Governmental Competition in Road Charging and Capacity Choice

Barry Ubbels
Erik Verhoef*

Department of Spatial Economics, Vrije Universiteit Amsterdam.

* Tinbergen Institute.

Governmental Competition in Road Charging and Capacity Choice [#]

Barry Ubbels[*] and Erik T. Verhoef[**]
Department of Spatial Economics
VU University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Email: bub@nea.nl / everhoef@feweb.vu.nl

Abstract

This paper studies policy interactions between an urban and a regional government, both controlling one link of a two-link serial road network, where regional drivers may use both roads and urban drivers use the urban road only. Both governments set capacity and toll on one link, in a two-stage game where tolls are set after capacities have been committed to, and try to maximize social surplus for their own population. We use a simulation model to investigate the welfare consequences of the various possible game-theoretical set-ups. We find that governmental competition may be rather harmful to aggregate social surplus, compared to first-best policies. The main determinant of social welfare is not which exact type of game is played between the two governments, but much more whether there is cooperation (leading to first-best) or competition between them. Only of secondary importance is the question who is leading in the price stage (if there is a leader). Sensitivity analysis suggests that the relative performance for most game situations improves when demand becomes more elastic, and remain insensitive with respect to the unit cost of capacity expansions.

[*] Barry Ubbels was at VU University when working on this paper. He is currently at NEA Transport Research and Training Ltd., P.O. Box 1969, 2280 DZ Rijswijk, The Netherlands.

[**] Corresponding author. Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam.

1.   Introduction

Road pricing is receiving increasing interest as a means to reduce traffic-related problems such as congestion and pollution. Practical schemes typically operate at limited spatial scales. Examples are individual toll roads or express lanes, and local area-wide schemes such as Singapore's Electronic Road Pricing scheme and the London Congestion Charge. Most analytical studies of road pricing assume that a public regulator is interested in maximizing some aggregate welfare measure, typically social surplus, and sets its policy instruments accordingly. But given the relatively small spatial scale at which most schemes operate, it seems worthwhile to reconsider road pricing in the light of fiscal competition, where different governments may pursue the interests of different populations, and to explore the positive and normative implications of such competition.

A small but growing literature has done this, and this paper aims to contribute to that literature by looking at a specific set-up that is designed to capture the essential mechanisms and interactions that become relevant when an urban government sets a congestion toll that is to be paid not only by local inhabitants, but also by people from surrounding regions who may, for example, commute to the city considered. Some of the more elementary and predictable mechanisms at work in such a setting became clear in the referendum on the Stockholm congestion charge (Stockholmsförsöket, 2006). The scheme was supported primarily by inhabitants of Stockholm itself, who may expect to pay the (cordon) charge less often than others but to benefit from the scheme's revenues, while the scheme was opposed primarily from people outside the central Stockholm area, for whom the opposite holds.

We develop a simple two-link serial roads network model, with a 'regional' road and an 'urban' road, where 'regional' drivers travelling to the city use both roads, and 'urban' drivers use the urban road only. We take a long-run perspective where both governments set capacity and toll on their links, in a two-stage set-up where tolls are set after capacities have been committed to. In doing so, both governments are assumed to try to maximize social surplus for their own population only, while optimal coordination of policies forms the first-best benchmark. Our model is thus representative for cases where a city can employ its own road pricing scheme, and adjust road capacity accordingly, and where – although the regions in the hinterland have the same instruments for their own roads – an asymmetry occurs because people from outside the city far more often travel within the city's boundaries that that city residents do in the regional hinterland. The implied danger of tax-exporting toll setting by the city plays an important role in debates on the question of whether road pricing schemes should be primarily local policies, or a national responsibility.

Despite the simplicity of our set-up, the analytics of the two-stage model appeared to be too tedious to allow for the derivation of insightful transparent analytical results, and we therefore use a simulation model to investigate the welfare consequences of the various possible game-theoretical structures. We consider various such structures, with Nash and Stackelberg behaviour possible in both stages of the game, and – with Stackelberg behaviour

– both the city and the region considered in the role of the 'leader'. Our approach is therefore not to make any *a priori* assumptions on which game structure would be more plausible in reality, but rather we test them all to see whether it actually matters what the exact structure is. This approach turns out to be rewarding in the sense that we find that results of different game structures tend to come in certain clusters – so that, within a cluster, the exact structure is immaterial for the eventual welfare properties of the resulting equilibrium.

The paper starts off with a concise review of the relevant literature in Section 2 below. Section 3 introduces the model, while numerical results are presented in Section 4. Section 5 presents sensitivity analyses, and Section 6 concludes.

2.   Prior literature

A well-known fundamental insight in economics states that under otherwise first-best circumstances, efficient prices should reflect marginal costs. In the context of road pricing this translates into the well-known Pigouvian tax rule, which states that the toll on a congested road be equated to the marginal external cost. Such a pricing scheme is often referred to as 'first-best' because it maximizes efficiency, given that efficiency is also maximized in all other relevant markets in the system considered – an important assumption that is, in fact, often left implicit. But the literature on road pricing has been extended in various directions, including many cases where such first-best pricing is either not feasible because the tax instrument itself is not optimal, or it is not efficient because there are other market failures to be considered, besides congestion or other external costs on the road under consideration. In such cases, second-best pricing becomes relevant. A recent review of the rapidly growing literature on second-best road pricing is provided by Small and Verhoef (2007).

An important type of second-best pricing, directly relevant for our analysis, occurs when tolls are only implemented on single lanes or roads, with alternative routes being unpriced. Imposition of so-called 'quasi first-best pricing' on a tolled road, implying that the toll is simply set equal to marginal external costs on that road, is then not welfare maximising. Second-best tolls instead take spill-overs upon unpriced capacity into account (Lévy-Lambert, 1968; Marchand, 1968). Welfare gains from second-best tolls with unpriced substitutes are generally found to be rather low (*e.g.*, Liu and McDonald, 1998), but become higher when allowing for heterogeneity of travellers (Verhoef and Small, 2004) or for the dynamics of departure time adjustments (Braid, 1996; De Palma and Lindsey, 2000).

This two-route network has also been used to study the impact of different ownership regimes. Verhoef, Nijkamp and Rietveld (1996) considered, for instance, two private ownership regimes; one where one of the routes is private and the other has free access, and a second situation where a private monopoly controls both routes. They find that revenue maximising tolling on two routes may actually sometimes lead to a more efficient usage of road space than does second-best optimal one-route tolling. Hence, it may be more efficient to have a monopolist controlling the entire network, rather than just a part of it. De Palma and Lindsey (2000) were among the first to consider strategic competition in this line of research,

on a similar two-route network. They focused on the efficiency of private toll roads versus free access but also versus public toll road pricing, employing a dynamic model of bottleneck congestion. Their results show that two competing private roads can yield most of the potential efficiency gains from first-best pricing, at least if neither road has a dominant fraction of total capacity.

Authors	Year	Parallel (P) / Serial (S) / (with transit (T))	Pricing (P) / Capacity (C)	Public control (G) / Private control (P) / Free of toll (F)	Single operator (S) / Duopoly (D) / More than two (M)
Levy-Lambert	1968	P	P	GF	S
Marchand	1968	P	P	GF	S
DeVany and Saving	1980	P	PC	P	M
De Palma and Leruth	1989	P	PC	P	D
McDonald	1995	P	P	GF	S
Viton	1995	P	P	PF	S
Braid	1996	P	P	GF	S
Verhoef, Nijkamp and Rietveld	1996	P	P	GPF	S
Liu and McDonald	1998	P	P	GPF	S
Liu and McDonald	1999	P	P	GF	S
De Palma and Lindsey	2000	P	P	GPF	D
Levinson	2000	S	P	G	M
Small and Yan	2001	P	P	GPF	S
Verhoef	2002	Generalized	P	G	S
Engel, Fischer and Galetovic	2004	P	P	P	M
Verhoef and Rouwendal	2004	PS	PC	GF	S
Verhoef and Small	2004	PS	P	GPF	S
De Borger, Proost and Van Dender	2005	P (T)	P	G	D
Van Dender	2005	P	P	P	D
De Borger and Van Dender	2006	P	PC	P	D
Proost and Sen	2006	No explicit network	P	G	D
De Borger, Dunkerley and Proost	2007	S (T)	PC	G	D
Verhoef	2007	PS	PC	GPF	S
Small and Verhoef (Ch. 6)	2007	PS	P	P	M
This paper		S	PC	G	D

Table 1. An overview of the recent economic literature on road pricing and network interactions (by year of publication)

The economic literature on road pricing and network interactions has developed rapidly over the past decades, producing analyses on various types of networks, with various types and combinations of operators (*i.e.*, private and public), using various types of instruments (*i.e.*, tolls and capacities), under various types of competitive conditions (*i.e.*, the number of

operators simultaneously active on the network). Table 1 attempts to provide a systematic overview of this literature, by classifying the various contributions according to a number of dimensions. The table focuses on road pricing, and therefore ignores the narrowly related literature on bi-modal competition such as rail-road (*e.g.* Arnott and Yan, 2000), as well as that on airport congestion pricing (*e.g.* Brueckner, 2002; Pels and Verhoef, 2004).

