Int Tax Public Finance (2016) 23:934–959
DOI 10.1007/s10797-015-9388-6

ORIGINAL PAPER

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

Bruno De Borger[1] · **Stef Proost**[2]

Published online: 23 December 2015

Abstract This paper studies the political economy of pricing and investment for excludable and congestible public goods in a federal state. Although the model applies to many local congestible public facilities (such as libraries, museums and public swimming facilities), our main motivation is the problem of providing and pricing road infrastructure in federal states. The two-region model we develop allows for spill-overs between regions, it takes into account congestion, and it captures demand heterogeneity both between and within regions. Regional decisions are taken by majority voting; federal decisions are taken by a minimum winning coalition in a legislature of regionally elected representatives. 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 decision making and decentralized decision making yield the same socially undesirable outcome, with prices that are much too high. Third, the performance of decentralized supply is strongly enhanced by local self-financing rules; this prevents potential exploitation of users within regions.

Electronic supplementary material The online version of this article (doi:10.1007/s10797-015-9388-6) contains supplementary material, which is available to authorized users.

✉ Bruno De Borger
bruno.deborger@uantwerpen.be

Stef Proost
stef.proost@kuleuven.be

1 University of Antwerp, Antwerp, Belgium

2 KULeuven, Leuven, Belgium

Keywords Congestible local public goods · Pricing · Capacity decisions · Fiscal federalism

JEL Classification D62 · H23 · R41 · R48

1 Introduction

In this paper, we develop a political economy model to compare federal and local decision making on user prices and capacity investment for congestible local public goods in federal states.[1] The model allows for interregional spill-overs, it takes into account congestion, and it captures user heterogeneity both between and within regions. Although the framework is more generally applicable to local congestible public facilities such as libraries, public transport, museums and public swimming facilities, one of our main motivations is the problem of providing and pricing road infrastructure in federal states.[2]

In most countries, decision-making responsibilities for the provision and financing of urban and regional roads are divided between different levels of government. Although practices widely differ across countries, they 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 the externalities associated with the use of transport services. This is remarkable, because congestion, accidents and some types of pollution are typically highly localized problems. Instead, many countries rely on pricing instruments that are 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 USA, however, fuel taxes do differ between states, although this differentiation does not necessarily match local conditions. Second, another common characteristic of transport policy making is the 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 USA (see, for example, Levinson 2002) or the Trans-European networks in the EU (see Proost et al. 2014), it also applies to urban or regional projects with rather localized benefits (Knight 2004).

[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] Although our focus will be on pricing and capacity of transport infrastructure decisions, it is clear that the problem of dividing responsibilities over different levels of government also exists for many other public services. Convincing evidence is provided in a recent study by the OECD (see Blöchliger 2008); he analyzes how decision authority is divided across levels of government for a number of public services (including education, hospitals, public transport and nursing homes) in different countries. 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.

These observations raise a number of questions, related to the division of authority between different government levels and to the choice of particular policy instruments. For example, under what conditions are the outcomes of decentralized political decision making socially preferable to those of centralization? Do these conditions provide a justification for the observation that some pricing and capacity decisions for road infrastructure are taken at the central level, whereas others are largely decentralized? 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 in cities, 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 demand 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 by a minimum winning coalition in a legislature of regionally elected representatives.

A brief description of the main results follows. 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, unlike in the paper of Besley and Coate (2003), symmetric regions and zero spill-overs are neither necessary nor sufficient to produce first-best outcomes. On the one hand, decentralization implies a welfare loss even without spill-overs, unless all voters in both regions are users. On the other hand, decentralization can be first-best even in the presence of spill-overs: heterogeneity within regions can compensate for the existence of spill-overs. Third, 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. Fourth, we find that imposing self-financing for transport infrastructure on the individual regions strongly enhances the performance of centralized decision making. In fact, under a mild additional assumption, it allows to attain the first-best social optimum. As will become clear, one of the main advantages of a regional self-financing rule is that it protects users of the infrastructure from being exploited by non-users. Although one should be careful with 'explaining' the real world on the basis of very stylized theoretical models, the analysis of this paper contributes to understanding why some decisions are taken at the central level, whereas others are not, and why the use of user fees for road infrastructure is not widespread, whereas parking charges are.

The paper builds upon several strands of literature. First, it relates to the literature on the provision of local public goods. Oates' (1972) decentralization theorem argued that, when regions are heterogeneous, the advantages of decentralization (accounting for taste differences) have to be traded off against the disadvantages (regions ignore spill-overs). However, when applied to infrastructure decisions, the decentralization theorem did not receive much empirical support (see, for example, Hulten and Schwab 1997). Moreover, the 'second-generation' literature on fiscal federalism (e.g., Persson and Tabellini 2000; Lockwood 2002; 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 not only theoretically unattractive, it was found to be inconsistent with empirical evidence (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), Lorz and Willmann (2005) and Hatfield and Padro i Miquel (2012).

