

Priority Rules Versus Scarcity Premiums in Rail Markets

Achim I. Czerny

WHU – Otto Beisheim
School of Management
achim.czerny@whu.edu

Kay Mitusch

Universität Karlsruhe
mitusch@iww.uni-karlsruhe.de

Andreas Tanner

Technische Universität Berlin
ant@wip.tu-berlin.de

Excellence in
Management
Education

WHU – Working Paper
Series in Economics

WP 10/03

Priority Rules Versus Scarcity Premiums in Rail Markets

Achim I. Czerny*

WHU – Otto Beisheim School of Management
achim.czerny@whu.edu

Kay Mitusch

Universität Karlsruhe
mitusch@iww.uni-karlsruhe.de

Andreas Tanner

Technische Universität Berlin
ant@wip.tu-berlin.de

April 21, 2010

Abstract

The regulation of rail network access is a key component of the EU policy that aims to strengthen rail markets. Two specific regulations are proposed: (i) a priority for long-distance (freight) services and (ii) a scarcity premium. Based on a congested network with two rail links, numerical simulations demonstrate that total surplus can be greater under the priority rule, which depends on the network charge per train-kilometer. Consumer surplus, on the other hand, is always greater under the priority rule, while fixed network-cost recovery is easier to achieve if a scarcity premium exists.

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

Keywords: Rail markets, priority rules, scarcity premiums, congestion, network allocation.

*We thank Felix Höfler for helpful suggestions.

1 Introduction

The current framework for the regulation of access to incumbents' rail networks 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 can occur that the demand for certain segments of the rail network exceeds the capacity limit 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, capacity allocation may also be based on a priority rule. In its White Paper on Transport 2010, the European Commission promotes a specific priority rule that favors freight over passenger services. This priority rule aims to shift long-distance freight traffic from road to rail.

The main contribution of this paper is to identify the policy implications of regulation regimes that include a priority for long-distance (freight) services or a scarcity premium. Since it is difficult to derive general analytical results, our findings are mainly based on numerical simulations. We find that a priority for long-distance services can increase total surplus if the network charge per train-kilometer reaches great levels. This is because the network charge is greater for long-distance services, which creates a disadvantage for long- relative to short-distance services that can, at least partly, be compensated by a priority for long-distance services. By contrast, a scarcity premium can increase total surplus if the charge per train-kilometer is at low levels because its effect on long-distance services is of less importance in this scenario. The simulation results are clear-cut with respect to consumer surplus;

consumer surplus is always greater under the priority for long-distance services. This indicates that, in reality, the introduction of a scarcity premium may have a strong negative effect on rail customers because it is supposed to increase the rail service charges. Fixed network cost recovery may however be much easier to achieve with a scarcity premium. Overall, we find that no regime dominates the other one in all respects. Hence, there is no clear-cut ranking between a priority rule or a scarcity premium from a policy viewpoint.

Our model includes a simple network with two rail links; one link is congested due to given network conditions. Congestion implies delays that are costly for customers (missed business meetings or interruptions in the production process) and train operating companies (penalty payments to customers and overtime premiums to employees). Since delays are increasing in the amount of services on a given network, an upper limit for services (capacity limit), which we consider as given, is used to control delays and delay costs.¹ Note that, in our framework, congestion occurs in situations where services are below the capacity limit, which is in contrast to the EU's notion of congestion.² This is because we follow a stationary-state congestion approach that is common for congestion models related to air and road transport markets and that, in our opinion, provides a realistic picture of rail markets as well.³

Moreover, we consider a vertical structure with a monopolistic infrastructure provider and train operating companies (TOCs) under perfect com-

¹Capacity limits are used to control delays and also for safety reasons.

²See Nash (2005) for a discussion of congestion, congestion charges, scarcity, and scarcity charges or, respectively, scarcity premiums.

³Basso 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.

petition. For many situations at hand, this is an unrealistic assumption. However, there are some market segments, in particular in the area of rail freight services, where the assumption of a competitive market is quite realistic. For example, in Germany the long distance rail freight markets from the North Sea ports (Hamburg, Bremen) to the cities in the South-West of the country and to Italy are served by many different companies competing for customers. Note that the concept of perfect competition between TOCs is useful to abstract away from the effect of market structure on the internalization of self-imposed delays, which has been extensively studied in the area of air transport.⁴ On the demand side, there is one group of customers that demands short-distance rail services, while a second group demands long-distance rail services.

Finally, we consider two regulation regimes for the allocation of limited network capacity. In any of these two regimes the network charge includes a given component that is linear proportional to travel distances, which is similar to the simple network charges per train-kilometer that are common in Europe (Nash, 2005). However, if for the given network charge demand exceeds the capacity limit, one regime favors long-distance services over short-distance services, while the other regime exactly balances demand and capacity supply by the introduction of a scarcity premium.

Other economists considered alternative ways to design railway auctions in more general capacity limited (but uncongested) networks (Brewer and Plott, 1996; Nilsson, 1999; Parkes, 2001; Parkes and Ungar, 2001 and Nilsson, 2002; for an overview see Borndörfer et al., 2006). However, to our knowledge, there is no paper that investigates a priority for long-distance services and

⁴The issue of self-internalization in congestible and oligopolistic air transport markets has been investigated by Brueckner (2002), Zhang and Zhang (2006), Baake and Mitusch (2007), Daniel and Harback (2008) and Brueckner and van Dender (2008).

a scarcity premium from the policy perspective available until now. We therefore believe that our analysis can contribute to a better understanding of the economic effects that are implied by the current European rail policy.

The structure of this paper is as follows. Section 2 presents the basic model set-up. Section 3 considers rail services when priority is attached to long-distance services, while Section 4 considers rail services under a scarcity premium. Section 5 compares outcomes under regimes I and regime II from a policy perspective. Section 6 offers conclusions.

