Allocation of congested rail network capacity: priority rules versus scarcity premiums

Achim I. Czerny*
WHU – Otto Beisheim School of Management
achim.czerny@whu.edu

Kay Mitusch
Berlin Institute of Technology
km@wip.tu-berlin.de

Andreas Tanner
Berlin Institute of Technology
ant@wip.tu-berlin.de

Berlin, February 2, 2009

Abstract

We consider a vertically integrated rail service provider, a congested, capacity limited network, and two customers. One customer demands short- and one long-distance services. The supply is determined by a regulator choosing capacity limits, service charges, and allocation regimes. Regimes can be of two types: (i) a priority rule 'revenue maximization' and (ii) a scarcity premium. Our key results are based on a Monte Carlo simulation. We find that no regime dominates the other one in all respects. In particular, if total surplus is relevant and service charges are low or consumer surplus is relevant, revenue maximization should be preferred.

JEL Classification: D42, D45, H45, L92, R41, R48.

Keywords: Rail network, congestion, service charge, capacity limit, revenue maximization, scarcity premium.

*We thank Felix Höffler for helpful suggestions.

1 Introduction

Rail transport is considered to be more environmentally friendly than road or air transport. For this reason, EU policy aims to increase the use of rail services by supporting competition in the rail industry. However, many national markets inside the EU are dominated by incumbent and vertically integrated rail service providers. Furthermore, duplication of rail networks, as a means to enhance competition between rail network providers, is expensive and not a relevant policy option. Instead, EU policy tries to increase competition between train operating companies on given rail networks. An essential element of this strategy is the regulation of newcomers' access to incumbents' networks.

The basic framework for access regulation has been set by the EU Directive 2001/14 "... on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure ...." According to that directive, rail network providers should serve train operating companies with capacity in return for a minimum access charge. It may be that the demand for certain segments of the rail network exceeds capacity for given minimum access charges; following the EU's diction, these segments are congested. In the case of a congested network segment, the directive proposes the use of a scarcity premium. However, if a scarcity premium is not used, allocation of congested capacity may also be based on priority rules.

(Nash, 2005) provides a survey on rail network charges in Europe. He finds that simple charges per train-kilometer, which may be differentiated by type of traction, weight, speed, and axleload of the train, are most common. Scarcity premiums are rarely used, but some countries move towards this direction. For instance, there is a congestion charge per train-kilometer determined by delay costs in Great Britain, and an extra charge (utilization factor) for heavily used rail links in Germany. Priority rules are also applied in Germany: in a situation with excess demand, priority is provided to trains that increase revenues of the rail network provider (Mitusch and Tanner, 2008).

In this paper, we explore the policy implications of moving from priority rules towards scarcity premiums (as proposed by EU policy makers). In particular, we explore the effects of such a policy change on total surplus, consumer surplus, and the rail network providers' revenues. Different priority rules can be applied in practice, for instance, priority can be provided to trains that increase revenues of the rail network provider (as in Germany), to passenger over freight traffic, or scheduled over non-scheduled traffic. To compare outcomes under scarcity premiums and priority rules we, however, focus on the first option, which we call 'revenue maximization'.

We consider a congested network including two rail links and follow a stationary-state congestion approach, which is commonly used to analyze road or air transport markets.[1] Furthermore, we measure congestion in terms of schedule stability (a schedule is stable if one delayed train would cause no delays to other trains (Goverde, 2007)). Delays are costly for customers (e.g., missed business meetings or interruptions in the production process) and rail service providers (e.g., penalty payments to customers and overtime premiums to employees) and determined by the size of the rail network and the amount of rail services. Furthermore, delays are increasing in the amount of services on a given network, and thus an upper limit for services on a given rail network (capacity limit) also determines an upper limit for delays and delay costs.[2] Note that in this framework congestion is possible even though

[1](Basso and Zhang, 2008) and (Brueckner and van Dender, 2008) use steady-state congestion models in a context of airport markets. (Mun and Ahn, 2008) and (Verhoef, 2008) use steady-state congestion models in a context of road transport markets.

[2]Capacity limits are used to control delays and also for safety reasons.

demand does not exceed the capacity limit, which is in contrast to the EU's notion of congestion.[3]

Moreover, we consider a monopolistic, vertically integrated rail service provider with zero variable costs and positive fixed network-costs. The rail service provider serves two (representative) customers. One customer demands short-distance services and the other one long-distance services. The behavior of the rail service provider is exogenously determined by a regulator who fixes service charges (which are linear in train-kilometers), the capacity limit, and allocation regimes. Regimes can be of two types: (i) a service charge, no scarcity premium, and revenue maximization, and (ii) a service charge and a scarcity premium, which balances demand and capacity supply. We consider a two-stage game in which customers, first, report demand for short- and long-distance services to the rail service provider. Second, the rail service provider determines services according to regimes (i) or (ii). As a benchmark, we also consider "optimal services" that maximize total surplus or, respectively, consumer surplus for given service charges, given scarcity premium, and given capacity limit.

We find that none of the two regimes considered is likely to imply optimal services, no matter whether total surplus or consumer surplus is relevant from a policy viewpoint. Furthermore, it is difficult to directly compare outcomes of regimes (i) and (ii) in a general setting. For this reason, we take resort to a Monte Carlo simulation. The simulation demonstrates that the effect of regimes on total surplus is ambiguous: if access charges are low, total surplus is greater under revenue maximization (and no scarcity premium), and if access charges are high, total surplus is greater under a scarcity premium. Why is total surplus greater under revenue maximization in some situations?

[3]See (Nash, 2005) for a discussion of congestion, congestion charges, scarcity, and scarcity charges or, respectively, scarcity premiums.

Service charges that depend on train-kilometers are greater for long- than for short-distance services, which is to the disadvantage of long-distance services. Clearly, this disadvantage is more relevant if service charges are high. On the other hand, we show that revenue maximization favors long- over short-distance services, which can neutralize the disadvantage from service charges for long-distance services and increases total surplus in some situations. In contrast, a scarcity premium is charged to both, short- and long-distance services, and can not neutralize the disadvantage from service charges for long-distance services. However, if service charges are low, a scarcity premium increases total surplus, because distortions from service charges are less relevant and a scarcity premium favors customers who value services most.

Our simulation results are clear-cut regarding consumer surplus: consumer surplus is greater under revenue maximization than under a scarcity premium. This indicates that, from a customers' viewpoint, revenue maximization should be favored. On the other hand, a scarcity premium increases the rail network provider's revenues and, hence, contributes to cost recovery.

Overall, we find that no regime dominates the other one in all respects. Hence, there is no clear ranking between scarcity premiums or priority criteria from a policy viewpoint. If total surplus is relevant, service charges are low, and cost-recovery is required, a scarcity premium should be the right choice. If total surplus is relevant and service charges are high or consumer surplus is relevant, it should be revenue maximization. It is, however, difficult to understand whether access charges are low or high in practice, which adds to the risk of choosing the wrong regime.

