Access Pricing for Mixed Users

Etienne Billette de Villemeur
University of Toulouse (IDEI & GREMAQ).

First Version (Please do not quote)

Abstract

The major part of the rail network is used for both passenger and freight traffic. I exhibit the socially optimal access pricing structure when the heterogeneity of users is taken into account. Users differ in their willingness to pay but also in the maintenance cost they induce for the authority in charge of the network. Tracks are an essential and congestible facility. Different regimes are thus considered depending on the presence or not of congestion. When there is congestion, the difference in speeds (that governs the magnitude of the negative externality imposed on the others) appears to be an important determinant of the access prices. Along the paper, the socially optimal price structure is compared to the price structure that would prevail if the authority in charge of the network is profit maximising. The most striking result of the paper is that, in all cases, optimal pricing formulae do *not* refer to average costs when heterogeneity is taken into account.

J. E. L. Classification Numbers: D42, L10, L92.

Keywords: Rail industry, access-pricing, congestion.

1 Introduction

Tracks constitute an essential facility for the rail transportation operators and access pricing is the *crux* of a good working of a liberalised industry. Clearly, we are far from the best use of the infrastructure. In Europe, the average speed of goods international transport is 18 km/h ![1] There is thus no surprise in the observed decline of rail's share among the transportation modes.

A sound pricing scheme should however take into account that the major part of the rail network is used for both passenger and freight traffic. I thus consider a model where users may differ in their willingness to pay but also in the maintenance cost they induce for the authority in charge of the network. The socially optimal access pricing structure, with and without subsidies (First and Second-best), as well as the profit-maximising price structure are exhibited. The central issue consists in knowing whether the infrastructure manager is able to discriminate across users. We consider two polar cases: perfect discrimination and no discrimination and compare both for each of the scenarii considered.

Tracks is an essential but also a congestible facility. Congestion affects indeed 20% of the 16000 km of the European rail network.[2] I thus examine how the price structure should be modified to in order to take this aspect into account. Again the pricing scheme depends on the information available. If no congestion characteristics but the speed is known to the network administration, the tariffs should have a (standard) peak-load pricing component as already proposed by Vickrey (1963). However, such a bold approach does

[1]Source: European Comission White Paper: "European Tranport Policy for 2010: Time to decide" (2001).

[2]Source: European Comission, *ibidem*

account for the fact that, in the rail sector, congestion depends on the exact pattern of services. In a simplified two-types model, a pricing formula is proposed that reckon the fact that, an heterogeneous pattern of trains use a greater slot capacity than a set of trains with identical characteristics.

2 The model

The users of the rail network are characterised by a (vector of) parameter(s) $\boldsymbol{\theta}$ that accounts for all their characteristics. They potentially differ in the benefits they derive from the use of the infrastructure (hence in their willingness to pay for it) and in the costs they impose both on the owner of the infrastructure (the maintenance costs) and on the other users (the congestion effects). A key issue of the problem is naturally the discrimination across users that may be impossible for informational or legal reasons. Two polar situations will be considered: perfect discrimination and no discrimination. In order to identify the working of the different mechanism at hand, I assume first that there is no congestion.

Let $\pi_{\boldsymbol{\theta}}(x)$ denote the benefits derived from running x trains of characteristics $\boldsymbol{\theta}$. If $p_{\boldsymbol{\theta}}$ denotes the access-price for type-$\boldsymbol{\theta}$ trains, the ensuing demand for network access from this type of trains is given by

$$x_{\boldsymbol{\theta}}(p_{\boldsymbol{\theta}}) = \arg\max_{X} \{\pi_{\boldsymbol{\theta}}(x) - p_{\boldsymbol{\theta}} x\}. \tag{1}$$

In what follows, the total demand will be denoted

$$X(p) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} x_{\boldsymbol{\theta}}(p)\, dG(\boldsymbol{\theta}).$$

For the infrastructure owner, a train of type $\boldsymbol{\theta}$ induces a cost denoted $c_{\boldsymbol{\theta}}$. Her benefits thus write

$$\Pi = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (p_{\boldsymbol{\theta}} - c_{\boldsymbol{\theta}})\, x_{\boldsymbol{\theta}}(p_{\boldsymbol{\theta}})\, dG(\boldsymbol{\theta}) - F, \tag{2}$$

where F denotes the fixed costs of network maintenance and $G(\boldsymbol{\theta})$ denotes the distribution function over the space of characteristics Θ.

