

The interaction between tolls and capacity investment in serial and parallel transport networks

Bruno De Borger, Fay Dunkerley and Stef Proost*

Abstract

The purpose of this paper is to compare the interaction between pricing and capacity decisions on simple serial and parallel transport networks. When individual links of the network are operated by different regional or national authorities, toll and capacity competition is likely to result. Moreover, the problem is potentially complicated by the presence of both local and transit demand on each link of the network. We bring together and extend the recent literature on the topic and, using both theory and numerical simulation techniques, provide a careful comparison of toll and capacity interaction on serial and parallel network structures. First, we show that there is more tax exporting in serial transport corridors than on competing parallel road networks. Second, the inability to toll transit has quite dramatic negative welfare effects on parallel networks. On the contrary, in serial transport corridors it may actually be undesirable to allow the tolling of transit at all. Third, if the links are exclusively used by transit transport, toll and capacity decisions are independent in serial networks. This does not generally hold in the presence of local transport. Moreover, it contrasts with a parallel setting where regional authorities compete for transit; in that case, regional investment in capacity leads to lower Nash equilibrium tolls.

Keywords: congestion pricing, transport investment, transit traffic
JEL: H23, H71, R41, R48

*De Borger: Department of Economics, University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium (bruno.deborger@ua.ac.be); Dunkerley: Department of Economics, Catholic University of Leuven, Naamsestraat 69, B-3000 Leuven (fay.dunkerley@econ.kuleuven.ac.be); Proost: Department of Economics, Catholic University of Leuven, Naamsestraat 69, B-3000 Leuven, and CORE, 1348, Louvain-La-Neuve, Belgium (stef.proost@econ.kuleuven.ac.be). The first author is grateful to the University of Antwerp for support through its TOP-program; the second and third author benefited from financial support of the FUNDING consortium (6 th Framework research Program of the EC (6.2)) on the Funding of the Trans European Networks (www.econ.kuleuven.be/funding). This paper developed out of preliminary work presented at the Third Conference on Railroad Industry Structure, Competition and Investment, Stockholm, October 2005. We are grateful to Guido Friebel and Marcus Ksoll for useful discussions on the issue. Finally, the final version of the paper benefited from the detailed comments of an anonymous referee.

1. Introduction

Congestion is a serious problem in many countries worldwide. Apart from a variety of other measures, economists have long advocated the use of pricing policies to tackle this problem. Moreover, it has been recognized that in the long-run pricing can be accompanied by investment strategies to alleviate congestion. However, implementing pricing and investment policies on realistic transport networks leads to a number of potential complications. First, since different links (highways, roads, railroads,...) of a network may be under the jurisdiction of different governments and most links are used both by local transport and by through traffic (transit), the fear exists that competition for transit toll revenues may induce governments or operators to exploit transit transport by imposing high tolls. Second, governments may strategically invest to capture transit toll revenues. Third, when tolling transit is for some reason not feasible, regions may be reluctant to invest in capacity because the benefits accrue to a significant extent to foreigners. In sum, the possibility of strategic behaviour by governments and the interaction of local and transit transport raises a number of questions about toll and capacity choices on transport networks: (i) How does investment in infrastructure capacity affect the pricing behaviour of governments; (ii) What are the welfare effects of toll and capacity competition for transit; (iii) To what extent do the outcomes of this competition between governments depend on the structure of the transport network.

The purpose of this paper is to study the interaction between pricing and capacity decisions on simple networks, where individual links of the network are operated by different governments. We do so by bringing together and extending the recent literature on the topic. As real-world networks are highly complex, we focus in this paper on two extreme network structures. Both are highly simplified representations of realistic networks, but they capture the main ingredients of the interactions between local and transit traffic for joint toll and capacity competition between regions. The first one is a parallel network structure in which long distance transit traffic has a choice between different jurisdictions' networks. For example, there are two main routes from South-Central Europe (Switzerland, Austria, Italy) to the north (Belgium, Netherlands, etc.), one through France, the other via Germany. Another example is the transalpine crossing between Germany and Italy, where the main links pass either through Austria or through Switzerland. In both examples,

transit has a choice of routes and it interacts with local traffic in each country. Note that, with minor adjustments, the choice between two modes that connect a given origin and destination fits within this framework as well. For example, freight connections between ports such as Antwerp or Rotterdam and the Ruhr in Germany have a choice between road, rail or inland waterways. Some transport between Finland and Germany has the choice between shipping (through Kiel) and road modes (via Sweden and Denmark).

The second network structure we consider is a serial transport corridor, which provides a more realistic representation of many road and rail systems. Both the Trans European Networks (basically a border-crossing rail and highway system) in Europe and the interstate highway system in the US fit this setting of serial transport corridors. Moreover, a serial setting applies to inter-modal freight trips where the transfer facility (ports, airports, freight terminal) and the upstream or downstream infrastructure is controlled by different governments. The possibility of strategic behaviour in the case of a serial corridor has been noted several times before. In the case of railroads, for example, EU Directive 2001/14 has explicitly argued that coordination between countries is needed in order to avoid the negative effects of the lack of harmonisation of different charging systems used by member states. Moreover, Nash (2005) finds some evidence of tax exporting behaviour in an analysis of European infrastructure charges.

The issue of optimal pricing and investment decisions on simple transport networks has been studied before. First, various studies have considered parallel network structures. For example, different aspects of pricing of congestible parallel roads have been studied by, e.g., Braid (1986), Verhoef et al. (1996), De Palma and Lindsey (2000), McDonald and Liu (1999), Small and Yan (2001), and Van Dender (2005). As far as we know, the only study to analyze the problem within the context of toll competition between governments is De Borger, Proost and Van Dender (2005); they do so for fixed capacity, however. Both De Palma and Leruth (1989) and De Borger and Van Dender (2006) study two-stage games in capacities and prices for congested facilities, but they do not consider the interaction between local and transit traffic and they do not look at issues of tax and capacity competition. Second, several studies address strategic behaviour in serial transport corridors. One study looks specifically at tax exporting in such a setting, but ignores capacity decisions (Levinson (2001)). More recently, De Borger, Dunkerley and Proost (2006a) do

consider pricing and capacity investment in a two-stage game for a serial network, and illustrate the welfare effects for various sets of tolling instruments. Finally, Bassanini and Pouyet (2005) study the non-cooperative choice of financing system (i.e., does the system allow subsidies to be paid out of general tax revenues) by two national infrastructure managers who maximize welfare in their country while covering network costs. Agrell and Pouyet (2006) extend this work, focusing on countries' incentives to improve investment efficiency.

In this paper, we extend and integrate earlier findings on tax and capacity games between welfare maximising governments in both serial and parallel networks. Although some of the results of the current paper have been reported separately in some of the studies referred to above, our focus here is on the differences in the nature and extent of toll and capacity competition between regions, depending on the structure of the network. The comparison between serial and parallel networks discussed and numerically illustrated in this paper yields several new insights that have important policy implications. In both types of network structure, fiscal and expenditure externalities give rise to strategic pricing and investment behaviour by the various governments involved. However, the Nash equilibrium tolls and capacity levels differ drastically between network settings and the welfare implications of particular policies are sometimes diametrically opposed. For example, the desirability of taxing transit at all is highly dependent on the network structure. Throughout, we assume that countries maximise a welfare function consisting of local consumer surplus and tax revenues from local and transit traffic, and we study strategic tolling by individual countries under various tolling schemes. First, we assume that local traffic and transit can be tolled separately. Second, we look at the case where only uniform tolls are possible or acceptable. Third, we consider the case where only local traffic can be tolled.