Our contribution is to develop a simple model of a congested rail network and to explore how priority rules and scarcity premiums affect shortand long-distance rail services from a policy viewpoint depending on service charges. Although both regimes considered are not likely to imply optimal services, they do represent regimes that are currently applied and should become more relevant in the future. In contrast, other economists considered auction designs and optimal pricing in the context of rail markets and in more general capacity-limited (but non-congested) networks (see, e.g., (Brewer and Plott, 1996), (Nilsson, 1999), (Parkes, 2001), (Parkes and Ungar, 2001), and (Nilsson, 2002)); for an overview over this strand of literature see (Borndörfer et al., 2006). For a general overview on the relevant theory of package auctions see (Milgrom, 2007).

The structure of this paper is as follows. Section 2 presents the basic model set-up. Section 4 considers rail services under revenue maximization, and Section 5 considers services under a scarcity premium. Section 6 considers the outcome of regimes in terms of total surplus and consumer surplus. Section 7 provides the results of a Monte Carlo simulation. Section 8 offers conclusions.

2 The basic model of a congested rail network

We consider a serial rail network that includes three cities A, B, and C and two rail links connecting cities A and B as well as B and C. Distances between A and B as well as B and C are normalized to one. The rail network is used to offer rail services, which includes transportation of passengers or freight between cities. Rail services going into different directions can be operated independently. It is, therefore, sufficient to focus on one direction. Figure 1 illustrates the rail network.

$$A \xrightarrow{\text{congested}} B \xrightarrow{\text{not congested}} C$$

Figure 1: Three cities, A, B, and C, and a simple rail network including a congested rail link connecting cities A and B and an uncongested rail link connecting cities B and C.

There are two (representative) customers B and C. Customer B demands short-distance services that include rail services between cities A and B, and customer C demands long-distance services that include rail services between cities A and C (from A to C via B). The travel distance for long-distance services is twice as large as the travel distance for short-distance services. The amount of short-distance services is denoted by $q_B \geq 0$ and the amount of long-distance services by $q_C \geq 0$ (subscripts indicate customers and destinations). Customers' pay-offs are

$$B_x(q_x) = a_x\, q_x - \frac{b_x\, q_x^2}{2} \tag{1}$$

for all $x \in \{B, C\}$ with $a, b > 0$. For simplicity, we do not consider short-distance services between cities B and C.

We follow a stationary-state congestion approach and assume that delays occur *only* on the rail link connecting cities A and B, because of given rail network conditions (see Figure 1). Average delay costs (total delay costs divided by the total amount of short- and long-distance services), denoted by Ψ, are determined by the total amount of short- and long-distance services, i.e.,

$$\Psi(q_B, q_C) = q_B + q_C, \tag{2}$$

implying convex delay costs

$$(q_B + q_C)\,\Psi(q_B, q_C) = (q_B + q_C)^2. \tag{3}$$

Delay costs are borne by customers B and C, who consider average delay costs Ψ as given.[4] To control delays (and for safety reasons) there is an upper limit for the total amount of rail services (capacity limit), denoted by $\bar{q} > 0$, such that

$$q_B + q_C \leq \bar{q} \tag{4}$$

must be satisfied.

There is one monopolistic and vertically integrated rail transport service provider. The service provider offers short- and long-distance services with zero variable costs and positive, fixed network costs, denoted by F. Service charges are linear in transport distances, and the service charge per distance unit is $p \geq 0$. Hence, the service charge for short-distance services is p, and the one for long-distance services is $2p$, because distances between cities A and B as well as B and C are normalized to 1. The rail service provider might also charge a scarcity premium per distance unit, denoted by $\gamma \geq 0$. The scarcity premium should balance rail service demand and capacity supply and is only charged for the use of the congested rail link connecting cities A and B. Furthermore, the capacity limit $\bar{q}$, service charge p, and the choice of allocation regimes is exogenously determined.

[4]Given average delay costs are typically assumed in the context of road transport markets and atomistic demand structures. There is debate about the adequacy of this assumption in the context of airports, because airline markets are oligopolistic, see (Daniel and Harback, 2008) and (Brueckner and van Dender, 2008). We, however, leave it for future research to relax the assumption of given average delay costs in the context of rail services.

In a first step, it is useful to determine the equilibrium demand for short- and long-distance services as functions of p (ignoring the capacity limit $\bar{q}$ and the scarcity premium γ). Allocation regimes are considered thereafter.

3 Demand for short- and long-distance services as functions of p

Customers benefit from short- or long-distance rail services, but they also experience delay costs and pay service charges. Furthermore, delay costs depend on the total amount of services, which creates a strategic connection between customers B and C. Ignoring the capacity limit, B's best response to a given amount of long-distance services, q_C, is

$$q_B^r(q_C) = \arg\max_{q_B} B_B(q_B) - q_B\,(p + \Psi(q_B, q_C)) \text{ s.t. } q_B \geq 0, \tag{5}$$

and C's best response to a given amount of short-distance services, q_B, is

$$q_C^r(q_B) = \arg\max_{q_C} B_C(q_C) - q_C\,(2\,p + \Psi(q_B, q_C)) \text{ s.t. } q_C \geq 0. \tag{6}$$

Denote the equilibrium demand for rail services q_x as a function of p by $D_x(p)$, and also denote two critical levels of p by

$$p_B = \frac{a_B\,(1 + b_C) - a_C}{b_C - 1} \text{ and } p_C = \frac{a_C\,(1 + b_B) - a_B}{1 + 2\,b_B}. \tag{7}$$

Proposition 1 *If capacity limits do not exist, best responses (5) and (6) imply a unique set of equilibrium demands for short- and long-distance services* $D_x(p)$ *with* $x \in \{B, C\}$ *and*

$$\frac{\partial D_x(p)}{\partial a_x} \geq 0 \textit{ and } \frac{\partial D_x(p)}{\partial a_y} \leq 0 \tag{8}$$

for all $x \in \{B, C\}$ *and* $y \neq x$*. Furthermore, if* $b_C \leq 1$*,*

$$\frac{\partial D_B(p)}{\partial p} = \begin{cases} \geq 0 & \textit{for} \quad p \in [p_B, p_C] \\ 0 & \textit{for} \quad p < p_B \\ \leq 0 & \textit{for} \quad p > p_C, \end{cases} \tag{9}$$

and if $b_C > 1$*,* $\partial D_B(p)/\partial p \leq 0$*. In contrast,* $\partial D_C(p)/\partial p \leq 0$ *always holds.*

Proof See Appendix A. ■

The comparative statics results (8) are selected to demonstrate the interdependencies in (equilibrium) demand for short- and long-distance services in our model of a congested rail network. They demonstrate that a decreasing demand for short-distance services increases the demand for long-distance services, and vice-versa, because $\partial D_x(p)/\partial a_y \leq 0$ holds. The intuition is, less short-distance services imply less delays and delay costs, which increases the demand for long-distance services. They also demonstrate that the relationship between the demand for short-distance services, $D_B(p)$, and p is ambiguous. Note that a change of p affects long-distance services twice as much as short-distance services; if p is increasing, a strong reduction in long-distance services and delays as well as an increasing demand for short-distance services can be the consequence. In contrast, the relationship between demand for long-distance services, $D_C(p)$, and p is always negative in our setting.