A first dimension distinguished in Table 1 concerns the type of network. Besides the exceptions of Verhoef (2002) who studies generalized networks of undetermined size and shape, Proost and Sen (2006) who look at vertical competition between two governments without modelling an explicit network, and Verhoef and Small (2004) and Verhoef and Rouwendal (2004) who use a three-link network with serial and parallel links, most papers look at interactions between either exclusively parallel links, or exclusively serial links. Networks with parallel links appear to have been far more popular in this research field, for no apparent empirical reason.

Next, most contributions take capacities as fixed and focus on toll setting. Besides two contributions in the 1980's, only recently has capacity choice been brought into the analysis, thus allowing a long-run perspective. When combined with price setting, it is habitual to assume – like we will do below – that strategic decision making can be characterized as a two-stage process, where prices are set after capacities have been committed to.

A third dimension concerns the types of ownership of links that are assumed to occur. Overall, there appears to be a rather balanced mix of analyses that consider public ownership (*i.e.*, governments who seek to maximize social surplus); private ownership (*i.e.*, firms who try to maximize profits); roads that are free of toll; or various combinations of these.

A final distinction is between models that have only a single operator active in the network, possibly 'competing' against free roads, and models with strategic interactions between operators – either two, in case of a duopoly, or more than two.

It seems impossible to summarize the findings from such a broad literature in a few, generally supported conclusions. Tautologically, when public regulators are involved, policy coordination is, in terms of overall efficiency impacts, preferable to competition between governments. Because 'foreign' users on a jurisdiction's road(s) only matter for local welfare in that they may bring in toll revenues and hinder local travellers (on the own road or elsewhere in the network), a government's behaviour may resemble that of a profit-maximizer with respect to its tolling of foreign travellers. For private operators, Small and Verhoef (2007) show how the main insights from Economides and Salop (1992) carry over to congested networks: an increasing number of parallel private competitors (suppliers of substitutes) seems to bring tolls closer to the efficient level, while an increasing number of serial competitors (suppliers of complements) has the opposite effect. Insofar as perverse incentives for public operators are fed by the desire to raise toll revenues from foreign users, one would expect a similar regularity for governments.

Otherwise, if anything, a joint but somewhat unsatisfactory conclusion from the studies mentioned in Table 1 would be that the eventual (welfare) implications from road

pricing with network interactions can be expected to depend rather heavily on the details of the network structure; on elasticities of demand functions and the distribution of demand over the network; on the nature of congestion and its distribution over the network; on the distribution of operators and toll-free roads over the network; on the type of operators involved; and on the type of instruments they have available. In other words: at the current state of knowledge, it seems dangerous if not impossible to predict, from earlier studies, what will be the approximate outcome for a certain configuration.

This provides justification for our study, but also makes it important to relate our attempt to the prior studies reviewed in Table 1. As explained, ours will be a two-link serial network, with a regional road and an urban road, where regional drivers may use both roads, and urban drivers use the urban road only. Two public regulators strategically set capacities and then tolls, in a two-stage game.

In terms of characteristics indicated in Table 1, this makes the analysis closest to that by De Borger, Dunkerley and Proost (2007). They consider a symmetric set-up where two links in series are both controlled by a different government, and traffic consists of local travellers and transit. Consistent with the findings of Economides and Salop (1992) and Small and Verhoef (2007) for private operators, equilibrium transit tolls are found to inefficiently high. There are two major differences between the current analysis and that of De Borger, Dunkerley and Proost (2007). The first is that they study a purely symmetric case with horizontal competition between the two governments. Our analysis introduces asymmetry and vertical tax competition, where the city government charges regional drivers who are also taxed in the region, while the regional government cannot charge urban residents. The second difference is that De Borger, Dunkerley and Proost (2007) consider only Nash behaviour of the two governments, while we will also give a systematic treatment of the various games that become possible when Stackelberg behaviour is introduced.

Somewhat less close to ours is the model of Proost and Sen (2006). They consider a one-stage game with only price instruments: parking for the urban government and a cordon toll for the regional government. Our analysis therefore differs in that we consider a two-stage game, with capacities set in the first stage. Moreover, the regional government in Proost and Sen (2006) is also concerned with the urban residents' welfare: the city is part of the region. As a result, the welfare loss of non-coordination between governments is limited in their model. Our region only includes the hinterland, making the competition between governments 'sharper' and the relative welfare losses of non-coordination larger.

Probably 'third-closest' is the model of Levinson (2000), also looking at strategic interactions between governments on a serial network. But Levinson keeps capacity fixed, and focuses on the choice of revenue raising mechanisms (tax versus toll) in absence of congestion and other external costs. He finds that larger regions are more likely to tax than smaller regions, who are relatively better of with (cordon) tolls at their borders.

Finally, the already mentioned result of Small and Verhoef (2007) for serial roads, namely that revenue-maximization leads to overcharging, will be found to be relevant also for

the present set-up, in that the city government will have an incentive to overcharge regional drivers, who are used as a source of revenues by the city.

        The other contributions in Table 1 will not individually be compared to ours, for reasons of space and because they are simply too different in set-up to warrant a detailed comparison; either because the network – often with parallel links – is too different, or because strategic interactions between regulators are absent.

3.   The model

Figure 1 depicts the spatial structure of our model. There are two serial roads, denoted $i=\{1,2\}$, where road 1 is under the control of the city's government and road 2 of the region's government. Both roads are congestible; so that the generalized price depends on the use levels N_i. The city's road 1 is used by the city's residents ('group C'), and by commuters who travel from the region to the city ('group $R1$'): $N_1 = N_C + N_{R1}$. The region's road is used by regional travellers only; besides commuters $R1$, these could be people who travel in the region only and therefore only use road 2 ('group $R2$'): $N_2 = N_{R1} + N_{R2}$. Reverse commuting (living in the city, travelling to the region) is therefore ignored. Drivers are identical in all respects except for their origin-destination pair and implied route (*i.e.*, the distinction between groups C, $R1$ and $R2$), and, within each group, their willingness to pay to make a trip. We abstract from dynamics and consider stationary traffic conditions.

        A government controlling a road can set its capacity, cap_i, and its toll, τ_i. The government keeps all toll revenues levied on its road, but also carries its full capacity cost. (We therefore ignore intergovernmental grants, which are in practice frequently used to make governments share the cost of road construction.) Under these conditions, both governments are assumed to maximise social surplus for their inhabitants, defined as total (Marshallian) benefits from travelling minus total user costs (including time costs) incurred, minus the capacity costs, plus the jurisdiction's net result on inter-jurisdictional toll payments.
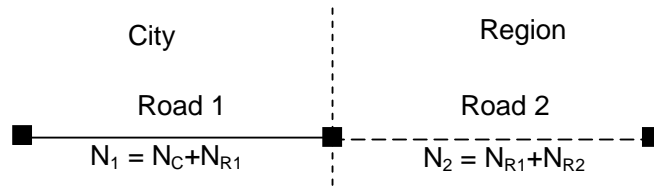

Figure 1. The spatial structure of the model

We then face the choice of whether to develop a model with general demand and cost functions, of undetermined functional form, and try to derive general analytical results for these, or to use specific functional forms that allow for numerical analysis. The former strategy has the advantage of producing results that are more generally valid, but the disadvantage that the marginal conditions that one may derive still give limited insight into relative welfare differences between various equilibria of interest. It is for that reason common to do both. Unfortunately, we were unable to derive transparent and insightful

analytical results for two-stage games with general demand and cost functions. This motivated us to concentrate mainly on numerical analysis in this paper. We will therefore not engage in deriving theoretical results on properties of reaction functions. Only the first-best problem and two second-best cases, where the 'other' government's capacity and toll are both kept fixed when setting one's own toll and capacity, produce digestible analytical results. These we present in the Appendix. We emphasize that those results do not refer to equilibria for any of the sequential two-stage games that are of primary interest for this paper. Nevertheless, the second-best cases discussed in the Appendix provide some deeper insight into the basic forces at work in our model, which makes it worth presenting them. In particular, the toll rules for the two second-best cases are, in fact, relevant also for some of the games considered below, as they reflect the toll rules that will be set in the second stage of a game by a regulator who takes the toll and capacity of the competing government as given. (Because capacities, in the various games, are always set in the first stage and are therefore not set taking tolls as given, a similar statement cannot be made on the investment rules.)