Second, our model is related to the literature on pricing and investment of transport services. One line of research has focused on distortionary taxes and subsidies in the transport sector, and another analyzed the role of congestion and other externalities for pricing, investment and cost recovery in a single-region setting[3]. In multi-regional models with spill-overs—in the sense of foreign users of domestic transport infrastructure—it has been shown that decentralized decisions may imply large welfare losses, depending on the pricing instruments used and the nature of the network structure between regions (De Borger et al. 2005, 2007; Ubbels and Verhoef 2008). Most recently, Brueckner (2015) studies a model of multi-jurisdictional monocentric cities that include both transport and land markets. He shows that decentralized capacity choices (made by individual zones within the city) generate the social optimum, despite the presence of spill-overs[4]. Finally, a number of papers explicitly dealing with the political economy of transport decision making have appeared. 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). An exception is Knight (2004). He uses a legislative bargaining framework to explain the allocation of highway funds in the USA, showing that elected representatives may use their political power at the federal level to favor their own region, and the empirical results support his prediction.

As argued above, our paper is inspired by the second-generation literature on fiscal federalism. It is the first to compare the properties of decentralized and centralized decision making for congestible local public goods that can be priced. Introducing

[3] See, among many others, Calthrop et al. (2010), Gutiérrez-i-Puigarnau and van Ommeren (2011), Kidokoro (2006), Small and Verhoef (2007).

[4] We return to this finding in Sect. 4 below.

pricing decisions implies that we focus on a different problem than the problem of the appropriate level of providing ‘pure’ local public goods, as studied by Lockwood (2002) and Besley and Coate (2003). User pricing regulates use of the congestible infrastructure, and it contributes to financing the capacity supplied. This leads to two central insights. First, in the absence of restrictions on federal decision making, decentralized decisions on pricing are often better in terms of expected welfare, even in the presence of spill-overs. Second, the superiority of decentralized decision making continues to hold when capacity decisions are added to the analysis. Imposing a formal budget restriction on regional authorities may even force decentralized decisions to become first-best.

Structure of the paper is the following. In the next section, we describe the model. In Sect. 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 Sect. 4, we extend the analysis to cover joint pricing and capacity decisions, and we ask whether self-financing rules improve the efficiency of centralized and decentralized decision making. Policy implications of our findings are summarized in Sect. 5. A final section concludes and suggests avenues for further research.

2 Model setting

In this section, we describe the model. We use a setting with two regions, indexed $i = 1, 2$.[5] 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 the infrastructure. Although not crucial for the qualitative results of the paper, this binary setup with users and non-users facilitates both the derivation of clear-cut results and the presentation of the results in a transparent way.[6] Throughout the paper, we will further interpret the ‘use of infrastructure’ as road use, so that non-users can be thought of as people that do not have access to a car or that, for whatever reason, prefer other means of transport. Users make two types of trips: trips in the home region and trips in the other region. To simplify the exposition without affecting the qualitative insights to be derived from the model, we assume that the demand for both types of trips is independent.[7]

[5] Adding more symmetric regions does not add new insights. We return to this issue in the conclusion.

[6] The model could be generalized by having two groups of users within each region, intensive and less intensive users. However, in terms of policy implications, we would obtain the same qualitative results, in the sense that the inefficiency becomes proportional to the difference in user intensity. A model with a continuum of users with different intensity could also be worked out, but then the conflict within a region would depend on the shape of the distribution. In both cases, the derivation of results becomes much more complex. The advantage of the binary case is that the results depend only on the intensity of use of one group.

[7] 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

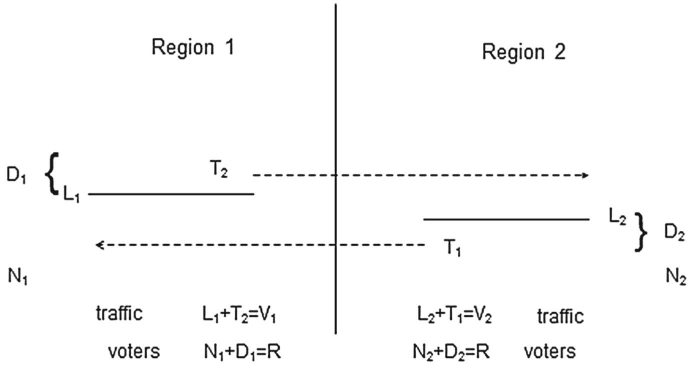

Fig. 1 Model setting

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.

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

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

Footnote 7 continued
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 et al. 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.

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, demand functions of local users and users that live in the other region are proportional. This allows 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 $\theta \quad (0 \leq \theta \leq 1)$ will play an important role in the analysis; it is the share of trips or kilometers that users from a given region make 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 is the sum of the money costs of car use and the time cost of the trip; it is assumed to be linearly rising in the volume–capacity ratio:

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

Adding 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 is responsible for the costs of road capacity. They balance their budget via head taxes and/or via user prices on road use. The rental cost of current capacity K^0 amounts to ρK^0.

