are uniform and demand parameters are the same in both regions by assumption. A similar expression holds when the representative from the other region is decisive.

Using these insights, we can rewrite the toll rules for both regions, after simple algebra, as

$$\tau^u(1)=\frac{\beta V^u(1)}{K}+\left[\frac{(1+\theta_1-\theta_2)}{\eta_1}-1\right]\frac{V^u(1)}{\frac{\partial V^u(1)}{\partial\tau}}$$

$$\tau^u(2)=\frac{\beta V^u(2)}{K}+\left[\frac{(1-\theta_1+\theta_2)}{\eta_1}-1\right]\frac{V^u(2)}{\frac{\partial V^u(2)}{\partial\tau}}$$

(A2.3)

Here the notation $\tau^u(i), V^u(i)$ refers, respectively, to the uniform user price the representative from region *i* wants to charge in both regions, and the resulting volumes.

Assuming linear demand, and equating the generalized price and generalized cost in both regions, we can solve for the volumes and, hence, for the uniform toll. We find after straightforward algebra

$$V^u(1)=\frac{a-\alpha}{2A-X^u(1)};V^u(2)=\frac{a-\alpha}{2A-X^u(2)} \qquad \text{(A2.4)}$$

Here, *A* is as defined above, and

$$X^u(1) = \left(\frac{1+\theta_1-\theta_2}{\eta_1}\right)b;\ X^u(2) = \left(\frac{1-\theta_1+\theta_2}{\eta_2}\right)b \tag{A2.5}$$

The toll as function of the parameters only can be obtained by substituting (A2.4) in (A2.3), and using the linear specification of the demand function (1). We find

$$\begin{aligned} \tau^u(1) &= \left[\frac{a-\alpha}{2A-X^u(1)}\right]\left[A-X^u(1)\right] \\ \tau^u(2) &= \left[\frac{a-\alpha}{2A-X^u(2)}\right]\left[A-X^u(2)\right] \end{aligned} \tag{A2.6}$$

Finally, welfare can be found as (based on the general welfare expression (A1.1)):

$$\begin{aligned} W^u(1) &= \left[\frac{a-\alpha}{2A-X^u(1)}\right]^2\left[A+\frac{1}{2}b-X^u(1)\right] \\ W^u(2) &= \left[\frac{a-\alpha}{2A-X^u(2)}\right]^2\left[A+\frac{1}{2}b-X^u(2)\right] \end{aligned} \tag{A2.7}$$

Given these results, we want to know whether

$$W^u(1)+W^u(2) < or > 0.5\left[W_1^c(1)+W_2^c(1)\right]+0.5\left[W_1^c(2)+W_2^c(2)\right]$$

To analyze this question, note from (A1.8) that

$$\bar{X}^c(1) = 0.5\left[X_1^c(1)+X_2^c(1)\right] = \frac{b}{\eta_1}\left(1+\theta_1-\theta_2\right)$$

Comparison with (A2.5) gives the result

$$X^u(1) = \bar{X}^c(1)$$

An analogous argument yields

$$X^u(2) = \bar{X}^c(2)$$

Although the result we are interested in holds in general, it will be helpful to first show it under the assumption of symmetry. Then we have

$$X^u(1) = \bar{X}^c(1) = X^u(2) = \bar{X}^c(2) = \frac{b}{\eta} \geq X^{FB} = b \tag{A2.8}$$

Using the same argument as in Appendix 1 it then follows immediately that

$$\bar{\tau}^c \leq \tau^u < \tau^{FB}$$

and, because of concavity of welfare in tolls

$$E(W^c) < W^u < W^{FB}$$

In the more general case without symmetry, (A2.8) no longer necessarily holds. For example, it may be the case that for one region

$$X^{u}(i) < X^{FB} = b$$

However, relation (A2.8) still holds on average. To see this, note:

$$0.5\left[X^{u}(1)+X^{u}(2)\right] = 0.5\left[\bar{X}^{c}(1)+\bar{X}^{c}(2)\right] = 0.5\left[\frac{(1+\theta_1-\theta_2)}{\eta_1}+\frac{(1-\theta_1+\theta_2)}{\eta_2}\right]b$$

It is easy to show that, given the restrictions on the parameters

$$0 \le \theta_i \le 1;\, 0.5 \le \eta_i \le 1$$

we necessarily have

$$0.5\left[\frac{(1+\theta_1-\theta_2)}{\eta_1}+\frac{(1-\theta_1+\theta_2)}{\eta_2}\right]b \ge X^{FB} = b$$

A simple graphical analysis, similar to Figure A1, can be used to show that this condition is sufficient for uniform welfare to never be below expected welfare under a standard minimum winning coalition.