2.1 Social Optimum

2.1.1 Perfect discrimination

In this section, we analyse the first-best allocation, that is the allocation that maximizes social welfare (the sum of consumer surplus and firms' profits). At this stage the company is not required to break-even. We thus implicitly assume that (i) perfect discrimination is possible, *i.e.* all the characteristics $\boldsymbol{\theta}$ are observable,.and that (ii) fixed costs can be financed without efficiency costs through a subsidy financed from the general budget. Such a solution is theoretical and usually not considered to be realistic. Nevertheless it provides us with an interesting benchmark.

Total surplus can be expressed as follows:

$$W = \int_{\boldsymbol{\theta}} \left[\pi_{\boldsymbol{\theta}}(x_{\boldsymbol{\theta}}) - c_{\boldsymbol{\theta}} x_{\boldsymbol{\theta}}\right] dG(\boldsymbol{\theta}) - F, \tag{3}$$

where fixed costs and maintenance costs are substracted from the sum of the users' benefits. Differentiating (3) with respect to $x_{\boldsymbol{\theta}}$ yields the following first-order condition:

$$\pi'_{\boldsymbol{\theta}}(x_{\boldsymbol{\theta}}) = c_{\boldsymbol{\theta}}. \tag{4}$$

Equation (4) evidences that the optimal number of type-$\boldsymbol{\theta}$ trains is such that the marginal benefits of a train for the operator exactly equates the marginal costs for the infrastructure administrator. Given equation (1) defining the demand for access, the optimal allocation leads to the optimal pricing formula:

$$p_{\theta} = c_{\theta}. \tag{5}$$

Expression (5) show that price discrimination is necessary for the first-best allocation to be decentralized. Interestingly, this is also a sufficient condition for an efficient setting of the characteristics of transportation services. If the equation holds, the (only) users are those for whom the access-price is smaller than the firms' marginal benefits .

A consequence of this (optimal) pricing policy is however that the company does not break-even. More precisely, sales will cover only variable costs and the deficit will amount to the fixed costs F. As a result, the first-best solution is not feasible if the infrastructure administrator faces a break-even constraint. One should then consider a second-best solution where prices are set above marginal cost in order to cover all costs. This issue is addressed below.

2.1.2 No price discrimination

Before to turn to the study of the second-best allocation, we now analyse the first-best allocation when price discrimination is not possible. In this case, the total surplus still writes as (3). However the number of type-$\boldsymbol{\theta}$ users $x_{\boldsymbol{\theta}}$ must now obey equation (1) with $p_{\boldsymbol{\theta}} = p$ and there is only one instrument left, namely the price. Differentiating (3) with respect to p yields the following first-order condition:

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\pi'_{\boldsymbol{\theta}}\left(x_{\boldsymbol{\theta}}\right) - c_{\boldsymbol{\theta}}\right] x'_{\boldsymbol{\theta}}\left(p\right) dG\left(\boldsymbol{\theta}\right) = 0. \tag{6}$$

By using equation (1), equation (6) can be rewritten as

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[p - c_{\boldsymbol{\theta}}\right] \alpha_{\theta}\epsilon_{\boldsymbol{\theta}} dG\left(\boldsymbol{\theta}\right) = 0. \tag{7}$$

where $\epsilon_{\boldsymbol{\theta}}$ denotes the (absolute value of) the price elasticity of type-$\boldsymbol{\theta}$ demand for access

$$\epsilon_{\boldsymbol{\theta}} = \frac{p}{x_{\boldsymbol{\theta}}\left(p\right)}\left(\frac{-dx_{\boldsymbol{\theta}}}{dp}\right),$$

and α_θ the share of type-$\boldsymbol{\theta}$ users

$$\alpha_\theta = \frac{x_{\boldsymbol{\theta}}(p)}{X(p)}.$$

Interestingly enough, the "demand side" enters now into the definition of the optimal allocation. The reason for this derives from the fact that, as evidenced by equation (5), productive efficiency is not possible without price discrimination. Indeed, "low cost" users are priced above the cost they really impose on the infrastructure and some users may not get access while they would be able to derive benefits that exceed their cost. Similarly, "high cost" users are priced below the cost they impose on the infrastructure and some may get access while they impose a higher cost on the infrastructure than the price they pay for it. In minimising the distortions, however, one should not only take into account the difference between price and marginal cost. It should also determines how much of the demand is diverted from over-pricing and how much of the demand comes from under-pricing. How the demand is affected by a change in price is precisely what is measured by the price elasticity. When equation (7) holds true, the overall distortion (due to the absence of price discrimination) is minimum.