The results of this paper include the following. First, if the network were exclusively used by transit transport, we show that toll and capacity decisions are independent in serial networks; in a parallel setting, however, it is shown that extra investment in capacity in a given region leads to lower Nash equilibrium tolls in both regions. The former result does not generalize to a setting with both local and transit demand, the latter does. Second, the nature and extent of competition in capacity and tolls differs strongly between network types. For example, in absolute values reaction functions for transit tolls are much more responsive to tolls abroad on serial than on

parallel networks. The policy implication of this finding is that, *ceteris paribus*, one expects much more tax exporting behaviour, and hence higher toll levels, in serial transport corridors than on competing parallel road networks. Third, the inability to toll transit has quite dramatic negative welfare effects on parallel networks, partly because it strongly reduces the incentives to invest. On the contrary, in serial transport corridors it may actually be undesirable to toll transit. Again, this has a clear policy implication. It implies that it may not be wise for the EU to allow individual countries to independently decide on toll levels on transit traffic that passes through their jurisdiction.

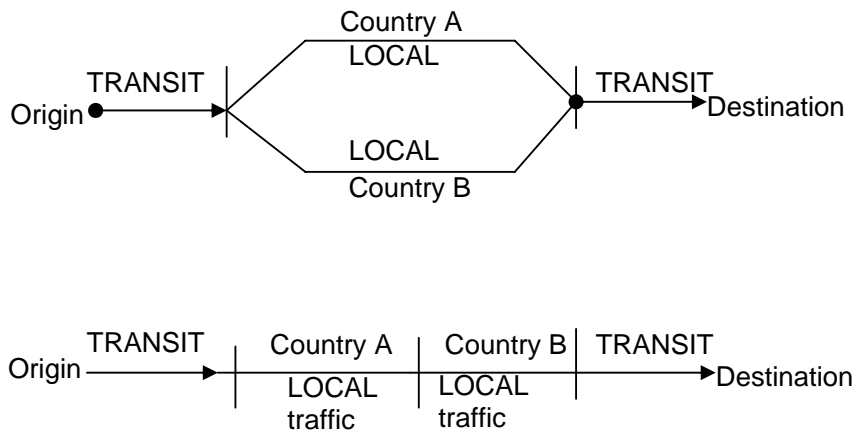
The structure of the paper is as follows. In Section 2 we introduce the notation, the networks analyzed and the model used. We then look at the simplified case of zero local demand to get some preliminary intuition on the problem of competition for transit in serial and parallel settings (Section 3). Theoretical results obtained for the more general case with local and transit demand are summarized in Section 4. Numerical illustrations and a careful comparison of the welfare effects of tax and capacity competition in different networks are given in Section 5. Finally, Section 6 offers some tentative policy conclusions.

2. Model description

The models we used for the study of tax competition on simple transport networks can be summarized as follows. First, two very stylized types of network structure have been considered; they are illustrated on Figure 1. As argued in the introduction, we distinguish parallel and serial network structures as very simple descriptions of real-world toll and capacity competition problems. The first case arises when different parallel links can be used to make a particular trip; each link is used by both local traffic and transit traffic (through traffic) and is operated by a different authority (say, a government). Moreover, transit traffic has a choice between the different parallel links, so that governments compete for transit tax revenues. The second prototype network arises when transit has to use a route that sequentially runs through the territory of different governments. The existence of these ‘transport corridors’ leads to a different type of tax competition: transit no longer has any route choice, but the same transit traffic may sequentially be taxed by each of the

governments operating the serial links. Note that in the two network types origins and destinations are assumed to lie outside the network.

Figure 1 Parallel (upper part) versus serial (lower part) competition



We assume that governments are interested in maximizing a (local) social welfare function that reflects two concerns, viz. (i) the travel conditions of its local users and the associated welfare, and (ii) total tax revenues on the link it controls. The assumption that transit traffic has its origin and destination outside the two-link network implies that the two governments are not interested in the transport costs and the welfare of transit. Finally, we assume that all traffic flows are uniformly distributed over time and are equal in both directions, allowing us to focus on one representative unit period and one direction.

To model both parallel and serial settings, we proceed as follows. We denote the network links running through the territory of governments A and B by appropriate subscripts. We assume each link carries local traffic and transit traffic. Local traffic uses only the local link. Transit traffic chooses one of the links (parallel case) or passes through the two links. The capacity of each link can be augmented through investments; however, once capacity is chosen for a given link it is potentially congestible.

Demand for local transport in regions A and B is represented by the strictly downward sloping and twice differentiable inverse demand functions $P_A^Y(Y_A)$ and $P_B^Y(Y_B)$, respectively, where Y_A and Y_B are the local flows on both links. As is common in the transport literature, the prices $P_i^j(\cdot)$ are generalised prices including

resource costs, time costs and toll payments. Similarly, overall demand for transit traffic is described by the strictly downward sloping inverse demand function $P^X(X)$, where X is the transit traffic flow. Importantly, the treatment of transit differs for parallel and serial settings. We have the following definitions:

$$\text{Parallel links} \quad X = X_A + X_B$$

$$\text{Serial links} \quad X = X_A = X_B$$

In the case of parallel links, total demand for transit is ‘distributed’ over the two alternatives; with serial links, all transit passes through both regions A and B.

We now turn to the cost side. Although other tolling regimes will be considered (see below), here we formulate cost functions for the case of differentiated tolls between local transport and transit traffic. In that case, the generalised user cost functions for local use of links A and B are given by, respectively:

$$g_A^Y = C_A(V_A R_A) + t_A.$$

$$g_B^Y = C_B(V_B R_B) + t_B.$$

Here, the $C_i(\cdot)$ are the time plus resource costs on link i , and R_i is the inverse of capacity. The user cost function is twice differentiable and strictly increasing in $V_i R_i$, the total traffic volume relative to capacity. Making time costs a function of volume-capacity ratio is a common practice in transport economics (see, e.g., Verhoef et al. (1996)). The t_i are the tolls on local transport. Similarly, the generalised user cost for transit through region i ($i=A,B$), denoted as g_i^X , equals the sum of the time and resource costs of travel plus the transit tolls, denoted τ_i , in both A and B:

$$g_A^X = C_A(V_A R_A) + \tau_A$$

$$g_B^X = C_B(V_B R_B) + \tau_B$$

The transport user equilibrium is defined by equating generalized prices and generalized costs. In the parallel case, it is assumed that from the viewpoint of transit the two routes are perfect substitutes; moreover, we focus on internal solutions throughout, i.e., we exclude the case where one link is not used at all. Under those conditions the transport user equilibria for the serial and parallel networks can be summarized as follows:

$$\begin{aligned}
\text{Serial network} \quad P_A^Y(Y_A) &= g_A^Y \\
P_B^Y(Y_B) &= g_B^Y \\
P^X(X) &= g_A^X + g_B^X
\end{aligned} \tag{1}$$

$$\begin{aligned}
\text{Parallel network} \quad P_A^Y(Y_A) &= g_A^Y \\
P_B^Y(Y_B) &= g_B^Y \\
P^X(X) &= g_A^X = g_B^X
\end{aligned} \tag{2}$$

The user equilibrium requires generalized prices and generalized cost to be equal for all traffic types. The equilibrium conditions (1) and (2) can be solved for the demands for local and transit traffic as a function of taxes and capacities in both countries. In the case of a serial network structure, solving (1) and using the definitions of the generalized user costs given before, yields the demand functions¹:

$$\begin{aligned}
&Y_A^r(t_A, t_B, \tau_A + \tau_B, R_A, R_B), Y_B^r(t_A, t_B, \tau_A + \tau_B, R_A, R_B) \\
&X^r(t_A, t_B, \tau_A + \tau_B, R_A, R_B)
\end{aligned}$$

Similarly, solving (2) implies demand functions:

$$\begin{aligned}
&Y_A^r(t_A, t_B, \tau_A, \tau_B, R_A, R_B), Y_B^r(t_A, t_B, \tau_A, \tau_B, R_A, R_B) \\
&X_A^r(t_A, t_B, \tau_A, \tau_B, R_A, R_B), X_B^r(t_A, t_B, \tau_A, \tau_B, R_A, R_B)
\end{aligned}$$

Note the difference. The serial structure implies demands that depend on local tolls and on the total toll paid by transit. The parallel structure has demands that depend on local tolls and on the individual tolls for transit in A and B.