3 Pricing the use of existing capacity

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, we assume simple majority voting. When preferences satisfy certain conditions, the median voter is decisive. For centralized decision making, our basic model setting follows Besley and Coate (2003). They suggest that a legislature of locally elected representatives makes the decisions by forming a minimum winning coalition. Although our assumptions on the political decision-making process are inspired by the dominant assumptions in the literature, they may have a bearing on the results. Do note, however, that the purpose of our model is to study how to improve the efficiency of political decisions on pricing and capacity provision for congestible local public goods. The inefficiencies come from two sources: non-users having to pay for the supply of capacity and regions having to pay for the supply in other regions. This dual heterogeneity (users versus non-users, two regions) is at the same time the minimum structure to generate these inefficiencies and the structure that generates the largest inefficiencies.[8]

Before moving to political economy issues, 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

$$\underset{\tau}{\text{Max}} \left\{ \int_0^V P(V)\mathrm{d}V - V \cdot g(V) \right\} + \tau V - \rho K^0$$

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 Sect. 4).

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:

$$\tau = \frac{\beta V}{K^0} \tag{4}$$

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 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

[8] One can think of several alternatives to the assumptions we make. First, assuming a minimum winning coalition mechanism at both the local and federal level would generate the same results. The reason is that, if within a region users have a majority, they would form a minimum winning coalition. Second, as long as the regional representative is selected by local majority voting, assuming majority voting at both the local and federal levels would also generate the same results. Of course, the political decision-making structure assumed does become more important for the results if one considers more complex preference heterogeneity. For example, consider a structure with highly intensive users, less intensive users and non-users; alternatively, one might specify a two-dimensional structure where decisions on two different local public goods are to be made. In those cases, the simple majority rule does no longer generate an equilibrium, and one has to resort to another type of model (e.g., probabilistic voting, models that explicitly study the coalition formation process). Deriving clear-cut results in these cases is much more difficult.

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 so that the decisive policy maker is a member of this group. We assume that he will choose the user price level that maximizes the following objective function:

$$\underset{\tau}{Max}\, \frac{\theta}{D}\left\{\int_0^V P(V)\mathrm{d}V - V\cdot g(V)\right\} + \frac{\tau V - \rho K^0}{R} \tag{5}$$

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. Formulation (5) implies that the policy maker, as representative of a large group of homogenous individuals, understands the problem like a social planner, except that he adapts the objective function to his own interests.

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

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

Straightforward algebra then produces the following user price rule

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

In this expression, the superscript ‘d’ stands for the ‘decentralized’ case, and 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, $0.5 < \eta \leq 1$.

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 price a user wants is determined by two forces, tax exporting behavior and unwillingness to share net revenues with non-users. 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). Moreover, note two other implications of (6). First, 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. Our modeling framework remains perfectly valid to analyze such subsidies, but they are not the focus of this paper. We assume throughout the paper that all user fees are nonnegative. 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 $0 \leq \eta < 0.5$? They will then opt for the revenue-maximizing user price: they do not pay but do share in the excess revenues. 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:

$$\tau^{\mathrm{d}} = \frac{\beta V}{K^0} - \frac{V}{\frac{\partial V}{\partial \tau}} \qquad (7)$$

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[9].

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. 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

$$\underset{\tau_1,\tau_2}{\text{Max}} \frac{\theta_1}{D_1}\left\{\int_0^{V_1} P(V_1)\mathrm{d}V_1 - V_1\cdot g(V_1)\right\} + \frac{1-\theta_2}{D_1}\left\{\int_0^{V_2} P(V_2)\mathrm{d}V_2 - V_2\cdot g(V_2)\right\} + \frac{\tau_1 V_1 - \rho K_1^0 + \tau_2 V_2 - \rho K_2^0}{2R}$$

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, and the second term is his net 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 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:

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

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

The notation $\tau_i^{\mathrm{c}}(j)$ 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 use the

[9] Alternatively, one could assume that each region delegates both users and non-users to the federal parliament (for example, proportional representation), and let one representative from the federal parliament randomly be elected as agenda setter. He/she then forms a minimum winning coalition with other members of the legislature. This approach would have required to analyze many more possible coalition formations. We therefore followed Besley and Coate (2003) and opted for a setup in which each region delegates one representative (elected by majority voting) to the federal parliament. Also note that we implicitly assume the central decision maker has perfect information on preferences and costs in both regions. This can be justified in our model, as the use of the local public good is observable (road use), and it can be priced.

infrastructure), (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 (see (9)).

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

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

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

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). This type of behavior, whereby policy makers favor the own region at the expense of others, has been empirically documented (see, for example, Knight 2004).