Finally, an analogous exercise shows that all the above results remain valid when drivers have a majority in one region only. The only difference is that in that case the non driver wants a revenue maximizing toll, but this holds both under decentralized and uniform decision making. Hence, results are the same.

To conclude this appendix, observe that the condition

$$\theta_i \ge \eta_i$$

is again sufficient for welfare under decentralized decisions to exceed welfare under centralization with a uniformity constraint. To see this, $\theta_i \ge \eta_i$ immediately implies

$$\tau^{u}(1) < \tau_1^{d} < \tau_1^{FB}; \quad \tau^{u}(2) < \tau_2^{d} < \tau_2^{FB}$$

Given that the welfare function is concave in tolls this shows

$$W^{u} < W^{d}$$

If, however,

$$\theta_i < \eta_i$$

then we have $\tau^{u}(1) < \tau_1^{FB} < \tau_1^{d}; \quad \tau^{u}(2) < \tau_2^{FB} < \tau_2^{d}$. It is then easily seen that for sufficiently large spillovers and a sufficiently large driver majority uniformity may well be better than decentralization.

Appendix 3: centralized decisions by legislative bargaining

First, assume drivers have a majority in both regions. Then the objective function under legislative bargaining can be written as -- assuming equal bargaining power of elected representatives -- as the simple sum of their individual objectives. Outcomes therefore solve the following problem

$$\underset{\tau_1,\tau_2}{\text{Max}} \ \frac{\theta_1}{D_1}\{\int_0^{V_1} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\tau_1 V_1 + \tau_2 V_2}{2R}$$
$$+\frac{\theta_2}{D_2}\{\int_0^{V_2} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{1-\theta_1}{D_2}\{\int_0^{V_1} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{\tau_1 V_1 + \tau_2 V_2}{2R}$$

Following the same logic as before, we easily derive the tax rules

$$\tau_1^b = \frac{\beta V_1^b}{K_1^0} + \{\frac{\theta_1}{\eta_1} + \frac{(1-\theta_1)}{\eta_2} - 1\}\frac{V_1^b}{\frac{\partial V_1}{\partial \tau_1}} \tag{A3.1}$$

$$\tau_2^b = \frac{\beta V_2^b}{K_2^0} + \{\frac{\theta_2}{\eta_2} + \frac{(1-\theta_2)}{\eta_1} - 1\}\frac{V_2^b}{\frac{\partial V_1}{\partial \tau_1}} \tag{A3.2}$$

Here we have denoted the tolls and volumes under bargaining in the two regions as

τ_i^b, V_i^b,

respectively. Volumes, welfare and tolls can be derived easily, using analogous methods as before. We find:

$$V_1^b = \frac{a-\alpha}{2A - X_1^b}; \quad V_2^b = \frac{a-\alpha}{2A - X_2^b} \tag{A3.3}$$

$$W_1^b = \left[\frac{a-\alpha}{2A - X_1^b}\right]^2 \left[A + \frac{1}{2}b - X_1^b\right]; \quad W_2^b = \left[\frac{a-\alpha}{2A - X_2^b}\right]^2 \left[A + \frac{1}{2}b - X_2^b\right] \tag{A3.4}$$

$$\tau_1^b = \left[\frac{a-\alpha}{2A - X_1^b}\right]\left[A - X_1^b\right]; \qquad \tau_2^b = \left[\frac{a-\alpha}{2A - X_2^b}\right]\left[A - X_2^b\right] \tag{A3.5}$$

In these expressions ,

$$X_1^b = \left(\frac{\theta_1}{\eta_1} + \frac{1-\theta_1}{\eta_2}\right)b$$
$$X_2^b = \left(\frac{\theta_2}{\eta_2} + \frac{1-\theta_2}{\eta_1}\right)b \tag{A3.6}$$

We want to compare the welfare performance of bargaining and decisions by a minimum winning coalition under a toll uniformity restriction. Hence, we want to find out whether

$$\left[W^u(1)+W^u(2)\right] < or > W_1^b + W_2^b$$

First, consider the symmetric case for simplicity. We then have, see (A2.5) and (A2.6)

$$X^u(1) = X^u(2) = X_1^b = X_2^b = \frac{1}{\eta} b$$

Hence,

$$\tau^u(1) = \tau^u(2) = \tau_1^b = \tau_2^b$$

It follows

$$W^u(1) = W^u(2) = W_1^b = W_2^b \quad \rightarrow \quad W^u(1) + W^u(2) = W_1^b + W_2^b$$

We conclude that symmetry implies that uniform pricing and bargaining lead to the same welfare. Moreover, welfare is the same in both regions.