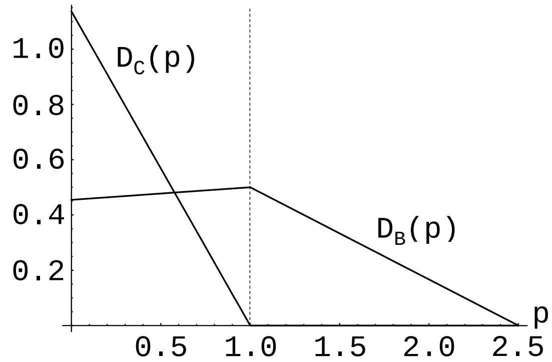

Figure 2: The equilibrium demand for short-distance services, $D_B(p)$, and long-distance services, $D_C(p)$. Parameters: $a_B = a_C = 5/2, b_B = 2, b_C = 4/5$.

Figure 2 illustrates (equilibrium) demand for short-distance services, $D_B(p)$, and for long-distance services, $D_C(p)$. Parameter values are $a_B = 5/2, a_C = 5/2, b_B = 2, b_C = 4/5$ (we further refer to this specific numerical example in the remainder). Observe that $b_C = 4/5 < 1$ and, hence, $\partial D_B(p)/\partial p > 0$ is possible. Demands are separated in two parts by a dashed vertical line. On the left hand side of the dashed line the slope of the demand for short-distance services is positive, and on the right hand side it is negative. In contrast, the slope of the demand for long-distance services is always negative. We include capacity limit $\bar{q}$ and consider rail services under revenue maximization in the next section.

4 Rail services under revenue maximization (no scarcity premium)

Total demand, $D_B(p) + D_C(p)$, can exceed the capacity limit $\bar{q}$. In this section, we assume that the rail network provider chooses services to maximize revenue from service charges in the case of excess demand (which we call regime m). A scarcity premium is not used under regime m.[5] We model regime m as a two-stage game:

Stage 1: Customers B and C observe p and report demand for short-distance services, denoted by $D_B^m(p)$, and demand for long-distance services, denoted by $D_C^m(p)$, to the rail service provider.

Stage 2: The rail service provider determines services such that, under given demand reports, $D_B^m(p)$ and $D_C^m(p)$, and a given capacity limit, $\bar{q}$, the revenue raised by service charges is maximized. Under regime m, services are

$$(q_B^m, q_C^m) = \arg\max_{q_B, q_C} p\,(q_B + 2\,q_C) \text{ s.t.} \tag{10}$$

$$q_x \in [0, D_x^m(p)] \text{ for all } x \in \{B, C\} \text{ and } q_B + q_C \leq \bar{q}. \tag{11}$$

We obtain the following equilibrium results:

Proposition 2 *Under regime m, short-distance services are*

$$q_B^m = \min\{D_B(p), \bar{q} - q_C^m\} \tag{12}$$

[5] A positive scarcity premium is considered in the following section.

and long-distance services are

$$q_C^m = \min\left\{\max\left\{\frac{a_C - 2\,p - \bar{q}}{b_C}, D_C(p)\right\}, \bar{q}\right\}. \tag{13}$$

in equilibrium.

Proof The total amount of services is $q_B + 2\,q_C$, and $2\,\bar{q}$ is an upper limit for total services under a given capacity limit $\bar{q}$. This implies that $2p\,\bar{q}$ is an upper limit for revenues ($\gamma = 0$ given). The upper limit can, however, only be reached if $q_C = \bar{q}$ is satisfied. It follows, revenue maximization favors long- over short-distance services in stage 2 of the game, and long-distance services are provided with capacity, which can be at the expense of short-distance services. The capacity constraint, therefore, splits up into one constraint for short-distance services, $q_B \leq \bar{q} - q_C$, and another one for long-distance services, $q_C \leq \bar{q}$.

Consequently, in stage 1 of the game, B's best responses are

$$q_B^{r,m}(q_C) = \arg\max_{q_B} B_B(q_B) - q_B\,(p + \Psi(q_B, q_C)) \text{ s.t. } q_B \in [0, \bar{q} - q_C] \tag{14}$$

and C's best responses are

$$q_C^{r,m}(q_B) = \arg\max_{q_C} B_C(q_C) - q_C\,(2\,p + \Psi(q_B, q_C)) \text{ s.t. } q_C \in [0, \bar{q}]. \tag{15}$$

Observe that the relevant upper limit for q_C is $\bar{q}$ (not $\bar{q} - q_B$), and recall that B and C consider average delay costs as given (by assumption). The first-order conditions that correspond to best responses (14) and (15) are

$$B_B'(q_B^{r,m}) - \Psi(q_B^{r,m}, q_C) - p - \phi_B^m + \mu_B^m = 0 \tag{16}$$

	$\phi_B^m = 0$	$\phi_B^m > 0$
$\phi_C^m = 0$	$D_B(p), D_C(p)$	$\bar{q} - \max\{0, \frac{a_C - 2p - \bar{q}}{b_C}\}, \max\{0, \frac{a_C - 2p - \bar{q}}{b_C}\}$
$\phi_C^m > 0$	$0, \bar{q}$	

Table 1: Short- and long-distance services under regime m. Expressions inside the cells and on the left-hand side refer to short- and the ones on the right-hand side to long-distance services.

and

$$B'_C(q_C^{r,m}) - \Psi(q_B, q_C^{r,m}) - 2\,p - \phi_C^m + \mu_C^m = 0, \tag{17}$$

and μ_x and ϕ_x are Lagrange-multipliers that refer to the non-negativity or, respectively, capacity constraints. Simultaneously solving conditions (16) and (17) leads to equilibrium demand reports $D_B^m(p)$ and $D_C^m(p)$. Furthermore, it holds $D_B^m(p) = q_B^m$ and $D_C^m(p) = q_C^m$, because customers B and C correctly anticipate the revenue maximization rule in stage 2. Equilibrium results are summarized in Table 1 (expressions inside the cells and on the left-hand side refer to short- and the ones on the right-hand side to long-distance services). Finally, Proposition 2 reproduces the solutions in Table 1. ■

Figure 3 illustrates rail services under regime m depending on p and for a given capacity limit $\bar{q} = 1/2$ (parameter values are equal to the ones used in Figure 2, which illustrates demand functions). Demand functions, $D_x(p)$, are indicated by dashed lines. In this instance, the capacity limit is binding for all $p < 1$. Observe, $q_C^m > D_C(p)$ is possible, because long-distance services are favored under regime m. If the capacity constraint is binding, a change in long-distance services can change short-distance services by the same total amount (case $\phi_B > 0, \phi_C = 0$), consequently, average delay costs would remain unaffected. For this reason, the demand for long-distance services

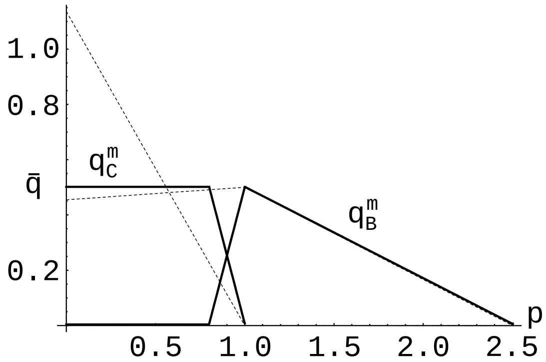

Figure 3: Short-distance services q_B^m and long-distance services q_C^m depending on p with capacity limits $\bar{q} = 1/2$. Dashed lines indicate demand functions $D_B(p)$ and $D_C(p)$.

is greater compared to a situation in which an increasing demand for long-distance services also increases the total amount of rail services and, hence, average delay costs.