We use simple linear demand and cost functions, primarily for reasons of tractability (a secondary defence for the user cost function is that the reduced form average cost function for the basic bottleneck model is also linear; see also Arnott, De Palma and Lindsey, 1993). The inverse demand functions can thus be written as:

$$D_j = d_j - a_j \cdot N_j \qquad j = \{C, R1, R2\} \tag{1}$$

where d_j and a_j are non-negative parameters.

Generalized average user cost functions, reflecting all costs incurred by each user including the value of travel time, are assumed to be homogeneous of degree zero in use and capacity, and linear for a given capacity, so can be written as:

$$c_i = k_i + b_i \cdot \frac{N_i}{cap_i} \qquad i = \{1,2\} \tag{2}$$

where also k_i and b_i are non-negative parameters, the former reflecting free-flow generalized cost and the latter the effect of congestion upon generalized cost.

For tractability, we assume that there are neutral scale economies in road capacity construction so that a road's aggregate capacity costs is linear in capacity:

$$C_i^{cap} = p_{cap} \cdot cap_i \qquad i = \{1,2\} \tag{3}$$

where p_{cap} is the unit price of capacity, assumed equal between the two roads. Note that the neutral scale-economies cost structure implied by (2) and (3) implies that our road network qualifies for the application of the Mohring-Harwitz (1962) result on exact self-financing of optimally designed and priced roads (see also below). Small and Verhoef (2007) provide a derivation and further discussion of this result, and also review the empirical evidence on the economies of scale in highway capacity cost. They conclude that mild scale economies are

likely to apply for networks in major cities, which may disappear altogether in very large cities. Assuming neutral scale economies therefore seems reasonable.

The generalized prize p_i, faced by users of road i, adds the toll τ_i to the generalized average cost c_i of equation (2):

$$p_i(\cdot) = c_i(\cdot) + \tau_i \qquad i = \{1,2\} \tag{4}$$

The specific network structure implies that the following equilibrium conditions apply:[1]

$$\begin{cases} D_C(N_C) = p_c \equiv p_1(N_1) \\ D_{R1}(N_{R1}) = p_{R1} \equiv p_1(N_1) + p_2(N_2) \\ D_{R2}(N_{R2}) = p_{R2} \equiv p_2(N_2) \end{cases}$$

with :

$$N_1 = N_C + N_{R1}$$
$$N_2 = N_{R1} + N_{R2}$$

(5)

We take total social surplus W as the relevant measure of welfare, and as the objective that a government seeks to maximize. It is equal to the 'variable' social surplus (the benefits B as given by the relevant area under the demand curve(s)[2], minus total user costs), minus the total capacity costs, plus the net result on inter-jurisdictional toll payments. We can define it for the city (W_C), for the region (W_R) and at the aggregate level ($W \equiv W_C + W_R$):

$$W_C = B_C - N_C \cdot c_1 + N_{R1} \cdot \tau_1 - p_{cap} \cdot cap_1 \tag{6a}$$

$$W_R = B_{R1} + B_{R2} - N_{R1} \cdot c_1 - (N_{R1} + N_{R2}) \cdot c_2 - N_{R1} \cdot \tau_1 - p_{cap} \cdot cap_2 \tag{6b}$$

$$W = B_C + B_{R1} + B_{R2} - (N_C + N_{R1}) \cdot c_1 - (N_{R1} + N_{R2}) \cdot c_2 - p_{cap} \cdot (cap_1 + cap_2) \tag{6c}$$

Tolls paid by commuters ($R1$) on road 1 are a transfer between jurisdictions, and thus increase the city surplus and reduces that in the region. All other toll payments cancel out in local welfare functions, because they constitute a transfer from local residents to the local authorities. For aggregate welfare W, also the transfer from the region's commuters to the city's government of course cancels.

As stated, the Appendix derives general analytical expressions for toll rules and investment rules for three cases: the first-best case; the case where the city optimizes, and the regional toll and capacity are given; and the reverse case where the region optimizes, while the city's toll and capacity are given. The insights from the Appendix can be summarized as follows. The first-best equilibrium involves tolls that are equal to marginal external costs, and a conventional investment rule that equates marginal costs of capacity expansion to the marginal benefits. When either government sets their instruments in isolation, the investment

---

[1] Corner solutions, where the demand of at least one of the groups is reduced to zero, will not be considered.

[2] The mathematical operation is: $B_j(N_j) = \int_0^{N_j} D(n)\,dn$.

rule does not change: given the level of road use, capacity is set at the efficient level. Because the levels of road use will be different than in the first-best equilibrium, the equilibrium capacity level will, of course, still be different from the first-best level. Both governments have an incentive to set the toll above the road's marginal external cost. For the city this extracts additional toll revenues from regional users. For the region, this is meant to (imperfectly) internalise the congestion externality that $R1$ drivers impose upon one another on the city's road, link 1.

These same forces will also be at work in the two-stage games to be analyzed in the next section, although the two-stage nature of the games will normally cause the capacity rules of the Appendix not to apply exactly, while the second-best toll rules presented are only representative of a regulator exhibiting Nash behaviour. As indicated, the numerical analysis of the various games reflects the comparative static impacts of toll and investment rules upon possible equilibria, but is of course less general than our analytical results in the Appendix, because they pertain to an assumed set of demand and cost functions. A sensitivity analyses, in Section 5, aims to compensate for this.

4.   Numerical results

4.1.   *Calibration of the 'base case', and first-best and second-best equilibria*

For the 'base case' of our numerical model, the following parameter values were chosen: $a = 0.6$; $d_C = d_{R2} = 140$; $d_{R1} = 280$; $k_1 = k_2 = 20$; $\beta = 20$; $p_{cap} = 2$; $cap_1 = 500$; and $cap_2 = 500$. The base-case equilibrium leads to a reasonable demand elasticity of –0.4 for each group, at an equilibrium use of 167 for groups C and $R2$, 333 for $R1$, and equilibrium travel costs twice the 'free-flow' levels. These parameter values were otherwise not motivated by any desire to represent a realistic situation; also note that, because we are free to choose units of traffic flow and prices, only relative indicators of the equilibrium (elasticities, relative travel times, *etc.*) matter. The equality between equilibrium demands for groups $R2$ and C was motivated by the desire to have the example as symmetric as possible.

Table 2 shows the detailed numerical results for the various scenarios under study. Because the units are arbitrary, we have normalized these results and express them as a fraction of first-best values (which, itself, are given between brackets in the row 'First-best').

In the first-best optimum, the symmetry of the network causes the tolls and capacities to be equal for both roads. The toll amounts to 6.32, the marginal external costs in the optimum. The capacities are about 3.4 times as high as the initial levels. This is, of course, rather extreme. It is a direct consequence of the relative level of p_{cap} that we have chosen, and could therefore easily have been avoided. We have, however, chosen this parameterisation so as to create sufficient disparity between the initial equilibrium capacities and the optimum, so that relative differences between various options can easily be observed. Although it cannot be seen directly from Table 2, the underlying results confirm that the optimal network is exactly self-financing, in accordance with the Mohring and Harwitz (1962) result.

In the second-best case where the city sets the toll and capacity on road 1 while the region remains passive, we find a substantially higher social surplus for the city (W_C) than in the first-best situation. In contrast, the welfare of the region (W_R) has decreased considerably. Aggregate welfare (W) is lower than in the base-case equilibrium, so that the relative efficiency indicator ω is negative: $\omega = -0.19$.[3] The toll is equal to 78.11, more than 12 times as high as the first-best toll. The corresponding toll rule in the Appendix shows that the city adds a positive mark-up to the conventional Pigouvian toll, to extract toll revenues from the regional commuters (N_{R1}). A countervailing force in doing so is that the toll is also paid by the city residents, but initially, only 1/3 of intra-city traffic concerns city residents. Moreover, toll payments by city residents, although distorting prices within the city, in themselves only constitute transfers within the city. The number of commuters decreases significantly, as well as the welfare level in the region, and so does the number of local drivers (N_C). Capacity is set considerably lower than in the optimum, but the ratio demand/capacity is the same, because the first-best investment rule still applies.

The other second-best situation is when the region optimises while the city does not change capacity and toll. The region's second-best toll also exceeds the first-best toll, but not by as much as in the previous case. Because the toll internalizes congestion for groups $R1$ and $R2$ (although imperfectly), it will increase the welfare of the region. Since the welfare of the city remains almost constant (relative to the base-case situation), also overall welfare increases: $\omega = 0.48$. This more favourable result is to a large extent explained by the fact that the region cannot tax city residents, and hence will not raise the toll with the purpose of extracting revenues from non-inhabitants. Furthermore, the toll internalises part of the congestion externality in the city: namely, insofar as it is imposed by regional residents on themselves. The asymmetry in the spatial set-up thus translates into rather diverging welfare impacts from the two second-best policies where one government optimizes while the other remains passive. The city's incentive to raise revenues from regional drivers has potentially strong negative impacts upon overall efficiency.