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

One important observation can be made right at the outset: depending on parameter values, both decentralization and centralization may lead to the highest welfare. If users have a majority in both regions, all combinations for which $\theta_i = \eta_i$ 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 ($\theta_1 = \theta_2 = 0.5$; $\eta_1 = \eta_2 = 1$).This is not surprising: if all voters are users 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. Decentralized decisions, however, would imply user prices exceeding first-best (see (6)).

In what follows, 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 afterward.

3.3.1 Pricing 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 'pricing' 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 user 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:

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

$$\begin{aligned} \text{C(1)} \qquad & \tau_1^{\mathrm{c}}(1) = \frac{\beta V}{K^0} + \left\{1 - \frac{2\theta}{\eta}\right\}\left(\frac{V}{-\frac{\partial V}{\partial \tau_1}}\right); \\ & \tau_2^{\mathrm{c}}(1) = \frac{\beta V}{K^0} + \left\{1 - \frac{2(1-\theta)}{\eta}\right\}\left(\frac{V}{-\frac{\partial V}{\partial \tau_2}}\right) \\ \text{C(2)} \qquad & \tau_1^{\mathrm{c}}(2) = \frac{\beta V}{K^0} + \left\{1 - \frac{2(1-\theta)}{\eta}\right\}\left(\frac{V}{-\frac{\partial V}{\partial \tau_1}}\right); \\ & \tau_2^{\mathrm{c}}(2) = \frac{\beta V}{K^0} + \left\{1 - \frac{2\theta}{\eta}\right\}\left(\frac{V}{-\frac{\partial V}{\partial \tau_2}}\right) \end{aligned} \tag{14b}$$

In Figs. 2 and 3, we illustrate the role of the parameters for the position of the different equilibria. Due to the symmetry assumption, the first-best (FB) and decentralized (D) pricing equilibria are both situated on the 45° line; the equilibria under centralized decisions (C(1), C(2)) 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 equilibria relative to the 45° line depends on the level of spillovers, whereas their position relative to the first-best largely depends on the voting shares.

In Fig. 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

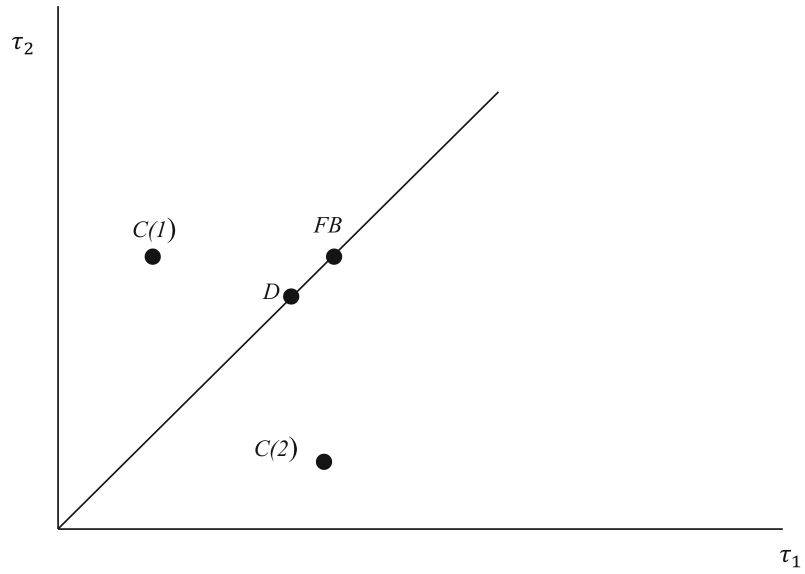

Fig. 2 Comparing tolling equilibria (case $1 > \theta > \eta > 0.5$)

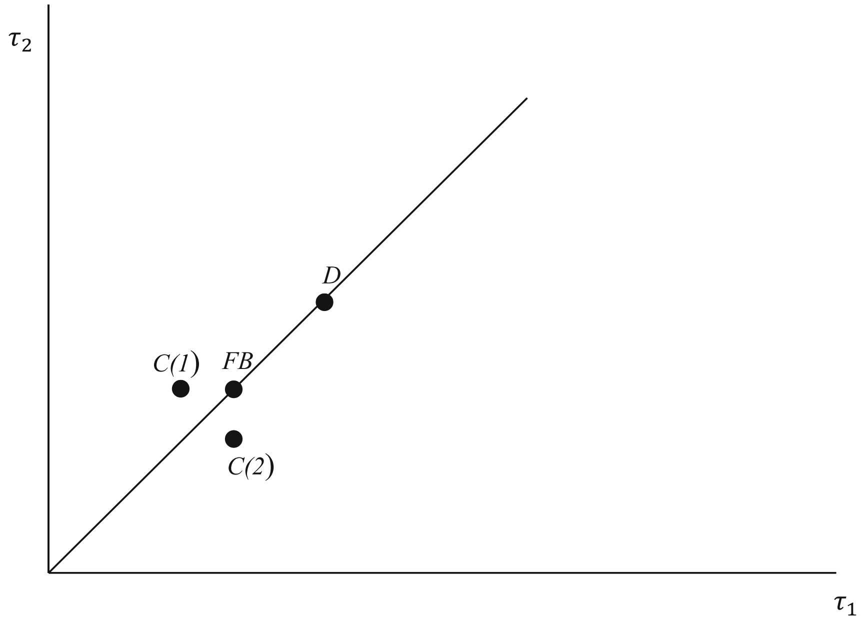

Fig. 3 Comparing tolling equilibria (η large, θ close to 0.5)