Note that equations (16) and (17) determine the amount of short- and long-distance services, but they ignore rationing of individuals inside customer groups. The rationing of individuals inside customer groups will, however, become an issue when we consider total surplus (rail service provider's profit plus consumer surplus) or consumer surplus.

5 Rail services under a scarcity premium

Under regime m, revenue maximization is applied to determine rail services in a situation with excess demand. Excess demand occurs if for a given value of p, the total demand for rail services exceeds the capacity limit $\bar{q}$. A possibility to avoid excess demand is to introduce a scarcity premium $\gamma \geq 0$

that is charged in addition to service charges (which we call regime a). Under regime a, the total charge for short-distance services is $p+\gamma$ and the total charge for long-distance services is $2p+\gamma$. Regime a is also modeled as a two-stage game:

Stage 1: Customers B and C observe the service charge p and report demand for short- and long-distance services as functions of the scarcity premium $\gamma \geq 0$, denoted by $D_B^a(p,\gamma)$ and $D_C^a(p,\gamma)$, to the rail service provider.

Stage 2: The rail service provider chooses $\gamma \geq 0$ to balance the demand for short- and long-distance services and the capacity limit. A positive scarcity premium γ^a is

$$\gamma^a \in \{\gamma : D_B^a(p,\gamma) + D_C^a(p,\gamma) = \bar{q}\}. \tag{18}$$

If excess demand is not present $\gamma^a = 0$ holds. Services are

$$(q_B^a, q_C^a) = (D_B^a(p,\gamma^a), D_C^a(p,\gamma^a)). \tag{19}$$

We obtain the following equilibrium results:

Proposition 3 *Under regime a, the scarcity premium is*

$$\begin{aligned}
\gamma^a \quad = \quad & \max\{0, && (20)\\
& \frac{a_C b_B + a_B b_C - b_C(p+\bar{q}) - b_B(2p+\bar{q}+b_C\bar{q})}{b_B+b_C}, && (21)\\
& a_B - p - (1+b_B)\,\bar{q}, && (22)\\
& a_C - 2p - (1+b_C)\,\bar{q}\}\,, && (23)
\end{aligned}$$

short-distance services are

$$q_B^a = \begin{cases} D_B(p) & for \quad \gamma^a = 0 \\ \dfrac{a_B - a_C + p + b_C \bar{q}}{b_B + b_C} & for \quad \gamma^a = (21) \\ \bar{q} & for \quad \gamma^a = (22), \end{cases} \tag{24}$$

and long-distance services are

$$q_C^a = \begin{cases} D_C(p) & for \quad \gamma^a = 0 \\ \dfrac{a_C - a_B - p + b_B \bar{q}}{b_B + b_C} & for \quad \gamma^a = (21) \\ \bar{q} & for \quad \gamma^a = (23). \end{cases} \tag{25}$$

Furthermore, $\gamma^a > 0$ *implies*

$$\frac{\partial(p+\gamma^a)}{\partial p} \leq 0 \; and \; \frac{\partial(2p+\gamma^a)}{\partial p} \geq 0. \tag{26}$$

Proof See Appendix B. ■

Figures 4 and 5 are based on the same numerical instance as Figures 2 and 3, which illustrate demands $D_x(p)$ or, respectively, rail services under regime m. Figure 4 illustrates total charges (sum of service charges and scarcity premium) depending on p and demonstrates the negative or, respectively, positive relationships between total charges for short- and long-distance services and p. Figure 5 illustrates rail services under regimes m (dashed lines) and a (solid lines). Short-distance services are extended at the expense of long-distance services under regime a compared to regime m (i.e., $\gamma^a > 0 \Rightarrow q_B^a > q_B^m$ and $q_C^a < q_C^m$ in this instance).

The question is, what is better, regime m and revenue maximization or regime a and a scarcity premium? We will address this question from a policy viewpoint in the following section.

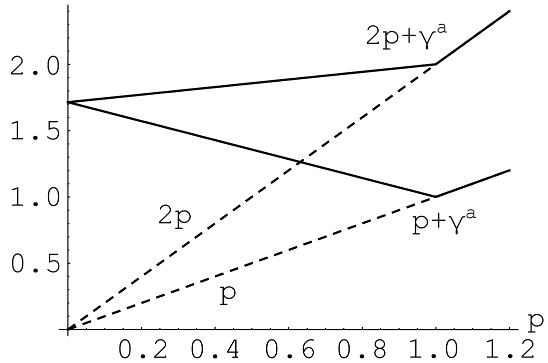

Figure 4: Total service charges for short- and long-distance services depending on p under regime a. Capacity limit is $\bar{q} = 1/2$.

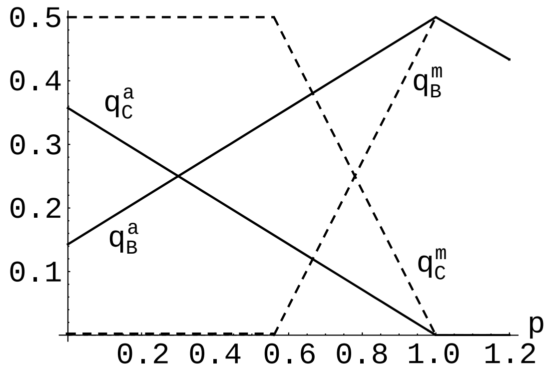

Figure 5: Short- and long-distance services under regimes m (dashed lines) and a (solid lines) depending on p. Capacity limit is $\bar{q} = 1/2$.

6 Revenue maximization versus scarcity premium from a policy viewpoint

Different measures can be applied to evaluate rail services. From a policy viewpoint, total surplus or consumer surplus are relevant measures, which we use to compare outcomes under regimes m and a. We will also touch the issue of cost recovery.

The rail service provider's revenues are

$$R(q_B, q_C) = (p + \gamma)\, q_B + (2p + \gamma)\, q_C, \tag{27}$$

consumer surplus is

$$S(q_B, q_C) = \sum_{y \in \{B,C\}} [B_y(q_y) - q_y\, \Psi(q_B, q_C)] - R(q_B, q_C), \tag{28}$$

and the regulator's objective function is

$$V(q_B, q_C) = S(q_B, q_C) + \beta R(q_B, q_C) \tag{29}$$

with $\beta \in \{0, 1\}$.[6] If $\beta = 1$, consumer surplus and profits are provided with the same weight and, thus, total surplus is relevant. If $\beta = 0$, consumer surplus is relevant.

[6] The objective function (29) is similar to the one considered by Baron and Myerson (Baron and Myerson, 1982) except that we do not include subsidies or taxes.

For benchmarking purposes, we also consider 'optimal services' that maximize the regulator's objective function, $V(q_B, q_C)$, for given p, γ, and $\bar{q}$. Optimal services are

$$(q_B^*, q_C^*) = \arg\max_{q_B, q_C} V(q_B, q_C) \text{ s.t. } q_B + q_C \leq \bar{q} \text{ and } q_B, q_C \geq 0. \tag{30}$$

The corresponding first-order conditions are

$$B_B'(q_B^*) - \underbrace{\left(\Psi(q_B^*, q_C^*) + (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_B}\right)}_{\text{marginal delay costs}} - (1-\beta)\,(p+\gamma) - \lambda^* + \mu_B^* = 0 \tag{31}$$

and

$$B_C'(q_C^*) - \left(\Psi(q_B^*, q_C^*) + (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_C}\right) - (1-\beta)\,(2\,p+\gamma) - \lambda^* + \mu_C^* = 0, \tag{32}$$

and λ and μ_x are Lagrange-multipliers that refer to the capacity constraint or, respectively, non-negativity constraints.