Next turn to asymmetric regions. Note that there are two other sets of conditions such that bargaining and uniformity produce the same overall federal welfare, although regional welfare will differ across regions. If

$$\frac{\theta_1}{\eta_1} = \frac{\theta_2}{\eta_2}$$

then (A2.5)-(A2.6) and (A3.5)-(A3.6) imply

$$\tau^u(1) = \tau_2^b$$
$$\tau^u(2) = \tau_1^b$$

Moreover, if

$$\frac{1-\theta_1}{\eta_2} = \frac{1-\theta_2}{\eta_1}$$

the same expressions yield

$$\tau^u(1) = \tau_1^b$$
$$\tau^u(2) = \tau_1^b$$

In both cases, we have

$$W^u(1) + W^u(2) = W_1^b + W_2^b$$

To study the relative welfare performance of the two systems under asymmetry, note that we have, again using (A2.6) and (A3.6):

$$X^u(1) + X^u(2) = X_1^b + X_2^b \qquad \text{(A3.7)}$$

We can now show that the following two sets of joint conditions are sufficient for bargaining to yield higher welfare than imposing a uniformity constraint on centralized MWC-decision making:

$$\begin{aligned} &\frac{\theta_1}{\eta_1} > \frac{\theta_2}{\eta_2} \quad \textit{and} \quad \frac{1-\theta_1}{\eta_2} < \frac{1-\theta_2}{\eta_1} \\ &\frac{\theta_1}{\eta_1} < \frac{\theta_2}{\eta_2} \quad \textit{and} \quad \frac{1-\theta_1}{\eta_2} > \frac{1-\theta_2}{\eta_1} \end{aligned} \tag{A3.8}$$

To see this, take the first set of inequality conditions as an example. They imply (using (A2.5) and (A3.6)):

$$\begin{aligned} &X_1^b > X^u(2) \\ &X^u(1) > X_2^b \\ &X_2^b > X^u(2) \end{aligned}$$

Together with (A3.7) this implies that the bargained tolls are 'in between' the two bargained tolls. Given concavity on the welfare function this implies

$$W^u(1) + W^u(2) < W_1^b + W_2^b \,.$$

It will be instructive to illustrate an example of this case graphically, see Figure A2. The relative position of the X's on the lower panel (together with constraint (A3.7)) produces the relative toll levels on the horizontal axis of the upper panel. It then immediately follows from the concavity of the welfare function that bargaining yields higher total welfare than a minimum winning coalition under uniformity restrictions.