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, Fig. 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 3 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-users, and tolls will be higher than first-best (see 14a). The centralized solutions wanted by the representatives from the two regions will 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 in Fig. 3 suggest that centralization will yield higher welfare than decentralized decisions. We return to this observation below.

3.3.2 Welfare comparison: centralized versus decentralized decisions

Assuming a risk-neutral definition of welfare under centralization, we want to know whether

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

In this expression, W_i^{d} is welfare in region i under decentralization, and $W_i^{\mathrm{c}}(j)$ 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 of the electronic supplementary material.

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 of the electronic supplementary material 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,

Table 1 Comparing decentralized and centralized decision making: pricing a given capacity in the case of zero spill-overs and symmetric regions

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

The meaning of the price regimes HIGH, LOW and LOWLOW is explained in the text

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 of the electronic supplementary material. 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 (i) whenever non-users decide on user prices, they ignore the interests of users and prices will be inefficiently high; (ii) when users decide on prices, they will choose too low prices to avoid too much redistribution to non-users. Although this happens in both the centralized and decentralized cases, it is more pronounced in the centralized case because prices in the non-decisive region are always set very high. 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.

Introducing spill-overs 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 drivers have a majority in both regions. First, in Appendix 1 of the electronic supplementary material, we 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 Fig. 2 above. Second, if previous condition does not hold so that $\theta_i < \eta_i \quad (i = 1, 2)$, centralized decisions may outperform decentralized decisions. In this case, large spill-overs imply that with decentralized decisions, the interests of users from other regions are not captured in making decisions. In view of the discussion surrounding Fig. 3, however, we show in Appendix 1 of the electronic supplementary material 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. This is the case if driver majorities are large, and in any given region, use of the infrastructure comes about equally from local users and from people from outside the region. Intuitively, the former condition means that revenue sharing is not an issue, and 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.

Consider the case where users have a majority in one region only. We then again find that a sufficient condition for decentralization to be better is that $\theta_i \geq \eta_i$ (Appendix 1 of the electronic supplementary material). Finally, if non-users have a majority in both regions, the two political regimes give the same outcome, and parameters do not make a difference. Centralized and decentralized decisions will ignore the interests of all road users, both local and 'foreign.' This yields outcomes that are equally undesirable, leading to much too high user prices everywhere.

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 of the electronic supplementary material. The numerical exercise reported here is based on the following inverse demand function in each region i ($i = 1, 2$):

$$P_i = 1.2 - 0.0001 \times V_i.$$

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 example, we calculate the relative welfare performance of the two systems. We focus on the symmetric ($\theta_1 = \theta_2$, $\eta_1 = \eta_2$) case[10]. This has the advantage that the results can be easily graphically summarized, see Fig. 4. On the horizontal axis, we show the share of users in the region, and 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 users and users 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,

[10] The general asymmetric ($\theta_1 \neq \theta_2$, $\eta_1 \neq \eta_2$) case is analyzed in the working paper.

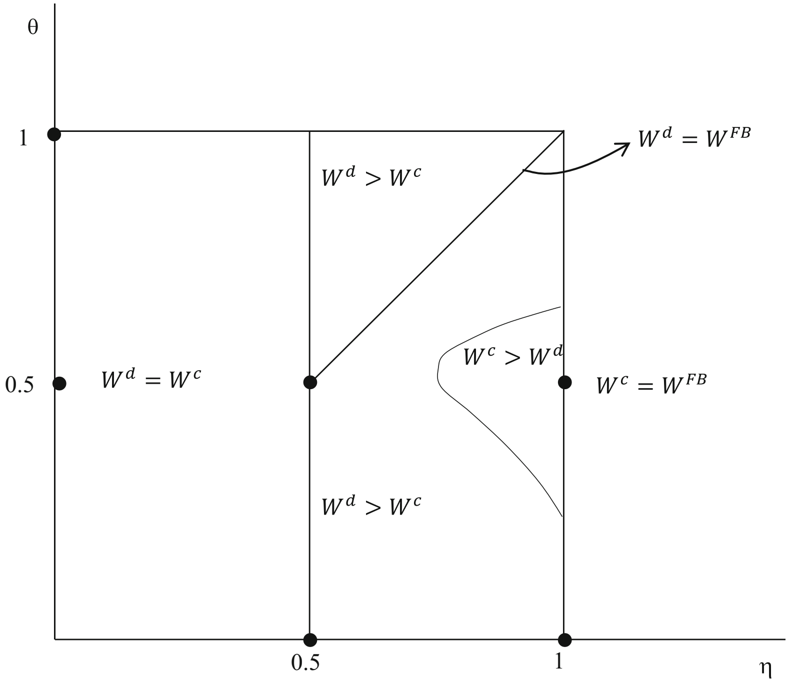

Fig. 4 Welfare comparison: decentralization versus centralization (symmetric case)

only when drivers have a large majority and spill-overs are close to $\theta_i = 0.5$ does centralization yield the highest welfare.