2 The Model

We consider a railway line that connects three cities A, B, and C. The distances between A and B as well as B and C are normalized to one. The infrastructure is used to provide rail services (passenger or freight) between cities. Rail services going into different directions can be operated independently. We can therefore concentrate on one direction. Figure 1 illustrates.

We start by describing the demand side. There are customers of type 1 and 2. Type 1 customers demand short-distance services between cities A and B, while type 2 customers demand long-distance services between cities A and C (via city B). Short-distance services may include passenger transport and long-distance services freight transport. Since travel distances between cities are normalized to one, the travel distance for long-distance services is exactly twice as large as the one for short-distance services. We do not consider short-distance services between B and C to reduce the complexity of the analysis. The main insights are however unaffected by this simplification.

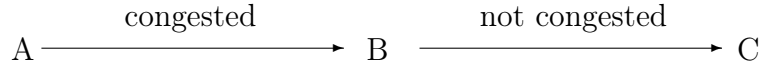


Figure 1: A railway line connecting three cities, A, B, and C. By assumption, the AB segment is congested, while the BC segment is uncongested even at a price of zero.

The amount of short-distance services is denoted by $q_1 \geq 0$ and the amount of long-distance services by $q_2 \geq 0$. Setting aside congestion cost, type i 's inverse demand is

$$P_i = a_i - b_i q_i \quad (1)$$

with $a_i, b_i > 0$ and $i = 1, 2$.

We follow a stationary-state congestion approach and assume that delays *only* occur on the AB segment 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 linear proportional to the total amount of short- and long-distance services with

$$\Psi = q_1 + q_2, \quad (2)$$

which implies convex delay costs

$$(q_1 + q_2) \Psi = (q_1 + q_2)^2. \quad (3)$$

Delay costs are borne by customers, and customers consider Ψ as given.⁵ To control delays (and for safety reasons) there is an upper limit for the total amount of rail services. This capacity limit is denoted by $\bar{q} > 0$ and implies

$$q_1 + q_2 \leq \bar{q}. \quad (4)$$

Turning to the supply side, we take a simplified approach that allows us to analyze the effects of network regulation in a direct way. Our aim is to model the joint effects of regulation concerning (i) network access charges and (ii) the allocation rule for granting access to scarce parts of the network (bottlenecks). In any of the regulatory regimes considered here, the regulated normal network access charge is linear proportional to the travel distance and given. The charge per distance unit is denoted by $p \geq 0$. Therefore, the charge for short-distance services is p and that for long-distance services is $2p$ because the distances between cities are normalized to 1. Fixed network costs are denoted by $F > 0$. The marginal costs of the infrastructure provider and TOCs are constant and normalized to zero. Then, the customers' rail charge would coincide with p for short-distance services and $2p$ for long-distance services due to our assumption of perfect competition between TOCs and given that no other infrastructure charge besides p exists. However, depending on the regulatory regime under consideration, an additional infrastructure charge can exist. In this paper, we concentrate on two regulatory regimes denoted by I and II.

Under regime I there is only the price variable p and the allocation rule that allocates scarce parts of the network to long-distance services first. In

⁵In reality, delays will also increase costs of rail companies due to increased energy and staff costs or penalty payments to customers. The assumption that delays are borne by customers is however without loss of generality.

contrast, regulatory regime II allows to charge a special scarcity premium per distance unit for the congested line segment only. This surcharge is denoted by $\gamma \geq 0$. Thus, under regime II, the price for short-distance services is $p + \gamma$ and the price for long distance services is $2p + \gamma$. Taking the spirit of a scarcity premium seriously, we assume that the premium γ is set by the regulator to balance rail service demand and capacity on the congested line segment. As a consequence, no other allocation rule is required, since the surcharge will drive out any excess demand on the AB segment.

3 Regime I: Priority to Long-Distance Services

In a first step, it is useful to ignore \bar{q} (and γ) and determine the equilibrium demands for short- and long-distance services depending on p . Allocation is considered thereafter.

3.1 The demands under a non-binding capacity limit

Customers pay a service charge but they also bear the delay cost Ψ . So, there is a ‘full price’ for rail services denoted by ρ_i^I with $\rho_1^I = p + \Psi$ and $\rho_2^I = 2p + \Psi$.⁶ Letting D_i denote the equilibrium demand of type- i customers in a scenario without a binding capacity limit, it is useful to distinguish three cases: (i) $D_1, D_2 > 0$, (ii) $D_1 \geq 0$ and $D_2 = 0$ and (iii) $D_2 \geq 0$ and $D_1 = 0$. These demands are piecewise defined in p , and the relevant parts are separated by the following two critical levels of p ,

$$p_1 = \frac{a_1(1+b_2) - a_2}{b_2 - 1} \quad \text{and} \quad p_2 = \frac{a_2(1+b_1) - a_1}{1 + 2b_1}. \quad (5)$$

⁶Oum, Zhang and Zhang (2004) introduced the concept of a full price in a context with air traffic delay.

Letting ΔX denote $X|_{y=y''} - X|_{y=y'}$ and Δy denote $y'' - y'$ with $y'' \neq y'$, we obtain the following results.

Lemma 1 *If a capacity limit is absent ($\bar{q} \rightarrow \infty$), a unique set of equilibrium demands (D_1, D_2) exists with*

$$\frac{\Delta D_i}{\Delta a_i} \geq 0 \text{ and } \frac{\Delta D_i}{\Delta a_j} \leq 0 \quad (6)$$

for all $i = 1, 2$ and $j \neq i$. Moreover, if $b_2 \leq 1$,

$$\frac{\Delta D_1}{\Delta p} = \begin{cases} \geq 0 & \text{for } p \in [p_1, p_2] \\ 0 & \text{for } p < p_1 \\ \leq 0 & \text{for } p > p_2, \end{cases} \quad (7)$$

and if $b_2 > 1$, $\Delta D_1/\Delta p \leq 0$ for all $p \geq 0$, while $\Delta D_2/\Delta p \leq 0$ always holds true.

