TI 2006-036/3
Tinbergen Institute Discussion Paper

Governmental Competition in Road Charging and Capacity Choice

Barry Ubbels
*Erik Verhoef**

Department of Spatial Economics, Vrije Universiteit Amsterdam.

* Tinbergen Institute.

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Governmental competition in road charging and capacity choice[1]

Barry Ubbels
Erik Verhoef
Department of Spatial Economics
Free University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
Email: bubbels@feweb.vu.nl

Abstract
In this study we have analysed policy interactions between an urban and a regional government which have different objectives (welfare of its own citizens) and two policy instruments (toll and capacity) available. Using a simulation model, we investigated the welfare consequences of the various regimes that result when both governments compete, and take sequential decisions on prices and capacities. We find that competition between governments may not be very beneficial to overall welfare in society compared with one central government. It appears that the tendency of tax exporting is very strong in this setting where commuters have to pay road tolls set by the city government. The main issue is not which exact type of game is played between the two actors, 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). Sensitivity analysis suggests that the performance for most game situations improves when demand becomes more elastic. When the price of road investment changes, the performance relative to the optimal situation remains more or less equal for all cases.

Keywords: competition, governments, tolls, capacities

1. Introduction

One of the best-known policy prescriptions from economic theory is that, to reach an efficient allocation of resources, prices should be set equal to marginal costs. However, this policy prescription holds exactly only under first-best conditions, and even then only when distributional considerations are ignored, or lump-sum redistribution is possible. In reality, first-best conditions are never completely met, i.e. there are always additional constraints, apart from

[1] This research was carried out within the NWO/Connekt VEV project on “A Multidisciplinay Study of Pricing Policies in Transport - An Economic Perspective”; nr. 014-34-351. Financial support is gratefully acknowledged.

the inevitable technological ones. These additional constraints typically make it necessary to amend the simple rule that prices should be set equal to marginal costs.

This paper focuses on one particular reason for the presence of second-best constraints: the fact that authorities of different jurisdictions pursue their own objectives and use their own policy instruments. Such interactions between different governments (e.g. federal level versus regions or cities) introduce the potential for vertical or horizontal tax competition[2]. The chosen instrument values may be optimal for the population represented by the decision maker (for instance, the inhabitants of a region), but may at the same time be suboptimal when also considering the interest of people elsewhere (in different regions). In such cases, the values of the policy variables determined by the one authority are beyond direct control of other authorities, implying a second-best constraint.

It is obvious that different types of institutional relationships may be relevant in the field of transport policy competition. This paper addresses competition in road capacity and road tolling between a city and a region. Many transport problems have a strong local or urban dimension. This may explain the increasing interest in giving sufficient possibilities for regional/local authorities to deal with these issues: at least they know local circumstances best. So far, regions or cities usually have had some form of regulation within their control (e.g. parking), but their possibilities to implement other forms of transport (pricing) policy are generally limited. Nevertheless, it seems that initiatives with greater policy responsibilities at a lower level may be gaining ground, and can be very successful in addressing local transport problems. Good examples are value pricing projects in the US, and the London congestion charging scheme. One difficulty that often arises with such schemes is the involvement of multiple levels of government, which typically can differ in their objectives and powers. For example, local authorities may assign lower weights to the welfare of non-residents than do national or regional governments. The cost of funds may differ, because of differences in the types and levels of taxes that governments rely on for revenue. And control of various pricing and non-pricing policies may be divided up between levels of governments. All these differences create the potential for conflict, as well as strategic behaviour in the form of tax exporting, tax competition, etc.

In this paper we study the welfare consequences of strategic interaction between an urban government and a region, both having two different policy instruments: tolls, and investment in road capacity. We will use a small two-link network that enables us to study these choices in the simplest possible network configuration. The paper starts with a discussion of the previous literature relevant in the context of our analysis. Section 3 introduces the model, and presents the simulation results for different situations (ranging from the first-best optimum as a useful benchmark to various non-cooperative game situations). This section also presents some sensitivity analyses. Section 4 concludes.

2. The literature

In recent years, considerable theoretical research has been undertaken into the effectiveness of urban transport pricing. It has been shown that traditional forms of pricing policies in road transport, such as vehicle and fuel taxes, will often be too crude to achieve anything near the welfare gains that a more targeted and refined system of road pricing would yield. There is a

[2] Vertical tax competition refers to a situation in which governments at different hierarchical levels compete, while horizontal tax competition results when governments at the same level compete.

well-developed transport economic theory to determine appropriate prices, including justifications for subsidies in public transport, and the introduction of point-of-use pricing to encourage greater efficiency. Marginal cost pricing is the benchmark that achieves maximum social welfare under otherwise first-best conditions.

But the literature on optimal pricing of road use in the presence of congestion has been extended in various directions, including optimal and non-optimal tolling for different network configurations. A common constraint encountered in practical situations is that tolls are only implemented on single lanes or roads, with travel on alternative routes being unpriced. An example is the SR 91 in California (Liu and McDonald, 1998). Imposition of 'quasi first-best pricing' on the toll road (simply ignoring spillovers to the unpriced parallel road) is then not welfare maximising (Lévy-Lambert, 1968; Marchand, 1968); governments should instead apply second-best pricing. Welfare gains from second-best tolls with unpriced substitutes are generally found to be rather low (see, e.g., Liu and McDonald, 1999), but become higher when allowing for a heterogeneous population (Verhoef and Small, 2004) or for the dynamics of departure time adjustments (e.g. Braid, 1996). De Palma and Lindsey (2000) focused on allocative efficiency of private toll roads versus free access and public toll road pricing on a similar network, allowing for dynamics of peak congestion. Their study suggests that two competing private roads can yield most of the potential efficiency gains from first-best pricing if neither road has a dominant fraction of total capacity. A single private road competing with a free-access road tends to be most efficient if the two roads have approximately equal capacities and if the private road does not suffer a significant travel time disadvantage.

These type of networks have also been used to study the impact of different ownership regimes. Verhoef et al. (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. It is shown that revenue maximising tolling on two routes may actually 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. More recently, road capacity too has been endogenised in this type of analyses. Yang and Meng (2000) looked at the selection of the capacity and toll charge of a new road (in a build-operate-transfer (BOT) framework) and the evaluation of the relevant benefits to the private investor, the road users, and society as a whole. Ubbels and Verhoef (2004) and Verhoef (2005) focused on the auctioning of concessions to privately operate the road, and considered various criteria ("indicators") that a government may use to realise more satisfying bids (in terms of tolls and road capacity) from interested firms. It was shown that the design of the auction has a considerable impact on the resulting welfare gains. De Borger and Van Dender (2005) studied the duopolistic interaction between facilities subject to congestion (such as airports or roads) that supply perfect substitutes and make sequential decisions on capacities and prices. The situation of competition (Nash equilibria) is compared with the results of a monopoly and first-best outcomes. They show that price competition between duopolists is beneficial for consumers, but introducing capacity competition is harmful. The duopolist situation offers lower service quality (defined as the inverse of time costs) than the monopolist, who does provide the socially-optimal quality level.

Tax competition between different levels of government may also occur (for a review, see De Borger and Proost (2004)). Optimal transport pricing with multiple governments can become complicated, because many vertical and horizontal fiscal externalities occur simultaneously, with spillovers of congestion and environmental externalities (see Markusen et al., 1995; Wilson, 1999). First, the importance of international transport, including pure transit flows, in some countries implies that a substantial share of locally-generated externalities is due to foreigners. To the extent that these flows can be taxed, this may induce tax-exporting behaviour. Second, international transport implies that the tax base of transport services is to a large extent mobile between countries. This may be the basis for inefficient tax competition. Third, vertical externalities arise because different levels of government (e.g. city and national governments) may be interested in taxing the same base, for example, in an effort to tackle pollution or congestion. Alternatively, they may actually use different instruments to deal with the same problem. Fourth, some transport externalities generate international spillovers (e.g. global warming, acid rain, etc.), which should be appropriately accounted for. Finally, apart from tax externalities, transport generates various expenditure externalities: investment in infrastructure has benefit spillovers to foreigners, local road investments affect federal fuel tax revenues, etc. It appears that literature on these kinds of tax competition issues in transport (we restrict ourselves to transport and do not include the public finance literature) is fairly limited (De Borger and Proost, 2004). But some studies addressing tax (in terms of road pricing) competition in transport have been found and will be discussed in the following subsection.

2.1 Horizontal linkages

Horizontal tax competition refers to a situation where governments at the same hierarchical level compete; for instance, Austria and Switzerland compete for toll revenues from traffic between Germany and Italy. This example, in which road users have a choice of routes and where both routes are priced by a different government, is called parallel tax competition. Only a few models look explicitly at parallel tax competition between two governments.

De Borger et al. (2005) analyse tax competition between countries that each maximise the surplus of local users plus tax revenues in controlling local traffic and through-traffic (similar to the situation of Austria and Switzerland). Three different pricing systems were considered: one with toll discrimination between local traffic and transit traffic; with only uniform tolls; and one system with tolls on local drivers only. The results suggest that the welfare effects of introducing transit tolls are large, but that differentiation of tolls between the different types of traffic as compared with uniform tolling does not yield large welfare differences. Countries may decide to cooperate on toll setting, but this leads only to small welfare gains in comparison with non-cooperative transit tolling (Nash equilibrium with uniform tolls).