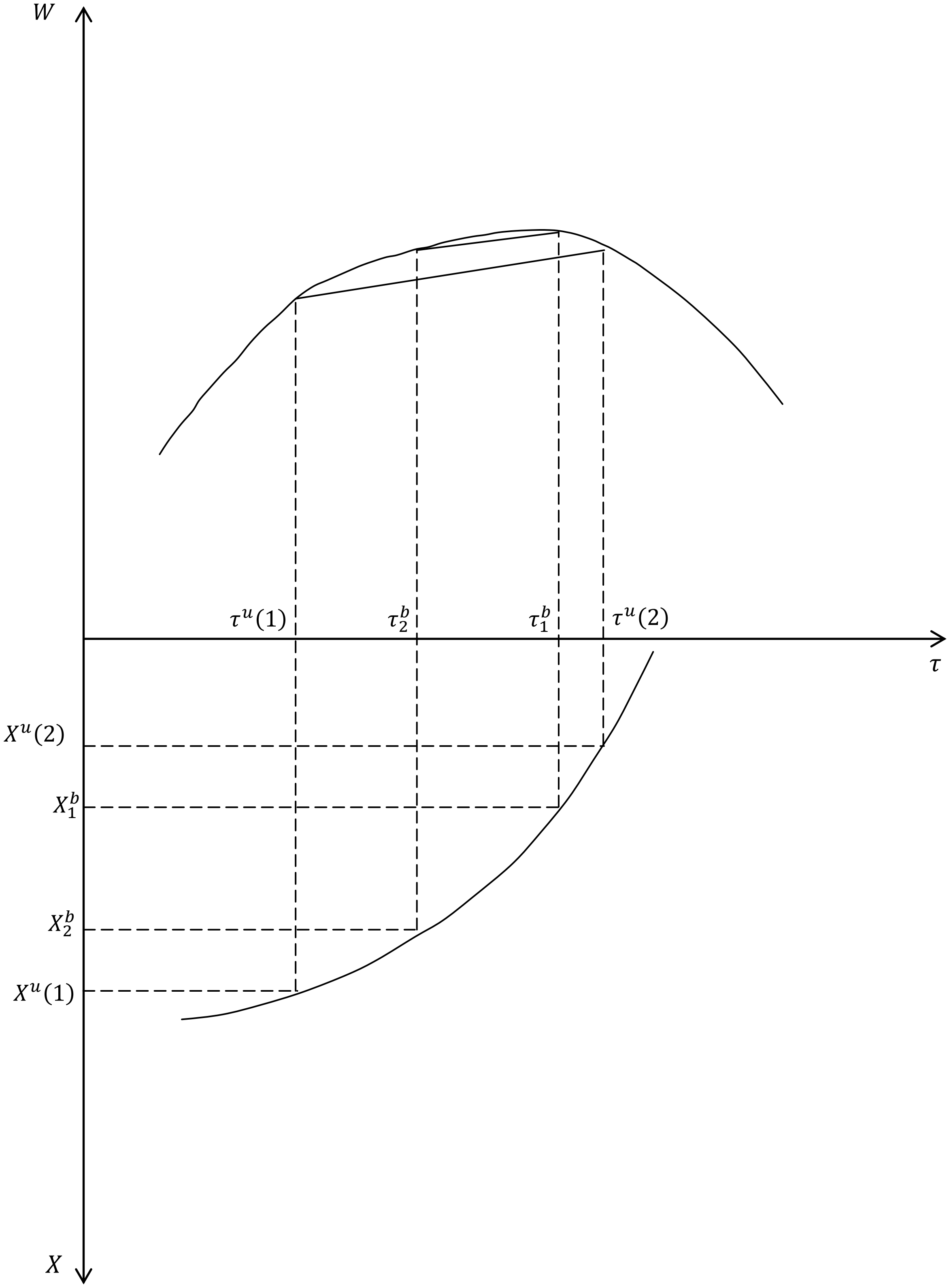


Figure A2: Bargaining versus a minimum winning coalition under uniformity

A similar result holds for the second set of inequality restrictions. If, however, one of the following sets of conditions hold

$$\frac{\theta_1}{\eta_1} < \frac{\theta_2}{\eta_2} \quad and \quad \frac{1-\theta_1}{\eta_2} < \frac{1-\theta_2}{\eta_1}$$
$$\frac{\theta_1}{\eta_1} > \frac{\theta_2}{\eta_2} \quad and \quad \frac{1-\theta_1}{\eta_2} > \frac{1-\theta_2}{\eta_1} \tag{A3.9}$$

then uniformity is necessarily better than bargaining. The first set of conditions imply

$$X_1^b < X^u(1)$$
$$X^u(2) < X_2^b$$
$$X_1^b < X_2^b$$

Together with (A3.7) we now have that the two uniform tolls will now in between the negotiated tolls. We find

$$W^u(1) + W^u(2) > W_1^b + W_2^b$$

Uniformity is better than bargaining. A similar story applies to the second set of inequalities.

Loosely speaking, bargaining will certainly be better if the driver majorities are close to being equal; in that case, (A3.8) automatically holds so that bargaining is better. If there are large differences in user majorities and differences in spill-overs are of the opposite sign, then uniform prices may be better. Numerical analysis suggests, see the main body of the paper, that for most plausible parameter configurations, bargaining is better.