Proof See Appendix A. ■

The comparative-static results in (6) demonstrate the relationship between delay and demands. They show that a decreasing demand for short-distance services would increase the demand for long-distance services and vice-versa ($\Delta D_i/\Delta a_j \leq 0$). This is because a reduced demand in one market segment reduces delays, which increases the demand in the other market segment.

Lemma 1 also shows that the relationship between p and D_1 is ambiguous, while the relationship between p and D_2 is clear-cut and negative. This is because a change of p hits long-distance traffic twice as much as short-distance traffic. Type 2 customers will thus react quite strong to an increase of p . Then, since delays have decreased, it might even happen that the demand for short-distance services increases (though p has increased).

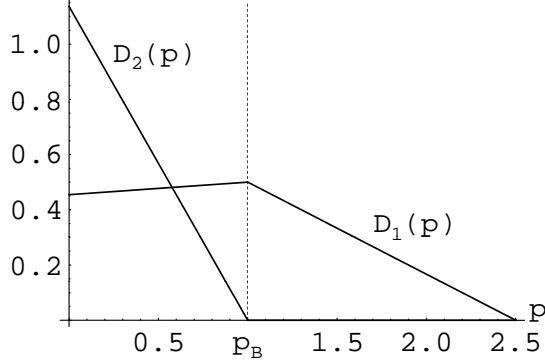


Figure 2: Demands D_1 and D_2 . Parameters: $a_1 = a_2 = 5/2, b_1 = 2, b_2 = 4/5$.

Figure 2 illustrates the equilibrium demands D_1 and D_2 when capacity constraints do not exist. The demands are separated into two parts, indicated by a dashed vertical line at $p = p_1$. Parameter values are $a_1 = a_2 = 5/2, b_1 = 2, b_2 = 4/5$ (this specific numerical instance is used throughout the paper to illustrate outcomes under regimes I and II). Observe that $\Delta D_1/\Delta p > 0$ holds for all $p \leq p_1$ in this instance because $b_2 = 4/5 < 1$ (see Lemma 1). In contrast, the slope of the demand for long-distance services, $\Delta D_2/\Delta p$, is always negative. In the next section we introduce the capacity limit \bar{q} and consider railway allocation under regime I, which gives priority to long-distance services.

3.2 Rail services under regime I

Total demand $D_1 + D_2$ can exceed the capacity limit \bar{q} . In this section, we consider regime I and assume that, in the case of excess demand ($D_1 + D_2 >$

\bar{q}), the rail network provider prefers long-distance services over short-distance services. Regime I includes two-stages.

Stage 1: Customers of type 1 and 2 report their demand for short-distance services and long-distance services for a given price p to the TOCs. The aggregated demand reports are denoted by D_i^I . Then, TOCs forward reports (D_1^I, D_2^I) to the rail network provider.

Stage 2: For given demand reports D_i^I and a given capacity limit \bar{q} , the rail network provider allocates capacity to type-2 customers first. The remaining capacity is then allocated to type-1 customers.

We assume that customers correctly anticipate the priority rule in the second stage and that they do not overstate demand in the sense that they do not demand services that are unavailable under the given priority rule and capacity constraint. For example, $D_i^I > \bar{q}$ will not occur under this assumption. Letting q_i^I denote the allocation under regime I, $q_i^I = D_i^I$ therefore always holds. This is to reduce complexity but does not limit the validity of our main insights.

Note that the demand report D_2^I can be greater than D_2 . This is because the priority rule imposes different constraints on the specific amounts of long- and short-distance services. The constraint is $q_2 \leq \bar{q}$ for long-distance services. By contrast, the constraint is $q_1 \leq \bar{q} - q_2$ for short-distance services. Hence, short-distance services exist if and only if $D_2^I < \bar{q}$. The crucial element here is that, when the capacity constraint binds, an increase of long-distance services reduces short-distance services by exactly the same amount. And, in this situation, long-distance services can be increased but average delay costs Ψ remain constant. This is different in a situation where the capacity constraint is not binding and where an increase of long-distance services will

not reduce short-distance services by exactly the same amount. This implies the following allocations.

Lemma 2 (i) *Under regime I, long-distance services are*

$$q_2^I = \min \left\{ \max \left\{ \frac{a_2 - 2p - \bar{q}}{b_2}, D_2 \right\}, \bar{q} \right\}. \quad (8)$$

and short-distance services are

$$q_1^I = \min \{ D_1, \bar{q} - q_2^I \}. \quad (9)$$

(ii) *The sign of $\Delta q_1^I / \Delta p$ is unclear, while $\Delta q_2^I / \Delta p \leq 0$ is always true.*

Proof See Appendix B ■

Recall that full prices are determined by p and Ψ . Since a change in p affects the demand for long-distance services and Ψ , the relationship between q_1^I and p is ambiguous.

Figure 3 illustrates rail services under regime I, q_i^I , depending on p and for a given capacity limit $\bar{q} = 1/2$ (all other parameter values are equal to the ones used in Figure 2). Demand functions in a situation with $\bar{q} \rightarrow \infty$, D_i , are indicated by dashed lines. In this instance, the capacity limit is binding for all $p < 1$. Observe that constellations with $q_2^I > D_2$ exist for $p < 1$. This is because, in this situation, Ψ is unaffected by an increase of long-distance services, which increases the demand for long-distance services relative to a situation without a binding capacity limit. For this reason, $D_2^I > D_2$ can happen when the capacity constraint becomes binding.⁷

⁷This is related to the results obtained by Brueckner (2009) and Basso and Zhang (2010), who found that the first-best optimal price for airport slots would be greater than the first-best optimal congestion charge.