4.2.   *Non-cooperative game equilibria*

We now turn to strategic interactions between the two governments when setting tolls and road capacities in a non-cooperative fashion. In general, such games could be assumed to have a Nash structure, where each government takes the other's behaviour as given, or a Stackelberg structure, where one of the two governments acts as leader and chooses its policy instruments taking the other's response into account, while the follower takes the leader's choices as given. Moreover, because we model the choice of capacity and tolls as a two-stage game, with capacity set in the first stage and tolls set in the second stage when capacities have been committed to, Nash and Stackelberg behaviour could in fact apply at different stages of

---

[3] This indicator ω is defined as the difference between welfare in the situation under study and base-case welfare, divided by the difference between first-best welfare and base-case welfare.

the same game, and either the city or the region may lead in a Stackelberg stage. Even if we exclude, on the grounds of implausibility, the possibility of reversed Stackelberg leadership between stages (*i.e.* where the region leads in the one stage and the city in the other), this still leaves seven different situations to be considered. A first is the 'Full Nash' equilibrium, in which both prices and capacities are set in a non-cooperative Nash game. Next there are three different cases with city leadership in the first, the second, or both stages; and finally the same three situations with the region as the leader.

Faced by so many possible equilibria, one could choose to narrow down the options by asking which is more plausible. This question, however, is not straightforward to answer in general, because it may depend on local circumstances such as the relative sizes of the city and the surrounding region(s), on institutional aspects and the flexibility in policy making for both governments, *etc.* We therefore decided to follow the alternative strategy, and report on all seven possibilities. This has the advantage of giving insight into the question of to what extent the exact structure of the game in fact matters for the eventual outcomes. We will coin the different games such that Nash/Stack means Nash behaviour in the short run (the toll stage, but the second stage of the game) and Stackelberg in the long run (the capacity stage, which is the first stage of the game).

Full Nash

In the first stage of the Full Nash equilibrium, both governments choose capacity as the best reply to the other's selected capacity, knowing that given these capacities, a Nash equilibrium in tolls will apply in the second stage. Figure 1 represents this graphically. The thinner lines give iso-surplus contours in the capacity-capacity space, and connect capacity combinations yielding equal local surpluses, *given* that a Nash price game will be played once the capacities are set. The solid contours refer to the city and represent a higher surplus when moving to the right. The dashed contours refer to the region and represent a higher surplus when moving up.

Next, the thicker lines give the reaction functions: the best response (in terms of capacity) given the capacity set by the other government, and given (again) that a Nash price game will be played once the capacities are set. The solid city's reaction function is therefore found as the connection between exactly vertically-sloped points of the city's various iso-surplus functions. The region's reaction function, in a similar fashion, connects exactly horizontally-sloped points of the region's various iso-surplus functions. The intersection of the two reaction functions then defines the Full Nash equilibrium.

The numerical results in Table 2 show that the Full Nash situation results in a small relative (overall) welfare gain ($\omega = 0.07$) compared with the base-case equilibrium. The city again has the incentive to extract toll revenues from the non-residents, leading to excessively high tolls. Regional tolls are somewhat lower than in the second-best 'region' situation, because congestion in the city has reduced. The capacities in both jurisdictions are adjusted more or less proportionally to equilibrium link flows compared to the first-best situation, but not exactly so because capacity choice has become a strategic instrument in the game's first

stage. This confirms that the capacity rules of equations (A.4b) and (A.6b), which would imply equi-proportional changes in flow and capacity, do not apply in the equilibria for sequential games.
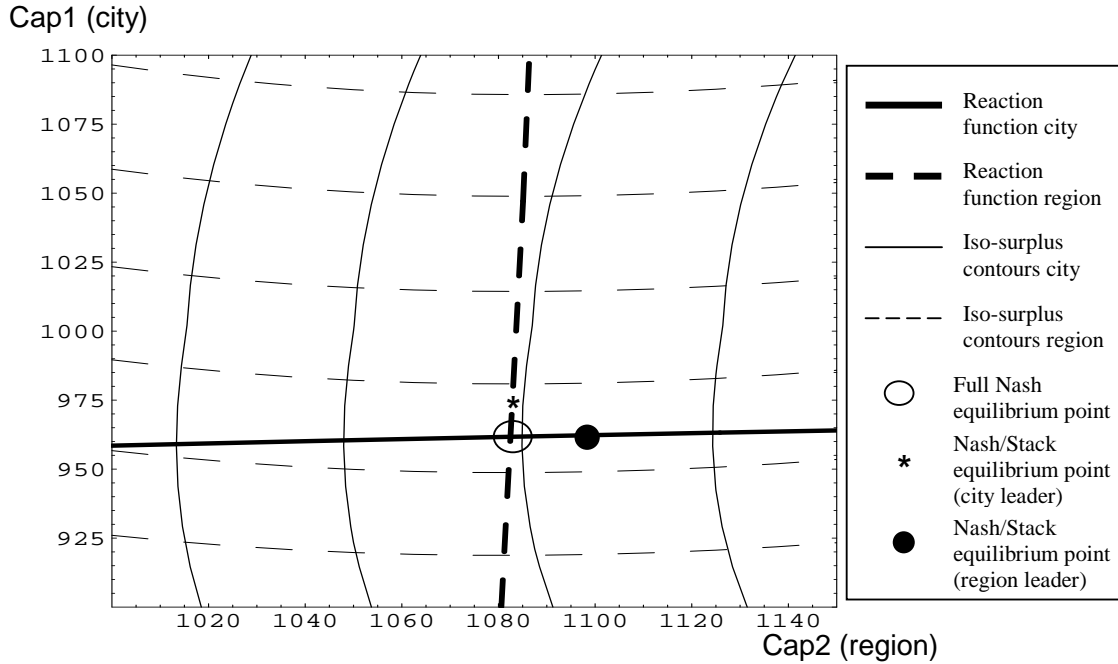
Cap1 (city)

Figure 2. Reaction functions and iso-surplus contours for Nash toll games

City leadership

City leadership can occur in a Nash/Stack setting, in Stack/Nash, and in Stack/Stack. The first of these has a Nash toll stage and can therefore be analyzed using the same reaction functions and iso-surplus contours as used for Full Nash, shown in Figure 2.

When the city is the leader in Nash/Stack, it searches the highest city surplus level W_C along the region's reaction function. This is the point where the region's reaction function is tangent to the highest achievable iso-surplus curve of the city. It is the Nash/Stack equilibrium point (city leader) indicated with a star in Figure 2. The increase in capacity of the region is negligible compared with the Full Nash equilibrium. The change in capacity chosen by the leading city is somewhat larger, but is still also modest. There is a small increase in social surplus for the city compared to Full Nash (which is lost in rounding in Table 2, but visible in Figure 2). Interestingly, also the follower benefits from the higher level of capacity: both jurisdictions end up on a higher iso-surplus contour in Figure 2. However, the welfare changes are very small: given that the pricing game entails Nash competition, the nature of the capacity game appears to be less important for the final outcome.

When the toll stage is characterised by Stackelberg behaviour, another set of iso-surplus contours and (hence) reaction functions applies. Figure 3 shows the set when the city leads. Two equilibria are of interest: Stack/Nash where the capacity stage entails a Nash game, and Stack/Stack where both instruments are set in a Stackelberg way. The

corresponding equilibria are again found as the intersection of the reaction functions (Stack/Nash), and as the point of tangency between the region's reaction function and the highest achievable city's iso-surplus contour (Stack/Stack). Both games have slightly higher city tolls than what was found under Nash pricing, slightly higher capacities, and a slightly higher welfare level for the city compared with the Nash pricing games. The welfare of the region decreases slightly in this situation compared with the various Nash-pricing equilibria. But the relative differences with Nash tolling games are small.
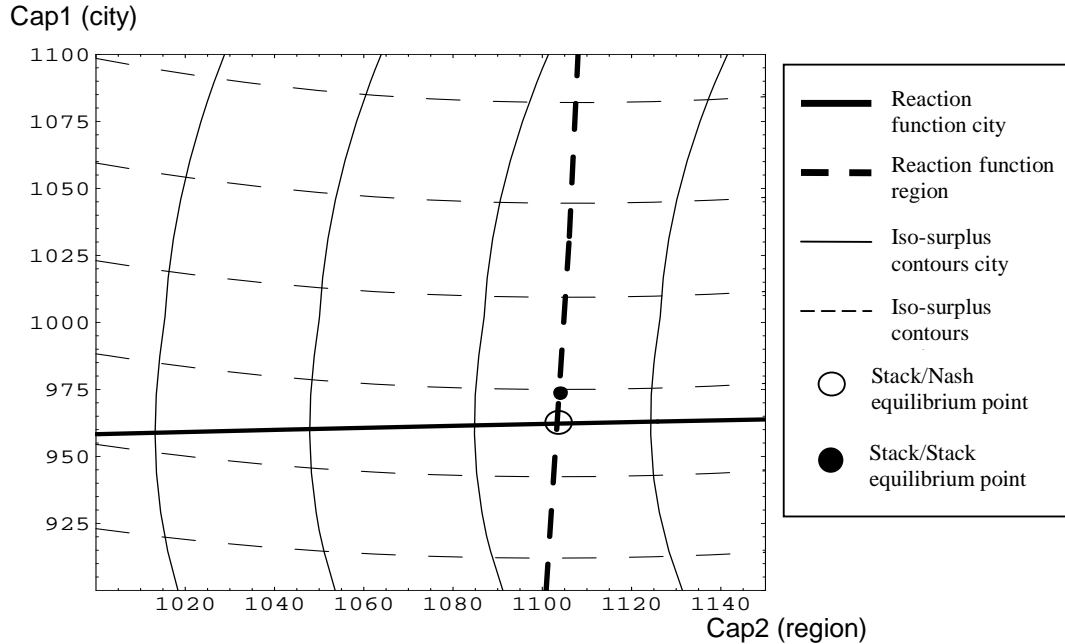