If $\beta = 1$, total surplus is relevant and optimal services do not depend on p or γ (insert $\beta = 1$ into (31) and (32) for a test). This is because changes in p or γ change the distribution of surplus but not total surplus (as long as services are unaffected). Note that $\partial\Psi/\partial q_B = \partial\Psi/\partial q_C$ holds. Therefore, $\beta = 1$ and non-binding non-negativity constraints ($\mu_B^* = \mu_C^* = 0$) together imply that services are optimal if marginal pay-offs are equal to marginal congestion costs (which is a standard result); $B_B'(q_B^*) = B_C'(q_C^*)$ directly follows.

If $\beta = 0$, the rail network operator maximizes consumer surplus. In contrast to total surplus, consumer surplus is determined by p and γ. Furthermore, the total service charge for long-distance services exceeds the one for short-distance services by amount p; $\beta = 0$ and $\mu_B^* = \mu_C^* = 0$, consequently, imply that services are optimal if condition $B_B'(q_B^*) + p = B_C'(q_C^*)$ is satisfied (deduct (32) from (31) and rearrange for a test).

Excess demand is possible under regime m, and total surplus or consumer surplus depends on the amount of short- and long-distance services but also on the rationing rule for individuals inside each customer group. In the following, we apply the efficient-rationing rule, which implies that customers who value services most are served first. Then, the outcome under regime m is determined by $V(q_B^m, q_C^m)$ with $\gamma = 0$. We find that regime m can imply optimal services (q_B^*, q_C^*):

Proposition 4 *If $\gamma = 0$, $V(q_B^m, q_C^m) = V(q_B^*, q_C^*)$ is possible for all $\beta \in \{0, 1\}$, for all $p \geq 0$, and for all cases including $\mu_B^* = \mu_C^* = 0$, $\mu_B^* > 0$ and $\mu_C^* = 0$, and $\mu_B^* = 0$ and $\mu_C^* > 0$.*

Proof See Appendix C. ■

The key finding of Proposition 4 is: the preferred treatment of long-distance services under regime m can neutralize the relative disadvantage for long-distance services from service charges. As a consequence, optimal services are possible. Moreover, optimal services are possible for *all* the cases considered (the different cases are formed by the relevant combinations of $\beta = 0$, $\beta = 1$, $p = 0$, $p > 0$, $\mu_B^* = \mu_C^* = 0$, $\mu_B^* > 0$ and $\mu_C^* = 0$, $\mu_B^* = 0$ and $\mu_C^* > 0$). We now compare optimal services and services under regime a from a policy viewpoint.

Proposition 5 *If $\gamma = \gamma^a, \ldots$*
$\ldots \beta = 1$, and $\mu_B^* = \mu_C^* = 0$, $V(q_B^a, q_C^a) < V(q_B^*, q_C^*)$ for all $p > 0$.
$\ldots \beta = 1$, and $\mu_B^* = \mu_C^* = 0$ and $p = 0$ or $\mu_x^* > 0$ and $\mu_y^* = 0$ $(y \neq x)$, $V(q_B^a, q_C^a) = V(q_B^*, q_C^*)$ is possible.

$\ldots \beta = 0$, $V(q_B^a, q_C^a) < V(q_B^*, q_C^*)$ *for all* $p \geq 0$.

Proof See Appendix D. ■

The key finding of Proposition 5 is: regime a can lead to optimal services in some cases, but optimal services are not possible in a number of other cases. This is in contrast to regime m, where optimal services are possible for all the cases considered. There are two reasons why regime a fails to imply optimal services in some cases.

First, regime a can not neutralize the relative disadvantage of long-distance services from service charges, because the scarcity premium is charged to both long- and short-distance services. If total surplus is relevant and p is strictly positive, long-distance services are, therefore, too low (implying $B_B'(q_B^a) < B_C'(q_C^a)$); except in extreme cases, namely, if either $q_C^a = q_C^* = 0$ or $q_C^a = q_C^* = \bar{q}$ holds true (implying $\mu_x^* > 0$ and $\mu_y^* = 0$). If total surplus is relevant and $p = 0$, it is also possible that regime a leads to optimal services (because $B_B'(q_B^a) < B_C'(q_C^a)$ is consistent with $B_B'(q_B^*) + p = B_C'(q_C^*)$).

Second, externalities exist. The demand for rail services is determined by marginal pay-offs, $B_x'(q_x)$, the total service charge, $p + \gamma^a$ or $2p + \gamma^a$, and average delay costs Ψ (see first-order conditions (40) and (41) in Appendix B). In contrast, if consumer surplus is relevant from a policy perspective, optimal services are determined by marginal pay-offs and marginal delay costs (see (31) and (32)). Note that average delay costs only cover a share of marginal delay costs (see (31)), and the difference between marginal delay costs and average delay costs, which is $(q_B + q_C)\partial\Psi/\partial q_B$ in our setting, determines external marginal delay costs. Hence, external marginal delay costs are present, and the amount of services is too large and does not maximize consumer surplus under regime a. We consider the access charge p and capacity limit $\bar{q}$

as given; however, if consumer surplus is relevant from a policy viewpoint, a change of $\bar{q}$ can not solve the externality problem. This is in contrast to the case where total surplus is relevant.

In general, we find that regimes m and a are both unlikely to imply optimal services, because optimal services are only reached under specific parameter constellations. There are, however, differences that complicate direct comparison of regimes, but affect total and consumer surplus. In particular, long-distance services are favored under regime m and not under regime a, and γ is different under regimes m and a (0 versus γ^a). To illustrate and compare total and consumer surplus under regimes m and a we consider a Monte Carlo simulation in the following section.

7 Monte Carlo simulation

The Monte Carlo simulation is based on a sample of 25 pairs of pay-off functions for short- and long-distance services, $(B_B(q_B), B_C(q_C))$. Each pair requires the choice of 4 parameters (a_B, a_C, b_B, and b_C). Parameters are drawn from a random process that follows a uniform distribution with support $[0, 3]$. Pairs of pay-off functions leading to no difference in services, i.e., parameter constellations leading to $q_x^m = q_x^a$ for all $x \in \{B, C\}$ and all $p \geq 0$, are sorted out. Furthermore, $\bar{q} = 1/2$ and $F = 1/4$.

Figure 6 depicts the aggregated total surplus under regimes m and a, denoted by $\sum(S^z + R^z - F)$ with $z \in \{m, a\}$, and the aggregated consumer surplus, denoted by $\sum S^z$, depending on p (aggregation refers to the sample of 12 pairs of pay-off functions). Outcomes under regime m are indicated by dashed lines and the ones under regime a by solid lines.