Equation (7) gives rise to the optimal ("first-best") pricing rule:

$$p = \widetilde{c} = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\alpha_\theta \frac{\epsilon_{\boldsymbol{\theta}}}{\epsilon_X}\right) c_{\boldsymbol{\theta}} dG(\boldsymbol{\theta}), \tag{8}$$

where ϵ_X the (absolute value of) of the *total* demand price elasticity

$$\epsilon_X = \frac{p}{X(p)} \left(\frac{-dX(p)}{dp}\right).$$

It evidences that, when price discrimination is not possible, the usual "marginal cost pricing" rule writes with a cost $\widetilde{c}$ that differs from the average marginal cost

$$\overline{c} = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \alpha_\theta c_{\boldsymbol{\theta}} dG(\boldsymbol{\theta}).$$

More specifically, if “high cost” users tend to be associated with relatively low values of elasticity, the cost $\widetilde{c}$ will tend to be lower than $\overline{c}$ leading the infrastructure owner not to cover even the variable costs. The converse will happen if “high cost” users tend to be associated with relatively high values of elasticity. Given the heterogeneity of the marginal costs of infrastructure usage reported in the literature (See *e.g.* Gaudry and Quinet 2003), a good evaluation of the reference cost $\widetilde{c}$ appears of crucial importance for the assessment of the pricing policy.

Note that, as for prefect discrimination case, there are no reasons for which the infrastructure owner would be able to cover the fixed costs and break-even.

2.2 Transportation services with a profit maximising monopolist

The first-best allocation has been computed by considering social welfare and by fully ignoring the issue of profitability. We now turn to the reverse situation by considering the choices made by a profit-maximising infrastructure manager. Again, two cases are to be considered, depending on whether there is price discrimination or not.

If perfect discrimination is possible, the infrastructure owner maximizes its profits for each type $\boldsymbol{\theta}$ of train. By deriving (2) with respect to $p_{\boldsymbol{\theta}}$ and re-arranging terms, we obtain the standard Lerner formula:

$$\frac{p_{\boldsymbol{\theta}} - c_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}} = \frac{1}{\epsilon_{\boldsymbol{\theta}}}. \tag{9}$$

Although simple, the pricing rule (9) conveys an interesting implication. If the infrastructure manager aims at maximizing profits and is able to discriminate across users, then the price ordering may not reflect the cost ordering. It may well be the case that a “high cost” type with a very elastic demand is charged a lower access price than a "low cost" type with a less elastic demand. Again, given the estimated range of price-elasticities for different categories of travellers[3] that has to be reflected in the elasticities by type of train $\epsilon_{\boldsymbol{\theta}}$, this is much more than a thought experiment. In particular it would be no wonder if long-distance trains would be asked to pay a higher (access) fee/km than short-distance (or inner-city) services, despite the later are typically retained to induce higher maintenance costs.

If price discrimination is *not* possible, the pricing rule appears *a priori* to be less straightforward. Imposing $p_{\boldsymbol{\theta}} = p$ and deriving the profits (2) with respect to this unique price gives rise to the FOC:

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[x_{\boldsymbol{\theta}}(p) + (p - c_{\boldsymbol{\theta}})\, x'_{\boldsymbol{\theta}}(p)\right] dG(\boldsymbol{\theta}) = 0.$$

This equation can nevertheless be rewritten under the familiar form

$$\frac{p - \widetilde{c}}{p} = \frac{1}{\epsilon_X}, \tag{10}$$

where the "marginal cost " $\widetilde{c}$ is the optimal ("first-best" without discrimination) price defined by equation (8). Interestingly enough, this is not the average marginal costs $\overline{c}$ that should be considered in order to compute the profit-maximising prices. Again, given the observed heterogeneity in the rail track wear-and-tear costs, the point appears to be empirically relevant.
Finally, note that the profits raised are lower without discrimination than the one that will prevail if the pricing rule (9) can be implemented. Assessing the magnitude of the loss is obviously a challenging problem. It goes however beyond the scope of this paper and the issue is left for future research.

[3]See, *e.g.* "Progress in Rail Reform", a report by the Australian Productivity Commission (1999). The price elasticities for CityRail in New South Wales are reported to vary from -0.08 for a single ticket user to -0.53 for a customer buying a travel pass.

2.3 Second-best

We now turn to the so-called second-best problem, where social welfare is maximised under the constraint that the users pay for the cost they induce so that there is no need for subsidies.