Figure 3. Reaction functions and iso-surplus contours for city leadership in the toll stage

What causes these differences between Nash tolling and Stackelberg tolling to be so small? It reflects that the region's toll response to the city's toll decision is relatively unimportant to the city. This unimportance stems from two facts. First, the region's toll is relatively small compared with the city's toll, so that changes in the region's toll (even when significant in a relative sense) are still rather unimportant to the city. Secondly, the region's toll aims to internalise the region's commuter congestion, both on the region's and the city's road. Because a change in the city's toll only affects congestion for some of the region's travellers, and only for a part of their trip, also relative changes of the region's toll in response to city toll changes will be limited. This makes it relatively unimportant for the city to take this response into account, instead of treating the region's toll as being fixed.

Region leadership

Also for region leadership we consider Nash/Stack, Stack/Nash, and Stack/Stack. The first of these can again be analyzed using Figure 2. The Nash/Stack equilibria in Figure 2 reveal an interesting result: both jurisdictions would prefer the other government to lead in a

Stackelberg capacity stage, when followed by a Nash toll stage, above leading themselves – which in turn is preferred over a Nash capacity stage. Such seemingly counterintuitive results are not that rare in game theory (see Dowrick, 1986).

Otherwise, the differences between equilibria with Stackelberg leadership in the capacity stage and Nash in the pricing stage terms of capacity are small. Leadership for the region in the toll stage, however, changes outcomes more significantly. Figure 4 shows the relevant contours and reaction functions.
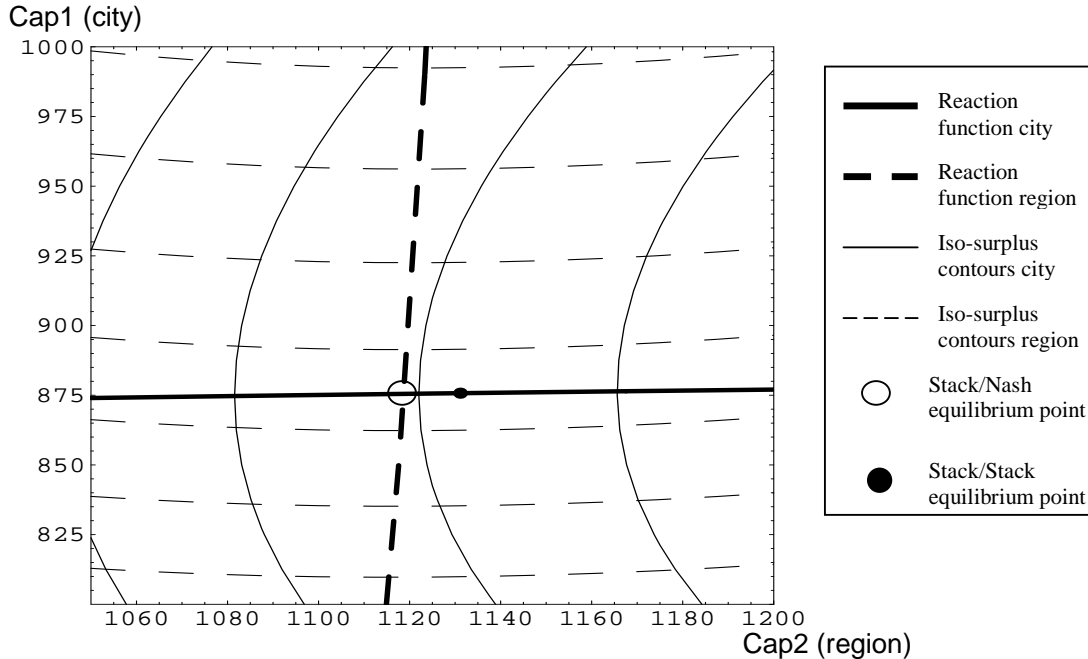

Figure 4. Reaction functions and iso-surplus contours for region leadership in the toll stage

The regional toll is now around three times as high as in the other scenarios. The city toll is somewhat lower. As a consequence, commuters are less inclined to travel, leading to less pressure on road space in the city. Capacity chosen by the city is therefore lower than in the other games, while that of the region is somewhat higher. The main reason why the region, when leading the price stage, increases its toll is that, by doing so, it can discourage its commuters from travelling to the city and, hence, 'losing' toll revenues to the other government. Again, the nature of the game in the capacity stage is less important for the eventual outcome, as soon as we know that the region leads the toll game: the Stack/Stack and Stack/Nash equilibrium points are also in Figure 4 relatively close. Note that, although toll leadership is attractive to the region, it is not beneficial for overall welfare, given the negative values for ω reported in Table 2.

Therefore, whereas price leadership of the city leads to only small changes compared with Nash price behaviour, the differences are bigger when leadership of the region is at stake. The explanation mirrors the one given earlier. The relatively high toll levels in the city,

and the direct losses for the region stemming from this, make it worthwhile for the region to adapt their own toll with the purpose of affecting the city's toll.

Overview

Figure 5 summarizes the differences in capacities (left panel) and tolls (right panel) for the different scenarios discussed above. While the assumed symmetry in the numerical example leads to identical toll levels and capacity in the base case and in the first-best equilibrium, the figure shows that all other scnarios produce asymmetric outcomes. Independent of the type of game, we find that in the game equilibria, capacities are below first-best levels and tolls are above first-best levels, while the region has a higher capacity and a lower toll than the city. The interpretation has been given above.
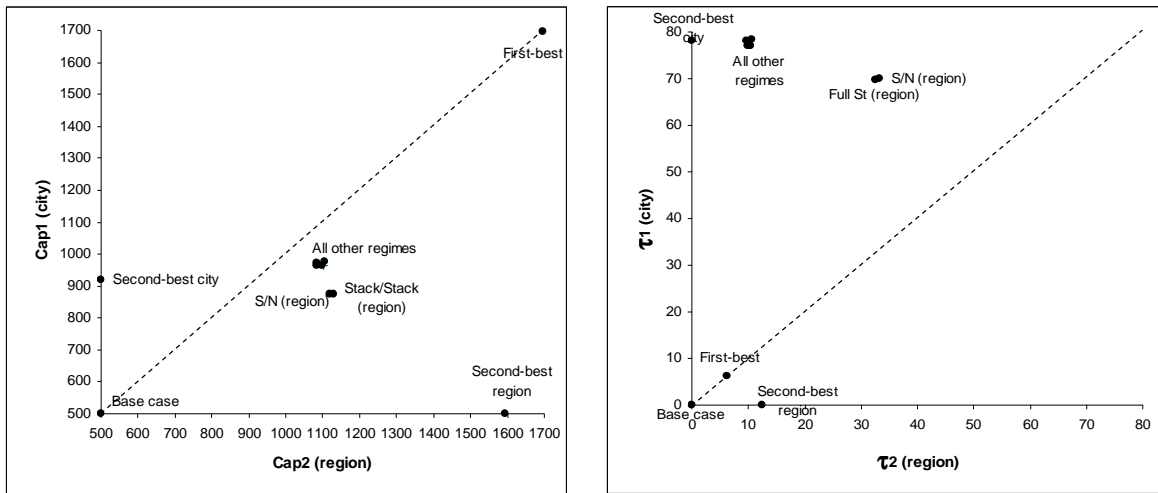

Figure 5. Equilibria capacity levels (left panel) and tolls (right panel) for the different scenarios

Next, the differences between the game theoretic equilibria are relatively small. More precisely, there are two clusters of equilibria: one cluster in which the region leads in the price stage, and one cluster that encompasses all other regimes. The differences within the clusters are so small that the dots in Figure 5 cannot even be distinguished graphically. But also the two clusters are relatively close, compared to first-best tolls and capacities. This suggests that the main issue is not which exact type of game is played between the two governments, but much more whether there is cooperation (leading to first-best) or competition between governments, where of secondary importance is the question who is leading in the price stage (if there is a leader). Leadership in the capacity stage is nearly without consequences in our numerical model.