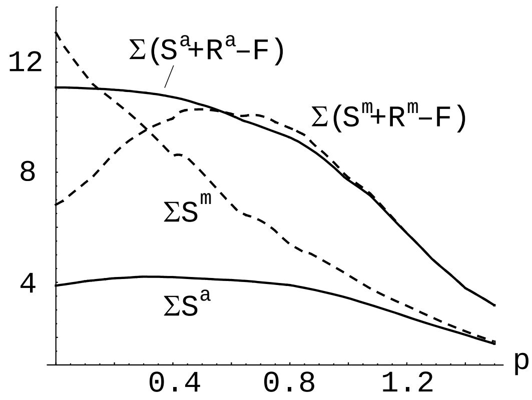

Figure 6: Aggregated total surplus, $\sum(S^z + R^z - F)$ with $z \in \{m, a\}$, and aggregated consumer surplus, $\sum S^z$, under regimes m (dashed lines) and a (solid lines) depending on p. Furthermore, $\bar{q} = 1/2$ and $F = 1/4$. The figure is based on a random sample of 25 pairs of pay-off functions for short- and long-distance services.

Simulation results are ambiguous regarding aggregated total surplus: if p is low, aggregated total surplus is greater under regime m (i.e., $\sum(S^m + R^m - F) > \sum(S^a + R^a - F)$), and if p is high, aggregated total surplus is greater under regime a (i.e. $\sum(S^m + R^m - F) < \sum(S^a + R^a - F)$). Note that if p is high, the relative disadvantage for long-distance services due to service charges is most relevant, and, as a consequence, the preferred treatment of long-distance services under regime m is most relevant, too. If p is low, aggregated total surplus is greater under a scarcity premium, because distortions from service charges are less relevant and a scarcity premium favors customers who value services most. It is, however, difficult to understand whether service charges are high or low in practice, which creates a risk of choosing the wrong regime.

The effect of regime a on consumer surplus is ambiguous in theory. On the one hand, a scarcity premium implies that customers who value services most are provided with capacity, which increases consumer surplus. On the other hand, it increases total service charges, which reduces consumer surplus. Despite ambiguous theoretical results, the simulation indicates that aggregated consumer surplus is greater under regime m than under regime a (i.e., $\sum S^m \geq \sum S^a$). Moreover, it is possible that aggregated consumer surplus changes while aggregated total surplus remains constant (right-hand side of Figure 6). This is because the scarcity premium does not necessarily change services at all, but increases total service charges. Altogether, this indicates that customers are better-off under regime m.

We can also use Figure 6 to analyze revenues and cost recovery under regimes m and a. Note that profits can be determined by the difference between aggregated total surplus and aggregated consumer surplus, and if the curve that depicts aggregated total surplus intersects the one that depicts aggregated consumer surplus, aggregated profits are zero. The theoretical effect of γ^a on revenues is ambiguous, because it increases total service charges but reduces the total amount of services. Our simulation, however, indicates that revenues are greater under regime a than under regime m. Observe that under regime a, aggregated total surplus is greater than aggregated consumer surplus, which implies strictly positive profits (inside the range of p considered in Figure 6). Furthermore, a reduction of p increases aggregated profits further, i.e., the loss of revenues raised by p due to a reduction of p is more than compensated by additional revenues raised by γ^a. The picture changes under regime m, because an intersection between the curve depicting aggregated total surplus and aggregated consumer surplus exists. Profits are negative on the left-hand side of the intersection point and positive on the right-hand side. The simulation, therefore, indicates that cost recovery is easier to reach under regime a than under regime m.

8 Conclusions

In this paper, we developed a simple model of a congested and capacity limited network with two links, which is used to offer short- and long-distance services. We considered two regimes to allocate limited network capacity: (i) 'revenue maximization' and no scarcity premium, and (ii) a scarcity premium that balances demand and capacity. As a benchmark, we also considered 'optimal services' that maximize total surplus or, respectively, consumer surplus for a given capacity limit, service charge, and scarcity premium.

Our key results were the following. We found that none of the two regimes is likely to imply optimal services, because this would require special parameter constellations that are hardly relevant in reality. Furthermore, based on a Monte Carlo simulation, we found that if service charges are high, total surplus is greater under revenue maximization than under a scarcity premium and vice-versa. It is, however, difficult to understand whether service charges are low or high in reality, which creates a risk of choosing the wrong regime. The simulation also indicated that consumer surplus is always greater under revenue maximization than under a scarcity premium. In contrast, cost recovery is easier to reach under a scarcity premium.

The current key results are based on a highly stylized model, which includes several simplifying assumptions; some of them seem to be critical and others not. The non-critical assumptions include linear model specifications (linear demand and linear average congestion costs), steady-state congestion, a vertically integrated rail service provider, and a network with only two raillinks. The critical assumptions include efficient rationing, no 'misreporting' of customer demand, the absence of intermodal competition, and given network conditions, capacity limits, and service charges. First, we discuss the non-critical assumptions, and, second, the critical ones.

A change to non-linear model specifications will most likely not affect our key results, because they do not depend on the shape of functions. The same holds true for a change towards a dynamic model structure, which includes peak and off-peak periods. With peak and off-peak periods customers, who are not served during peak hours, can change to off-peak periods. Although this provides a more exact picture of reality, we do not expect that this would change our key results. Note that our model of a vertically integrated rail service provider is similar to a model that includes a monopolistic rail network provider and train operating companies under perfect competition. This is because profits of train operating companies would be equal to zero, and therefore vertical separation would not affect total service charges in this case (as long as total costs are unaffected by separation). Moreover, the consideration of networks that include a large number of rail links and services, complicates the calculation of revenue maximizing services and scarcity premiums, but there is no obvious reason why this should change our key results.

We now turn to the model assumptions that we consider as critical for our key results. Total and consumer surplus strongly depends on the rationing of individuals inside customer groups, and efficient rationing maximizes total and consumer surplus for given amounts of short- and long-distance services. Under revenue maximization, our results are actually based on efficient rationing and, therefore, they are likely to overstate total and consumer surplus, because efficient rationing may not always be achieved in reality.

Both regimes considered can be affected by 'misreporting' in reality. For instance, customers might collude and understate their 'true' demand to reduce scarcity premiums. On the other hand, customers might report short- as long-distance services to receive a preferred treatment under revenue maximization. If customer reports are not binding (i.e., if customers are not obliged to make full use of provided services), this could also lead to a loss of revenues, because service charges are lower for short- than for long-distance services. The effect of misreporting on total and consumer surplus is, however, difficult to predict.

Another critical aspect, which we ignored in this paper, is the existence of intermodal competition. In reality, customers can switch to other modes of transport such as road, air, or inland water transportation in cases they are not served by rail service providers or in cases where total rail service charges are excessive. Hence, to obtain a better understanding of outcomes under different rail service allocation regimes it would be useful to take intermodal competition into consideration.

Finally, in this paper we consider the network conditions, capacity limit, and service charge as exogenous. In reality, it is, however, possible to build new capacity in order to reduce congestion. Furthermore, capacity limits and service charges might be chosen differently depending on the allocation regime and depending on whether total or consumer surplus is relevant from a policy viewpoint. It would, therefore, be useful to extend the current analysis and consider network conditions, capacity limits, and service charges as endogenous variables.

Overall, the research provided in this paper deals with relevant elements of rail transport markets and provides theoretical evidence on the principles underlying the relative outcomes under priority rules or scarcity premiums from a policy viewpoint. However, future research is required in order to obtain a more complete picture.