Second, if there is bargaining between one region where drivers have a majority (say, region 1) and a region where a non-driver is elected as representative (say, region 2) then the objective function is

$$\underset{\tau_1,\tau_2}{\text{Max}} \ \frac{\theta_1}{D_1}\{\int_0^{V_1} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\tau_1 V_1 + \tau_2 V_2}{2R}$$
$$+ \frac{\tau_1 V_1 + \tau_2 V_2}{2R}$$

We find the tax rules

$$\tau_1 = \frac{\beta V_1}{K_1^0} + \{\frac{\theta_1}{\eta_1} - 1\}\frac{V_1}{\frac{\partial V_1}{\partial \tau_1}}$$

$$\tau_2 = \frac{\beta V_2}{K_2^0} + \{\frac{1-\theta_2}{\eta_2} - 1\}\frac{V_2}{\frac{\partial V_2}{\partial \tau_2}}$$

where it should be noted that now $\eta_2 < 0.5$.

Under standard centralized decisions, the driver from region 1 wants the very low toll (if spill overs are limited) in his own region and a high toll in the other region, but the non-driver of region 2 wants the revenue maximizing toll everywhere. Bargaining leads to a mixture of these wishes. The outcome depends. For example, if there are no spillovers, the outcome is a high toll in region 2, because both representatives now want the revenue maximizing toll for this region. However, if there are large spillovers, the elected representative from region 1 wants a low toll in region 2, whereas the person from region 2 still wants a high toll. The outcome then depends on the relative strength of these two tendencies. Small spill-overs imply that decentralization outperforms centralized decisions, but for high spill-overs the opposite may hold.

Third, bargaining between <u>two regions where non-drivers have a majority</u> gives revenue maximizing charges everywhere, as it does under a uniformity restriction.

Appendix 4: Will regions agree to transfer decision power to the federal level?

We first look at the case where drivers have a majority in both regions, then we analyze the case where in one region non-users have a majority. The case where non-drivers have a majority in both regions is not treated because it is trivial.

<u>*Drivers have a majority in both regions*</u>

Let us define the welfare of a driver of region 1 when decentral political decisions are made. Given spill-overs, the driver enjoys a benefit of driving in region 2 as well as in his own region 1; moreover, he shares in the toll revenues collected in region 1, but gets nothing from the revenues in region 2. His total welfare is therefore

$$W^d(\textit{driver region}\,1) =$$
$$\frac{\theta_1}{D_1}\{\int_0^{V_1^d} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2^d} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\tau_1^d V_1 - \rho K_1^0}{R} \quad \text{(A4.1)}$$

In this expression the toll is in both regions the decentralized toll; the volumes are those consistent with these tolls. Following the same methods that have been used several times before, we can rewrite (A4.1) as

$$W^d(\textit{driver region}\,1) = \frac{\theta_1}{D_1}\frac{b}{2}\left(V_1^d\right)^2 + \frac{1-\theta_2}{D_1}\frac{b}{2}\left(V_2^d\right)^2 + \frac{\tau_1^d V_1^d}{R}$$

Assuming symmetry for simplicity ($\theta_1 = \theta_2 = \theta; D_1 = D_2 = D; \tau_1^d = \tau_2^d = \tau^d; V_1^d = V_2^d = V^d$), we obtain

$$W^d(driver\ region\,1) = \frac{1}{D}\frac{b}{2}\left(V^d\right)^2 + \frac{\tau^d V^d}{R} \qquad \text{(A4.2)}$$

Multiplying both sides by R and using (A1.3-A1.4-A1.5), this can easily be shown to imply the following

$$R*W^d(driver\ region\,1) = \left(\frac{a-\alpha}{2A-\frac{b\theta}{\eta}}\right)^2\left[A+\frac{b}{\eta}\left(\frac{1}{2}-\theta\right)\right] \qquad \text{(A4.3)}$$

Now turn to the total (expected) welfare of this same driver when uniform pricing decisions are taken at the central level. He benefits from driving in regions 1 as well as 2. Moreover, he now receives his share of the joint toll revenues in both regions. However, the uniform toll levels will depend on who is in charge at the central level. If he is in charge then his welfare is

$$\frac{\theta_1}{D_1}\{\int_0^{V_1^u(1)} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2^u(1)} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\left(\tau^u(1)\right)\left(V_1^u(1)+V_2^u(1)\right)}{2R}$$

If his driving colleague from region 2 is decisive at the central level his welfare is

$$\frac{\theta_1}{D_1}\{\int_0^{V_1^u(2)} P(V_1)dV_1 - V_1.g(V_1)\} + \frac{1-\theta_2}{D_1}\{\int_0^{V_2^u(2)} P(V_2)dV_2 - V_2.g(V_2)\} + \frac{\left(\tau^u(2)\right)\left(V_1^u(2)+V_2^u(2)\right)}{2R}$$