Denote by $L = W + \lambda \Pi$ the Lagrangian expression associated with this problem while λ is the multiplier of the break-even constraint. In the case of perfect discrimination, deriving L with respect to $p_{\boldsymbol{\theta}}$, we obtain the following first order conditions:

$$\frac{\partial L}{\partial p_{\boldsymbol{\theta}}} = \left\{ \pi'_{\boldsymbol{\theta}} \left[x_{\boldsymbol{\theta}} \left(p_{\boldsymbol{\theta}} \right) \right] - c_{\boldsymbol{\theta}} \right\} x'_{\boldsymbol{\theta}} \left(p_{\boldsymbol{\theta}} \right) + \lambda \left\{ x_{\boldsymbol{\theta}} \left(p_{\boldsymbol{\theta}} \right) + \left(p_{\boldsymbol{\theta}} - c_{\boldsymbol{\theta}} \right) x'_{\boldsymbol{\theta}} \left(p_{\boldsymbol{\theta}} \right) \right\} = 0$$

all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By using equation (1), which defines the access demand, together with the various notations introduced above, it can be simplified to obtain:

$$\frac{p_{\boldsymbol{\theta}} - c_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}} = \frac{\lambda}{1 + \lambda} \frac{1}{\epsilon_{\boldsymbol{\theta}}}. \tag{11}$$

Equation (11) shows that, as long as perfect discrimination is possible, all types of train can be considered in a separate manner. The rule that governs the setting of prices at the second-best is the standard Ramsey formula. Since the distortion that follows from a price set above the marginal cost increases with the elasticity of demand, the mark-up should be inversely related to this price elasticity. It is set in such a way that the overall distortion is minimised and the firm can recover all its costs. The magnitude of the distortion is measured by the shadow price λ which is the only parameters that does not depend only on the type $\boldsymbol{\theta}$ but on the overall distribution of characteristics in the population.

As for the case where profit-maximisation is the objective of infrastructure manager, equation (11) implies that a user inducing a higher cost should not necessarily be charged a higher price.[4] In other words, even when the

[4] although it is less likely since $\lambda / (1 + \lambda) < 1$.

maximisation of social welfare is the goal, the very fact of taking into account the sustainability of the infrastructure breaks the direct relationship between price and marginal costs.

If price discrimination is *not* possible, imposing $p_{\boldsymbol{\theta}} = p$ and deriving the Lagrangian L with respect to this unique price gives rise to the FOC:

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{ \left[\pi'_{\boldsymbol{\theta}} \left[x_{\boldsymbol{\theta}} \left(p \right) \right] - c_{\boldsymbol{\theta}} \right] x'_{\boldsymbol{\theta}} \left(p \right) + \lambda \left[x_{\boldsymbol{\theta}} \left(p \right) + \left(p - c_{\boldsymbol{\theta}} \right) x'_{\boldsymbol{\theta}} \left(p \right) \right] \right\} dG \left(\boldsymbol{\theta} \right) = 0.$$

Again, this condition gives rise to a familiar Ramsey formula :

$$\frac{p - \widetilde{c}}{p} = \frac{\lambda}{1+\lambda} \frac{1}{\epsilon_X}, \tag{12}$$

where the "marginal cost " $\widetilde{c}$ is defined by equation (8). Note that the Lagrange multiplier obtained here is *higher* than the one that appears in equation (11). The lack of price discrimination has thus a *double* negative impact on welfare. First, the infrastructure owner cannot differentiate across types (hence costs) with induces welfare losses that are already present in the first-best allocation. Second, it is more difficult to raise profits, thus the necessary price distortion appears to be more important without price discrimination.[5]

3 Effect of congestion

Peak-load pricing is a well-known recommendation of transportation economists.[6] Its precise display when users are heterogeneous is less obvious, especially if heterogeneity may also concerns the congestion characteristics.

[5] This does not mean however that the access price without discrimination will be necessarily higher than the average price with prefect discrimination since there is no straightforward relationship neither between the "marginal cost" $\widetilde{c}$ and the average marginal cost $\overline{c}$, nor between the elasticity of the aggregate demand ϵ_X and the average elasticity.

[6] See, among others, Vickrey (1963) and Arnott and Small (1994)

In what follows, we tackle the issue of congestion by considering again the problems addressed above with the additional constraint

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} t_{\boldsymbol{\theta}}(H)\, x_{\boldsymbol{\theta}}(p_\theta)\, dG(\boldsymbol{\theta}) \leq T, \tag{13}$$

where $t_{\boldsymbol{\theta}}(H)$ is the minimum delay that should separate the access of a type $\boldsymbol{\theta}$-train from the rest of the traffic and T a constant characterising the "capacity" of the track where congestion actually occurs. Although the rail infrastructure is a complex network of tracks, few of them are critical for an efficient management. Part of the tracks constitute indeed true bottlenecks that condition the correct working of the whole system. I thus ignore the precise structure of the network to concentrate on pricing formula for a single but congested track. The complexity of the problem finds its origin in the multiple sources of heterogeneity. In addition to the differences already introduced above, trains may differ in the duration of their use of the tracks (a direct consequence of the speed differences). However, the delay $t_{\boldsymbol{\theta}}$ depends more generally on the precise distribution of types along the day, namely H. Indeed similar trains may follow with a reduced delay while different trains has to be separated by a extended one. In order to isolate the various phenomena at hand, I first assume that $t_{\boldsymbol{\theta}}(H)$ is independent of both $\boldsymbol{\theta}$ and H to highlight the effects of the limited capacity of the track; The effect of the differences in speeds and in congestion are studied afterwords.