As argued above, the analysis of tax and capacity competition is studied for several tax regimes. They are summarized in Table 2. We distinguish between:

- (i) Different tolls on local and transit traffic; this is the case explained above, where we used τ_i and t_i for the toll on transit and local demand in region i , respectively ($i=A,B$).

¹ These are “reduced” demand functions indicated with a superscript r where the effects of government control parameters (tolls and capacity) on the congestion levels via transit and local demands are integrated.

- (ii) Uniform tolls on local and transit transport. Uniform tolls are denoted $\theta_i = \tau_i = t_i$.
- (iii) The case where transit remains un-tolled.

TYPE OF TOLLING	DESCRIPTION
Toll differentiation	Local users are tolled differently than transit users
Uniform toll	Local users and transit users pay the same toll
Local traffic only	Only local users are tolled

Table 2: Different types of tolling²

Finally, except otherwise noted, we assume that the governments are interested in maximizing a welfare function that consists of the consumer surplus for its local users plus all tax revenues generated on local and transit demand, net of investment costs associated with capacity provision. As an example, in the case of differentiated tolls the objective function for region A is given by:

$$W_A = \int_0^{Y_A} (P_A^Y(y))dy - g_A^Y Y_A + t_A Y_A + \tau_A X - K_A \frac{1}{R_A} \quad (3)$$

where K_A is the unit cost of capacity expansion. This specification implies that we assume constant returns to scale in capacity extension throughout. Note that the same objective function is used for serial and parallel networks. The difference is situated in the demand functions (see above). Finally, note that its generalisation to the cases of uniform tolls or local tolls only is obvious.

3. Toll and capacity competition in parallel and serial networks: the case of zero local demand.

To set the stage for the more general tax and capacity competition on simple transport networks, we start the analysis by considering a simplified case, viz. the case of zero local transport. In that case, there is no interaction between local and transit traffic, and the complexity of tolling and capacity interaction between regions is

² Note that the three types of tolling regimes have potential policy relevance in a federal structure such as, e.g., the European Union. For discussion of the tolling regimes, see De Borger, Proost and Van Dender (2005).

easier to analyze. Moreover, to analyze the two-stage Nash game it will be instructive to introduce simple functional forms for demand and cost functions³. Specifically, we assume all demand and cost functions are linear. Demands are given by:

$$\begin{aligned}
 P^X(X) &= a - bX \\
 P_A^Y(Y_A) &= c_A - d_A Y_A \\
 P_B^Y(Y_B) &= c_B - d_B Y_B \\
 &\text{with } a, b, c_A, d_A, c_B, d_B > 0
 \end{aligned}$$

Cost functions for transport time (and resources) are specified as:

$$\begin{aligned}
 C_A(X + Y_A) &= \alpha_A + \beta_A R_A(X + Y_A) \\
 C_B(X + Y_B) &= \alpha_B + \beta_B R_B(X + Y_B) \\
 &\text{where } \alpha_i, \beta_i > 0, i = A, B
 \end{aligned}$$

Note that demands and costs are linear in generalized prices and volume-capacity ratios, respectively. In the zero local demand case $Y_A = Y_B = 0$.

Despite its simplicity, the analysis in this section is useful to get some preliminary insights. Some of these are interesting in their own right, and they have not been derived in the literature before. It highlights the crucial role of the network structure for toll and capacity competition. We first deal with the serial case, then study the parallel one. Surprisingly, in the former case there is a simple closed form solution where pricing and capacity Nash equilibria are in fact totally independent. On the contrary, in the parallel case even the simple problem modelled here generates complex interactions between the capacity and pricing stage of the game, and it implies the possibility of multiple equilibria at the capacity stage. These findings will be useful to interpret the more general cases with both local and transit transport, as discussed later in the paper (see Section 4).

3.1. Zero local demand: the serial case

Consider the simple case where there is no local transport on either of the two serial links. In fact, this may have some policy-relevance for small countries, where local transport on part of the network is almost negligible (e.g., the highway passing through Luxemburg; some rail connections). Under those conditions, noting that

³ Note that these specifications will be used for all tax regimes considered as well, see below.

transit demand is by definition equal in both regions, the cost functions for transport time (and resources) in regions A and B reduce to:

$$\begin{aligned} C_A(X) &= \alpha_A + \beta_A R_A X \\ C_B(X) &= \alpha_B + \beta_B R_B X \end{aligned}$$

Using the equilibrium conditions for transit we then easily show that the reduced form transit demand (i.e., the demand for transit as a function of tolls and capacities in both regions) is given by:

$$X^r = \frac{(a - \alpha_A - \alpha_B) - \tau_A - \tau_B}{N}, \text{ where } N = b + \beta_A R_A + \beta_B R_B \quad (4)$$

Note that the objective in this simplified case consists of maximizing the transit tax revenues minus capacity costs for each region by an appropriate choice of tax and capacity. Indeed, if local demand is zero, the objective function (3) for region A reduces to:

$$W_A = \tau_A X - K_A \frac{1}{R_A}$$

We solve the two-stage price-capacity game by backwards induction. First consider the pricing game for given capacities. The first-order condition with respect to the price in region A is:

$$\tau_A \frac{\partial X^r}{\partial \tau_A} + X^r = 0$$

Solving for the tax rate in A immediately yields:

$$\tau_A = \beta_A R_A X + (b + \beta_B R_B) X$$

Note that $\beta_A R_A X$ can be interpreted as the marginal external cost of congestion in region A. Then the optimal pricing rule (5) shows that the toll always exceeds the local marginal external cost. In fact, it implies more than that: the toll in region A covers more than the marginal external congestion cost in A plus the one in B. This phenomenon is similar to the issue of double marginalization in industrial organization (Tirole (1993)), Bresnahan and Reiss (1985)). It suggests that one expects relatively high tolls on transit in serial networks. Importantly, also note another implication. The optimal toll in A is equally affected by congestion in A and B: the effect of an exogenous increase in $\beta_A R_A X$ is the same as for an increase in $\beta_B R_B X$. This follows from the serial network setting. Higher congestion in either A

or B raises the generalized cost of the complete trip and hence has the same effect on transit demand. It therefore triggers the same price response in a given region.

Substituting the expression derived for X^r , equation (4), in the toll rule (5) and solving for the tax reaction function, we find:

$$\tau_A = \frac{a - \alpha_A - \alpha_B}{2} - \frac{1}{2} \tau_B \quad (6)$$

Two issues stand out. First, the reaction function is downward sloping in the toll charged by the other region. In particular, it implies that if one region raises its toll by one euro, then the overall toll on the whole trajectory increases by precisely 0.5 euro. This is a well known result in the vertical integration literature in industrial organization, where a cost increase at the downstream level is only partially reflected in final output prices. Second and surprisingly, note that the reaction function is independent of capacities and of congestion: it neither depends on the R_i , nor on the β_i . The reason is that regional congestion affects the toll in both regions equally; as a consequence, it does not affect the interaction in tolling behavior between the two regions.

Using the analogous expression for the tax rate in B and solving for the Nash equilibrium yields:

$$\tau_A^{NE} = \tau_B^{NE} = \frac{a - \alpha_A - \alpha_B}{3} \quad (7)$$

The structure of this Nash equilibrium closely resembles the standard private duopoly result (Tirole (1993)). It shows that the standard ‘double marginalization’ result still holds in the presence of congestion. Moreover, it has powerful implications. It means that the only Nash equilibrium in tolls: (i) is symmetric, even if the free-flow cost parameters differ; (ii) is independent of capacities, and (iii) is independent of the slope of the congestion function, so that tolls are not used to control local congestion.