Another study compares the outcomes of tax policies of Belgium and the rest of Europe in a model with both domestic and international freight transport flows and domestic passenger transport (De Borger et al. (2004)). They used a numerical optimisation model to determine optimal pricing policies (consistent with EU legislation: no discrimination between domestic and international freight transport) for four situations: a reference situation reflecting unchanged policies; a federal optimum; a local optimum for one individual country; and the Nash equilibrium solution. The local optimum for Belgium and the Nash equilibrium demonstrated the inefficiency of tax exporting behaviour: taxes on freight flows were found to be substantially higher than in the federal optimum, depending on the share of international flows in countries. The local optimum for Belgium (maximise welfare of Belgian residents) involved higher welfare than both the federal optimum and the Nash equilibrium outcome. The former was due to tax exporting (taxes on freight transport largely exceed marginal external costs), the latter to the absence of reactions by the rest of Europe.

A second type of horizontal tax competition can be referred to as serial tax competition. Traffic using a route that sequentially runs through the territory of different governments can be taxed by each of the governments. These countries may apply individual tolling instruments on their part of the network, with potentially substantial welfare losses as a result. Despite its importance, serial tax competition in transport has not been given much attention in the literature (De Borger and Proost, 2004). One exception is the empirical work of Levinson (2001). He examines the question why some US States impose tolls while others rely more heavily on fuel and other taxes. The share of highway revenues from tolls is explained by the share of non-residential workers, policies of neighbouring States, historical factors, and the population. The analysis confirms theory in the sense that jurisdictions are more likely to opt for toll financing (instead of fuel or other taxes) when the share of non-residential drivers is large. Obviously, tolls become more attractive because they allow price discrimination and tax exporting. It is also suggests that decentralising financial responsibilities and creating smaller jurisdictions (the greater the share of non-local traffic) increases the likelihood of tolling.

2.2 Vertical linkages

Vertical tax competition refers to competition between different hierarchical levels of government. This issue too is relevant in practice, but again it has hardly been addressed in the literature. Most transport users are not only taxed or subsidised via various instruments, but typically these different taxes are set by different levels of government. Road users, for instance, may face fuel taxes determined by national governments and parking fees or cordon tolls implemented by cities or regions. The different responsibilities for transport policy instruments induces a number of complicated interactions between governments because of overlap of tax bases, differences in objectives between governments, spillovers of externalities, etc. A number of reasons can be identified to explain why the resulting tax schemes will generally be suboptimal (De Borger and Proost, 2004):

- Overlapping tax bases may create fiscal externalities: an increase in national fuel taxation reduces demand for transport, including local traffic, and therefore affects revenues from the local government (e.g. toll revenues, public transport revenues). These side effects are often ignored by the national government in setting fuel taxes, yielding too high national taxes;
- Tax exporting by the city government: a city will take more care of its own residents than of commuters. This may lead to excessively high congestion charges for commuters from an overall welfare viewpoint;
- Externality spillovers: local authorities only care about externalities imposed on local residents; this induces them to set local taxes too low;
- The use of imperfect and different instruments by different governments: local authorities only have a few instruments available, such as cordon tolls and parking charges, to control externalities.

The welfare effects of a combination of different pricing instruments (in many cases to internalise external costs) have been studied extensively in the literature. For instance, Calthrop et al. (2000) used a numerical simulation model of an urban transport market to examine the efficiency gains from various parking policies with and without a simple cordon system. They show that the pricing of parking and road use needs to be simultaneously determined, and that second-best pricing of all parking spaces produces higher welfare gains than the use of a single-ring cordon scheme, though marginally lower than the combination of a cordon charge with resource-cost pricing of parking places.

These studies show the efficiency effects of various pricing instruments available to governments, but do not address the interaction effects when different levels of government are responsible for different pricing instruments that affect users from different jurisdictions. One study that does analyse a policy game between two different levels of governments with two different pricing policies in transport is a case study described in the EU project MC-ICAM (MC-ICAM, 2004). This analysis uses a quasi-general equilibrium model (TRENEN) with two policy instruments: a parking fee set by the city to be paid by all drivers in the city, and a cordon toll set by the region only to be paid by the commuters. Toll revenues are redistributed by the region to commuters only, while parking fee revenues are redistributed to urban citizens. The city maximises the welfare of the inhabitants only, whereas the region maximises a weighted sum of the welfare of its urban citizens and its commuters (where the weights correspond to their relative numbers in the total population). In this setting three different solutions have been analysed:

- A centralised solution, in which the region chooses both pricing instruments simultaneously and maximises the aggregate welfare of commuters and urban citizens;
- A non-cooperative Nash equilibrium, in which each government takes the other’s choice as given;
- A Stackelberg equilibrium, in which the region acts as leader and chooses its policy instrument (cordon toll) first.

The results show that an efficiency loss occurs in the policy game situations because the city overcharges for parking in order to export taxes to commuters who are outside its jurisdiction (but not unlimited because city inhabitants also pay the parking fee). The region responds by setting a toll lower than the first-best level, because it does not want to discourage commuting too much. The non-cooperative Nash and Stackelberg equilibria achieve most of the welfare improvements (respectively, 89% and 92%) when compared with the (optimal) centralised solution. The reasons for these fairly high welfare levels, given by the authors, are that both governments are assumed to maximise welfare rather than revenues, and that parking fees and cordon tolls are substitutes. Further improvements can be achieved by changing the sharing rules for tax revenue in favour of the city inhabitants, although this comes at a cost of greater inequity to commuters.

2.3 Our approach

Although the literature on tax competition in transport is still relatively small, a number of cases have already been considered, and it is important to emphasise how our paper differs from earlier work. We will deal with competition between a region and a city that differ in their objectives and instruments, similar to the case study in MC-ICAM (2004). However, we bring road capacity into the analysis as a policy variable that is available to both government levels besides the road charges that they both can set. Parking fees, in contrast, are not included because they would be indistinguishable from the road tax on the city’s infrastructure in our simple network.

The set-up appears to be relevant for current transport policies. Many urban areas often experience congestion problems, and it is no longer considered straightforward (in so far as it has been) that a national government should decide how to deal with these. Decentralisation has given more powers to lower levels of government, which have more knowledge of the local situation. When local authorities decide on their policy approach, a realistic set of instruments would include capacity enlargement besides road pricing. We therefore study the interaction between a region and a city that are free to choose the capacity of the roads and road tolls, and focus on the welfare consequences of their behaviour under various game-theoretic settings.

3. A modelling framework to study government competition

We consider a simple serial type of road network with two links, one of which is controlled by the regional authority and the other by the city. This road network is used by three different types of drivers. *Commuters* live in the region and work in the city, and therefore use both the regional road and the infrastructure of the city. *Regional drivers* live in the region and never enter the city, and therefore only use the regional road. And *city drivers* live in the city and never use the regional road. Reverse commuting (living in the city, travelling to the region) is therefore ignored. Both roads are subject to congestion.

Commuters therefore pay both tolls when levied, while regional and city drivers only pay the toll set by their own government. The two levels of government are assumed to receive the full revenues of the tax instruments they control, but they also pay for their own (road) capacity. Each government maximises the welfare of its inhabitants.

Various settings of interest can be identified. A first one involves the global optimum, where both tolls and both capacities are set so as to maximise the aggregate surplus (in the region and city together). Next, without policy coordination, two equilibrium concepts may, in principle, be relevant: a Nash equilibrium, in which both actors take the other's choices as given, and a Stackelberg equilibrium, where a leader takes the follower's responses into account, while the follower takes the leader's action as given.

Our analysis of toll and capacity decisions considers these as the results from a sequential, two-stage, game in capacities and prices. In the first stage both governments make capacity choices, and in the second stage toll competition takes place. This appears realistic because also in practice tolls can be varied in the short run, after capacities have been set. The two-stage nature of the game gives rise to various combinations of Nash and Stackelberg games in the various stages, theoretically even including reversed leadership in the two stages. We will consider a few of these; namely, those combinations that appear more realistic.

Since we are interested in the relative efficiency of these non-cooperative situations, we compare the outcomes with the first-best optimal solution in this setting, i.e., the global optimum mentioned before. To express equilibrium welfare levels in relative terms, we will, of course, also need to specify a reference equilibrium in our numerical exercises.

3.1 Model formulation

We consider what is probably the simplest possible set-up to study the problem just described. The network consists of two serial links, connecting the regional (suburban) area with the centre of a city. It is assumed that the pricing and capacity choice of each link is the responsibility of a different government. In the base-case equilibrium, both links have equal capacities and zero tolls. Commuters from the region (group R_1) are assumed to enter the city, and consequently use both links. Regional drivers (R_2) and city drivers (C) only use the link in their own jurisdiction.