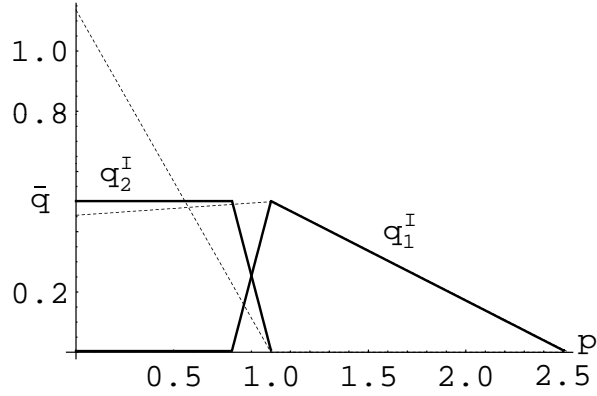


Figure 3: Services q_i^I for $\bar{q} = 1/2$. Dashed lines indicate demands D_i .

Note that short-distance services in (9) and long-distance services in (8) determine the amount of short- and long-distance services, but they ignore rationing of individuals inside customer segments. The rationing of individuals inside customer segments will however become an issue when we consider total surplus (profits plus consumer surplus) or consumer surplus in Section 5.

4 Regime II: Scarcity Premium

Regime I includes a priority rule that favors long-distance services over short-distance services in a situation with excess demand, and excess demand occurs if, for a given price p , $D_1 + D_2$ exceeds \bar{q} . One possibility to avoid excess demand from the beginning is to introduce a scarcity premium $\gamma \geq 0$ that is charged in addition to the service charge, which is the case under regime II. The charge for short-distance services then becomes $p + \gamma$ and that for long-distance services becomes $2p + \gamma$. Since customers pay the service charge

but also bear delay cost Ψ , demand is determined by the ‘full prices’ denoted by ρ_i^{II} with $\rho_1^{II} = p + \gamma + \Psi$ and $\rho_2^{II} = 2p + \gamma + \Psi$. Regime II also includes two stages:

Stage 1: Customers of type 1 and 2 report their demand for short-distance services and long-distance services for a given p and depending on γ to the TOCs. The aggregated demand reports are denoted by $D_i^{II}(\gamma)$. Then, the TOCs forward reports $(D_1^{II}(\gamma), D_2^{II}(\gamma))$ to the rail network provider.

Stage 2: The rail service provider chooses $\gamma \geq 0$ to balance the total demand and the capacity limit, which implies the scarcity premium

$$\gamma^{II} \begin{cases} = 0 & \text{for } D_1^{II}(0) + D_2^{II}(0) \leq \bar{q} \\ \in \{\gamma : D_1^{II}(\gamma) + D_2^{II}(\gamma) = \bar{q}\} & \text{for } D_1^{II}(0) + D_2^{II}(0) > \bar{q}. \end{cases} \quad (10)$$

It is useful to distinguish between four cases:

- (1) $D_1^{II}(0) + D_2^{II}(0) < \bar{q}$,
- (2) $D_1^{II}(\gamma^{II}) + D_2^{II}(\gamma^{II}) = \bar{q}$ and $D_1^{II}(\gamma^{II}), D_2^{II}(\gamma^{II}) > 0$,
- (3) $D_1^{II}(\gamma^{II}) = \bar{q}$ and $D_2^{II}(\gamma^{II}) = 0$, and
- (4) $D_1^{II}(\gamma^{II}) = 0$ and $D_2^{II}(\gamma^{II}) = \bar{q}$.

These cases imply the following outcomes.

Lemma 3 (i) *Under regime II, the scarcity premium is*

$$\gamma^{II} = \max \left\{ 0, \frac{a_2 b_1 + a_1 b_2 - b_2(p + \bar{q}) - b_1(2p + \bar{q} + b_2 \bar{q})}{b_1 + b_2} \right\}, \quad (11)$$

$$a_1 - p - (1 + b_1) \bar{q}, \quad (12)$$

$$a_2 - 2p - (1 + b_2) \bar{q} \}, \quad (13)$$

short-distance services are

$$q_1^{II} = \begin{cases} D_1 & \text{for } \gamma^{II} = 0 \\ \frac{a_1 - a_2 + p + b_2 \bar{q}}{b_1 + b_2} & \text{for } \gamma^{II} = (11) \\ \bar{q} & \text{for } \gamma^{II} = (12) \\ 0 & \text{for } \gamma^{II} = (13), \end{cases} \quad (14)$$

and long-distance services are

$$q_2^{II} = \begin{cases} D_2 & \text{for } \gamma^{II} = 0 \\ \frac{a_2 - a_1 - p + b_1 \bar{q}}{b_1 + b_2} & \text{for } \gamma^{II} = (11) \\ 0 & \text{for } \gamma^{II} = (12) \\ \bar{q} & \text{for } \gamma^{II} = (13). \end{cases} \quad (15)$$

(ii) $\gamma^{II} > 0$ implies

$$\frac{\Delta(p + \gamma^{II})}{\Delta p} \leq 0 \quad \text{and} \quad \frac{\Delta(2p + \gamma^{II})}{\Delta p} \geq 0. \quad (16)$$

(iii) The sign of $\Delta q_1^{II}/\Delta p$ is unclear, while $\Delta q_2^{II}/\Delta p \leq 0$ holds true.

Proof See Appendix C. ■

Observe that $\gamma^{II} > 0$ implies a negative relationship between the short-distance charge and p , while the relationship between the long-distance charge and p is positive due to Lemma 3. This is because a change of p hits type-2 customers twice as much as type-1 customers.