In order to avoid a multiplication of cases, we shall study simultaneously, first-best, second-best and profit-maximisation[7]. It is indeed possible to introduce a general problem which Lagrangian writes

$$\begin{aligned} L &= \int_{\boldsymbol{\theta}} \left[\pi_{\boldsymbol{\theta}}(x_{\boldsymbol{\theta}}) - c_{\boldsymbol{\theta}} x_{\boldsymbol{\theta}}\right] dG(\boldsymbol{\theta}) - F + \lambda \left\{ \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (p_{\boldsymbol{\theta}} - c_{\boldsymbol{\theta}})\, x_{\boldsymbol{\theta}}(p_{\boldsymbol{\theta}})\, dG(\boldsymbol{\theta}) - F \right\} \\ &\quad + (1+\lambda)\,\mu \left\{ T - \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} t_{\boldsymbol{\theta}}(H)\, x_{\boldsymbol{\theta}}(p_{\boldsymbol{\theta}})\, dG(\boldsymbol{\theta}) \right\}, \end{aligned}$$

[7] The same approach is adopted by Laffont and Tirole (1993).

where λ is the shadow price of the break-even constraint and μ the shadow price of congestion. When the first-best allocation is considered, the multiplier λ should be set to zero. When profit-maximisation is under scrutiny, one take the limit when λ goes to infinity. The factor $(1+\lambda)$ that multiplies the shadow price μ is introduced as to insure that congestion is taken into account also in this limit cases. After this technical preliminary, we now go to the results, according to different assumptions regarding the congestion characteristics $t_{\boldsymbol{\theta}}(H) \equiv t$.

3.1 Identical congestion characteristics

In this section, I assume that all the services have the same congestion caracteristics: $t_{\boldsymbol{\theta}}(H) \equiv t$.

If prefect discrimination is possible, the optimal pricing formula writes:

$$p_{\boldsymbol{\theta}} = c_{\boldsymbol{\theta}} + \frac{\lambda}{1+\lambda} \frac{p_{\boldsymbol{\theta}}}{\epsilon_{\boldsymbol{\theta}}} + \mu t. \tag{14}$$

In addition to the marginal cost and the Ramsey markup already present in the previous formulae, equation (14) contains a third term, namely μt that reflects the consequences of congestion for the optimal price. Interestingly enough, as long as the congestion characteristics do no differ, the optimal price is uniformly increased as to meet the capacity of the link. No difference should be made across users. Even if some appears to be more profitable than others the cost of congestion is distributed equally.

If there is no discrimination, the optimal pricing formula writes:

$$p = \widetilde{c} + \frac{\lambda}{1+\lambda} \frac{p}{\epsilon_X} + \mu t. \tag{15}$$

Again, the treatment of congestion appears again as a completely separated problem in (15). This does not come as a surprise since it was already the case when prefect discrimination was assumed to be possible.

3.2 Pricing of congestion when duration of usage may differ

In this section I study how congestion should be priced if the trains have different speeds. In this case, their impact on congestion also differ. However, for the time being, the precise sequence of trains H is assumed not to be relevant, so that $t_{\boldsymbol{\theta}}(H) \equiv t_{\boldsymbol{\theta}}$.

If prefect discrimination is possible, the optimal pricing formula writes simply :

$$p_{\boldsymbol{\theta}} = c_{\boldsymbol{\theta}} + \frac{\lambda}{1+\lambda} \frac{p_{\boldsymbol{\theta}}}{\epsilon_{\boldsymbol{\theta}}} + \mu t_{\theta}. \tag{16}$$

Equation (16) states simply that congestion should be priced proportionally to duration of usage as measured by the parameter t_{θ}. Note however that the shadow price of congestion μ depends on the whole distribution of characteristics[8].