Both governments are assumed to maximise social surplus for their inhabitants, defined as total (Marshallian) benefits minus total user costs (including time costs), minus the capacity costs.
The model contains three different demand functions (one for each group), and two average-user cost functions: c_1 (N_C , N_{R1}; cap_1) for city road 1, and c_2 (N_{R1}, N_{R2}; cap_2) for regional road 2. At any interior equilibrium, average costs for the full trip, plus tolls, should be equal to marginal benefits for each type of user[3]. While Appendix B will derive first-order conditions for general demand and cost functions, of unspecified functional forms, we will be using linear functions in the numerical model. The inverse demand functions D_j can then be written as:

$$D_j = d_j - a_j * N_j \qquad j = C,R_1,R_2 \qquad (1)$$

where d_j and a_j are parameters, and N_j gives traffic flow for group j.
Next, for both links i, the average social cost (c_i) consists of a free-flow cost component k_i, and a congestion cost component, which is assumed to be proportional to total road usage on that link (N_i, with $N_1 = N_C + N_{R1}$ and $N_2 = N_{R1} + N_{R2}$), and inversely related to its capacity (cap_i):

$$c_i = k_i + b_i * N_i/cap_i \qquad i = 1,2 \qquad (2)$$

where k_i and b_i are parameters.
The generalised price p_i for a link adds the toll t_i to the average cost:

$$p_i = c_i + t_i \qquad i = 1,2 \qquad (3)$$

The generalised prices p_j for the three groups are then as follows:

$$\begin{aligned} p_{R1} &= p_1 + p_2 \\ p_{R2} &= p_2 \\ p_C &= p_1 \end{aligned} \qquad (4)$$

Next, operational profits (π) for the city or the region are defined as the difference between revenues from the tolls and the net costs for road provision (capacity costs). We thus neglect other costs for the government, such as maintenance and costs involved with toll collection (or assume that these are proportional to capacity). The capacity costs are proportional to capacity (cap_i) via a fixed price per capacity unit (p_{cap}, equal for both links)[4], while the revenues are the product of the toll and traffic demand. Profit thus becomes:

$$\pi_i = t_i * N_i - cap_i * p_{cap} \qquad i = 1,2 \qquad (5)$$

Total social surplus (W, the sum of surpluses in the city and the region) is our measure of welfare, and is equal to the 'variable' social surplus (the benefits B as given by the relevant area under the demand curve, minus total user costs), minus the total capacity costs:

[3] Corner solutions, where the demand of at least one of the groups is reduced to zero, will not be considered.
[4] Note that the average cost function implied by (2) and (3) and the constancy of the unit price for capacity 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).

$$W = B - N_C*c_1 - N_{R2}*c_2 - N_{R1}*(c_1+c_2) - (cap_1+ cap_2)* p_{cap} \quad (6)$$

W can be decomposed into city and regional welfare as follows:

$$W_C = B_C - N_C*c_1 + N_{R1}*t_1 - cap_1* p_{cap} \quad (7a)$$
$$W_R = B_{R1} + B_{R2} - N_{R1}*(c_1+c_2) - N_{R2}*c_2 - N_{R1}*t_1 - cap_2* p_{cap} \quad (7b)$$

Note that tolls paid by commuters (residents of the region) to make use of the city link are a transfer between jurisdictions and thus increase the welfare of the city and reduce the welfare levels of the region. All other toll payments cancel out in local welfare functions, because they constitute a transfer from local residents to the local authorities. All parameters are non-negative, and we will only consider interior equilibria, where both links are at least marginally used, and all OD pairs are at least marginally active.

The (three) user equilibrium conditions are then:

$$D_j = p_j. \qquad j = C,R_1,R_2 \quad (8)$$

Appendix B derives analytical expressions for toll rules and investment rules for three cases: namely, the first-best case; the case where the city optimises, taking the regional toll and capacity as given; and the case where the region optimises, taking the city's toll and capacity as given. In this section, we will proceed by presenting comparative static results for the different games under consideration. These reflect the impacts of toll and investment rules upon possible equilibria, but are of course less general than our analytical results because they pertain to an assumed set of demand and cost functions. To compensate for the latter disadvantage, sensitivity analyses will be provided in Section 3.5.

The insights from Appendix B can be summarised 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 *rule* does not change: given that 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 capacity *level* will, of course, be different from the first-best level. Both governments would have an incentive to set the toll above the 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 R_1 drivers impose upon one another on the city's road, link 1. We will now turn to the simulation results to see how these forces affect the eventual equilibrium.

3.2 Numerical Example

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; β = 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 of both city drivers and regional drivers, twice as many commuters, and equilibrium travel costs twice the 'free-flow' levels. The equality between equilibrium demands for groups R_2 and C was motivated by the desire to have the example symmetric in as many aspects as possible. Appendix C shows the detailed numerical results for the various scenarios under study. These parameter values were otherwise not motivated by any desire to represent a realistic situation. Before turning to the game equilibrium situations, we will first discuss a few benchmark situations.

Global first-best optimum

The first-best social optimum (maximum welfare) in this network involves an optimisation of both the capacity of the links and all tolls (first-best pricing requires tolling on all links of the network). The optimal road price, to be set by both governments, equals 6.32 (see Appendix C), bridging the gap between marginal private costs of 26.32 and marginal social costs of 32.64. In Appendix B we derive the first-best expression for both tolls. The first-best toll equals the marginal external (congestion) cost for both roads. Most welfare gains are realised by the region, a situation which can be explained by the large number of commuters (experiencing congestion) who live in the region (the welfare gains per initial traveller are higher for the city).

The total capacity of both roads together is about 3.4 times as high as the initial capacity. This is, of course, rather extreme. It is a direct consequence of the level of p_{cap} that we have chosen, relative to the other parameters, 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. The zero profit result for the provision of both links confirms the Mohring and Harwitz (1962) result of optimal investment. The revenues from an optimal toll will just cover the costs of the facility provider as long as there are no economies or diseconomies of scale in facility capacity and the facility provider is investing optimally (see Appendix B for a sketch of a proof).

Second-best optima: one government optimises

For the evaluation of the various situations to be considered, it is useful to know the (welfare) properties of the second-best solutions where only one government charges a toll for its link and optimises its capacity, with the purpose of maximising local welfare only, while the other region does not respond and sticks to the initial choices. There are two such second-best situations. Appendix B shows the first-order expressions for tolls and capacity in both cases.

The second-best situation in which the *city* sets the toll and capacity so as to achieve maximum welfare for its residents (N_C) leads to a considerably higher welfare level for the city compared with the first-best situation. In contrast, the welfare of the region is decreased considerably. Total welfare is decreased when compared with the base case equilibrium, leading to a negative value for the relative efficiency indicator ω [5]. The toll is equal to 78.11, more than 12 times as high as the first-best toll. The city has an incentive to set such a high toll because the toll revenues extracted from the regional commuters (N_{R1}) increase accordingly. This is of course constrained by the fact that the fees are also paid by the city residents, but note that, initially, only 1/3 of intra-city traffic concerns city residents. Moreover, toll payments by city residents, although distorting prices in the city, in themselves only constitute transfers within the city. The number of commuters decreases significantly, as well as the welfare levels in the region. The city link is also used less intensively by city drivers compared with the base-case situation, because of the

[5] 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.

excessive toll. The first-order expression of the toll set by the city under these conditions consists of the first-best toll with a positive mark-up (see Appendix B). Capacity chosen by the city is considerably lower compared with optimal pricing, but the ratio capacity/demand has not changed, because the first-best rule for optimising capacity still applies.
The other second-best situation is when the *region* optimises under the assumption that the city does not change capacity and toll. It appears that the demand for transport is not very different from the base-case and first-best situations. Appendix B shows that the region will apply a capacity rule similar to the first-best rule, which means that the capacity of link 2 will also not deviate much from optimal investment. This, and a toll that internalises congestion for groups R_1 and R_2 (the former imperfectly), will increase the welfare of the region. Since the welfare of the city remains almost constant (relative to the base-case situation), overall welfare will also increase (ω = 0.48). The higher relative efficiency 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, it internalises part of the congestion externality in the city: namely, insofar as it is imposed by regional residents on themselves.

3.3 Non-cooperative game equilibria

We will now consider the numerical outcomes when governments compete in order to achieve their own objectives when choosing tolls and road capacities. Different situations will be considered, corresponding with different types of game theoretic settings. Two different plausible equilibrium concepts of a non-cooperative game can be distinguished in such a setting: a Nash equilibrium and a Stackelberg equilibrium. In a Nash equilibrium, each government takes the other's choice as given. A Stackelberg game assumes that one of the two governments acts as leader and chooses its policy instruments taking the other's response into account, while the other government responds while assuming that the leader's choice will not be affected.
The present set-up is somewhat more complicated than that which we have just described, because we model the choice of capacity and tolls as a two-stage game. The choice of capacity is made first, followed by the decision to set the toll in the second stage. As a result, different types of behaviour could be assumed for each stage (Nash versus Stackelberg), with two types of leadership (i.e. the city or the region). This leaves seven different study situations when excluding reversed leadership between stages. First we analyse the Full Nash situation, in which both prices and capacities result from a non-cooperative Nash game. Then we consider three different cases with city leadership in the first, second, or both stages, followed by the same three situations with the region as the leader.

3.3.1 Full Nash equilibrium

In what we will call the Full Nash equilibrium, each government takes the other's choice as given in both stages. The city government chooses capacity that is the best reply to a given capacity selected by the region, while the regional government will select capacity as its best reply to the city government's selected capacity. Given these capacity choices, tolls are set in a similar way resulting in a stable Nash equilibrium in prices and capacities. Figure 1 shows 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 function: 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 absolutely vertically-sloped points of the city's various iso-surplus functions. The region's reaction function, in a similar fashion, connects absolutely horizontally-sloped points of the region's various iso-surplus functions. Finally, the intersection of the two reaction functions then defines the Full Nash equilibrium.

Appendix C shows that the Nash situation results in a small (overall) welfare gain (ω = 0.07) compared with the base-case equilibrium. The city again has the incentive to extract toll revenues from the non-residents (commuters), leading to excessively high tolls. Regional tolls are somewhat lower than 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 because capacity choice has become a strategic instrument in the game's first stage.