A Proof of Proposition 1

The following three (equilibrium) demand cases are possible:

(I) $D_B(p), D_C(p) > 0$,

(II) $D_B(p) \geq 0$ and $D_C(p) = 0$, and

(III) $D_B(p) = 0$ and $D_C(p) \geq 0$.

We consider the demand cases (I)-(III) one by one. Best responses (5) and (6) imply the following:
In case (I), demands are

$$D_B(p) = \frac{a_B - a_C + a_B\, b_C - p\,(b_C - 1)}{b_B + b_C + b_B\, b_C} \tag{33}$$

and

$$D_C(p) = \frac{a_C - a_B + a_C\, b_B - p\,(1 + 2\, b_B)}{b_B + b_C + b_B\, b_C}. \tag{34}$$

In case (II), demands are

$$D_B(p) = \max\left\{0, \frac{a_B - p}{1 + b_B}\right\} \text{ and } D_C(p) = 0. \tag{35}$$

In case (III), demands are

$$D_B(p) = 0 \text{ and } D_C(p) = \max\left\{0, \frac{a_C - 2\,p}{1 + b_C}\right\}. \tag{36}$$

Furthermore, it is useful to denote a critical level of a_B by

$$\tilde{a}_B = \frac{a_C}{1 + b_C}, \tag{37}$$

and to distinguish between $b_C > 1$, $b_C < 1$, and $b_C = 1$. The following table shows the relevance of cases (I)-(III) depending on b_C and p:

	(I)	(II)	(III)
$b_C > 1$	$p \leq p_B, p_C$	$p > p_C = \min\{p_C, p_B\}$	$p > p_B = \min\{p_B, p_C\}$
$b_C < 1$	$p \in (p_B, p_C]$	$p > p_C \geq p_B$	$p \leq p_B \leq p_C$ or $p_B > p_C$
$b_C = 1$	$a_B > \tilde{a}_B,\ p < p_C$	$a_B > \tilde{a}_B,\ p \geq p_C$	$a_B \leq \tilde{a}_B$

It is straightforward to obtain the comparative static results based on demand expressions (33)-(36).

B Proof of Proposition 3

Under regime a it holds the following. In stage 1, B's best responses are

$$q_B^{r,a}(q_C) = \arg\max_{q_B} B_B(q_B) - q_B\,(p + \gamma + \Psi(q_B, q_C)) \text{ s.t. } q_B \geq 0, \qquad (38)$$

and C's best responses are

$$q_C^{r,a}(q_B) = \arg\max_{q_C} B_C(q_C) - q_C\,(2\,p + \gamma + \Psi(q_B, q_C)) \text{ s.t. } q_C \geq 0. \qquad (39)$$

The first-order conditions that correspond to best responses (38) and (39) are

$$B'_B(q_B^{r,a}) - p - \gamma - \Psi(q_B^{r,a}, q_C)) + \mu_B^a = 0 \qquad (40)$$

and

$$B'_C(q_C^{r,a}) - 2\,p - \gamma - \Psi(q_B, q_C^{r,a})) + \mu_C^a = 0 \qquad (41)$$

where μ_x are Lagrange-multiplies that refer to non-negativity constraints. Simultaneously solving conditions (40) and (41) leads to (equilibrium) demand reports $D_B^a(p,\gamma)$ and $D_C^a(p,\gamma)$.

Demand reports, $D_x^a(p,\gamma)$, can form three demand cases that are similar to cases (I)-(III) for demand functions $D_x(p)$ (see the proof of Proposition 1 in Appendix A), which should also be distinguished here. Furthermore, it is useful to distinguish $\gamma = 0$ and $\gamma > 0$. Hence, there are altogether six cases to consider. For each case, we determine equilibrium demand reports $D_x^a(p,\gamma)$, scarcity premium γ^a, and services $q_x^a = D_x^a(p,\gamma^a)$ in the following. If $\gamma = 0$, $q_B^a = D_B^a(p,0) = D_x(p)$ for all $x \in \{B,C\}$ (which covers three relevant cases already).

If $\gamma > 0$, the following holds. In case (I), demand reports are

$$D_B^a(p,\gamma) = \frac{a_B(1+b_C) - a_C + p - b_C(p+\gamma)}{b_B + b_C + b_B b_C} \tag{42}$$

and

$$D_C^a(p,\gamma) = \frac{a_C(1+b_B) - a_B - p - b_B(2p+\gamma)}{b_B + b_C + b_B b_C}, \tag{43}$$

in combination with (18), demand reports imply scarcity premium

$$\gamma^a = \frac{a_C b_B + a_B b_C - b_C(p+\bar{q}) - b_B(2p + \bar{q} + b_C\bar{q})}{b_B + b_C}, \tag{44}$$

leading to services

$$q_B^a = \frac{a_B - a_C + p + b_C\bar{q}}{b_B + b_C} \text{ and } q_C^a = \frac{a_C - a_B - p + b_B\bar{q}}{b_B + b_C}. \tag{45}$$

In case (II), demand reports are

$$D_B^a(p,\gamma) = \max\left\{0, \frac{a_B - p - \gamma}{1 + b_B}\right\} \text{ and } D_C^a(p,\gamma) = 0, \tag{46}$$

in combination with (18), demand reports imply scarcity premium

$$\gamma^a = a_B - p - (1 + b_B)\,\bar{q}, \tag{47}$$

leading to services

$$q_B^a = \bar{q} \text{ and } q_C^a = 0. \tag{48}$$

In case (III), demand reports are

$$D_B^a(p, \gamma) = 0 \text{ and } D_C^a(p, \gamma) = \max\left\{0, \frac{a_C - 2p - \gamma}{1 + b_C}\right\}. \tag{49}$$

in combination with (18), demand reports imply scarcity premium

$$\gamma^a = a_C - 2p - (1 + b_C)\,\bar{q}, \tag{50}$$

leading to services

$$q_B^a = 0 \text{ and } q_C^a = \bar{q}. \tag{51}$$

Note that γ must be chosen such that $D_B(p) + D_C(p) \leq \bar{q}$ is satisfied for all cases (I)-(III). It follows that γ^a is determined by the maximum of zero, (44), (47), and (50), and quantities are determined by $D_x(p)$, (45), (48), or (51), depending on which case actually determines γ^a.

Finally, differentiating (44), (47), and (50) with respect to p leads to

$$\frac{\partial \gamma^a}{\partial p} \in [-2, -1]. \tag{52}$$

The comparative statics results (26) directly follow, which completes the proof.

C Proof of Proposition 4

Rearranging (31) and (32) leads to

$$\mu_B^* = -B_B'(q_B^*) + \Psi(q_B^*, q_C^*) + (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_B} + (1-\beta)(p+\gamma) + \lambda^* \quad (53)$$

and

$$\mu_C^* = -B_C'(q_C^*) + \Psi(q_B^*, q_C^*) + (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_C} + (1-\beta)(2p+\gamma) + \lambda^*. \quad (54)$$

Furthermore, rearranging (16) and (17) leads to

$$\mu_B^m = -B_B'(q_B^m) + \Psi(q_B^m, q_C) + p + \phi_B^m \quad (55)$$

and

$$\mu_C^m = -B_C'(q_C^m) + \Psi(q_B^m, q_C^m) + 2p + \phi_C^m. \quad (56)$$

A scarcity premium is not used under regime m. The comparison of outcomes under regime m and under optimal services, (q_B^*, q_C^*), should therefore be based on $\gamma = 0$. Comparison of equations (53) and (55) as well as equations (54) and (56) shows:
If $\gamma = 0$ and $\mu_B^* = \mu_C^* = 0$, $(q_B^m, q_C^m) = (q_B^*, q_C^*)$ holds if conditions

$$\phi_B^m = (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_B} - p\beta + \lambda^* \quad (57)$$

and

$$\phi_C^m = (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_C} - 2p\beta + \lambda^* \quad (58)$$

are satisfied. Observe that $\phi_B^m \geq \phi_C^m$ and that $\mu_B^* = 0$ implies that $q_B^* \geq 0$ and $\phi_C^m = 0$.