When no discrimination is possible, the optimal pricing formula writes:

$$p = \widetilde{c} + \frac{\lambda}{1+\lambda} \frac{p}{\epsilon_X} + \mu \widetilde{t}, \tag{17}$$

where

$$\widetilde{t} = \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\alpha_{\theta} \frac{\epsilon_{\boldsymbol{\theta}}}{\epsilon_X} \right) t_{\boldsymbol{\theta}} dG(\boldsymbol{\theta}). \tag{18}$$

Equation (17) and (18) make it clear that, in reality, the pricing of congestion cannot be considered as a problem completely independent from the other users' characteristics. In absence of discrimination, the only constraint is that the (unique) price is sufficiently high for the traffic not to exceed the capacity of the link so that equation (17) may appear a bit artificial. It nevertheless highlights interesting consequences of this form of congestion pricing. First, the price increase needed to decrease congestion has an impact

[8] Just like it is the case for λ.

on the demand for access that depends on the price elasticity of each user. Thus, the (shadow) cost of congestion as measured by μ does not really depends on the average time of usage $\bar{t} = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} t_{\boldsymbol{\theta}} dG(\boldsymbol{\theta})$ but on the "virtual value" $\widetilde{t}$ specified by equation (18). As a result, if the slower services are those with a relatively higher price elasticity, $\widetilde{t} > \bar{t}$ and congestion can be "easily" solved. Conversely, if the slower services display a smaller than average price elasticity, $\widetilde{t} < \bar{t}$ and congestion is expensive to be solved *i.e.* will require an "higher than expected" increase in prices. Second, because the reduction in access demand is directly related to the price sensitivity of the different users, the first to drop are the users with the highest price elasticity and not necessarily the slower services or those that pay a higher fraction of the congestion costs.

The interesting problem lies however in the intermediate case where the (multidimensional) parameter $\boldsymbol{\theta}$ is partially observable. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ where $\boldsymbol{\alpha}$ is observable, $\boldsymbol{\beta}$ unobservable. Assume furthermore that the observable component is related to the sole congestion properties, while the unobservable component is related to the other aspects like price-elasticities and cost properties. The F.O.C. defining the optimal pricing formula writes:

$$p_{\boldsymbol{\alpha}} = \widetilde{c_{\boldsymbol{\alpha}}} + \frac{\lambda}{1+\lambda}\frac{p_{\boldsymbol{\alpha}}}{\epsilon_{X_{\alpha}}} + \mu t_{\alpha}. \tag{19}$$

where $X_{\boldsymbol{\alpha}}(p_{\boldsymbol{\alpha}}) = \int_{\boldsymbol{\beta}\in\mathbf{B}} x_{\boldsymbol{\theta}}(p_{\boldsymbol{\alpha}}) \, dG_{\alpha}(\boldsymbol{\beta})$ and, more generally, the index $\boldsymbol{\alpha}$ indicates the (partial) aggregation over the users of type $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ that share an identical observable component $\boldsymbol{\alpha}$. If the pair $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is made of independent parameters and no information indicates a systematic bias, $\widetilde{c_{\boldsymbol{\alpha}}} = \widetilde{c}$ and $\epsilon_{X_{\alpha}} = \epsilon_X$ so that

$$\frac{p_{\boldsymbol{\alpha}} - \widetilde{c}}{p_{\boldsymbol{\alpha}}} = \frac{\lambda}{1+\lambda}\frac{1}{\epsilon_X} + \mu\frac{t_{\alpha}}{p_{\boldsymbol{\alpha}}}. \tag{20}$$

In words, the markup is the sum of two components: the standard Ramsey markup, identical for all users and a congestion term that increases with the duration of network use as measure by t_α. Again, one has to keep in mind that μ depends on the whole distribution of types and that the first users to drop are not necessarily those with the higher ratio $(t_\alpha / p_{\boldsymbol{\alpha}})$.

3.3 "Perfect Congestion Pricing" : the two types case

In this section, I recognize explicitly that congestion depends on the exact sequence of trains H. In order to give a simple and explicit solution to this complex problem, attention is restricted to the two types case: $\boldsymbol{\Theta} = \{\underline{\theta}, \overline{\theta}\}$. Without any loss of generality, $\underline{\theta}$ is assumed to be the slow type.
Denotes t_θ^0 the minimum delay between to services of the (same) type $\boldsymbol{\theta}$. It is well known that $t_\theta^0 \le t_{\boldsymbol{\theta}}(H)$ and given the restriction to the two type case, I shall use the following notations:

$$\begin{aligned} t_{\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}} \rightarrow \underline{\boldsymbol{\theta}}\right) &= t_{\underline{\boldsymbol{\theta}}}^0 \quad \textit{and} \quad t_{\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}} \rightarrow \overline{\boldsymbol{\theta}}\right) = t_{\underline{\boldsymbol{\theta}}}^0 + \Delta, \\ t_{\overline{\boldsymbol{\theta}}}\left(\overline{\boldsymbol{\theta}} \rightarrow \overline{\boldsymbol{\theta}}\right) &= t_{\overline{\boldsymbol{\theta}}}^0 \quad \textit{and} \quad t_{\overline{\boldsymbol{\theta}}}\left(\overline{\boldsymbol{\theta}} \rightarrow \underline{\boldsymbol{\theta}}\right) = t_{\overline{\boldsymbol{\theta}}}^0 + \delta. \end{aligned}$$

In words, when shifting from one type of train to the other, there is a increase in the delay between services. This increase is especially important when a fast services follows a slow one $\Delta \gg \delta$.