Figures 4 and 5 illustrate the service charges and allocations under regime II. They are based on the same numerical instance as Figures 2 and 3, which illustrate demands D_i or, respectively, allocations under regime I. Figure 4 depicts the total service charges under regime II (sum of the service charge

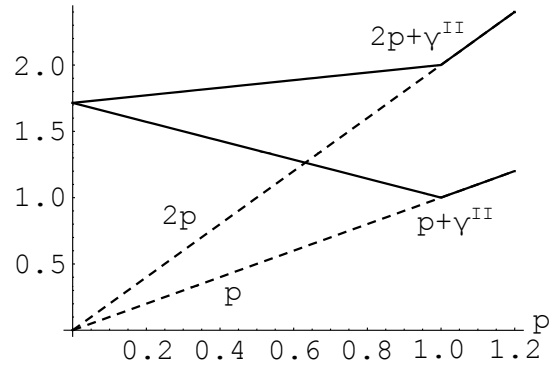


Figure 4: Total service charges under regime II with $\bar{q} = 1/2$.

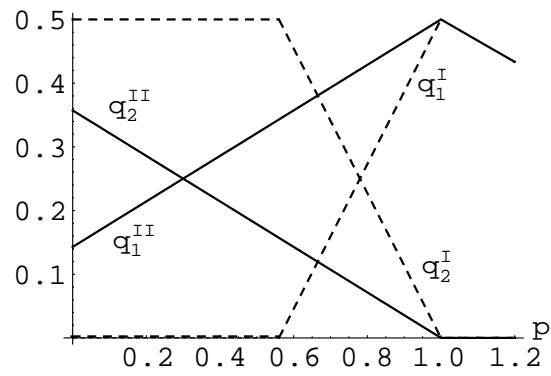


Figure 5: Allocations under regimes I (dashed lines) and II (solid lines) with $\bar{q} = 1/2$.

and scarcity premium) depending on p and illustrates the relationship between the service charges and p .

Figure 5 illustrates rail services under regimes I (dashed lines) and II (solid lines). It shows that, relative to regime I, short-distance services are extended at the expense of long-distance services under regime II (in this instance, $\gamma^{II} > 0 \Rightarrow q_1^{II} > q_1^I$ and $q_2^{II} < q_2^I$).

5 Regime I Versus Regime II

5.1 General considerations

We can find the following general relationship between services under regimes I and II:

Proposition 1 $q_1^I \leq q_1^{II}$ and $q_2^I \geq q_2^{II}$ holds true for all $p \geq 0$.

Proof A priority attached to long-distance services cannot increase the amount of short-distance services compared to a situation where a scarcity premium exists. On the other hand, the introduction of a scarcity premium can increase the amount of short-distance services but cannot increase the amount of long-distance services compared to a situation where priority is attached to long-distance services. ■

It hence turns out that the priority rule is effective in promoting long-distance services compared to a scarcity premium.

We now turn to the relative performance of regimes I and II. From a policy viewpoint, total surplus or consumer surplus are relevant measures to evaluate rail services, and we use both measures to compare allocations.

We also touch the issue of cost recovery, since cost recovery is of critical importance for many regulators.⁸

The rail service provider's revenue is always given by

$$R(q_1, q_2) = (p + \gamma) q_1 + (2p + \gamma) q_2 \quad (17)$$

with $\gamma = 0$ under regime I and $\gamma = \gamma^{II}$ under regime II. Excess demand may only occur under regime I, and total surplus or consumer surplus depends on the amount of short- and long-distance services but, in this situation, also on the rationing rule for individuals inside each customer group. In the following, we use the efficient-rationing rule, which implies that customers who value services most are served first. Under this rule, consumer surplus becomes

$$S^k(q_1, q_2) = \sum_{i=1}^2 \left(\int_0^{q_i} P_i(x_i) dx_i - \rho_i^k q_i \right) \quad (18)$$

for all $k \in \{I, II\}$. Furthermore, the regulator's objective function is

$$V^k(q_1, q_2) = S^k(q_1, q_2) + \beta R(q_1, q_2) \quad (19)$$

with $\beta \in \{0, 1\}$.⁹ If $\beta = 1$, consumer surplus and profits are provided with the same weight and thus total surplus is relevant. If $\beta = 0$, the regulator concentrates on consumer surplus.

It is difficult to derive general analytical results with respect to the relative performances of regimes 1 and 2. One can however find simple illustrative in-

⁸Given their complexity, Ramsey-optimal network charges are however beyond the scope of this paper.

⁹The objective function in (19) is similar to the one considered by (Baron and Myerson, 1982) except that we concentrate on the limiting cases $\beta \in \{0, 1\}$ (Baron and Myerson consider a continuous variation of $\beta \in [0, 1]$) and do not include subsidies or taxes.

stances to show that the effects of regimes I and II on total surplus, consumer surplus and revenues are highly ambiguous in all cases.

For instance, assume that there are two customers 1 and 2 where one demands a short-distance service with a benefit of 14 and the other customer demands a long-distance service with a benefit of 12. Only one customer can be served due to a capacity limit, and benefits are net of delay costs (which are equal for customers 1 and 2). Furthermore, take $p = 3$ as given. In this instance, customer 2 would be served under regime I where long-distance services are preferred over short-distance services. In contrast, a scarcity premium of $\gamma^{II} = 6$ resolves excess demand under regime II to the benefit of customer 1. In this instance, total surplus is greater under regime II. In contrast, consumer surplus is greater under regime I. Note that this result on consumer surplus would change if the benefit of the short-distance service would exceed a value of 15.

Consider a second instance where customer 1's benefit of the short-distance trip drops to 10, while customer 2's benefit of the long-distance trip remains constant at 12. In this instance, customer 2 is served under regime I, which maximizes total surplus. On the other hand, a scarcity premium of $\gamma^{II} = 6$ resolves excess demand to the benefit of customer 1 again, which now reduces total surplus and consumer surplus. Recall that the full price is always greater for long-distance services because p is charged twice in the case of long-distance services, which distorts allocation under regime II and leads to a loss of total surplus in this second instance.