3.3.2 City leadership

When introducing strategic leadership in our game-theoretic framework, we can in fact then distinguish three different situations depending on the stage(s) in which leadership is exercised. The first is the Stackelberg (capacity) Nash (pricing) situation; then, second, its mirror image is the Nash (capacity)/Stackelberg (price) situation, and the third possibility is the "Full" (both stages) Stackelberg equilibrium. Assuming that the city is the leader, prices in the first situation are set according to the properties of a Nash game (without any leadership), while the city takes account of the capacity chosen by the follower (region) when planning its own capacity level. The second situation is the opposite of the first: city leadership in setting tolls, while a first-stage Nash game characterises decisions on capacities. The last situation is a full Stackelberg solution in which both prices and capacities are set by the city, while anticipating correctly the reaction of the region for both instruments. Because there is no compelling reason why the one situation is more relevant than the other, we will consider them all, and determine in particular also whether there are large differences between them. We will coin the different games such that Nash/Stack means Nash behaviour in the short run (the second, price stage) and Stackelberg in the long run (the first, capacity stage).

For the two Nash/Stack games, the same iso-surplus contours and reaction functions are relevant as for the Full Nash game discussed earlier. When the city is the leader, it takes the reaction of the region into account (the region's reaction function is known to the city) when setting its own capacity. In fact, the city searches the highest surplus level for each point on the region's reaction function. This must be the point where the reaction function of the region and the iso-surplus curve of the city have equal slopes. This is the Nash/Stack equilibrium point (city leader) indicated in Figure 1. The increase in capacity of the region is very small compared with the full Nash equilibrium. The change in capacity chosen by the leader is somewhat larger. This explains the higher surplus level obtained by the leader caused by the information advantage for capacity setting only. But it is not only the leader that gains in terms of welfare; the follower also benefits from the higher level of capacity: both jurisdictions end up on a higher iso-surplus contour. However, the welfare changes are very small: given that the pricing game entails Nash competition, the nature of the capacity game appears less important for the final outcome.

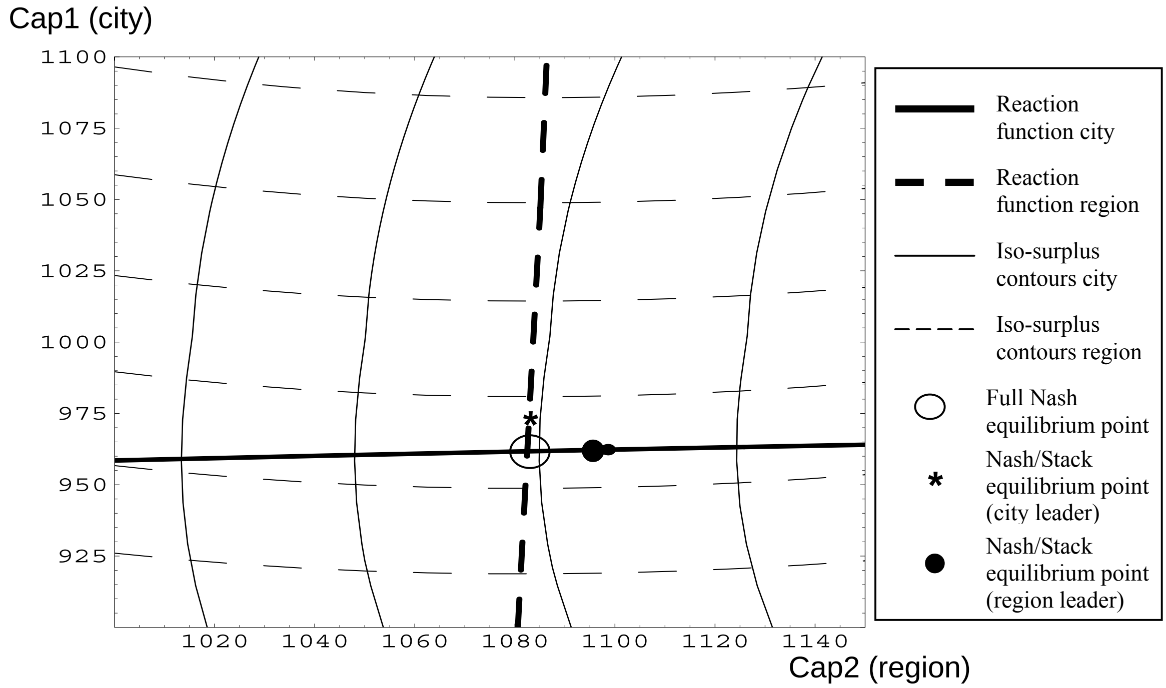

Figure 1: Reaction functions and iso-surplus contours for Nash pricing games

When the price stage is characterised by Stackelberg competition, another set of iso-surplus contours and (hence) reaction functions of course applies. Figure 2 illustrates the sets applying under city leadership in the price stage. Two equilibria are of interest: the Stack/Nash where capacities are chosen in a Nash way (where the reaction functions cross), and the Stack/Stack where both instruments are chosen in a Stackelberg way. It is now possible for the city to anticipate the toll response of the region. This leads to slightly higher city tolls than under Nash pricing, as well as 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. What causes the differences between Stackelberg pricing under city-leadership and Nash pricing to be so small, relatively speaking? 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 the region's toll (even when significant in a relative sense) are relatively 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.

Similar to the previous Stackelberg capacity game (with Nash prices), the city seeks a point on the reaction line of the region with slopes equal to its own iso-surplus curve. This is the Stack/Stack equilibrium point in Figure 2. The Stack/Nash equilibrium is again located at the intersection of the reaction functions.

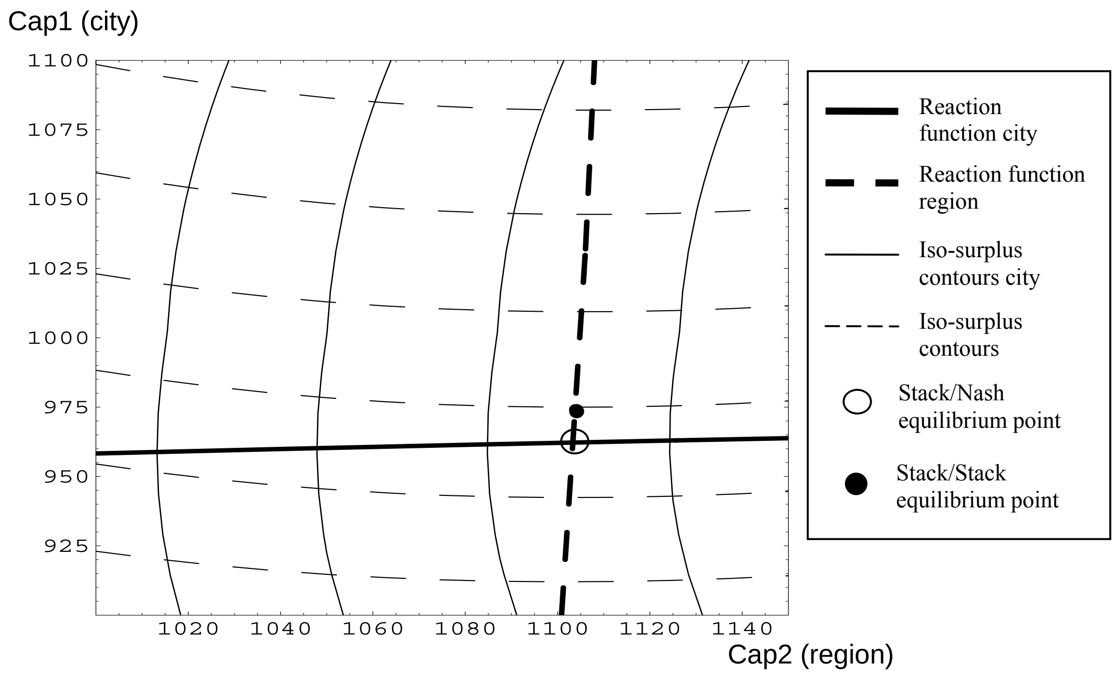

Figure 2: Reaction functions and iso-surplus contours for Stackelberg pricing games (city leadership)

3.3.3 Region leadership

Region leadership can also assume three forms. The earlier Figure 1 already illustrated the equilibrium situations with Nash pricing behaviour. The equilibrium outcome with the region as leader in the capacity game is therefore shown in this same figure (Nash/Stack point, region leader). It can be verified in Figure 1, and also in Appendix C, that both jurisdictions would prefer the other to lead in a Stackelberg capacity game, when followed by a Nash price game. Such seemingly counterintuitive results are not that rare in game theory (see Dowrick, 1986). Note that both jurisdictions prefer to lead in a Stack/Nash game when compared with Nash/Nash. While the differences between equilibria with Stackelberg leadership in the capacity stage and Nash in the pricing stage terms of capacity are very small, price leadership for the region does change outcomes. When setting its own toll, the region takes into account the incentive of the city to adapt its own toll. This leads to a regional toll that is around three times the toll under other games. 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 non-cooperative game situations, and that of the region is slightly 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 in 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 price game. The Stack/Stack and Stack/Nash equilibrium points are relatively close in Figure 3.
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.
These situations are shown in Figure 3. The reaction functions and surplus contours are again different from Figures 1 and 2. Surplus levels are higher for the region and lower for the city than in Figure 2. Price leadership may be attractive to the region, but it is not beneficial for overall welfare, given the negative value for ω mentioned in Appendix C. The combination of high toll levels with relatively small levels of capacity does not contribute to high welfare levels.