We now proceed to the first stage of the game, i.e., the game in capacities. The first order condition for optimal capacity choice in region A is:

$$\tau_A^{NE} \frac{\partial X^r}{\partial R_A} - X \frac{\partial \tau_A^{NE}}{\partial R_A} + \frac{K_A}{R_A^2} = 0 \quad (8)$$

The second term on the right hand side equals zero, because the taxes are independent of capacities. Working out the derivative of expression (4) with respect to inverse

capacity, substituting in (8), and using the definition of N, we find the capacity reaction function:

$$\psi(R_A, R_B) = \beta_A \tau_A \left[\frac{\tau_A + \tau_B - (a - \alpha_A - \alpha_B)}{(b + \beta_A R_A + \beta_B R_B)^2} \right] + \frac{K_A}{R_A^2} = 0 \quad (9)$$

This implicit function $\psi(R_A, R_B) = 0$ defines the reaction function of capacity in A with respect to capacity in B. Using the implicit function theorem, the slope of the reaction function can be written as, after simple algebra:

$$\frac{\partial R_A^R}{\partial R_B} = -\frac{1}{\psi_{R_A}} * \left[-2\beta_B \beta_A \tau_A \left(\frac{\tau_A + \tau_B - (a - \alpha_A - \alpha_B)}{(b + \beta_A R_A + \beta_B R_B)^3} \right) \right] \quad (10)$$

where $\psi_{R_A} < 0$ by the second order conditions for optimal capacity choice. It then follows that, given optimal taxes, the reaction functions are unambiguously positively sloped. To see this, it suffices to use the Nash equilibrium tax expressions derived above, so that $\tau_A + \tau_B - (a - \alpha_A - \alpha_B) < 0$ follows.

Positively sloped capacity reaction functions make sense: optimal capacity choice by A implies equality between marginal capacity costs and marginal revenue of capacity expansion. Now consider an increase in capacity in B. This certainly raises transit demand and hence (since taxes are independent of capacity) tax revenues in A. More importantly, however, given the demand function for transit derived above, it easily follows that the capacity change in B also raises the marginal revenue from an expansion at A. Given the constant marginal cost of capacity expansion, the increase in marginal revenue justifies a capacity expansion.

3.2. Zero local demand: the parallel case

The parallel case for zero local demand has been studied for revenue maximizing authorities by De Borger and Van Dender (2006) and De Palma and Proost (2006), although they do so in a somewhat different setting of competition between congestible private facilities. In our setting, cost functions for transport time (and resources) in regions A and B reduce to:

$$\begin{aligned} C_A(X_A) &= \alpha_A + \beta_A R_A X_A \\ C_B(X_B) &= \alpha_B + \beta_B R_B X_B \end{aligned}$$

Noting that $X = X_A + X_B$ the equilibrium conditions can be solved for the demand functions for transit via A and B, respectively. We find:

$$X_A^r = \frac{1}{M} [(a - \alpha_A - \tau_A)(b + \beta_B R_B) - b(a - \alpha_B - \tau_B)] \quad (11)$$

$$X_B = \frac{1}{M} [(a - \alpha_B - \tau_B)(b + \beta_A R_A) - b(a - \alpha_A - \tau_A)] \quad (12)$$

where $M = b(\beta_A R_A + \beta_B R_B) + \beta_A R_A \beta_B R_B$.

Consider the pricing game at given capacities. Solving the first-order condition for region A, $\tau_A \frac{\partial X_A^r}{\partial \tau_A} + X_A^r = 0$, leads to the following rule:

$$\tau_A = \beta_A R_A X_A + \frac{b \beta_B R_B X_A}{b + \beta_B R_B} \quad (13)$$

A similar expression is derived for B. Since the first term on the right hand side is the marginal external cost of congestion, this states that the toll will exceed the marginal congestion cost. This has been interpreted as saying that congestion generates power, in the sense that it allows revenue maximizing operators to raise tolls: in a parallel structure, a higher toll in A raises congestion in the competing region B, making region A more attractive (Verhoef et al. (1996), Van Dender (2005)). Note that investment and congestion now have different effects on tolling behaviour, depending on where the investment takes place.

Using the expression for X_A (see equation (4)) in (13), the toll reaction function for region A is readily obtained as:

$$\tau_A = \frac{Z_A}{2(b + \beta_B R_B)} + \frac{b}{2(b + \beta_B R_B)} \tau_B \quad (14)$$

where $Z_A = \beta_B R_B (a - \alpha_A) + b(\alpha_A - \alpha_B)$ is a function of demand and cost parameters. It follows that the reaction function is upward sloping; moreover, both its slope and its intercept explicitly depend on capacity in the competing region. This contrasts with the serial case analyzed before.

Solving the reaction function and its counterpart for region B for the Nash equilibrium yields (where Z_B is defined analogously as for region A):

$$\tau_A^{NE} = \frac{2Z_A(b^2 + M) + Z_B(b + \beta_B R_B)}{(3b^3 + 4M)(b + \beta_B R_B)} \quad (15)$$

$$\tau_B^{NE} = \frac{2Z_B(b^2 + M) + Z_A(b + \beta_A R_A)}{(3b^3 + 4M)(b + \beta_A R_A)} \quad (16)$$

Simple differentiation of the Nash equilibrium tolls with respect to the R_i shows that higher capacity in any given region induces both regions to reduce tolls. We have:

$$\frac{\partial \tau_A^{NE}}{\partial R_i} > 0, \quad i = A, B$$

In other words, a less congestible parallel network leads both competing authorities to reduce their tolls.

Unlike in the serial case, results for the capacity game are not straightforward. The dependency of tolls on capacities implies that the reaction functions in capacities implied by the first-order condition of the first-stage of the game, viz.

$$\tau_A^{NE} \frac{\partial X}{\partial R_A} - X \frac{\partial \tau_A^{NE}}{\partial R_A} + \frac{K_A}{R_A^2} = 0$$

are highly non-linear. It is shown in De Borger and Van Dender (2006) that this reaction function is plausibly downward sloping but that the nonlinearity implies the possibility of multiple equilibria. In fact, they show that asymmetric outcomes are more likely when unit capacity costs are low and/or structural transit demand is relatively inelastic. The interpretation of such an asymmetric equilibrium is quite intuitive. One region invests highly in capacity but also charges high tolls, so that congestion is low. The competing region provides much less infrastructure but also charges low tolls, so that congestion is much higher. Endogenously, toll and capacity competition induce regions to offer distinct packages, implying different ‘quality’ levels at different ‘prices’.

3.3. Zero local demands: summary and conclusion.

Tentative conclusions for the simple cases without local traffic are summarized in Table 3. First, toll reaction functions and capacity reaction functions have opposite signs in serial and parallel settings, as could be expected. Second, serial competition implies that toll and capacity decisions are independent, unlike for the parallel case. For the latter network structure, higher capacity in any given region induces both regions to reduce tolls. Third, in the serial network, double marginalization is likely to yield higher tolls on transit than in a parallel structure.

Finally, the parallel case may result in asymmetric equilibria even for the simple set-up considered above.