If $\gamma = 0$, $\mu_B^* = 0$ and $\mu_C^* > 0$, $(q_B^m, q_C^m) = (q_B^*, q_C^*)$ holds if conditions

$$\phi_B^m = q_B^* \frac{\partial \Psi}{\partial q_B} - p\beta + \lambda^* \tag{59}$$

and

$$B_C'(0) < \Psi(q_B^*, 0) + 2p \tag{60}$$

are satisfied. Notice, $\mu_C^* > 0 \Rightarrow \phi_C^* = 0$.

If $\gamma = 0$, $\mu_B^* > 0$ and $\mu_C^* = 0$, $(q_B^m, q_C^m) = (q_B^*, q_C^*)$ holds if conditions

$$B_B'(0) < \Psi(0, q_C^*) + p + \phi_B^m \tag{61}$$

and

$$\phi_C^m = q_C^* \frac{\partial \Psi}{\partial q_C} - 2p\beta + \lambda^* \tag{62}$$

are satisfied.

D Proof of Proposition 5

Regime a includes scarcity premium $\gamma^a \geq 0$. The comparison of outcomes under regime a and under optimal services, (q_B^*, q_C^*), should therefore be based on $\gamma = \gamma^a$. Total service charges under regime a and first-order conditions (40) and (41) imply

$$\mu_B^a = -B_B'(q_B^a) + \Psi(q_B^a, q_C^a) + p + \gamma^a \tag{63}$$

and

$$\mu_C^a = -B_C'(q_C^a) + \Psi(q_B^a, q_C^a) + 2p + \gamma^a. \tag{64}$$

Comparison of equations (53) and (63) as well as (54) and (64) shows:

If $\gamma = \gamma^a$, $\beta = 1$, and $\mu_B^* = \mu_C^* = 0$, $(q_B^a, q_C^a) = (q_B^*, q_C^*)$ holds if conditions

$$p\beta + \gamma^a = (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_B} + \lambda^* \tag{65}$$

and

$$2p\beta + \gamma^a = (q_B^* + q_C^*)\frac{\partial \Psi}{\partial q_C} + \lambda^* \tag{66}$$

are satisfied. Conditions (65) and (66) are satisfied if and only if $p = 0$ and $\gamma^a = (q_B^* + q_C^*)\partial\Psi/\partial q_B + \lambda^*$ (which holds for $\bar{q} \leq q_B^* + q_C^*$). Note, $(q_B^* + q_C^*)\partial\Psi/\partial q_B + \lambda^*$ is equal to external marginal congestion costs, which are determined by $\partial(q_B + q_C)\Psi/\partial q_x - \Psi$.

If $\gamma = \gamma^a$, $\beta = 1$, and $\mu_B^* > 0$ and $\mu_C^* = 0$, $(q_B^a, q_C^a) = (q_B^*, q_C^*) = (0, q_C^*)$ holds if conditions

$$2p + \gamma^a \geq q_C^* \frac{\partial \Psi}{\partial q_C} + \lambda^* \tag{67}$$

and

$$B_B'(0) < \Psi(0, q_C^*) + p + \gamma^a \tag{68}$$

are satisfied.

If $\gamma = \gamma^a$, $\beta = 1$, and $\mu_B^* = 0$, $\mu_C^* > 0$, $(q_B^a, q_C^a) = (q_B^*, q_C^*) = (q_B^*, 0)$ holds if conditions

$$p + \gamma^a \geq q_B^* \frac{\partial \Psi}{\partial q_B} \tag{69}$$

and

$$B_C'(0) < \Psi(q_B^*, 0) + 2p + \gamma^a \tag{70}$$

are satisfied.

If $\gamma = \gamma^a$, $\beta = 0$, and $\mu_x^* \geq \mu_y^* = 0$, $(q_B^a, q_C^a) > (q_B^*, q_C^*)$ follows, which is due to the existence of marginal external congestion costs, i.e., $(q_B^* + q_C^*)\partial\Psi/\partial q_B > 0$, and $B_x''(q_x) < 0$ for all $x \in \{B, C\}$ ($\lambda^* = 0$ in this case).

References

Baron, D. P. and Myerson, R. B. (1982). Regulating a monopolist with unknown costs. *Econometrica*, 50(4):911–930.

Basso, L. and Zhang, A. (2008). Congestible facility rivalry in vertical structures. *Journal of Urban Economics*, 1:218–237.

Borndörfer, R., Grötschel, M., Lukac, S., Mitusch, K., Schlechte, T., Schultz, S., and Tanner, A. (2006). An auctioning approach to railway slot allocation. *Competition and Regulation in Network Industries*, 1(2):163–196.

Brewer, P. J. and Plott, C. R. (1996). A Binary Conflict Ascending Price (BICAP) mechanism for the decentralized allocation of the right to use railroad tracks. *International Journal of Industrial Organization*, 14(6):857–886.

Brueckner, J. K. and van Dender, K. (2008). Atomistic congestion tolls at concentrated airports? seeking a unified view in the internalization debate. *Journal of Urban Economics*, 64(2):288–295.

Daniel, J. I. and Harback, K. T. (2008). (When) Do hub airlines internalize their self-imposed congestion delays? *Journal of Urban Economics*, 63(2):583–612.

Goverde, R. M. P. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B*, 41(2):179–201.

Milgrom, P. R. (2007). Package auctions and exchanges. *Econometrica*, 75(4):935–965.

Mitusch, K. and Tanner, A. (2008). Trassenzuteilung und Konfliktlösung nach der neuen EIBV: Anmerkungen aus volkswirtschaftlicher sicht.

Mun, S.-I. and Ahn, K. (2008). Road pricing in a serial network. *Journal of Transport Economics and Policy*, 42(3):367–395.

Nash, C. (2005). Rail infrastructure charges in europe. *Journal of Transport Economics and Policy*, 39(3):259–278.

Nilsson, J.-E. (1999). Allocation of track capacityExperimental evidence on the use of priority auctioning in the railway industry. *International Journal of Industrial Organization*, 17(8):1139–1162.

Nilsson, J.-E. (2002). Towards a welfare enhancing process to manage railway infrastructure access. *Transportation Research Part A*, 36(5):419–436.

Parkes, D. C. (2001). *Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency*. PhD thesis, University of Pennsylvania.

Parkes, D. C. and Ungar, L. H. (2001). An auction-based method for decentralized train scheduling. In *Proc. 5th International Conference on Autonomous Agents (AGENTS-01)*, pages 43–50.

Verhoef, E. T. (2008). Private roads: Auctions and competition in networks. *Journal of Transport Economics and Policy*, 42(3):463–493.