Uniformity implies $V_1^u(1) = V_2^u(1) = V^u(1); V_1^u(2) = V_2^u(2) = V^u(2)$. Moreover, next to the equalities given above, symmetry further implies $\tau^u(1) = \tau^u(2) = \tau^u; V^u(1) = V^u(2) = V^u$. Using these simplifications, and noting that there is a 50% probability that the driver from each region is decisive centrally, we employ straightforward algebra to show that expected welfare of the driver in region 1 can be written simply as follows:

$$E\left(W^u(driver\ region\,1)\right) = \frac{1}{D}\frac{b}{2}\left(V^u\right)^2 + \frac{\tau^u V^u}{R}$$

Finally, multiplying both sides by R, using (A2.4-A2.4-A2.6) and some straightforward algebra, we find

$$R*E\left(W^u(driver\ region\,1)\right)=\left(\frac{a-\alpha}{2A-\frac{b}{\eta}}\right)^2\left[A-\frac{1}{2}\frac{b}{\eta}\right] \tag{A4.4}$$

Comparison of (A4.3) and (A4.4) leads to some remarkable insights. First, if there are no spill-overs, welfare of a driver is identical under decentralized decisions and uniform central decisions. Second, when spill-overs do exist it easily follows that he will always prefer the centralized uniform outcome. To see this, note that the right hand side of (A4.3) and (A4.4) has the same structure. Differentiating the right hand side of (A4.3) with respect to the spill-over parameter we find

$$\frac{\partial\left[R*W^d(driver\ region\,1)\right]}{\partial\theta}=\left(\frac{a-\alpha}{2A-\frac{b\theta}{\eta}}\right)^2\left(\frac{2b}{\eta}\right)\left[A-\frac{\theta b}{\eta}\right]>0$$

This derivative is positive: the first two terms are positive, and the last term is also positive (assuming positive decentralized tolls, see (A1.4)-(A1.5)). As driver welfare is the same under decentralized and uniform central tolls when there are zero spill-overs and noting that an increase in spill-overs (a reduction in θ) reduces his welfare under decentralization, he will prefer uniform tolls.

The implication is powerful. It means that a driver will be willing to transfer decision making power to the central level if tolls are uniform. Since by assumption drivers have a majority in both regions, one expects regions to agree on central decision making subject to a uniformity constraint.

Drivers have a majority in only one region

The fact that a non-driver is in charge in one region and may become decisive centrally changes matters substantially. To fix ideas, let in region 1 drivers have a majority but in region 2 the median voter is a non-driver. We focus on the total welfare of this non-user from region 2 to make our most important point.

When decisions are made in a decentral way, he gets no benefit at all from the decisions in region 1: he does not drive and does not share in the toll revenues there. In region 2 his benefit is given by his share of the collected toll revenues. We can write his total welfare simply as

$$W^d(non-driver\ region\ 2)=\frac{\tau_2^d V_2^d}{R}$$

where the toll in region 2 is the revenue maximizing toll; the volume is the volume at this toll level. Multiplying by *R* and using (A1.14)-(A1.15) for toll and volume, we have

$$R*W^{d}(non-driver\ region\ 2)=\frac{(a-\alpha)^{2}}{4A} \tag{A4.5}$$

Now let decisions be taken at the central level subject to a uniformity constraint. If the driver from region 1 becomes decisive at the central level, the non-driver from region 2 gets his revenue share equal to

$$\frac{\tau^{u}(1)\left[2V^{u}(1)\right]}{2R}=\frac{\tau^{u}(1)\left[V^{u}(1)\right]}{R}$$

Here the toll is the one set by the representative from region 1 (who is a driver). Using (A2.4-A2.5-A2.6) this can be written as

$$\frac{1}{R}\left(\frac{a-\alpha}{2A-X^{u}(1)}\right)^{2}\left(A-X^{u}(1)\right)$$

where $X^{u}(1)=\frac{1+\theta_{1}-\theta_{2}}{\eta_{1}}b$. Similarly, if the representative from region 2 is himself decisive at the federal level, his benefit is

$$\frac{\tau^{u}(2)\left[2V^{u}(2)\right]}{2R}=\frac{\tau^{u}(2)\left[V^{u}(2)\right]}{R}=\frac{1}{R}\frac{(a-\alpha)^{2}}{4A}$$