	Parallel links	Serial links
Toll reaction functions	Upward sloping: $\frac{\partial \tau_A}{\partial \tau_B} > 0$	Downward sloping: $\frac{\partial \tau_A}{\partial \tau_B} < 0$
Impact of capacity increase in A on tolls	More capacity reduces tolls: $\frac{\partial \tau_A^{NE}}{\partial cap_A} < 0, \frac{\partial \tau_B^{NE}}{\partial cap_A} < 0$	Tolls independent of capacities: $\frac{\partial \tau_A^{NE}}{\partial cap_A} = 0, \frac{\partial \tau_B^{NE}}{\partial cap_A} = 0$
Capacity reaction functions	Downward sloping: $\frac{\partial R_A}{\partial R_B} < 0$	Upward sloping: $\frac{\partial R_A}{\partial R_B} > 0$

Table 3: Characteristics of reaction functions and Nash equilibrium in the case of zero local transport

4. Tax and capacity competition in transport networks: some general theoretical findings

In this section we extend and summarize the main theoretical findings on tax and capacity competition in parallel and serial networks for the more realistic case with both local and transit demand on each link. As we will see, many, but not all, of the results summarized in Table 3 for the case with zero local demand will continue to hold. The findings reported in this section come from different sources. First, optimal toll results at given capacities were derived for the parallel network case in De Borger, Proost, Van Dender (2005). Second, both toll and capacity competition in a serial corridor were studied in De Borger, Dunkerley and Proost (2006a). Third, for the purpose of the current paper we extended the analysis for the parallel case to incorporate capacity choices, using the same methodology as in the serial setting.

Since we limit the discussion here to a summary of the main findings, we refer to the papers mentioned for more details on their derivation. We proceed in several steps. We first briefly discuss the effects of tolls and capacities on the reduced-form

demands for local and transit transport. These demands are the solution of the user equilibrium conditions (1) and (2) as a function of all tolls and capacities in both regions. We then briefly discuss findings on the toll reaction functions in parallel and serial networks, and we report on what can be learned about strategic capacity choices.

4.1. The effect of tolls and capacities on the demand for local and transit transport

In tables 4a and 4b we summarize the results that describe the effects of tolls and capacity increases on the equilibrium demands for local and transit transport under both a parallel and serial setting (De Borger et al. (2005, 2006a)). Note that we assume interior solutions throughout; in the parallel case, this implies that both links are used in equilibrium.

Our findings are plausible and easily summarized. First, although all own price effects are obviously negative, cross price effects depend in an intuitive way on network structure; routes are substitutes or complements depending on network design. For example, a toll on transit in region A raises transit demand in a parallel setting because transit shifts from A to B. Moreover, this in turn raises congestion in B and hence reduces demand for local traffic in that region. In a serial setting, however, higher tolls on transit in A reduce transit demand throughout the corridor; the decline in congestion in B then raises demand for local transport in that region. Similarly, in a parallel setting, raising local tolls in A attracts transit to A and hence reduces transit demand in B. It therefore raises local demand in B because of declining congestion. In a serial network the same toll increase in A raises transit through B and hence reduces local demand there.

Second, consider the impact of capacity investments on demand. Again, results differ according to network design. Capacity investments in A raise demand for both transit and local demand in A, but the impact on demand in B depends on network structure. A serial setting yields more transit through B and hence less local demand there; a parallel setting shifts transit from B to A and raises local demand in B because of lower congestion.

	Effect on transit demand in A	Effect on transit demand in B	Effect on local demand in A	Effect on local demand in B
Toll on transit in A	<0	>0	>0	<0
Toll on local demand in A	>0	<0	<0	>0
Uniform toll on both local demand and transit in A	<0	>0	<0	<0
Capacity increase in A	>0	<0	>0	>0

Table 4a: Demand effects of tolls and capacity investment in parallel networks (effects of toll or capacity changes in one region on demand, holding all other tolls and capacities in both regions constant)

	Effect on transit demand in A and B	Effect on local demand in A	Effect on local demand in B
Toll on transit in A	<0	>0	>0
Toll on local demand in A	>0	<0	<0
Uniform toll on local demand and transit in A	<0	<0	>0
Capacity increase in A	>0	>0	<0

Table 4b: Demand effects of tolls and capacity investment in serial networks (effects of toll or capacity changes in one region on demand, holding all other tolls and capacities in both regions constant; note that transit demand in A equals that in B by definition)

4.2. Strategic tolling behaviour

In this sub-section we summarize what can be learnt about the tolling behaviour of countries under different network structures. We first describe the optimal toll rules in a given region, at given capacities and for given tolls set by the other region. Next we describe the characteristics of the toll reaction functions. We assume throughout that each regional government maximizes consumer surplus for its

local users plus all tax revenues generated on local and transit demand, net of investment costs associated with capacity provision. For example, in the case of differentiated tolls region A maximizes its objective function, given by (3), with respect to t_A, τ_A . Note that this optimization problem is the same for parallel and serial network settings; as argued before; what differs is the demand structure. Moreover, the generalisation to the cases of uniform tolls or local tolls only is obvious.

The optimal toll rules are derived in De Borger et al. (2005, 2006a). Interestingly, they are the same for parallel and serial settings, although (see below) they have very different practical implications depending on the network structure. The toll rules imply tax exporting behaviour: if regions can differentiate tolls between local and transit demand, then they will toll transit at a higher rate than local demand; if tolls are restricted to be uniform, then the optimal toll positively depends on the importance of transit. Moreover, the use of local tolls strongly depends on the instruments available. If transit can be tolled, then tolls on local traffic are set higher than the local marginal external congestion cost in order to reduce congestion on the local link and hence indirectly attract more transit (tax competition for transit). If, however, local tolls are the only instrument, then these tolls are set below local marginal external cost. The reason is that by doing so regions reduce transit demand; the latter generates congestion and does not contribute to welfare nor tax revenues.

There is some scarce empirical evidence that supports these theoretical predictions. For example, Nash (2005) reports very high rail rates in Switzerland and in former communist countries of the EU (e.g., Slovakia), pointing at double marginalization as a possible explanation⁴. His findings also suggest a relation between the share of transit and the level of infrastructure charges on European rail links. Moreover, he warns that levels of charges may be prohibitively high for international traffic involved in transit of a country.

The optimal toll rules for a given region implicitly describe the toll reaction functions. To illustrate the importance of network structure for strategic behaviour by regions in the simplest possible way, it is instructive to return to the case of linear demands and costs, presented in section 3 above. In the serial case, and focusing on

⁴ Also see Bassanini and Pouyet (2005). Interestingly, Nash (2005) argues that EU-Directive 2001/14 for rail is to be interpreted as not allowing discrimination by different operators for a given type of traffic, but that discrimination by one operator for different types of traffic does seem to be allowed. In other words, toll discrimination is not just theoretically interesting, but may have practical relevance.

differentiated tolls for local and transit transport, the reaction functions can then be shown to have the following structure (De Borger et al. (2006a)):

$$\tau_A = c_A^\tau - \left(\frac{1}{2}\right)\tau_B - \left(\frac{1}{2}z_1^B\right)t_B \quad (17)$$

$$t_A = c_A^t + \left(\frac{1}{2}L^A\right)\tau_B + \left(\frac{1}{2}z_1^B L^A\right)t_B \quad (18)$$

where the parameters c_A^τ , c_A^t , z_1^B and L^A are all functions of demand and cost parameters. Moreover, z_1^B (where $z_1^B < 0$ and less than one in absolute value) gives the effect of an exogenous increase in transit transport in region B on the demand for local transport in that region. Finally, $-1 < L^A < 0$. Note that in the absence of local demand these expressions are consistent with the results described in section 3. To see this, compare equation (17) with (6).

The interpretation of the effect of toll changes in region B on optimal tolls in A is then clear. We find that an increase in the transit tax in B induces region A to optimally reduce both its transit tax and the tax on local traffic. The higher tax on transit in B reduces transit demand and hence reduces congestion in A. The optimal response in A is therefore to reduce both taxes. Similarly, a higher local tax in B induces region A to optimally raise transit as well as local taxes in A. The higher local tax in B reduces congestion in B, and attracts more transit. This also raises congestion in A. Therefore, country A raises its tax rates on all traffic on its territory.