Clearly, in order to maximize the number of trains that may use the track in a given period, one shall attempt to minimize the extent of time "$T(H) = \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} t_{\boldsymbol{\theta}}(H)\, x_{\boldsymbol{\theta}}(p_\theta)\, dG(\boldsymbol{\theta})$". This minimum $T(H) = T^0$ is reached when all the services of the same type are gathered in which case

$$T(H) = X_{\underline{\boldsymbol{\theta}}}\left(p_{\underline{\boldsymbol{\theta}}}\right) t_{\underline{\boldsymbol{\theta}}}^0 + X_{\overline{\boldsymbol{\theta}}}\left(p_{\overline{\boldsymbol{\theta}}}\right) t_{\overline{\boldsymbol{\theta}}}^0 + \Delta + \delta.$$

Assuming such a grouping, the optimal pricing formulae do not differ from those considered just above, namely equation (16) if perfect discrimination is possible and equation (20) if no characteristics beyond those regarding congestion is observable. Note however that such a grouping is likely to generate important time costs for the network users since it requires a report or an anticipation with respect to their ideal time schedule.

If the costs of proceeding to a re-arrangement of the time schedule are considered to be excessive, both train types are likely to be widespread along the period T. Let the demand for access of the slow type trains $X_{\underline{\boldsymbol{\theta}}}(p_{\underline{\boldsymbol{\theta}}})$ be smaller than the demand of the fast type trains $X_{\overline{\boldsymbol{\theta}}}(p_{\overline{\boldsymbol{\theta}}})$. In this case, there are $X_{\underline{\boldsymbol{\theta}}}$ trains that insert themselves in groups type-$\overline{\boldsymbol{\theta}}$trains of size $X_{\overline{\boldsymbol{\theta}}}/X_{\underline{\boldsymbol{\theta}}}$.The necessary time for all the services to operate is now

$$T(H) = X_{\underline{\boldsymbol{\theta}}}(p_{\underline{\boldsymbol{\theta}}})\left(t^0_{\underline{\boldsymbol{\theta}}} + \Delta\right) + X_{\underline{\boldsymbol{\theta}}}(p_{\underline{\boldsymbol{\theta}}})\left(\frac{X_{\overline{\boldsymbol{\theta}}}(p_{\overline{\boldsymbol{\theta}}})}{X_{\underline{\boldsymbol{\theta}}}(p_{\underline{\boldsymbol{\theta}}})}t^0_{\overline{\boldsymbol{\theta}}} + \delta\right).$$

If prefect discrimination is possible, the optimal pricing formulae write now :

$$p_{\underline{\boldsymbol{\theta}}} = c_{\underline{\boldsymbol{\theta}}} + \frac{\lambda}{1+\lambda}\frac{p_{\underline{\boldsymbol{\theta}}}}{\epsilon_{\underline{\boldsymbol{\theta}}}} + \mu\left(t^0_{\underline{\boldsymbol{\theta}}} + \Delta + \delta\right) \tag{21}$$

$$p_{\overline{\boldsymbol{\theta}}} = c_{\overline{\boldsymbol{\theta}}} + \frac{\lambda}{1+\lambda}\frac{p_{\overline{\boldsymbol{\theta}}}}{\epsilon_{\overline{\boldsymbol{\theta}}}} + \mu t^0_{\overline{\boldsymbol{\theta}}} \tag{22}$$

If no discrimination is possible, the optimal pricing formula writes again as equation (17) above, namely

$$p = \widetilde{c} + \frac{\lambda}{1+\lambda}\frac{p}{\epsilon_X} + \mu\widetilde{t},$$

where, in the two type case $\widetilde{t}$ is given by

$$\widetilde{t} = \frac{1}{\epsilon_X}\left[\alpha_{\underline{\boldsymbol{\theta}}}\epsilon_{\underline{\boldsymbol{\theta}}}\left(t^0_{\underline{\boldsymbol{\theta}}} + \Delta + \delta\right) + \alpha_{\underline{\boldsymbol{\theta}}}\epsilon_{\overline{\boldsymbol{\theta}}}t^0_{\overline{\boldsymbol{\theta}}}\right]. \tag{23}$$