5.   Sensitivity analysis

The above results are of course likely to change with the parameter values chosen. To assess the robustness of our results, we have analyzed two types of effects. First, we will look at the

effect of changing the demand elasticity. While undoubtedly affecting the absolute impacts of different schemes, it is also of interest to see whether it affects their relative performance. Next, the impact of changes in p_{cap} will be considered. This is a means of controlling for the relative importance of congestion management, as opposed to strategic considerations, in the setting of tolls: a low p_{cap} makes investments cheap, and tolls will have a smaller role in congestion management.

Varying demand elasticity

Figure 6 shows how relative efficiency ω changes when equilibrium demand elasticities change from relatively elastic (left-hand side) to relatively inelastic (right-hand side) (this was achieved by simultaneously changing the parameters *a* and *d* for all demand functions, such that the same base equilibrium is obtained for every elasticity). While technically showing 9 indicators ω, only four clusters can be distinguished visually in Figure 6. This reflects that the relative closeness within the two clusters of games remains. In the figure, we use 'Group 2' to indicate the two games where the region leads in the toll stage, and 'Group 1' for all other games. The similarities identified for our base parameterization are therefore robust, and do not depend on the assumed demand elasticity.
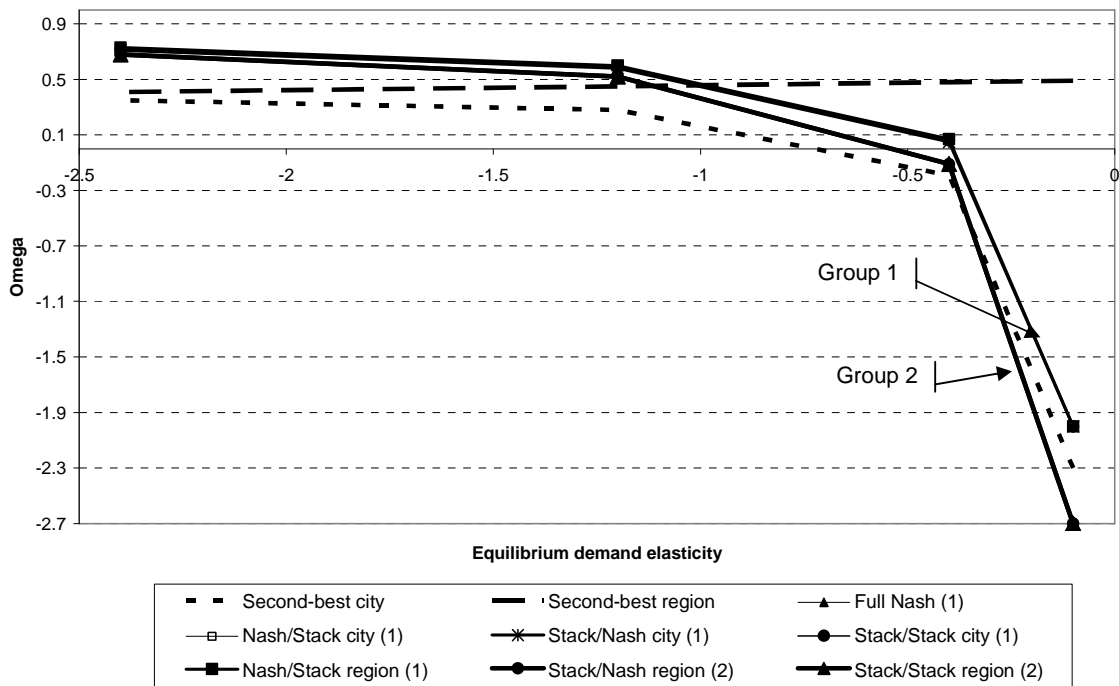
Figure 6. Indices of relative welfare improvement with varying demand elasticities

However, all schemes become less efficient when demand becomes less elastic. The 'Second-best city' scenario shows why: the city has a greater incentive to exploit its market power when demand becomes less elastic. The relative importance of the socially inefficient motive for tolling (revenue extraction from regional drivers by the city) rises compared with the

socially-efficient motive (congestion internalisation). Only the 'Second-best region', where the city's toll is fixed at zero, does not suffer from this inefficiency, as shown by the course of the associated curve.

Varying the price of road capacity

For the second sensitivity analysis, involving the unit price of road capacity p_{cap}, the relative welfare meausure ω unfortunately turned out to be a less instructive indicator. The reason is that, when the price of capacity increases, the absolute welfare gain from first-best regulation compared to the base equilibrium declines rapidly, and the patterns of ω are completely dominated by this change in the denominator. Therefore, we use a different relative welfare measure, namely aggregate social surplus as a fraction of its first-best level.

Figure 7 shows that, when the price of road capacity changes, this indicator remains rather constant for all scenarios. Note that the same two clusters of regimes emerge as in Figure 6. Although the price of capacity can be expected to influence the relative importance of congestion management, as opposed to strategic considerations, in the setting of tolls, this apparently again has little impact on the relative performance of the different scenarios.
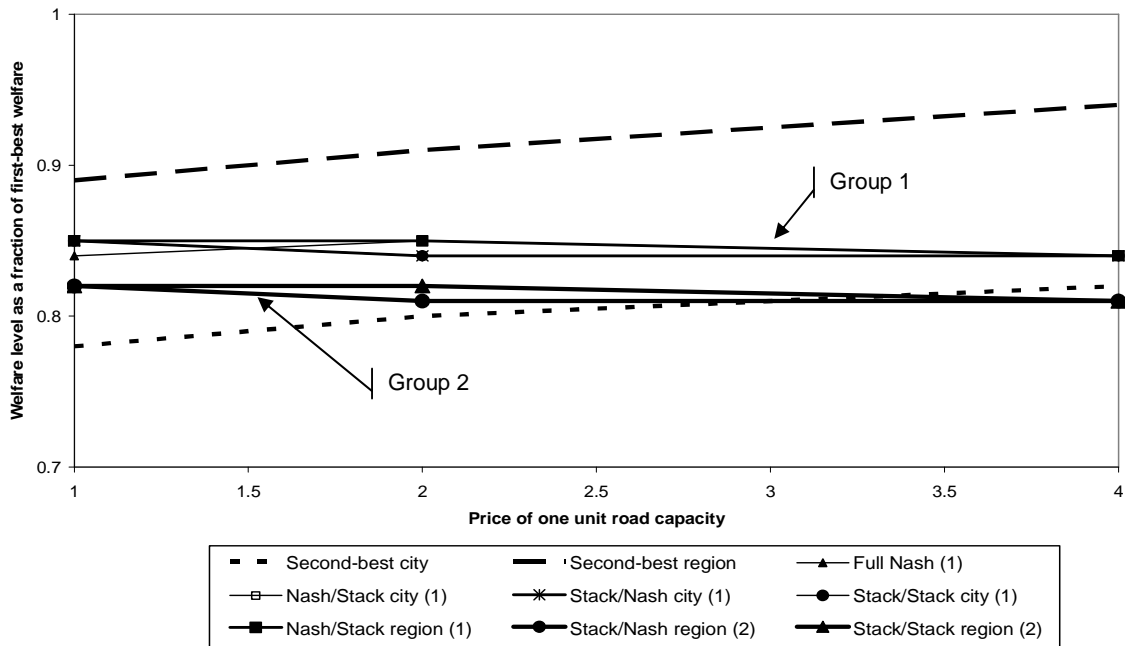
Figure 7. Welfare level for various situations as a fraction of first-best situation with varying prices of road capacity

6. Conclusion

This paper studied policy interactions between an urban and a regional government on a two-link serial road network, where both governments set capacity and toll on a road, in a two-stage game. The network was designed so as to capture the essential mechanisms and interactions that become relevant when an urban government sets a congestion toll that is to be paid not only by local inhabitants, but also by people from surrounding regions who may,

for example, commute to the city considered. We used a simulation model to investigate the welfare consequences of the various possible game-theoretical set-ups. Governmental competition was found to be rather harmful to aggregate social surplus, compared to first-best policies from a central government. The main determinant of social welfare is not which exact type of game is played between the two governments, but much more whether there is cooperation (leading to first-best outcomes) or competition between them. Only of secondary importance is the question who is leading in the price stage (if there is a leader).