Note that, in both instances, revenues are greater under regime II because $\gamma^{II} > p$ holds true in each instance. However, in reality we are supposed to find constellations where the revenues raised by γ would not be sufficient to cover the foregone revenues that could be raised by p . Hence, even with

respect to revenues, the ranking of regimes I and II is unclear from a theoretical viewpoint. To shed some light on the relative performance of regimes I and II, we present numerical simulations in the next section.

5.2 Numerical simulations

To obtain a more general picture about total surplus, consumer surplus and revenues under regimes I and II, we generated 50 pairs of inverse demands for short- and long-distance services. Each pair of inverse demands requires the choice of 4 parameters denoted by a_{1l}, b_{1l}, a_{2l} and b_{2l} where $l = 1, \dots, 50$ indicates the specific numerical instances. These parameters were drawn from a random process that follows a uniform distribution with support $[0, 3]$. Pairs where the allocation of rail capacity is equal under regimes I and II were sorted out to obtain a sharp picture of the differences between the regimes. Furthermore, we arbitrarily set $\bar{q} = 1/2$ and $F = 1/4$. The average amount of short- or long-distance services over all 50 observations is denoted by $\bar{q}_i^k = 1/50 \sum_{l=1}^{50} q_{il}^k$, the average consumer surplus is denoted by $\bar{S}^k = 1/50 \sum_{l=1}^{50} S^k(q_{1l}^k, q_{2l}^k)$, and the average total surplus is denoted by $\bar{T}S^k = 1/50 \sum_{l=1}^{50} [S^k(q_{1l}^k, q_{2l}^k) + R(q_{1l}^k, q_{2l}^k)]$. The outcomes are illustrated in Figures 6 and 7.

Figure 6 displays services under regime I (dashed lines) and regime II (solid lines) depending on p . Since regime I favors long-distance services over short-distance services, $\bar{q}_1^I \leq \bar{q}_1^{II}$ and $\bar{q}_2^I \geq \bar{q}_2^{II}$ always holds (see Proposition 1). Furthermore, there is a negative relationship between the amount of long-distance services and p under both regime I and regime II, while the relationship between short-distance services and p is unclear (see Lemmas 2 and 3).

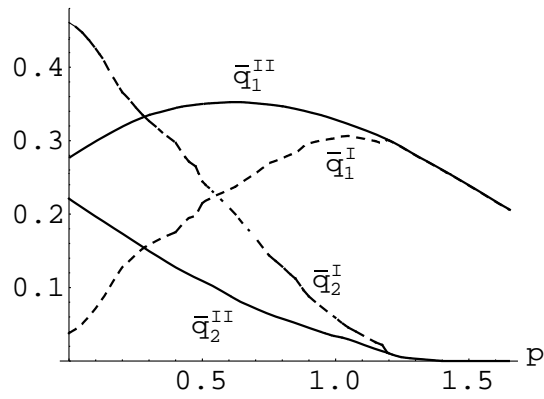


Figure 6: Services under regime I (dashed lines) and regime II (solid lines) with $\bar{q} = 1/2$ depending on p .

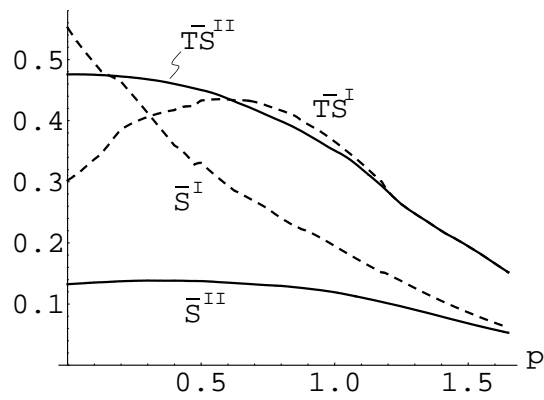


Figure 7: Average consumer surplus, \bar{S}^k , and average total surplus, $\bar{T}S^k$, under regime I (dashed lines) and regime II (solid lines) with $\bar{q} = 1/2$ and $F = 1/4$ depending on p .

Figure 7 displays the values of \bar{S}^k and $T\bar{S}^k$ that are associated with the rail services displayed in Figure 6. In the previous subsection, we used simple examples to demonstrate that the consumer-surplus effect of a scarcity premium is ambiguous in theory. On the one hand, a scarcity premium may imply that customers who exhibit a greater value for rail services are provided with capacity, which can increase consumer surplus. On the other hand, a scarcity premium increases the full price of traveling, which can work into the other direction and reduce consumer surplus. Despite these ambiguous theoretical effects, Figure 7 shows that consumer surplus can be consistently greater under regime I than under regime II ($\bar{S}^I \geq \bar{S}^{II}$ holds for all $p \geq 0$). Moreover, observe that, for great values of p , consumer surplus is greater under regime I than under regime II though the choice of regimes has no effect on the allocation of rail capacity in this area at all. Hence, the introduction of a scarcity premium may not always change services but can still increase the full price and reduce consumer surplus. Altogether, these numerical results indicate that customers are better-off under regime I.

By contrast, simulation results with respect to total surplus are ambiguous. If p takes great values, total surplus is greater under regime I, while total surplus is greater under regime II if p takes low values. Recall that p hits long-distance traffic twice as hard as short-distance traffic. This creates a relative disadvantage for long-distance services under regime II that depends on the level of p (the greater p , the greater is this disadvantage). Since this effect disappears when $p = 0$ holds true, total surplus reaches its maximum under regime II in exactly this situation. Under regime I, on the other hand, the relative disadvantage created by p is partly compensated by the priority for long-distance services, which provides an explanation for the ambiguous results with respect to total surplus.