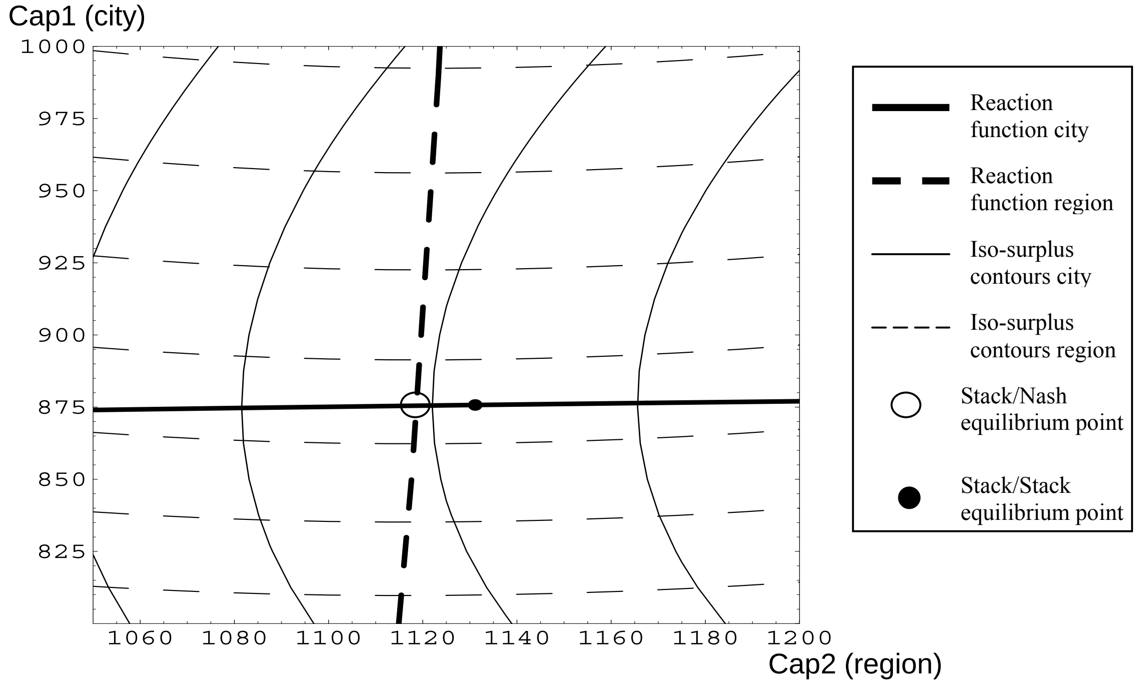

Figure 3: Reaction functions and iso-surplus contours for Stackelberg pricing games (region leadership)

3.3.4 Overview

Figure 4 illustrates the differences in capacities (left panel) and tolls for the different scenarios in our numerical example. While the assumed symmetry in the numerical example leads to identical toll levels and capacity in the first-best equilibrium, as well as in the base-case, the figure shows that all other regimes 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 is as given before.. 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 4 cannot even be distinguished graphically. But also the two clusters are relatively close, compared to the first-best choices of tolls and capacities. This suggests that the main issue is not which exact type of game is played between the two actors, 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.

The previous results show that competition between two different governments, in this setting, may not necessarily improve the welfare of society compared with a reasonably realistic initial situation, and, even if it does, the gains may be relatively small. The results depend heavily on the asymmetry that commuters should pay a toll levied by the city; while the opposite situation does not occur, this gives the city a tax-exporting instrument. The difference between both second-best situations illustrates this. In the non-cooperative game situations too, we find that the city has an incentive to set excessive tolls.

We find that Stackelberg leadership in one or both instruments improves the welfare of the leader when compared with a game with Nash properties. Being the leader in the toll stage is more important than 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).

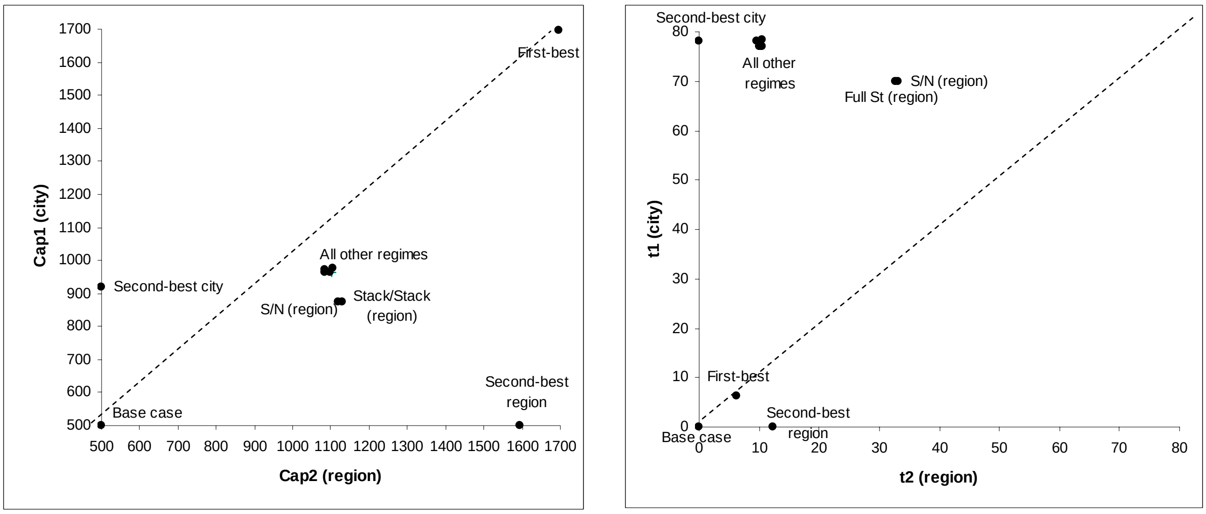

Figure 4: Overview of equilibria capacity levels (left panel) and tolls (right panel) for the different regimes

3.4 Sensitivity analysis

The results shown above are, of course, likely to change with the parameter values chosen. In order to consider the robustness of our results, we have analysed 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. We will summarise our findings by reporting the impacts on our efficiency indicator ω for changing the demand elasticity, and on a slightly different welfare indicator (the share of the optimal welfare) when changing the price of capacity.

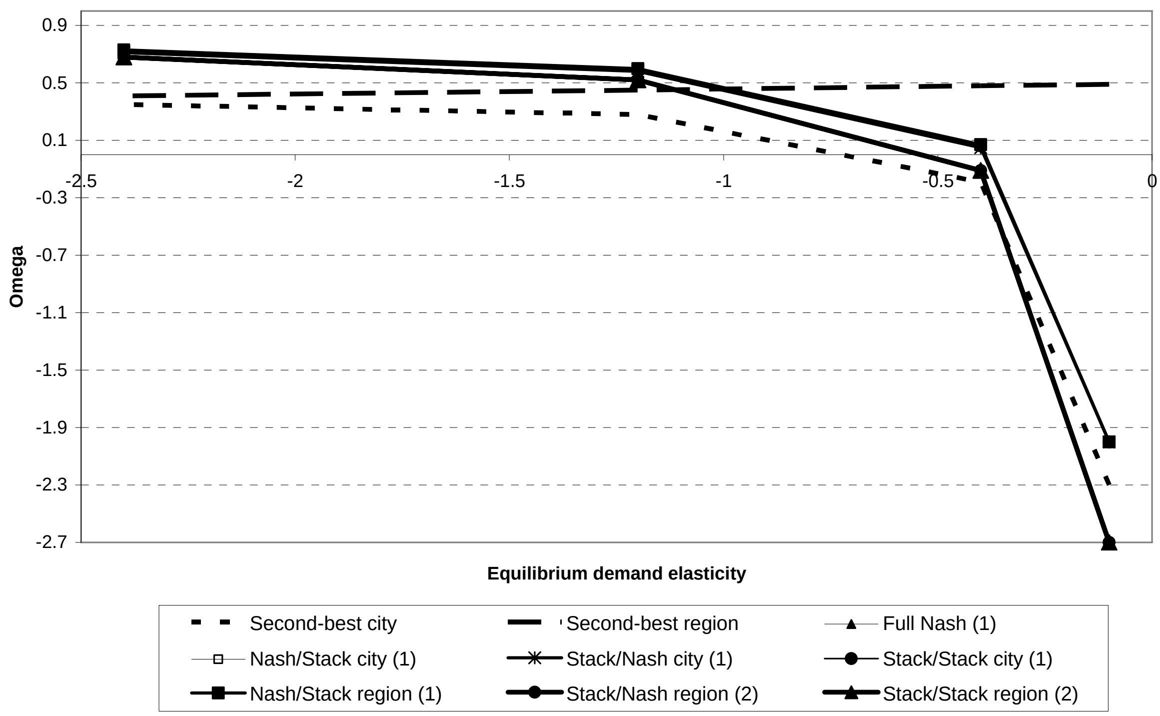

Figure 5: Indexes of relative welfare improvement with varying demand elasticities

Varying demand elasticity

Figure 5 shows the results in terms of ω when changing both the equilibrium demand elasticities from relatively elastic (left-hand side) to relatively inelastic (right-hand side) (by simultaneously changing the parameter values of a and d, for all demand functions, such that the same base equilibrium is obtained for every elasticity). While showing 9 indicators ω, only four clusters can be distinguished visually. This reflects that the relative closeness of the Full Nash, Nash/Stack city, Stack/Stack city, Stack/Nash city and Nash/Stack region remains. These cases all come under the heading "Group 1". The same holds for closeness within "Group 2", comprising the Stack/Stack region and Stack/Nash region. These similarities, already identified for our base parametrisation, are therefore robust and do not depend on the assumed demand elasticity. However, all schemes become less efficient when demand becomes less elastic. The "Second-best city" 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

Changing prices of road capacity (p_{cap}), our welfare indicator ω unfortunately becomes less insightful. The reason is that, when the price of capacity increases (when moving to the right), the difference between welfare in the optimal situation and the base case becomes rapidly smaller, because welfare levels in the base case remain relatively constant. It is then not so much the welfare level of the situation under study that determines the score of the welfare indicator, but rather the small number of the denominator. All *ω's*, consequently, reflect the changing difference between first-best and equilibrium rather the welfare in the region under consideration. Therefore, we simply decided to use the welfare of the situation under study as a fraction of the first-best welfare level.