Stackelberg leadership in one or both stages improves the welfare of the leader compared with a game with Nash properties. Leading in the toll stage is more important for a government than leading in the capacity game. Under Nash prices, a jurisdiction may in fact actually prefer the other government to lead in the capacity game, rather than leading themselves. But leading in the price stage may also be more important to one party than to the other. Factors that are of influence here are the Nash tolls set by the other government (the higher this toll, the more relevant it is to affect it), and the sensitivity of the other government's toll to one's own toll (the stronger this sensitivity, the more relevant it is to affect it).

Sensitivity analysis suggests that the relative performance for most game situations improves when demand becomes more elastic, and remain insensitive with respect to the unit cost of capacity expansions.

A dominant insight from the analysis is that a city, with a relatively large share of non-inhabitants on its road network, will have an incentive to set its tolls much too high. Are there ways to mitigate this distortion? One possible strategy would be to arrange the political institutions so as to constrain the city's monopoly power. For example, regional governments can be given an influence in decisions on toll setting that grows with the number of the region's inhabitants using the city's network. Or, the central government could require that a certain proportion of toll revenues collected be redistributed to the regional government. These and other possible arrangement offer important material for further research.

There are, of course, various other extensions worthy of further study. These include the expansion of the network beyond two links and allowing for multiple origin-destination pairs, and possibly also for trip diversion; the inclusion of more than two governments; the consideration of non-linear user and capacity cost and demand functions, *etc.* The fundamental forces at work seem to be so strong that such additions are unlikely to change the general pattern of overcharging that is found toll games like these. But it might be that that our conclusion on the relative closeness of the various games is less robust. This result seems certainly worthy of further exploration.

References

Arnott, Richard, André de Palma and Robin Lindsey (1993) "A structural model of peak-period congestion: A traffic bottleneck with elastic demand" *American Economic Review* **83** 161-179.

Arnott, Richard and An Yan (2000) "The two-mode problem: Second-best pricing and capacity" *Review of Urban and Regional Development Studies* **12** 170-199.

Braid, Ralph M. (1996) "Peak-load pricing of a transportation route with an unpriced substitute" *Journal of Urban Economics* **40** 179-197.

Brueckner, Jan K. (2002) "Airport congestion pricing when carriers have market power" *American Economic Review* **92** 1357-1375.

De Borger, Bruno, Fay Dunkerley and Stef Proost (2007) "Strategic investment and pricing decisions in a congested transport corridor" *Journal of Urban Economics* **62** 294-316.

De Borger, Bruno, Stef Proost and Kurt van Dender (2005) "Congestion and tax competition in a parallel network" *European Economic Review* **49** 2013-2040.

De Borger, Bruno and Kurt van Dender (2006) "Prices, capacities and service levels in a congestible Bertrand duopoly" *Journal of Urban Economics* **60** 264-283.

De Palma, André and Luc Leruth (1989) "Congestion and game in capacity: a duopoly analysis in the presence of network externalities" *Annales d'Economie et de Statistique* **15/16** 389-407.

De Palma, André and Robin Lindsey (2000) "Private roads: Competition under various ownership regimes" *Annals of Regional Science* **34** 13-35.

DeVany, Arthur S. and Thomas R. Saving (1980) "Competition and highway pricing for stochastic traffic" *Journal of Business* **53** 45-60.

Dowrick, Steve (1986) "Von Stackelberg and Cournot duopoly: choosing roles" *RAND Journal of Economics* **17** 251-260.

Economides, Nicholas and Steven C. Salop (1992) "Competition and integration among complements, and network market structure" *The Journal of Industrial Economics* **40** 105-123.

Engel, Eduardo, Ronald Fisher and Alexander Galetovic (2004) "Toll competition among congested roads" *Topics in Economic Analysis & Policy* **4** Article 4.

Levinson, David M. (2000) "Revenue choice on a serial network" *Journal of Transport Economics and Policy* **34** 69-98.

Lévy-Lambert, H. (1968) "Tarification des services à qualité variable: application aux péages de circulation" *Econometrica* **36** 564-574.

Liu, Louie Nan and John F. McDonald (1998) "Efficient congestion tolls in the presence of unpriced congestion: A peak and off-peak simulation model" *Journal of Urban Economics* **44** 352-366.

Liu, Louie Nan and John F. McDonald (1999) "Economic efficiency of second-best congestion pricing schemes in urban highway systems" *Transportation Research* **33B** 157-188.

Marchand, Maurice (1968) "A note on optimal tolls in an imperfect environment" *Econometrica* **36** 575-581.

McDonald, John F. (1995) "Urban highway congestion: an analysis of second-best tolls" *Transportation* **22** 353-369.

Mohring, Herbert and Mitchell Harwitz (1962) *Highway Benefits: An Analytical Framework* Evanston, IL: Northwestern University Press.

Pels, Eric and Erik T. Verhoef (2004) "The economics of airport congestion pricing" *Journal of Urban Economics* **55** 257-277.

Proost, Stef and Ahksaya Sen (2006) "Urban transport pricing reform with two levels of government: a case study of Brussels" *Transport Policy* **13** 127-139.

Small, Kenneth A. and Jia Yan (2001) "The value of 'value pricing' of roads: Second-best pricing and product differentiation" *Journal of Urban Economics* **49** 310-336.

Small, Kenneth A. and Erik T. Verhoef (2007) *The Economics of Urban Transportation* Routledge, London (forthcoming).

Stockholmsförsöket (2006) *Facts and Results from the Stockholm Trial – Final version – December 2006*, http://www.stockholmsforsoket.se/templates/page.aspx?id=12555 (accessed 2 May 2007).

Ubbels, Barry (2006) *Road Pricing: Effectiveness, Acceptance and Institutional Aspects*. PhD-thesis, VU University Amsterdam.

Van Dender, Kurt (2005) "Duopoly prices under congested access" *Journal of Regional Science* **45** 343-362.

Verhoef, Erik T. (2002) "Second-best congestion pricing in general static transportation networks with elastic demands" *Regional Science and Urban Economics* **32** 281-310.

Verhoef, Erik T. (2007) "Second-best road pricing through highway franchising" *Journal of Urban Economics* **62** 337-361.

Verhoef, Erik T., Peter Nijkamp and Piet Rietveld (1996) "Second-best congestion pricing: the case of an untolled alternative" *Journal of Urban Economics* **40** 279-302.

Verhoef, Erik T. and Jan Rouwendal (2004) "Pricing, capacity choice and financing in transportation networks" *Journal of Regional Science* **44** 405-435.

Verhoef, Erik T. and Kenneth A. Small (2004) "Product differentiation on roads: constrained congestion pricing with heterogeneous users" *Journal of Transport Economics and Policy* **38** 127-156.

Viton, Philip A. (1995) "Private roads" *Journal of Urban Economics* **37** 260-289.

Appendix. Analytical results for first-best and second-best toll and capacity choice

Although we did not succeed in expressing the equilibria for the various two-stage games discussed in the main text as transparent analytical solutions, we did obtain analytical solutions for three equilibria without strategic interactions. These are the optimal toll and investment rules for the first-best optimum, in which aggregate welfare W of equation (6c) is optimized by setting both tolls and both capacities optimally, and those for two second-best equilibria, namely those where a government sets its own toll and capacity so as to maximize local surplus, while the other government does not respond. Although these second-best tax and investment rules result from simultaneous optimization of tolls and capacities, and can therefore not be taken as representative for equilibria in sequential games, the rules are insightful in that they reveal the basic considerations faced by the governments, which will also be relevant in the games. This justifies their presentation in this appendix. Moreover, the toll rules for the two second-best cases are, in fact, relevant also for some of the games considered in the main text, as they reflect the toll rules that will be set in the second stage of a game by a regulator who takes the toll of the competing government as given. Because capacities, in the various games, are always set in the first stage and are therefore not set taking tolls as given, a similar statement cannot be made on the investment rules.

First-best tolls and capacities

The first-best toll and capacity can be found by solving the following Lagrangian, which has W of equation (6c) as its objective and the equilibrium conditions of (5) as constraints:

$$\Lambda = \int_0^{N_C} D_C(n)dn + \int_0^{N_{R1}} D_{R1}(n)dn + \int_0^{N_{R2}} D_{R2}(n)dn - (N_C + N_{R1}) \cdot c_1 - (N_{R1} + N_{R2}) \cdot c_2$$
$$- p_{cap} \cdot (cap_1 + cap_2) + \lambda_C \cdot (p_C - D_C) + \lambda_{R1} \cdot (p_{R1} - D_{R1}) + \lambda_{R2} \cdot (p_{R2} - D_{R2}) \qquad \text{(A.1)}$$

We determine the first-order conditions with respect to all N_j, both cap_i, both τ_i, and all λ_j, and find the following first-best toll rules (for a full derivation, see Ubbels, 2006):

$$\tau_1 = N_1 \cdot \frac{\partial c_1}{\partial N_1} \qquad \text{(A.2a)}$$

$$\tau_2 = N_2 \cdot \frac{\partial c_2}{\partial N_2} \qquad \text{(A.2b)}$$

and the following investment rules:

$$- N_1 \cdot \frac{\partial c_1}{\partial cap_1} = p_{cap} \qquad \text{(A.2c)}$$

$$- N_2 \cdot \frac{\partial c_2}{\partial cap_2} = p_{cap} \qquad \text{(A.2d)}$$

Both pairs of rules confirm expectations: the two first-best tolls are equal to marginal external costs, and the two investment rules indicate that road capacity be expanded up to the point where the marginal benefit of doing so (on the left-hand side) equals the marginal cost (on the right-hand side). Because $N \cdot \partial c(\cdot)/\partial N = -cap \cdot \partial c(\cdot)/\partial cap$, exact self-financing with tax rules (A.2ab) and investment rules (A.2cd) is easily established, confirming applicability of the conventional Mohring-Harwitz result.