Figure 7 can also be used to analyze revenues and cost recovery under regimes I and II. Note that average profits can be determined by the difference $\bar{V}^k - \bar{S}^k$. For example, if $\bar{V}^k - \bar{S}^k = 0$ holds true, revenues exactly cover the overall fixed costs $50F$. The effect of regimes I and II is, again, ambiguous in theory. Since regime II favors long-distance services over short-distance services, revenues raised by p are greater under regime I than under regime II. The scarcity premium, on the other hand, creates a new source for revenues under regime II. Now, the comparison of profits under regimes I and II in Figure 7 provides a clear-cut result, revenues are greater under regime II. This result is based on the observation that $\bar{V}^I - \bar{S}^I \leq \bar{V}^{II} - \bar{S}^{II}$ holds true for all $p \geq 0$. Clearly, for p close enough to 0, $\bar{V}^I - \bar{S}^I < 0$ will always be true. It is however interesting to observe that, under regime II, revenues reach their maximum when $p = 0$ is satisfied. So, in this instance, a reduction of p will always increase profits because of the additional revenues raised by an increase of γ^{II} . Overall, this suggests that a scarcity premium may provide a significant contribution to cost recovery in reality.

6 Conclusions

In this paper, we developed a simple model of a congested and capacity limited rail network with two links that is used to offer short- and long-distance services. We considered two regimes to allocate limited network capacity: (i) a priority attached to long-distance services and (ii) a scarcity premium that exactly balances demand and capacity. We found that a priority for long-distance services can increase total surplus if the network charge per train-kilometer reaches great levels. By contrast, a scarcity premium can increase total surplus if the charge per train-kilometer is at low levels. The

simulation results are clear-cut with respect to consumer surplus, which were always greater under the priority attached to long-distance services. It also turned out that fixed network cost recovery may be much easier to reach under a charging system that includes a scarcity premium.

The simulations are based on a stylized model, and some of the assumptions employed may be of critical importance for the results obtained. For example, total and consumer surplus clearly 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 a priority attached to long-distance services, efficient rationing may not always be achieved in reality. Our results are therefore likely to overstate total and consumer surplus obtained under a priority for long-distance services.

Note that both regimes considered can be affected by ‘misreporting’ in reality.¹⁰ For instance, customers might collude and understate their ‘true’ demand to reduce scarcity premiums. Customers might also (wrongly) declare short- as long-distance services to receive a preferred treatment when priority is attached to long-distance services. The overall effect of misreporting on total surplus, consumer surplus and profits 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 the rail service charge is excessive. Hence, to obtain a better understanding of outcomes under

¹⁰Brueckner and Verhoef (2010) analyze manipulable congestion tolls.

different access regulation regimes it would be important to take intermodal competition into consideration.

Finally, in this paper we considered the network conditions, the capacity limit and the network charge per train-kilometer as given. By contrast, it is possible to build new capacity in order to reduce congestion in reality. Furthermore, the choice of the capacity limits and network charges may depend on the regulation regimes applied and the regulator's objective function. It would therefore be useful to extend the current analysis and consider network conditions, capacity limits, and the network charge per train-kilometer 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 a priority rule or a scarcity premium from a policy viewpoint. Future research is however required in order to obtain a more complete picture.

A Proof of Lemma 1

(i) If $(D_1, D_2) > 0$ holds true, conditions

$$P_1 = \rho_1^I \quad \text{and} \quad P_2 = \rho_2^I. \quad (20)$$

must be satisfied. Solving simultaneously the conditions in (20) implies

$$D_1 = \frac{a_1 - a_2 + a_1 b_2 - p(b_2 - 1)}{b_1 + b_2 + b_1 b_2} \quad (21)$$

and

$$D_2 = \frac{a_2 - a_1 + a_2 b_1 - p(1 + 2b_1)}{b_1 + b_2 + b_1 b_2}. \quad (22)$$

(ii) If $D_1 \geq 0$ and $D_2 = 0$ holds true, conditions

$$P_1 = \rho_1 \quad \text{and} \quad P_2 \leq \rho_2^I \quad (23)$$

must be satisfied. Substituting q_2 by 0 in ρ_1 and solving the first condition in (23) implies

$$D_1 = \max \left\{ 0, \frac{a_1 - p}{1 + b_1} \right\}. \quad (24)$$

(iii) If $D_2 \geq 0$ and $D_1 = 0$ holds true, conditions

$$P_1 \leq \rho_1 \quad \text{and} \quad P_2 = \rho_2^I \quad (25)$$

must be satisfied. Substituting q_1 by 0 in ρ_2 and solving the second condition in (25) implies

$$D_1 = 0 \quad \text{and} \quad D_2 = \max \left\{ 0, \frac{a_2 - 2p}{1 + b_2} \right\}. \quad (26)$$

Furthermore, it is useful to denote a critical level of a_1 by $\tilde{a}_1 = a_2/2$, and to distinguish $b_2 > 1$, $b_2 < 1$, and $b_2 = 1$. The following table shows the relevance of cases (i)-(iii) depending on b_2 and p :

	$(D_1, D_2) > 0$	$D_1 \geq 0, D_2 = 0$	$D_1 = 0, D_2 \geq 0$
$b_2 > 1$	$p \leq p_1, p_2$	$p > p_2 = \min\{p_1, p_2\}$	$p > p_1 = \min\{p_1, p_2\}$
$b_2 < 1$	$p \in (p_1, p_2]$	$p > p_2 \geq p_1$	$p \leq p_1 \leq p_2$ or $p_1 > p_2$
$b_2 = 1$	$a_1 > \tilde{a}_1, p < p_2$	$a_1 > \tilde{a}_1, p \geq p_2$	$a_1 \leq \tilde{a}_1$

It is straightforward to obtain the comparative-static results based on the demands in (21)-(26).