The last equality follows from noting that he charges the revenue maximizing toll in both regions . Multiplying by *R* and noting that both regions have an equal probability of being decisive at the central level, we obtain

$$R*W^{u}(non-driver\ region\ 2)=\frac{1}{2}\left(\frac{a-\alpha}{2A-X^{u}(1)}\right)^{2}\left(A-X^{u}(1)\right)+\frac{1}{2}\frac{(a-\alpha)^{2}}{4A} \tag{A4.6}$$

Finally, compare (A4.5) and (A4.6). Both expressions are equal for $X^{u}(1)=0$. Noting that

$$\frac{\partial\left[\left(\frac{a-\alpha}{2A-X^{u}(1)}\right)^{2}\left(A-X^{u}(1)\right)\right]}{\partial\left(X^{u}(1)\right)}=\frac{(a-\alpha)^{2}}{\left[2A-X^{u}(1)\right]^{3}}\left(-X^{u}(1)\right)<0$$

for any positive *X*, the right hand side of (A4.6) is smaller than that of (A4.5). Hence, the non-driver of region 2 is always better off under decentralization.

Again, the implication is powerful. A non-driver will never be willing to transfer decision power to the central level.

Appendix 5. Pricing and capacity decisions

First we show the derivation of the first best. Suppose the region solves

$$\underset{\tau,\mathrm{K}}{\mathrm{Max}}\ \{\int_0^V P(V)dV - V.g(V)\} + \tau V - \rho K$$

Differentiation of the objective function with respect to the toll and using the equality between generalized price and generalized cost, we have

$$-V\frac{dg}{d\tau} + \tau\frac{dV}{d\tau} + V = 0 \qquad \text{(A5.1)}$$

Noting that $V(P),\, P = g(V) = \alpha + \frac{\beta}{K}V + \tau$, we easily derive

$$\frac{dV}{d\tau} = \frac{\frac{\partial V}{\partial \tau}}{1 - \frac{\beta}{K}\frac{\partial V}{\partial \tau}}; \quad \frac{dg}{d\tau} = \frac{1}{1 - \frac{\beta}{K}\frac{\partial V}{\partial \tau}} \qquad \text{(A5.2)}$$

Substitution of this expression in (A5.1) produces

$$\tau = \frac{\beta V}{K} \qquad \text{(A5.3)}$$

Differentiation of the objective function with respect to capacity (again using equality between generalized price and generalized cost) gives

$$-V\frac{dg}{dK} + \tau\frac{dV}{dK} - \rho = 0 \qquad \text{(A5.4)}$$

We further derive

$$\frac{dV}{dK} = \frac{-\frac{\beta V}{K^2}\frac{\partial V}{\partial \tau}}{1 - \frac{\beta}{K}\frac{\partial V}{\partial \tau}}; \quad \frac{dg}{dK} = \frac{-\frac{\beta V}{K^2}}{1 - \frac{\beta}{K}\frac{\partial V}{\partial \tau}} \qquad \text{(A5.5)}$$

Substituting (A5.5) in (A5.4), working out and using the toll rule (A5.3) gives immediately

$$\beta\left(\frac{V}{K}\right)^2 = \rho \qquad \text{(A5.6)}$$

Next, consider the problem of the regional representative that is a driver. She solves

$$\underset{\tau,\,K}{\text{Max}} \;\; \frac{\theta}{D}\{\int_0^V P(V)dV - V.g(V)\} + \frac{\tau V - \rho K}{R}$$

$$s.t. K \geq 0$$

The first-order conditions with respect to user price and capacity can be written

$$\frac{\theta}{D}\left(-V\frac{dg}{d\tau}\right) + \frac{1}{R}\left(\tau\frac{dV}{d\tau} + V\right) = 0 \tag{A5.7}$$

$$\frac{\theta}{D}\left(-V\frac{dg}{dK}\right) + \frac{1}{R}\left(\tau\frac{dV}{dK} - \rho\right) = 0 \tag{A5.8}$$

Using the expressions (A5.2) and (A5.5), working out, substituting the optimal user price expression in the first-order condition for optimal capacity, and noting the definition $\eta = \frac{D}{R}$, it follows

$$\tau = \frac{\beta V}{K} + \{\frac{\theta}{\eta} - 1\}\frac{V}{\frac{\partial V}{\partial \tau}}$$

$$\beta\left(\frac{V}{K}\right)^2 = \rho$$

The derivation of all other pricing and capacity reported in Section 5 of the paper follows the same procedure as explained here.