As argued before, the structure of the reaction functions bears some close resemblance to well known results in industrial organisation. For example, it implies that an increase in the transit toll in one region by one unit raises the total toll on transit for the whole trajectory by less than one unit. This well known result on the pricing behaviour of successive monopolies (see Bresnahan and Reiss (1985), Tirole (1993)) persists when adding local transport.

The structure of the reaction functions for the parallel links are quite similar. We derived elsewhere (De Borger et al. (2005)):

$$\tau_A = d_A^\tau - \left(\frac{1}{2}\delta\right)\tau_B - \left(\frac{1}{2}\delta z_1^B\right)t_B \quad (19)$$

$$t_A = d_A^t + \left(\frac{1}{2}\delta K^A\right)\tau_B + \left(\frac{1}{2}\delta z_1^B K^A\right)t_B \quad (20)$$

where the coefficients depend on demand and cost parameters and:

$$\delta < 0, \quad 0 < |\delta| < 1$$

$$-1 < K^A < 0$$

The interpretation of the effects of changes in foreign taxes on optimal local taxes in A is then clear. We find that an increase in the transit tax abroad induces country A to optimally adjust both its transit tax and the tax on local traffic upwards, but that the impact on the transit tax is larger than the effect on the local tax. Why is this the case? The higher tax on transit in B reduces transit there and raises transit demand in A. This increases local congestion in A. The optimal response in A is therefore to raise both taxes. Similarly, a higher local tax in B induces country A to optimally reduce transit as well as local taxes in A. The higher local tax in B reduces congestion in B, and makes B relatively more – and A relatively less – attractive to transit traffic. This also reduces both congestion and tax revenues in A. To compensate, country A raises its tax rate on local traffic; this reduces congestion but raises tax revenues.

What do we learn from these reaction function specifications? First, comparing parallel and serial cases, we see that the slopes of reaction functions are of opposite signs, as one would expect. Second, in setting transit tolls regions react more strongly to toll changes abroad in serial networks than in parallel settings. This follows from $|\delta| < 1$. It implies that one expects more tax competition and tax exporting behaviour on serial networks than on parallel networks. Third, the strongest interaction between regions is in the transit tolls. Changes in local tolls have much less effect on other regions. In fact, the strategic interaction in local tolls is almost negligible. Economically, it makes sense. Local tolls only affect local tolls abroad via their impact on congestion and the shift in transit to the other region but there is no direct tax competition as in the case of transit.

4.3. Capacity reaction functions

Unfortunately, few general theoretical results could be derived on the nature of capacity competition, largely due to the complexity of the dependency of Nash equilibrium tolls on capacities in both regions. This dependence was already observed for the parallel network structure in the case of zero local demand, studied in section 3. Moreover, in the serial setting, the independence of Nash equilibrium tolls and

capacities that was observed in the case of zero local demand does not hold when this assumption is relaxed.

However, the scarce theoretical results as well as findings based on numerical work (see, among others, De Borger et al. (2006a), De Borger and Van Dender (2006)) lead to the following predictions⁵. First, consider the effect of capacity changes at the first stage on the Nash equilibrium tolls at the second stage. In a serial setting, capacity increases in region A reduce Nash equilibrium tolls on both transit and local demand in A. In B it will lead to more congestion and therefore higher tolls. In a parallel network, we see the opposite. A capacity increase in A leads to toll reductions there because of lower congestion but congestion in B will also be reduced, hence reducing tolls. Second, consider capacity reaction functions. In serial networks, capacities are strategic complements: capacity reaction functions are plausibly upward sloping. More capacity in A raises congestion in B, inducing this region to raise capacity as well. Parallel settings imply, on the contrary, that capacities are likely to be strategic substitutes: capacity reaction functions are plausibly downward sloping. More capacity in a region attracts transit from the other region, reducing the capacity requirements in that region. So the predictions reported in Table 3 for the case of zero local demand are likely to generalize to the situation with both transit and local demand.

5. Numerical illustration

This section presents some illustrative results based on numerical simulation analysis that allow us to compare the nature and extent of toll and capacity competition on simple serial and parallel networks. We first describe the calibration of the numerical illustration (subsection 5.1). Then we proceed to discuss the price setting and investment behaviour in the serial and parallel case. We consecutively analyse the efficiency of the zero toll Nash equilibrium capacity choices (subsection 5.2), the desirability of allowing the tolling of transit by differentiated or uniform tolls (subsection 5.3), and the welfare effects if only local transport can be tolled (subsection 5.4). Finally, we report results for the coordinated solution that would be welfare maximizing from the viewpoint of a federal authority that coordinates the

⁵ Moreover, note that they are corroborated by the empirical findings reported below.

whole network and, hence, avoids toll and capacity competition between regions. Throughout the focus is on the importance of the different network structures for the results.

5.1 Calibration of the reference case

We have chosen a numerical example with a maximum of comparability between the parallel and the serial network. For the sake of clarity, we limit ourselves to the symmetric case with two identical regions. The calibration process is illustrated using Table 5. The model is initially calibrated for the serial case with zero tolls. As explained in more detail below, capacity is chosen such that it indeed reflects a Nash equilibrium for the zero toll case in each region. As is clear from the table, we use the same local demand functions (and values of time), the same congestion technology and the same cost of capacity in the parallel and serial case. However, since the transit flows pass through two regions in the serial case and only through one region in the parallel case, the transit demand functions differ for the two cases: this is necessary because we want the total flows inside each country to be comparable for identical capacities. Note that the calibration procedure implies that the first-best federal optimum is identical for the serial and parallel network structures (see below). This is a desirable feature in view of the comparative nature of this paper.

The calibration process starts with the reference data for the serial case given in the lower left part of table 5. We choose local and transit flows that have a similar order of magnitude; moreover, reference time costs are of the same order of magnitude as the non time costs. The level of congestion in the reference equilibrium was such that the time costs were 50% higher than at zero traffic. This yields a generalised cost and a time cost in the reference equilibrium, as well as two points for the congestion function, so that the intercept α and the slope (which at constant capacity equals βR) can be determined. To complete the calibration of the demand functions, we have chosen an elasticity of local demand equal to -0.3. Finally, reference capacity was fixed at 2000. Since R is inverse capacity this, together with the slope of the congestion function, determines β . As suggested before, to facilitate the comparison with the other regimes that will be studied, it is assumed that the chosen capacity (2000) in the zero toll serial case is indeed the Nash equilibrium for each of the regions. This is done by determining the cost of capacity extension K in

such a way that this indeed holds. This completes the calibration of the serial network case.

For the parallel case, we use the same parameters except for the transit demand function; it has a steeper slope in order to have comparable flows on the different segments of the network. The calibration leads to the reference demands, generalized prices and capacities reported in the lower right part of Table 5.