Figure 4 also clearly illustrates that, unlike in the paper of Besley and Coate (2003) that deals with pure local public goods, heterogeneity within regions implies that symmetric regions and zero spill-overs are neither necessary nor sufficient to produce first-best outcomes. On the one hand, decentralization implies a welfare loss even without spill-overs, unless all voters in both regions are users. On the other hand, decentralization can be first-best even in the presence of spill-overs: heterogeneity within regions can compensate for the existence of spill-overs.

We summarize the results of this section 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 decisions to perform better than centralized decision making is that either spill-overs are very limited or, when spill-overs are substantial, that 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 are neither necessary nor sufficient for decentralized decisions to yield first-best outcomes.

4 Capacity decisions and user pricing: centralization versus decentralization

In this section, we extend the model to capture capacity choices as well as pricing decisions. There are several reasons for doing so. It allows us to find out whether the finding of the previous section is robust when governments can both choose the level of capacity to be offered and how the use of that capacity is financed (through user fees or general tax revenues). Extending the model also allows us to formally study the implications of budgetary constraints. This is interesting, because we often observe central governments that leave decisions to regional authorities, but subject to a financing constraint that all or part of capacity costs should be covered by regional user charges. This raises the question why federal governments impose this institutional restriction, and what the implications are for welfare. To what extent do budgetary restrictions imposed on regional governments further improve the performance of decision making?[11]

Before proceeding, note that there is at least one reason why our model is better suited to discuss pricing solutions than to analyze capacity decisions. Capacity decisions for most congestible public goods (consider roads) have a long-lead time and are durable, often lasting for several decades. In principle, this would require a dynamic framework that extends over several time periods and covers several political terms. Moreover, it should 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). Although interesting, these issues will not be dealt with in the current paper. We focus on a few additional insights that can be derived within the static framework used before.

We start by analyzing the first-best at the regional level. Looking at an arbitrary region and ignoring the regional index for convenience, the region solves

$$\underset{\tau,\mathrm{K}}{\mathrm{Max}} \left\{ \int_0^V P(V)\mathrm{d}V - V \cdot g(V) \right\} + \tau V - \rho K \tag{15}$$

[11] One can also imagine other institutional constraints that improve decentralized solutions. For example, decentralization may in some cases 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: governments were required to charge the same prices to non-local users as to local ones. The use of a ‘non-discrimination principle’ in pricing policies is widespread in practice, both in federalized countries and in the EU. We will not discuss its efficiency effects here (see De Borger et al. 2005 for such an analysis).

Unsurprisingly, the optimal user price equals marginal external cost; furthermore, optimal capacity equates marginal benefit and marginal cost. We find

$$\tau = \frac{\beta V}{K}; \quad \beta \left(\frac{V}{K}\right)^2 = \rho \tag{16}$$

It easily follows from (16) that, at the optimum, we have $\tau V = \rho K$: the first-best gives exact cost recovery. Given the assumptions underlying our model—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).

Comparing decentralization with central decisions, we can be very brief. It is easy to show that the pricing rules are the same as in Sect. 3 above, and capacity rules are always the first-best rules (although, of course, the volumes are not first-best). It is clear then 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 Sect. 3. Intuitively, as long as the same party decides 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. As a consequence, as before, decentralization outperforms federal decision making under most plausible parameter combinations.

Of course, if somehow pricing and capacity choices are constrained, we may potentially obtain fundamentally new insights. One of the promising additional constraints to be considered is a self-financing constraint for capacity. This is interesting, because we often observe central governments that leave decisions to regional authorities, but subject to a financing constraint that all or part of capacity costs should be covered by regional user charges. In what follows, we therefore discuss the effect of a self-financing or earmarking constraint. Many variants of this constraint exist, but we will, for expositional reasons, focus on a simple and strict specification, requiring full cost recovery. We briefly discuss the implications of a partial cost recovery or co-financing restriction below.