Figure 6 shows that, when the price of road capacity changes, the performance of all scenarios remains rather constant. About 80% of the first-best welfare level is achieved in the various scenarios, with the second-best region situation performing relatively best for all prices of road capacity. With lower prices of capacity, congestion becomes more easily solved through capacity adjustments. The region reacts accordingly by setting a lower toll. The distortive impact of tolling by the city (caused by the city's extraction of toll revenues from regional drivers) remains important, but the city's toll levels will not change much. As a result, welfare fractions remain fairly unchanged. Note that the same two clusters of regimes emerge as in Figure 4.

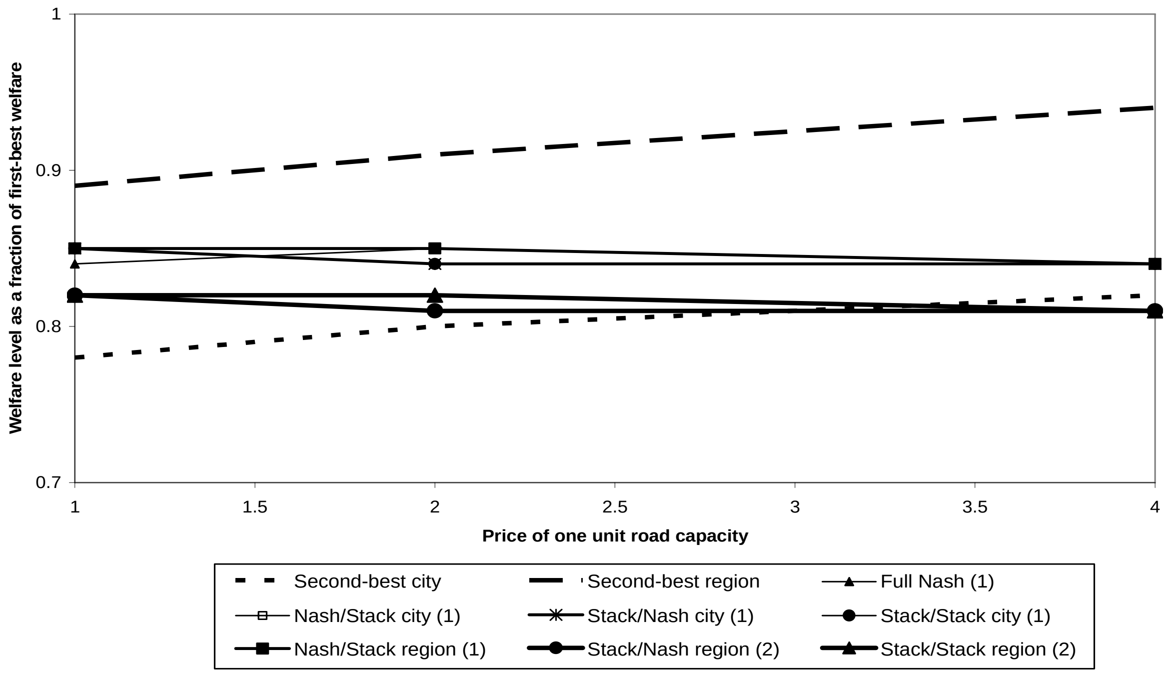

Figure 6: Welfare level for various situations as a fraction of first-best situation with varying prices of road capacity

4. Conclusions

In this study we have analysed policy interactions between an urban and a regional government with different objectives (namely maximization of their own citizens'welfare) and two policy instruments (toll and capacity) available. We considered a simple serial network with two roads. This setting is relevant in practice since many cities experience congestion problems, while it is often commuters residing in a region that suffer from congestion and who would pay city tolls. Using a simulation model, we investigated the welfare consequences of the various regimes that result when both governments compete, and take sequential decisions on prices and capacities with the aim to maximise their local social surplus. The first-best situation is a useful benchmark, and can be regarded as a situation of optimal instrument choice by a central government, or as the result from policy coordination.

Our analysis gives some useful insights into the joint pricing-capacity decisions with two governments involved. Competition between governments may not be very beneficial to overall welfare in society compared with one central government. It appears that the tendency of tax exporting is very strong in this setting where commuters have to pay road tolls set by the city government. The incentive to set excessive tolls is present in all scenarios where the city has control over this instrument.

We find that Stackelberg leadership in one or both instruments improves the welfare of the leader when compared with a game with Nash properties. Being the leader in the toll stage is more important than 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 the 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).

We found that, at least in our numerical model, the relative performance of the various game regimes is rather close for a variety of parameter ranges. 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 game equilibria. This suggests that the main issue is not which exact type of game is played between the two actors, 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.

The performance for most game situations improves when demand becomes more elastic. The city then has less incentive to exploit its market power. The relative importance of the socially-inefficient motive for tolling (revenue extraction from regional drivers by the city) decreases 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 the increased inefficiency when demand becomes more inelastic. The similarities between the different groups are robust and do not depend on the assumed demand elasticity. When the price of road investment changes, the performance relative to the optimal situation remains more or less equal for all cases.

Although it may seem attractive to national governments to deregulate and give powers to lower levels of government since they know local situations best, the outcomes of interaction may not always be promising. In our example, compared to the selected unpriced benchmark equilibrium, overall welfare is not helped very much, and in some cases may even decrease. Toll regulation for instance, although not considered, may seriously be considered as a useful tool to control one of the jurisdictions and improve welfare levels. This is only one among many other issues for further research. Another issue that seems relevant in this context is empirical research on the welfare effects of tax competition. This is only a simple illustrative numerical example, but it remains to be seen whether the costs of non-cooperative behaviour are substantial, in other plausible settings.

References

Braid, R.M, 1996, Peak Load Pricing of a Transportation Route with an Unpriced Substitute, *Journal of Urban Economics*, 40, pp. 179-197.

De Borger, B., C. Courcelle and D. Swysen, 2004, Optimal pricing of transport externalities in an international environment: some empirical results based on a numerical optimization model, *Regional Science and Urban Economics*, 34, pp. 163-201.

De Borger, B. S. Proost, and K. van Dender, 2005, Congestion and tax competition in a parallel network, *European Economic Review*, 49, pp. 2013-2040.

De Borger, B. and S. Proost, 2004, Vertical and horizontal competition in transport, paper presented at a STELLA meeting, Athens.

De Borger, B. and K. van Dender, 2005, Prices, capacities and service quality in a congestible Bertrand duopoly, paper presented at the ERSA 2005 conference in Amsterdam.

Dowrick, S., 1986, Von Stackelberg and Cournot Duopoly: Choosing Roles, *The RAND Journal of Economics*, 17 (2), pp. 251-260.

Calthrop, E., S. Proost and K. van Dender, 2000, Parking policies and road pricing, *Urban Studies*, 37 (1), pp. 63-76.

De Palma, A. and R. Lindsey, 2000, Private Toll Roads: Competition under Various Ownership Regimes, *The Annals of Regional Science*, 34, pp. 13-35.

Levinson, D., 2001, Why States toll, An empirical model of finance costs, *Journal of Transport Economics and Policy*, 35 (2), pp.223-238.

Lévy-Lambert, H., 1968, Tarification des services à qualité variable: application aux péages de circulation, *Econometrica,* 36 (3-4), pp. 564-574.

Liu, L.N. and J.F. McDonald, 1998, Efficient congestion tolls in the presence of unpriced congestion: a peak and off-peak simulation model, *Journal of Urban Economics,* 44 (3), pp. 352-366.

Liu, L.N. and J.F. McDonald, 1999, Economic efficiency of second-best pricing schemes in urban highway systems, *Transportation Research Part B*, 33, pp. 157-188.

Marchand, M., 1968, A note on optimal tolls in an imperfect environment, *Econometrica,* 36 (3-4), pp. 575-581.

Markusen, J. E. Morey, and N. Olewiler, 1995, Competition in Regional Environmental Policies when Plant Locations are endogenous, *Journal of Public Economics*, 56, pp. 55-73.

MC-ICAM consortium, 2004, Workpackage 7, project funded under the 6th framework RTD programme of the EU.

Mohring, H and M. Harwitz, 1962, *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston.