	Serial	Parallel
Demand function local	$P^Y = c - dY$, where $c=283.6$ and $d=0.17$	
Demand function transit	$P^X = a - bX$, where $a=567.1$ and $b=0.33$ $(X = X_A = X_B)$	$P^X = \frac{a}{2} - \frac{b}{4}(X_A + X_B)$
Congestion function	$C_i = \alpha + \beta R_i V_i$ for $i = A, B$, where $V_i = X_i + Y_i$, and $\alpha = 34.3$ and $\beta = 23.9$	
Cost of capacity	$K=18.7$	
Reference equilibrium = the zero toll equilibrium at optimal capacity	$Y = 1300$, $P^Y = 65$ $X = 1300$, $P^X = 130$ Capacity (=1/R) = 2000	$Y = 1206$, $P^Y = 81.3$ $X = 1206$, $P^X = 81.3$ Capacity (=1/R) = 1229

Table 5. Calibration of the numerical example (endogenous elements are in italic)

5.2 How efficient is capacity competition in the zero toll reference case?

The most relevant information on the results for different regimes and for the two different network structures are summarized in Table 6. First consider the no-toll case; as noted before, this case has been used for the calibration and serves as reference here. We see that there remain, both in the serial and parallel networks, important marginal external congestion costs that are not internalised. For example, in the serial case the local and global marginal external congestion costs equal 15.6 and 31.1, respectively. Note that the term ‘local’ refers to the marginal external congestion cost imposed on local users; ‘global’ refers to the overall cost imposed on local users and on transit. An important property of the serial case that was referred to in the theoretical sections of the paper but that is not apparent from this table, is that the capacity levels are strategic complements: whenever one region increases capacity, the other has an interest to follow in order to cope with the increased transit flow. In the parallel case, capacities in the two regions are strategic substitutes. We see that, although flows would be identical if capacity would be identical, the Nash equilibrium has a much lower level of capacity. Indeed, whenever a region (say A)

increases its capacity it attracts extra transit from the other region (say B). Making route A initially cheaper results in arbitrage over the network that produces strong disincentives to increase capacity in the first place.

5.3 Is allowing differentiated tolling welfare improving?

Economists have often advocated the use of tolling instruments to cope with non-internalized congestion. In principle, one can allow differentiated (between local and transit transport) or uniform tolling. Both cases are considered in Table 6, for the serial and parallel network structure. In the serial case, the Nash equilibrium results show that allowing regions to toll all transport on their network (whether by differentiated or uniform tolls) implies a decline in total welfare, i.e., it makes things worse compared to the no tolling case. Overall welfare decreases by 13% to 20% in the uniform and differentiated cases, respectively. Note that overall welfare reported in the table refers to the welfare of all network users; it includes both the welfare of local and of transit users. The reason for the welfare decline is related to the double marginalisation behaviour referred to before. It shows up because the two ‘monopolists’ do not coordinate their price setting of transit transport. As a consequence, we find very high margins on transit in the differentiated toll case; they are well above the marginal external congestion cost. In the uniform case it results in high tolls on all transport on the network. These high prices allow savings on capacity costs: optimal capacity is substantially lower than in the no toll reference situation. However, these savings do not compensate for the losses in consumer surplus, especially for transit users.

In the parallel case, we find the opposite results. Introducing tolling allows the low investment incentives of the no toll case to be overcome and this gives much higher capacity levels (about 2180 compared to 1229 in the zero toll reference). Optimal tolls are positive but, as predicted by the theory presented before, transit tolls are much lower than in the serial case; this holds both for differentiated and uniform

tolls. The consequence of much lower Nash equilibrium tolls on transit implies that in the parallel case overall welfare rises substantially⁶.

Note that, except for the toll levels themselves, differentiated tolling and uniform tolling generate very similar results in the parallel case. In the serial case, however, toll discrimination against transit is a more important problem; it can be mitigated somewhat by imposing uniform tolling.

5.4. Welfare effects of tolling local traffic only

Consider the results for the case where for whatever reason transit remains untolled. When only local traffic can be tolled, one obviously rules out tax exporting. Contrary to the cases where transit was tolled, this implies that on a serial network small welfare improvements are now attained compared to no tolling at all. The toll is slightly below local marginal external cost, and the toll somewhat reduces demand so that a lower capacity is optimal compared to the zero toll case (1945 compared to 2000). In the parallel case the welfare benefits are positive as well, because one can achieve a better use of the network by the local traffic and save some capacity costs. However, as only part of the traffic is actually controlled, the welfare gains that can be achieved remain very small. Also, observe that the optimal local toll is smaller than in the serial case. The reason is that the purpose of the local toll is to indirectly control transit as well as local traffic. Reducing transit by local tolls requires higher tolls in the serial case because transit demand through any given region is only affected via congestion increases. The reaction of transit is stronger in the parallel case because an alternative route is available, unlike in a serial corridor.

5.5. The ideal solution for a federal government: the first-best

Finally, we move away from tax and capacity competition between the two hypothetical regions. Instead, we assume that the whole network is under the control

⁶ Even if the capacity levels would be kept at the zero toll case (not shown in table 6), we found that allowing tolling would be beneficial. The gains of a better use of a given capacity are important, and abuse of a monopoly position by one region is limited by the Bertrand competition with the other region.

of one ‘federal’ government; it decides on tolls and capacity investments for the network by maximizing overall welfare for all users of the whole network.

The results are reported in the final two columns of Table 6. First, at this federal optimum tolls are set equal to the global marginal external congestion cost that takes account of the time losses imposed on both transit and on local traffic. Note that, although tolls can in principle be differentiated, there is no need to do so; the tolls on local and transit transport are equal at the optimum. Second, capacity levels are chosen simultaneously in each region such that the marginal cost of capacity extension equals the marginal benefit for all transport, transit and local. Third, note that except for rounding errors in the calculations, the optimum solutions for the parallel and serial networks are identical. This is due to the fact that the zero toll cases for both network types were calibrated using the same local demand functions, the same values of time, the same congestion functions and the same costs of extra capacity. Finally, given that the federal optimum yields identical tolls, capacities, demands and overall welfare levels for the two network types, the welfare improvement in the parallel case is much more important than on the serial network. This follows from the lower welfare level in the zero toll case for the parallel network.

Summarizing the first-best outcome, marginal external cost pricing and optimal capacity choice for the network as a whole yields much higher benefits in a parallel structure than in a serial setting. Note that the welfare improvements, even for the parallel case, seem rather modest: some 15% compared to the reference situation. This is however due to the fact that the reference itself was calibrated such that it corresponds to the Nash equilibrium with zero tolls but optimal capacity choice.

Table 6. Results from serial and parallel networks - symmetric model with 50% transit, tolls and capacity optimal

Transit share in No tolls system is 50%. The distinction between countries is eliminated because results are symmetric.

Variable	Unit	No tolls		Nash Equilibrium differentiation		Nash Equilibrium uniform		Nash Equilibrium local tolls only		Centralised-differentiation	
		serial	parallel	Serial	parallel	serial	parallel	serial	parallel	serial	parallel
Local demand	Trips	1300	1206	1219	1184	732	1131	1215	1149	1233	1232
Transit demand	Trips	1300	1206	396	1122	732	1131	1301	1207	1233	1232
Trip volume (country)	Trips	2600	2411	1616	2306	1465	2261	2516	2355	2467	2465
Local MEC	Euro/Trip	15,6	23,5	16,8	13,0	10,8	12,4	14,9	22,8	10,6	10,6
Global MEC	Euro/Trip	31,1	46,9	22,3	25,3	21,7	24,8	31,0	46,8	21,1	21,2
Local Toll	Euro/Trip	0	0,0	22,3	25,3	104,7	34,7	14,4	9,7	21,1	21,2
Transit Toll	Euro/Trip	0	0,0	160,4	35,6	104,7	34,7	0,0	0,0	21,1	21,2
Capacity	Trips	2000	1229	1732	2179	1618	2179	1945	1205	2791	2778
Local CS	Euro	141779	121929	124726	117517	45011	107257	123815	110729	127602	127403
Tax revenue (country)	Euro	0	0	90793	69938	153333	78410	17536	11129	52160	52318
Overall welfare	Euro	492348	441783	392658	504750	426209	504402	493956	442952	510409	510407
Welfare change compared to the case of no tolls	%	0	0	-20,25	14,25	-13,43	14,17	0,33	0,26	3,67	15,53

5.6. Summary of the numerical comparison

In Tables 7a and 7b we summarize the main implications of the numerical findings. We observe clear differences in the extent and the nature of tax competition (very severe in the serial case) and capacity competition (very severe in the parallel case). Moreover, welfare benefits differ according to network structure.