To study the impact of a formal cost recovery constraint, let us consider regional decisions under a federal restriction of full cost recovery. First, suppose drivers have a majority in the region. Consider the following maximization problem, where a multiplier λ is used for the budget constraint:

$$\begin{aligned} &\underset{\tau,K}{\text{Max}}\ \frac{\theta}{D}\left\{\int_0^V P(V)\mathrm{d}V - V\cdot g(V)\right\} + \frac{\tau V-\rho K}{R} \\ &\text{s.t.}\quad \frac{\tau V-\rho K}{R} = 0 \\ &\qquad K \geq 0 \end{aligned}$$

Assuming an internal solution for capacity, we find the following rules for optimal behavior:

$$\tau = \frac{\beta V}{K} + \left\{1 - \frac{\theta}{\eta}\frac{1}{1+\lambda}\right\}\left(\frac{V}{-\frac{\partial V}{\partial \tau}}\right) \tag{17}$$

$$\beta\left(\frac{V}{K}\right)^2 = \rho \tag{18}$$

The cost recovery constraint only affects the user pricing rule, which boils down to Ramsey pricing, and it leaves the capacity rule unaffected; it is still the first-best rule.

Interestingly, however, substituting (17)–(18) in the budget constraint and rearranging, the results simplify to:

$$\tau = \frac{\beta V}{K}; \quad \beta\left(\frac{V}{K}\right)^2 = \rho \tag{19}$$

This the first-best outcome. Ex ante, the fact that a regional cost recovery restriction produces the first-best was not obvious; ex post, it is not surprising. We know that the first-best yields exact cost recovery, so imposing this as a constraint yields the first-best[12].

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 17), 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 spill-overs. 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.

Note that we assumed that the federal political level imposes full cost recovery on the regional level. Of course, this is a rather extreme assumption. What is often observed in practice is that the federal level imposes a partial cost recovery, or co-financing, constraint on the regions. Although this obviously does not lead to the first-best, it is a welfare-improving measure that moves the equilibrium closer to the social optimum. This is easily shown by replacing the full cost recovery restriction $\tau V - \rho K = 0$ by a constraint requiring partial cost recovery, viz., $\tau V = z(\rho K)$, where $0 < z < 1$. Following the same steps as before, the result then easily follows.

What are the implications of the cost recovery constraint imposed on the regions 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

12 As mentioned in the introduction, Brueckner (2015) 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. Van der Loo and Proost (2013) showed a similar result for one region facing local and transit traffic but where the federal regulator cannot observe the level of congestion.

they select, among all feasible solutions, the best solution for the users. All one needs is that the minority of users has a very small weight in the objective function of the majority of non-users[13]. Seen from this perspective, the major advantage of the regional budget constraint is to protect users from exploitation by non-users.

Quite intuitively, it is easy to show that federal decisions subject to regional budgetary constraints yield identical outcomes as decentralized decision-making subject to the same cost recovery constraints. Interestingly, however, requiring a federal budget constraint (requiring cost recovery at the central level, but not at each regional level separately) on federal decisions does not imply the first-best; it does not even necessarily improve federal decision making in the absence of such constraint. Why not? One easily shows 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.

We summarize the results in the following proposition.

Proposition 2 *Choice of capacity and user prices.*

a. *When governments make decisions on both user prices and capacity, the ranking of political institutions is the same as in the case where user prices can be decided, but capacity is fixed*
b. *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.*
c. *A federal cost recovery constraint does not necessarily produce a first-best solution. It may not even improve the federal outcome.*
d. *One advantage of regional cost recovery restrictions is that they protect users from exploitation by non-users.*

5 Policy implications

Drawing too strong conclusions from our very stylized models is risky, but a number of insights do seem worth reporting.

First, 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

[13] 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.

is widespread in the center of cities (and even in many small municipalities)[14]. 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 zero or 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.

Second, the model offers some support for the often-observed organization of the decision-making and financing process of local infrastructure projects that are subject to substantial spill-overs. Federal governments are heavily involved in decision making and in financing of such projects, but co-financing by the local authorities is the rule rather than the exception. This is in fact welfare-improving: we showed that welfare is indeed strongly enhanced if the federal level imposes a cost recovery or co-financing constraint on the regions. The EU regulation stipulating that the revenues of the member states' distance charges cannot exceed the road infrastructure costs is a real-world application of the efficiency of imposing cost recovery rules (Van der Loo and Proost 2013).

Third, 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 are appropriate. In regional bus transport (used more for short trips), however, there are much less 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 and hospitals, 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.

[14] 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).

6 Conclusions and suggestions for further research

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; at the federal level, different cooperative and non-cooperative political decision-making systems were considered. Moreover, we extended the model to capture both capacity and pricing decisions and studied the effect of imposing cost recovery restrictions. The model allowed for heterogeneity within as well as between regions.

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 decision making and decentralized decision making yield the same outcomes; however, the outcome is very undesirable, with prices that are much higher than marginal social cost. 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.

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. 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).

Of course, it is clear that we made a series of restrictive assumptions that may have to be reconsidered in future work. We can distinguish between two types of assumptions; those that allow more flexibility within the same institutional framework, and assumptions that reconsider the institutional framework itself.

First, with respect to the former, it was assumed that the size of the groups of users and non-users was exogenously given and did not vary with user charges within the range of prices considered. Our model was further 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 federal restrictions on local decision making 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.