Ubbels, B. and E.T. Verhoef, 2004, Auctioning Concessions for Private Roads, *Tinbergen Institute discussion paper*, TI 2004-008/3, Amsterdam/Rotterdam.

Verhoef, E.T., 2005, Second-best road pricing through highway franchising, *Discussion paper TI,* 2005-082/3, Tinbergen Institute, Amsterdam-Rotterdam.

Verhoef, E.T., P. Nijkamp and P. Rietveld, 1996, Second-best congestion pricing: The case of an untolled alternative, *Journal of Urban Economics*, 40 (3), pp. 279-302.

Verhoef, E.T. and K.A. Small, 2004, Product differentiation on roads, *Journal of Transport Economics and Policy,* 38, 1, pp. 127-156.

Yang, H. and Q. Meng, 2000, Highway pricing and capacity choice in a road network under a build-operate-transfer scheme, *Transportation Research Part A*, 34, pp. 207-222.

Wilson, J., 1999, Theories of Tax Competition, *National Tax Journal*, 52, pp 269-304.

Appendix A: List of Symbols

N	Total number of road users
N_1	Number of road users on the city link
N_2	Number of road users on the regional link
N_C	Number of drivers only using the city link
N_{R1}	Number of commuters (using both links by going from region to city)
N_{R2}	Number of drivers only using the regional link
$D_j(N_j)$	Inverse demand function for group $j = \{C,R_1,R_2\}$
c_1	Average social (=marginal private) costs on the city link
c_2	Average social (=marginal private) costs on the regional link
cap_1	Capacity of the city link
cap_2	Capacity of the regional link
W	Total welfare
W_1	Welfare of city
W_2	Welfare of region
B_j	Benefits for group $j = \{C,R_1,R_2\}$
π_1	Profits for the city
π_2	Profits for the region
p_j	Generalised price for group $j = \{C,R_1,R_2\}$
t_1	Toll on city link
t_2	Toll on regional link
p_{cap}	Price of one unit road capacity
k_i	Average free-flow user cost on route i ($k_i > 0$)
b_i	Slope of the average cost function on route i ($b_i > 0$)
ω	Index of relative welfare improvement
a_j	Absolute value of the slope of the demand curve for group $j = \{C,R_1,R_2\}$ ($a_j > 0$)
d_j	Intersection of the demand curve with the vertical axis for group $j = \{C,R_1,R_2\}$ ($d_j > 0$)

Appendix B: Derivation of first-order conditions for toll and capacity for the first-best situation and two second-best situations

This appendix derives analytical expressions for toll rules and investment rules for three cases: first-best (Section B.2.1), for the city when taking the regional toll and capacity as given (in Section B.2.2), and for the region when taking the city's toll and capacity as given (in Section B.2.3).

B.2.1 First-best situation

The first-best toll and capacity can be found by solving the following Lagrangian:

$$\begin{aligned} L = & \int_0^{N_C} D_C(n)dn + \int_0^{N_{R1}} D_{R1}(n)dn + \int_0^{N_{R2}} D_{R2}(n)dn - N_C * c_1 - N_{R1} * (c_1 + c_2) - N_{R2} * c_2 \\ & - (cap_1 + cap_2) * p_{cap} + \lambda_C (p_C - D_C) + \lambda_{R1}(p_{R1} - D_{R1}) + \lambda_{R2}(p_{R2} - D_{R2}) \end{aligned} \quad \text{(A1)}$$

We determine the first-order conditions with respect to $N_C, N_{R1}, N_{R2}, cap_1, cap_2, t_1, t_2, \lambda_C, \lambda_{R1}$ and λ_{R2}:

$$\frac{\partial L}{\partial N_C} = D_C - c_1 - (N_C + N_{R1})\frac{\partial c_1}{\partial N_C} + \lambda_C\left(\frac{\partial c_1}{\partial N_C} - \frac{\partial D_C}{\partial N_C}\right) + \lambda_{R1}\frac{\partial c_1}{\partial N_C} = 0 \quad \text{(A2a)}$$

$$\begin{aligned} \frac{\partial L}{\partial N_{R1}} = & D_{R1} - (c_1 + c_2) - (N_{R1} + N_C)\frac{\partial c_1}{\partial N_{R1}} - (N_{R1} + N_{R2})\frac{\partial c_2}{\partial N_{R1}} + \lambda_C \frac{\partial c_1}{\partial N_{R1}} \\ & + \lambda_{R1}\left(\frac{\partial c_1}{\partial N_{R1}} + \frac{\partial c_2}{\partial N_{R1}} - \frac{\partial D_{R1}}{\partial N_{R1}}\right) + \lambda_{R2}\frac{\partial c_2}{\partial N_{R1}} = 0 \end{aligned} \quad \text{(A2b)}$$

$$\frac{\partial L}{\partial N_{R2}} = D_{R2} - c_2 - (N_{R1} + N_{R2})\frac{\partial c_2}{\partial N_{R2}} + \lambda_{R2}\left(\frac{\partial c_2}{\partial N_{R2}} - \frac{\partial D_{R2}}{\partial N_{R2}}\right) + \lambda_{R1}\frac{\partial c_2}{\partial N_{R2}} = 0 \quad \text{(A2c)}$$

$$\frac{\partial L}{\partial t_1} = \lambda_c + \lambda_{R1} = 0 \quad \text{(A2d)}$$

$$\frac{\partial L}{\partial t_2} = \lambda_{R1} + \lambda_{R2} = 0 \quad \text{(A2e)}$$

$$\frac{\partial L}{\partial cap_1} = -N_C \frac{\partial c_1}{\partial cap_1} - N_{R1}\frac{\partial c_1}{\partial cap_1} - p_{cap} + \lambda_C \frac{\partial c_1}{\partial cap_1} + \lambda_{R1}\frac{\partial c_1}{\partial cap_1} = 0 \quad \text{(A2f)}$$

$$\frac{\partial L}{\partial cap_2} = -N_{R2} \frac{\partial c_2}{\partial cap_2} - N_{R1}\frac{\partial c_2}{\partial cap_2} - p_{cap} + \lambda_{R2} \frac{\partial c_2}{\partial cap_2} + \lambda_{R1}\frac{\partial c_2}{\partial cap_2} = 0 \quad \text{(A2g)}$$

$$\frac{\partial L}{\partial \lambda_C} = c_1 + t_1 - D_C = 0 \quad \text{(A2h)}$$

$$\frac{\partial L}{\partial \lambda_{R1}} = c_1 + t_1 + c_2 + t_2 - D_{R1} = 0 \quad \text{(A2i)}$$

$$\frac{\partial L}{\partial \lambda_{R2}} = c_2 + t_2 - D_{R2} = 0 \quad \text{(A2j)}$$

We can eliminate the $\lambda \frac{\partial c}{\partial N}$ terms in (A2a)-(A2c). Defining link-flows $N_1=N_C + N_{R1}$, and $N_2=N_{R1} + N_{R2}$. This leaves:

$$t_1 = N_1 \frac{\partial c_1}{\partial N_1} + \lambda_C \frac{\partial D_C}{\partial N_C} \quad \text{(A3a)}$$

$$t_1 + t_2 = N_1 \frac{\partial c_1}{\partial N_1} + N_2 \frac{\partial c_2}{\partial N_2} + \lambda_{R1} \frac{\partial D_{R1}}{\partial N_{R1}} \quad \text{(A3b)}$$

$$t_2 = N_2 \frac{\partial c_2}{\partial N_2} + \lambda_{R2} \frac{\partial D_{R2}}{\partial N_{R2}} \quad \text{(A3c)}$$

$$\Rightarrow \lambda_C \frac{\partial D_C}{\partial N_C} + \lambda_{R2} \frac{\partial D_{R2}}{\partial N_{R2}} - \lambda_{R1} \frac{\partial D_{R1}}{\partial N_{R1}} = 0 \quad \text{(A3d)}$$

Since we had (A2d) and (A2e), we find that $\lambda_C = \lambda_{R1} = \lambda_{R2} = 0$.

This gives us the following first-best conditions for the optimal tolls:

$$t_1 = N_1 \frac{\partial c_1}{\partial N_1} \quad \text{(A4a)}$$

$$t_2 = N_2 \frac{\partial c_2}{\partial N_2} \quad \text{(A4b)}$$

Both optimal tolls equal marginal external (congestion) costs on both links, as expected.
The first order conditions for capacity are as follows (substituting A2d) and (A2e) in (A2f) and (A2g)):

$$-N_1 \frac{\partial c_1}{\partial cap_1} = p_{cap} \quad \text{(A4c)}$$

$$-N_2 \frac{\partial c_2}{\partial cap_2} = p_{cap} \quad \text{(A4d)}$$

These are again familiar first-best results, equating marginal cost of capacity expansion to marginal benefits on both routes. Because $N\frac{\partial c}{\partial N}=-cap\frac{\partial c}{\partial cap}$, exact self-financing with tax rules (A4a and A4b) and investment rules (A4c and A4d) is easily established, confirming applicability of the conventional Mohring-Harwitz result.