Table 7a: Summary of findings serial case (% changes are relative to the reference case without tolling)

	% Welfare change	% Capacity change	Transit toll	Conclusion
First-best federal optimum	+3.67	+40	Toll=MECC	Higher capacity and positive tolls
Nash toll discrimination	-20.25	-13	Toll much larger than MECC	Lower capacity and very high transit tolls; severe toll exporting (double marginalisation); tolling reduces welfare
Nash uniform toll	-13.43	-19	Toll much larger than MECC	Lower capacity and high uniform tolls; severe toll exporting (double marginalisation); tolling reduces welfare
Toll on local demand only	0.33	-3	Toll zero	Lower capacity because it mainly benefits transit

Table 7b: Summary of findings parallel network (% changes are relative to the reference case without any tolling)

	% Welfare change	% Capacity change	Transit toll	Conclusion
First-best federal optimum	+15.53	+125	Toll=MECC	Much higher capacity and positive tolls
Nash toll discrimination	+14.25	+77	Toll somewhat larger than MECC	Much higher capacity and higher transit tolls; tolling raises welfare
Nash uniform toll	+14.17	+77	Toll somewhat larger than MECC	Much higher capacity and higher transit tolls; tolling raises welfare
Toll on local demand only	0.26	-2	Toll zero	Lower capacity because it mainly benefits transit

6. Conclusions

The purpose of this paper was to provide a theoretical and numerical comparison of the toll and capacity competition to be expected on serial and parallel transport networks when regional governments set user charges and determine capacities. Although the networks were very simple, some interesting results could be derived. When deciding on transit tolls regions react more strongly to toll changes abroad on serial networks than in parallel settings. Moreover, tolls on transit are much higher in serial than in parallel settings, reflecting less elastic transit demand and double marginalisation. We further found that capacities are strategic complements in serial settings but substitutes in parallel networks. On a serial corridor, whenever one region increases capacity, the other has an interest to follow in order to cope with the increased transit flow. In the parallel case, we see the opposite, resulting in lower Nash equilibrium levels of capacity. Arbitrage between routes acts as a strong disincentive to increase capacity in this case.

The welfare effects of allowing tolling of transit are drastically different depending on the network structure. In the serial case, the results show that allowing regions to toll all transport on their network (whether by differentiated or uniform tolls) implies a decline in total welfare, i.e., it makes things worse compared to no tolling at all. Compared to the zero toll case, we found that overall welfare decreased by 13% to 20% for uniform and differentiated tolls, respectively. The reason for the welfare decline is related to the double marginalisation behaviour referred to before. It shows up because the two ‘monopolists’ do not coordinate their price setting of transit transport. As a consequence, we find very high margins on transit in the differentiated toll case; they are well above the marginal external congestion cost. These high prices allow savings on capacity costs: optimal capacity is substantially lower than in the no toll reference situation. However, these savings do not compensate for the losses in consumer surplus, especially for transit users. In the parallel case, we find the opposite results. Introducing tolling allows the low investment incentives to be overcome in comparison with a situation where regions do not toll transit at all, yielding higher capacity levels. In combination with much lower Nash equilibrium tolls on transit than in the serial setting, this implies that in the parallel case overall welfare substantially rises when regions are allowed to toll all traffic. Note that, if

only local transport can be tolled, there is a welfare increase on both types of networks. However, as only part of the traffic is actually controlled, the welfare gains that can be achieved remain very small.

We also found that, in a federal optimum in which one government decides on tolls and capacity investments for the whole network, tax and capacity competition can be avoided. At the federal optimum, tolls are set equal to the global marginal external congestion cost that takes account of the time losses imposed on both transit and on local traffic. Note that, although tolls can in principle be differentiated, there is no need to do so; the tolls on local and transit transport are equal at the optimum. Capacity would rise compared to the no toll equilibrium. However, marginal external cost pricing and optimal capacity choice for the network as a whole yields much higher benefits in a parallel structure than in a serial setting.

This paper has clear policy implications for the regulation of user fees by the European Commission and for the European policy on subsidising investments in infrastructure (e.g., the TEN-T axes). The Commission is actually capping the user tolls on motorways to the average infrastructure costs; the main motivation for this was the potential abuse of monopoly power. Our analysis suggests that it is indeed very damaging for European welfare if the EU were to let individual countries freely decide on tolls on the links of the serial TEN's they control. From this perspective restricting the tolls countries can charge is good policy. However, although restricting tolls seems indeed justified, the tolls should be allowed to reflect external congestion costs and not be based on average infrastructure costs. Moreover, our analysis of capacity decisions on transport networks where transit can not be tolled point to capacities being too low. This should be an important element in a federal infrastructure subsidy program. The results suggest that it might be a good idea to relate the provision of subsidies to the importance of transit flows and to the introduction of marginal social cost pricing.

We conclude our analysis by drawing attention to three caveats. First we assumed constant returns to scale in capacity extension. For some modes, such as roads and airports, this is considered reasonable; but for rail and inland waterways decreasing average costs are probably present. In the latter case, mainly the modelling of the capacity decisions needs to be adjusted and this may give rise to corner solutions (see De Borger, Dunkerley, Proost, 2006b). The second caveat is that we only examined proportional pricing solutions. It is well known that two-part pricing,

and more generally non linear pricing, can extract a larger share of the user surplus. Non-linear pricing exists (Eurovignettes or motorway vignette to cross Switzerland), but in our model setting it would require a specification of the different user classes and their demand functions to reach firm conclusions. Finally, we assumed a straightforward objective function for the regional government and a simple non cooperative behavioural setting to derive conclusions. It remains to be tested whether our propositions can be falsified.

References

- Agrell, P. and J. Pouyet, Regulatory competition in network interconnection pricing, Working paper, Département d'Economie, Ecole Polytechnique, Palaiseau, France.
- Bassanini, A. and J. Pouyet. 2005. "Strategic choice of financing systems in regulated and interconnected industries", *Journal of Public Economics* 89, 233-259.
- Braid, Ralph M. 1986. "Duopoly pricing of congested facilities," Columbia Department of Economics Working Paper No. 322
- Bresnahan, T. and P. Reiss. 1985. "Dealer and manufacturer margins", *Rand Journal of Economics* 16, 253-268.
- De Borger, B., Proost, S. and K. Van Dender. 2005. "Congestion and tax competition on a parallel network", *European Economic Review*, Vol 49/8, p 2013-2040
- De Borger, B., Dunkerley, F. and S. Proost. 2006a. "Strategic investment and pricing decisions in a congested transport corridor", under revision for *Journal of Urban Economics*.
- De Borger, B., Dunkerley, F. and S. Proost. 2006b. "Interstate rail versus road", Discussion Paper.
- De Borger, B. and K. Van Dender. 2006. "Prices, capacities and service qualities 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, A. and R. Lindsey. 2000. "Private toll roads: a dynamic equilibrium analysis", *Annals of Regional Science*, 34, 1, 13-35

- de Palma, A and S.Proost, 2006, "Congestion and the city", *Journal of Urban Economics* 60,
- Levinson, D. 2001 "Why States Toll – An empirical model of finance choice", *Journal of Transport Economics and Policy*, 35, 2, 223-238
- Nash, C. 2005. "Rail infrastructure charges in Europe", *Journal of Transport Economics and Policy* 39, 3, 259-278.
- Small, K. and J. Yan. 2001. "The value of "Value Pricing" of roads: second best pricing and product differentiation", *Journal of Urban Economics*, 49, 2, 310-336
- Tirole, Jean. 1993. *Theory of Industrial Organisation*. Boston, MIT Press.
- Van Dender, Kurt. 2005. "Duopoly prices under congested access," *Journal of Regional Science*, 45, 2, 343-362
- 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, 279-302