B Proof of Lemma 2

(i) Two constellations should be distinguished: (a) $D_1^I + D_2^I < \bar{q}$ and (b) $D_1^I + D_2^I \geq \bar{q}$. In (a), the capacity constraint is not binding and $(D_1^I, D_2^I) = (D_1, D_2)$ holds true. In (b), the capacity constraint is binding and an increase of the demand for long-distance services leads to a reduction of short-distance services by exactly the same amount. In this situation, the demand for long-distance services is obtained by substituting q_1 by $\bar{q} - q_2$ in the second equilibrium condition in (20), $P_2 = \rho_2^I$, and solving for q_2 . This gives

$$q_2^I = \frac{a_2 - 2p - \bar{q}}{b_2}. \quad (27)$$

The allocations in (8) and (9) follow.

(ii) Since the sign of $\Delta D_1/\Delta p$ is ambiguous due to Lemma 1, the sign of $\Delta q_1^I/\Delta p$ is also ambiguous. Furthermore, since $\Delta D_2/\Delta p \leq 0$ and the first derivative of the right-hand side in (27) with respect to p is also negative, $\Delta q_2^I/\Delta p \leq 0$ follows.

C Proof of Lemma 3

(i) If $\gamma = 0$, $D_i^{II}(0) = D_i$ for all $i = 1, 2$. If, in addition, $D_1^{II}(0) + D_2^{II}(0) \leq \bar{q}$, $q_i^{II} = D_i$ holds for all for all $i = 1, 2$.

If $\gamma > 0$ and $D_1^{II}(\gamma), D_2^{II}(\gamma) > 0$ holds true, conditions $P_1 = \rho_1^{II}$ and $P_2 = \rho_2^{II}$ must be satisfied. Solving these conditions simultaneously implies reports

$$D_1^{II}(\gamma) = \frac{a_1(1 + b_2) - a_2 + p - b_2(p + \gamma)}{b_1 + b_2 + b_1 b_2} \quad (28)$$

and

$$D_2^{II}(\gamma) = \frac{a_2(1 + b_1) - a_1 - p - b_2(2p + \gamma)}{b_1 + b_2 + b_1b_2}. \quad (29)$$

If, in addition, $D_1^{II}(0) + D_2^{II}(0) > \bar{q}$ holds true, the scarcity premium implied by (10) is

$$\gamma^{II} = \frac{a_2b_1 + a_1b_2 - b_2(p + \bar{q}) - b_1(2p + \bar{q} + b_2\bar{q})}{b_1 + b_2}, \quad (30)$$

leading to services

$$q_1^{II} = \frac{a_1 - a_2 + p + b_2\bar{q}}{b_1 + b_2} \text{ and } q_2^{II} = \frac{a_2 - a_1 - p + b_2\bar{q}}{b_1 + b_2}. \quad (31)$$

If $D_1^{II}(\gamma) > 0$ and $D_2^{II}(\gamma) = 0$, condition $P_1 = \rho_1^{II}$ must be satisfied. Solving this condition implies the report

$$D_1^{II}(\gamma) = \max \left\{ 0, \frac{a_1 - p - \gamma}{1 + b_1} \right\}. \quad (32)$$

If, in addition, $D_1^{II}(0) > \bar{q}$ holds true, the scarcity premium implied by (10) is

$$\gamma^{II} = a_1 - p - (1 + b_1)\bar{q}, \quad (33)$$

leading to services

$$q_1^{II} = \bar{q} \text{ and } q_2^{II} = 0. \quad (34)$$

If $D_1^{II}(\gamma) = 0$ and $D_2^{II}(\gamma) > 0$, condition $P_2 = \rho_2^{II}$ must be satisfied. Solving this condition implies the report

$$D_2^{II}(\gamma) = \max \left\{ 0, \frac{a_2 - 2p - \gamma}{1 + b_2} \right\}. \quad (35)$$

If, in addition, $D_2^{II}(0) > \bar{q}$ holds true, the scarcity premium implied by (10) is

$$\gamma^{II} = a_2 - 2p - (1 + b_2) \bar{q}, \quad (36)$$

leading to services

$$q_1^{II} = 0 \text{ and } q_2^{II} = \bar{q}. \quad (37)$$

Note that γ must be chosen such that $q_1^{II} + q_2^{II} \leq \bar{q}$ is always satisfied. It follows that γ^{II} is determined by the maximum of 0 or the right-hand sides of (30), (33) and (36), and quantities are determined by D_i or services in (31), (34) and (37), depending on which case actually determines γ^{II} .

(ii) Differentiating (30), (33), and (36) with respect to p leads to

$$\frac{\partial \gamma^{II}}{\partial p} \in [-2, -1]. \quad (38)$$

The comparative-static results in (16) follow.

(iii) Since the sign of $\Delta D_1 / \Delta p$ is ambiguous due to Lemma 1, the sign of $\Delta q_1^{II} / \Delta p$ is also ambiguous. Furthermore, since $\Delta D_2 / \Delta p \leq 0$ and the first derivative of long-distance services in (31) with respect to p is also negative, $\Delta q_2^{II} / \Delta p \leq 0$ follows.

References

- Baake, P. and Mitusch, K. (2007). Competition with congestible networks. *Journal of Economics*, 91(2):151–176.
- 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*, 61(2):218–237.
- Basso, L. and Zhang, A. (2010). Pricing vs. slot policies when airport profits matter, *Transportation Research, Part B*, 44(3):381–391.
- 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. (2002). Airport congestion when carriers have market power. *American Economic Review*, 92(5):1357–1375.
- Brueckner, J. K. (2009). Price vs. quantity based approaches to airport congestion management. *Journal of Public Economics*, 93(5-6):681–690.
- 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.
- Brueckner, J. K. and Verhoef, E. T. (2010). Manipulable congestion tolls. *Journal of Urban Economics*, 67(3):315–321.
- 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.

- 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.
- Oum, T. H., Zhang, A., and Zhang, Y. (2004). Alternative forms of economic regulation at airports. *Journal of Transport Economics and Policy*, 38(2):217–246.
- 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.
- Zhang, A. and Zhang, Y. (2006). Airport capacity and congestion when carriers have market power. *Journal of Urban Economics*, 60(2):229–247.