Second-best tolls and capacities: city optimizes, region does not respond

The appropriate Lagrangian now combines objective W_C from (6a) with the equilibrium conditions of (5) as constraints:

$$\Lambda = \int_0^{N_C} D_C(n)dn - N_C \cdot c_1 + N_{R1} \cdot \tau_1 - p_{cap} \cdot cap_1$$
$$+ \lambda_C \cdot (p_C - D_C) + \lambda_{R1} \cdot (p_{R1} - D_{R1}) + \lambda_{R2} \cdot (p_{R2} - D_{R2}) \tag{A.3}$$

We determine the first-order conditions with respect to all N_j, cap_1, τ_1, and all λ_j, and find the following first-best toll rule (again, for a full derivation, see Ubbels, 2006):

$$\tau_1 = N_1 \cdot \frac{\partial c_1}{\partial N_1} + \frac{\dfrac{\partial D_C}{\partial N_C} \cdot \left(-\dfrac{\partial D_{R1}}{\partial N_{R1}} \cdot \dfrac{\partial D_{R2}}{\partial N_{R2}} + \dfrac{\partial c_2}{\partial N_2} \cdot \left(\dfrac{\partial D_{R1}}{\partial N_{R1}} + \dfrac{\partial D_{R2}}{\partial N_{R2}} \right) \right) \cdot N_{R1}}{\left(\dfrac{\partial D_C}{\partial N_C} + \dfrac{\partial D_{R1}}{\partial N_{R1}} \right) \cdot \dfrac{\partial D_{R2}}{\partial N_{R2}} - \dfrac{\partial c_2}{\partial N_2} \cdot \left(\dfrac{\partial D_C}{\partial N_C} + \dfrac{\partial D_{R1}}{\partial N_{R1}} + \dfrac{\partial D_{R2}}{\partial N_{R2}} \right)} \tag{A.4a}$$

and the following capacity rule:

$$-N_1 \cdot \frac{\partial c_1}{\partial cap_1} = p_{cap} \tag{A.4b}$$

To start with the latter, the optimal investment rule is the same as it is in the first-best case; compare equation (A.2c). Note that the equilibrium capacity level will nevertheless normally be different, as the left-hand side will be evaluated for a different level of use N_1. The fact that the investment rule is the same may be surprising at first sight, but is consistent with earlier findings that show that the investment rule is the same for welfare-maximizing and profit-maximizing roads, as well as for intermediate cases such as Ramsey-pricing (Small and Verhoef, 2007). Because the city government would like to maximize surplus for the local road users (*C*) and profits from the regional users (*R1*), the result obtained makes perfect sense.

The toll rule of (A.4a) eludes easy understanding. The first term is still straightforward: it repeats the first-best expression and is equal to the marginal external congestion cost on the city's road. But the city will raise the toll beyond this level. The toll expression is the result of the city government compromising between two toll rules: the marginal external costs, which would be optimal for the own citizens when driving in isolation; and the marginal external costs plus a demand-related monopolistic mark-up, which

would be optimal if only regional drivers used the city's infrastructure. (Note that the second term is positive as long as demands are downward sloping.) The second term in (A6b) is so complex because the demand sensitivity group $R1$ depends, in part, also on cost and demand elasticities in the region, and therewith upon the slopes of the cost function for road 2 and the demand function for group $R2$.

Second-best tolls and capacities: region optimizes, city does not respond

The final Lagrangian combines objective W_R from (6b) with the equilibrium conditions of (5) as constraints:

$$
\begin{aligned}
\Lambda = & \int_0^{N_{R1}} D_{R1}(n)dn + \int_0^{N_{R2}} D_{R2}(n)dn - N_{R1} \cdot c_1 - \left(N_{R1} + N_{R2}\right) \cdot c_2 - N_{R1} \cdot \tau_1 - p_{cap} \cdot cap_2 \\
& + \lambda_C \cdot \left(p_C - D_C\right) + \lambda_{R1} \cdot \left(p_{R1} - D_{R1}\right) + \lambda_{R2} \cdot \left(p_{R2} - D_{R2}\right)
\end{aligned}
\tag{A.5}
$$

We now determine the first-order conditions with respect to all N_j, cap_2, τ_2, and all λ_j, and find the following first-best toll rule (also this derivation is given in Ubbels, 2006):

$$
\tau_2 = N_2 \cdot \frac{\partial c_2}{\partial N_{R1}} + \frac{-N_{R1} \cdot \dfrac{\partial c_1}{\partial N_1} \cdot \dfrac{\partial D_C}{\partial N_C} \cdot \dfrac{\partial D_{R2}}{\partial N_{R2}}}{-\dfrac{\partial D_C}{\partial N_C} \cdot \left(\dfrac{\partial D_{R1}}{\partial N_{R1}} + \dfrac{\partial D_{R2}}{\partial N_{R2}}\right) + \dfrac{\partial c_1}{\partial N_1} \cdot \left(\dfrac{\partial D_C}{\partial N_C} + \dfrac{\partial D_{R1}}{\partial N_{R1}} + \dfrac{\partial D_{R2}}{\partial N_{R2}}\right)}
\tag{A.6a}
$$

and the by now familiar capacity rule:

$$
-N_2 \cdot \frac{\partial c_2}{\partial cap_2} = p_{cap}
\tag{A.6b}
$$

The re-occurrence of the first-best investment rule should be no surprise. Again, understanding the toll rule of (A.6a) is more challenging. The first term is again equal to the marginal external congestion costs on the tolled road. The second term is positive. It reflects the region's attempt to also internalise the congestion externality that regional commuters impose upon one another on the city's road, link 1, as given by the first two terms in the numerator. The correction term captures substitution effects that will make this attempt less effective than it would be if link 1 were used by regional drivers alone.

Scenario		N_C	N_{R1}	N_{R2}	N_1	N_2	cap_1	cap_2	W	W_C	W_R	τ_1	τ_2	ω
Base equilibrium		0.93	0.93	0.93	0.93	0.93	0.29	0.29	0.83	0.76	0.85	0	0	0
First-best		1 (179)	1 (358)	1 (179)	1 (537)	1 (537)	1 (1697)	1 (1697)	1 (57621)	1 (9603)	1 (48018)	1 (6.32)	1 (6.32)	1
Second-best	City optimises	0.33	0.65	0.97	0.54	0.76	0.54	0.29	0.80	2.29	0.50	12.36	0	-0.19
	Region optimises	0.93	0.94	0.94	0.94	0.94	0.29	0.94	0.91	0.76	0.94	0	1.96	0.48
Full Nash		0.34	0.65	0.96	0.55	0.75	0.57	0.64	0.85	2.28	0.56	12.18	1.55	0.07
City leader	Nash prices, Stackelberg capacities	0.34	0.65	0.96	0.55	0.75	0.57	0.64	0.84	2.28	0.56	12.18	1.53	0.07
	Stackelberg prices, Nash capacities	0.33	0.65	0.96	0.54	0.75	0.57	0.65	0.84	2.28	0.55	12.34	1.52	0.05
	Full Stackelberg	0.33	0.65	0.96	0.54	0.75	0.57	0.65	0.84	2.28	0.55	12.32	1.52	0.06
Region leader	Nash prices, Stackelberg capacities	0.34	0.65	0.96	0.55	0.75	0.57	0.65	0.85	2.28	0.56	12.18	1.53	0.07
	Stackelberg prices, Nash capacities	0.41	0.58	0.75	0.52	0.64	0.52	0.66	0.81	2.03	0.57	11.04	5.19	-0.11
	Full Stackelberg	0.41	0.58	0.75	0.52	0.64	0.52	0.67	0.82	2.03	0.57	11.04	5.19	-0.11

Note: All results are relative to the first-best outcomes, numerical values in first-best are given between brackets.

Table 2. Numerical results (base-case parameters)