Second, one could reconsider or generalize the institutional framework itself. The dominant message of this paper was that, for the selected local and federal institutions, decentralized decision making tends to do better. An obvious question is then whether there exist institutions that improve centralized decision making? Such institutions exist in many countries and usually take the form of constitutional constraints. The most important ones observed in practice include imposing uniform centralized pricing and bargaining at the federal level. Uniform federal pricing avoids that the region in command at the federal level abuses its position to charge very high prices in the other region. Federal bargaining among regions has the potential for improving efficiency for the same reason, as exploitation of one region by another is again avoided. Whether these constitutional constraints tilt the balance again in favor of centralized decision making depends on the precise policy parameters. This is further explored in a companion paper (De Borger and Proost 2015).

Finally, one might explore the link between the model of this paper and the literature on partial expenditure decentralization, whereby different government levels share expenditure responsibility for public good provision. In a recent paper, Joanis (2014) develops a political agency model in which two government levels are involved in the provision of a local public good, and where voters are imperfectly informed about the contribution of each government to the public good. Extending the model of this paper to study the efficiency implications of partial expenditure decentralization in the case of regional roads seems an interesting avenue for further research.

Acknowledgments The paper was presented at the IEB conference The Political Economy of Decentralization (Barcelona, June 2013), at the conference of the International Association of Transport Economists (ITEA, Northwestern University, July 2013) and at the Universities of Antwerp, Leuven and UC-Irvine. For useful suggestions, we are grateful to Ami Glazer, Jan Brueckner, Federico Boffa and Amedeo Piolatto. Two referees and an editor of this journal provided extensive written comments that substantially improved the paper.

References

Besley, T., & Coate, S. (2003). Centralized versus decentralized provision of local public goods: A political economy approach. *Journal of Public Economics*, *87*, 2611–2637.

Blöchliger, H. (2008). Market mechanisms in public service provision, OECD network on fiscal relations across governments, working paper no 6, OECD.

Borck, R., & Wrede, M. (2005). Political economy of commuting subsidies. *Journal of Urban Economics*, *57*, 478–499.

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

Brueckner, J. (2015). Decentralized road investment and pricing in a congested, multi-jurisdictional city: Efficiency with spillovers. *National Tax Journal*, *68*, 838–854.

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

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

De Borger, B., & Proost, S. (2015). *Centralized versus decentralized pricing of road transport in federal states: The role of institutional constraints*. Mimeo.

De Borger, B., Dunkerley, F., & Proost, S. (2007). Strategic investment and pricing decisions in a congested transport corridor. *Journal of Urban Economics*, *62*, 294–316.

De Borger, B., Proost, S., & Van Dender, K. (2005). Congestion and tax competition on a parallel network. *European Economic Review*, *49*, 2013–2040.

De Palma, A., & Lindsey, R. (2007). Transport user charges and cost recovery. In A. De Palma, R. Lindsey, & S. Proost (Eds.), *Investment and the use of user price and tax revenues in the transport sector* (pp. 29–57). Amsterdam: Elsevier.

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

Gutiérrez-i-Puigarnau, E., & van Ommeren, J. (2011). Welfare effects of distortionary fringe benefits taxation: The case of employer-provided cars. *International Economic Review*, *52*(4), 1105–1122.

Hatfield, J., & Padro i Miquel, S. (2012). A political economy theory of partial decentralization. *Journal of the European Economic Association*, *10*(3), 605–633.

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

Joanis, M. (2014). Shared accountability and partial decentralization in local public good provision. *Journal of Development Economics*, *107*, 28–37.

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

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.

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

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

Lorz, O., & Willmann, G. (2005). On the endogenous allocation of decision powers in federal structures. *Journal of Urban Economics*, *57*(2), 242–257.

Mohring, H., & Harwitz, M. (1962). *Highway benefits: An analytical framework*. Evanston, IL: Northwestern University Press.

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

Oates, W. E. (2005). Toward a second-generation theory of fiscal federalism. *International Tax and Public Finance*, *12*(4), 349–373.

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

Persson, T., & Tabellini, G. (2000). *Political economics: Explaining economic policy*. Cambridge: MIT Press.

Proost, S., Dunkerley, F., Van der Loo, S., Adler, N., Bröcker, J., & Korzhenevych, A. (2014). Do the selected Transeuropean investment projects pass the cost benefit test? *Transportation*, *41*(1), 107–132.

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

Russo, A. (2013). Voting on road congestion policy. *Regional Science and Urban Economics*, *43*(5), 707–724.

Small, K., & Verhoef, E. (2007). The economics of urban transportation. London and New York: Routledge.

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

Van der Loo, S., & Proost, S. (2013). The European road pricing game: How to enforce optimal pricing in high-transit countries under asymmetric information. *Journal of Transport Economics and Policy*, *47*, 399–418.