B.2.2 Second-best: city optimises

The appropriate Lagrangian now reads as follows:

$$L=\int_0^{N_C} D_C(n)dn - N_C * c_1 + N_{R1} * t_1 - cap_1 * p_{cap} + \lambda_C\left(p_C - D_C\right) + \lambda_{R1}\left(p_{R1} - D_{R1}\right) + \lambda_{R2}\left(p_{R2} - D_{R2}\right) \quad \text{(A5a)}$$

We determine the first-order conditions with respect to $N_C, N_{R1}, N_{R2}, cap_1, t_1, \lambda_C, \lambda_{R1}$ and λ_{R2}:

$$\frac{\partial L}{\partial N_C}=D_C-c_1-N_C\frac{\partial c_1}{\partial N_C}+\lambda_C\left(\frac{\partial c_1}{\partial N_C}-\frac{\partial D_C}{\partial N_C}\right)+\lambda_{R1}\frac{\partial c_1}{\partial N_C}=0 \quad \text{(A5b)}$$

$$\frac{\partial L}{\partial N_{R1}}=-N_C\frac{\partial c_1}{\partial N_{R1}}+t_1+\lambda_C\frac{\partial c_1}{\partial N_{R1}}+\lambda_{R1}\left(\frac{\partial c_1}{\partial N_{R1}}+\frac{\partial c_2}{\partial N_{R1}}-\frac{\partial D_{R1}}{\partial N_{R1}}\right)+\lambda_{R2}\frac{\partial c_2}{\partial N_{R1}}=0 \quad \text{(A5c)}$$

$$\frac{\partial L}{\partial N_{R2}}=\lambda_{R2}\left(\frac{\partial c_2}{\partial N_{R2}}-\frac{\partial D_{R2}}{\partial N_{R2}}\right)+\lambda_{R1}\frac{\partial c_2}{\partial N_{R2}}=0 \quad \text{(A5d)}$$

$$\frac{\partial L}{\partial t_1}=N_{R1}+\lambda_c+\lambda_{R1}=0 \quad \text{(A5e)}$$

$$\frac{\partial L}{\partial cap_1}=-N_C\frac{\partial c_1}{\partial cap_1}-p_{cap}+(\lambda_C+\lambda_{R1})\frac{\partial c_1}{\partial cap_1}=0 \quad \text{(A5f)}$$

$$\frac{\partial L}{\partial \lambda_C}=c_1+t_1-D_C=0 \quad \text{(A5g)}$$

$$\frac{\partial L}{\partial \lambda_{R1}}=c_1+t_1+c_2+t_2-D_{R1}=0 \quad \text{(A5h)}$$

$$\frac{\partial L}{\partial \lambda_{R2}}=c_2+t_2-D_{R2}=0 \quad \text{(A5i)}$$

The first-order condition for capacity can relatively easy be derived by substituting (A5e) in (A5f):

$$-(N_C + N_{R1})\frac{\partial c_1}{\partial cap_1} = p_{cap} \tag{A6a}$$

This is equal to the first-best rule for capacity.

The derivation of the first-best expression for the toll set by the city is more tedious. Therefore, we only present the resulting expression:

$$t_1 = N_1\frac{\partial c_1}{\partial N_1} + \frac{\frac{\partial D_C}{\partial N_C}\left(-\frac{\partial D_{R1}}{\partial N_{R1}} * \frac{\partial D_{R2}}{\partial N_{R2}} + \frac{\partial c_2}{\partial N_2} * \left(\frac{\partial D_{R1}}{\partial N_{R1}} + \frac{\partial D_{R2}}{\partial N_{R2}}\right)\right) N_{R1}}{\left(\frac{\partial D_C}{\partial N_C} + \frac{\partial D_{R1}}{\partial N_{R1}}\right)\frac{\partial D_{R2}}{\partial N_{R2}} - \frac{\partial c_2}{\partial N_2}\left(\frac{\partial D_C}{\partial N_C} + \frac{\partial D_{R1}}{\partial N_{R1}} + \frac{\partial D_{R2}}{\partial N_{R2}}\right)} \tag{A6b}$$

The first term repeats the first-best expression and is equal to the marginal external congestion cost. But the city will raise the toll beyond this level. The term 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. The second term in (A6b) is so complex because the demand by group R_1 depends, in part, also on cost and demand elasticity in the region. The second term is namely positive (as long as demands are downward sloping).

B.2.3 Second-best; region optimises

The appropriate Lagrangian now reads as follows:

$$\begin{aligned} L = &\int_0^{N_{R1}} D_{R1}(n)dn + \int_0^{N_{R2}} D_{R2}(n)dn - N_{R1} * c_1 - N_{R1} * c_2 - N_{R2} * c_2 - N_{R1} * t_1 \\ &- cap_2 * p_{cap} + \lambda_C\left(p_C - D_C\right) + \lambda_{R1}\left(p_{R1} - D_{R1}\right) + \lambda_{R2}\left(p_{R2} - D_{R2}\right) \end{aligned} \tag{A7a}$$

We determine the first-order conditions with respect to $N_C, N_{R1}, N_{R2}, cap_2, t_2, \lambda_C, \lambda_{R1}$ and λ_{R2}:

$$\frac{\partial L}{\partial N_C} = -N_{R1}\frac{\partial c_1}{\partial N_C} + \lambda_C\left(\frac{\partial c_1}{\partial N_C} - \frac{\partial D_C}{\partial N_C}\right) + \lambda_{R1}\frac{\partial c_1}{\partial N_C} = 0 \tag{A7b}$$

$$\begin{aligned} \frac{\partial L}{\partial N_{R1}} = &D_{R1} - c_1 - N_{R1}\frac{\partial c_1}{\partial N_{R1}} - c_2 - (N_{R1} + N_{R2})\frac{\partial c_2}{\partial N_{R1}} - t_1 + \lambda_C\frac{\partial c_1}{\partial N_{R1}} \\ &+ \lambda_{R1}\left(\frac{\partial c_1}{\partial N_{R1}} + \frac{\partial c_2}{\partial N_{R1}} - \frac{\partial D_{R1}}{\partial N_{R1}}\right) + \lambda_{R2}\frac{\partial c_2}{\partial N_{R1}} = 0 \end{aligned} \tag{A7c}$$

$$\frac{\partial L}{\partial N_{R2}} = D_{R2} - c_2 - (N_{R1} + N_{R2})\frac{\partial c_2}{\partial N_{R2}} + \lambda_{R2}\left(\frac{\partial c_2}{\partial N_{R2}} - \frac{\partial D_{R2}}{\partial N_{R2}}\right) + \lambda_{R1}\frac{\partial c_2}{\partial N_{R2}} = 0 \qquad \text{(A7d)}$$

$$\frac{\partial L}{\partial t_2} = \lambda_{R2} + \lambda_{R1} = 0 \qquad \text{(A7e)}$$

$$\frac{\partial L}{\partial cap_2} = -(N_{R1} + N_{R2})\frac{\partial c_2}{\partial cap_2} - p_{cap} + (\lambda_{R1} + \lambda_{R2})\frac{\partial c_2}{\partial cap_2} = 0 \qquad \text{(A7f)}$$

$$\frac{\partial L}{\partial \lambda_C} = c_1 + t_1 - D_C = 0 \qquad \text{(A7g)}$$

$$\frac{\partial L}{\partial \lambda_{R1}} = c_1 + t_1 + c_2 + t_2 - D_{R1} = 0 \qquad \text{(A7h)}$$

$$\frac{\partial L}{\partial \lambda_{R2}} = c_2 + t_2 - D_{R2} = 0 \qquad \text{(A7i)}$$

The first order condition for capacity is, again, easy to derive and again equals the first-best rule for capacity (substituting (A7e) in (A7f)):

$$-N_2\frac{\partial c_2}{\partial cap_2} = p_{cap}$$

The derivation of the first-best expression for the toll set by the region is more tedious. Therefore, we only present the resulting expression:

$$t_2 = N_2\frac{\partial c_2}{\partial N_{R1}} + \frac{-N_{R1} * \frac{\partial c_1}{\partial N_1} * \frac{\partial D_C}{\partial N_C} * \frac{\partial D_{R2}}{\partial N_{R2}}}{-\frac{\partial D_C}{\partial N_C}\left(\frac{\partial D_{R1}}{\partial N_{R1}} + \frac{\partial D_{R2}}{\partial N_{R2}}\right) + \frac{\partial c_1}{\partial N_1}\left(\frac{\partial D_C}{\partial N_C} + \frac{\partial D_{R1}}{\partial N_{R1}} + \frac{\partial D_{R2}}{\partial N_{R2}}\right)}$$

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.

Appendix C: Numerical Results

The following table presents the numerical results for various scenarios relative to the first-best outcomes.

Scenario		N_C	N_{R1}	N_{R2}	cap_1	cap_2	W	W_C	W_R	t_1	t_2	ω
Base equilibrium		0.93	0.93	0.93	0.29	0.29	0.83	0.76	0.85	0	0	-
First-best		179	358	179	1697	1697	57621	9603	48018	6.32	6.32	1
Second best	**City optimises**	0.33	0.65	0.97	0.54	0.29	0.80	2.29	0.50	12.36	0	-0.19
	Region optimises	0.93	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	0.57	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	0.57	0.64	0.84	2.28	0.56	12.18	1.53	0.07
	Stackelberg prices, Nash capacities	0.33	0.65	0.96	0.57	0.65	0.84	2.28	0.55	12.34	1.52	0.05
	Full Stackelberg	0.33	0.65	0.96	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	0.57	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	0.66	0.81	2.03	0.57	11.04	5.19	-0.11
	Full Stackelberg	0.41	0.58	0.75	0.52	0.67	0.82	2.03	0.57	11.04	5.19	-0.11