Finally if there is discrimination in the sole congestion dimension, the optimal markups write

$$\frac{p_{\underline{\boldsymbol{\theta}}} - \widetilde{c}}{p_{\underline{\boldsymbol{\theta}}}} = \frac{\lambda}{1+\lambda}\frac{1}{\epsilon_X} + \frac{\mu}{p_{\underline{\boldsymbol{\theta}}}}\left(t^0_{\underline{\boldsymbol{\theta}}} + \Delta + \delta\right) \tag{24}$$

$$\frac{p_{\overline{\boldsymbol{\theta}}} - \widetilde{c}}{p_{\overline{\boldsymbol{\theta}}}} = \frac{\lambda}{1+\lambda}\frac{1}{\epsilon_X} + \frac{\mu}{p_{\overline{\boldsymbol{\theta}}}}t^0_{\overline{\boldsymbol{\theta}}}. \tag{25}$$

Equation (21)-(25) deserve a few comments. Whatever the context which is considered (perfect price discrimination, no discrimination, partial discrimination), the slow type is attributed the "merits" of the delay resulting from the non-homogeneity of the traffic. The delay associated to the type-$\underline{\boldsymbol{\theta}}$ trains is indeed $t^0_{\underline{\boldsymbol{\theta}}} + \Delta + \delta$. This does not comes from the characteristic of the type however; but rather on the fact that I assumed that *(i)* the $X_{\underline{\boldsymbol{\theta}}}$ trains are uniformly distributed over the whole period T and *(ii)* the type-$\overline{\boldsymbol{\theta}}$ trains are in a greater amount. As a result, the demand in type-$\underline{\boldsymbol{\theta}}$ services $X_{\underline{\boldsymbol{\theta}}}$ determines the number of train type changes over the period. Would the $X_{\overline{\boldsymbol{\theta}}}$ trains be uniformly distributed over the whole period T and the slower services represent a greater share of the traffic, the time penalty $\Delta+\delta$ should be attributed to the type-$\overline{\boldsymbol{\theta}}$ services.

More generally, it appears that a sound congestion pricing would require each service to pay for its exact impact on the comprehensive time of use of all users. If it follows or precede a service of the same type, a train should be charged $\mu t^0_{\boldsymbol{\theta}}$ for congestion. If it induces a shift of type (and back), it should also be charged for this, hence it will be asked to pay. $\mu \left(t^0_{\boldsymbol{\theta}} + \Delta + \delta\right)$ for congestion. Of course, the cost of changing the time schedule are likely to exceed $\mu \left(\Delta + \delta\right)$ in the short-run. However, such a menu of fares may induce a train company to reconsider its time schedule in the long run, providing the right incentive for a more efficient use of the (existing) network.

4 Conclusion

In this paper, optimal fares are proposed for a sound access pricing of the rail network. The issue, of crucial importance for the deregulation of the European industry, is considered in different contexts. The administration in charge of the infrastructure is successively assumed to be for-profit and not-for-profit, benefitting from subsidies and submitted to a break-even constraint. The heterogeneity of the users is explicitly taken into account and both perfect discrimination and no price discrimination are considered. One of the most striking result is that, while the pricing formulae appear to be "standard", they should not refer to the average value of the marginal cost $\overline{c}$ but to a weighted sum $\widetilde{c}$,with weights that depend on the price-elasticity of the users. Given the heterogeneity in both the marginal costs and the price-elasticties reported by the empirical literature, the implications are significant. The second part of the paper is devoted to the pricing of congestion. Of course, the impact of congestion pricing depends on the price-sensitivity of the users. Again, by taking explicitly heterogeneity into account, it has been possible to show that a correct congestion pricing should not be based on average values. To conclude this second part on congestion, I focus on the two-type case as to be able to take into account the specificity of the rail sector, namely the fact that the capacity usage depends on the specific pattern of trains.

This papers bears nevertheless a number of limits. As underlined by Coulthard, Matthews and Nash (2003), socially optimal pricing would require to take into account other elements such as environmental costs and reports to other modes. The issue of infrastructure enhancement has also been put aside. Finally the demand for slots and the pattern of trains is essentially taken as given. These issues are left for future research. I believe however that by tackling explicitly the issue of access-pricing and heterogeneity the paper provides significant insights into a problem that has been often overlooked.

5 References

Arnott, R., and K.A. Small, [1994]. “The economics of traffic congestion”, *American Scientist* **82**, 446–455.

Coulthard, S., Matthews, B. and C. Nash [2003]. “Rail Track charges in Great Britain - the issue of charging capacity”, *mimeo*, Institute for Transport Studies, University of Leeds

European Commission [2001], *European Tranport Policy for 2010: Time to decide,* White Paper.

Gaudry, M. and E. Quinet, [2003]. “Rail Track Wear-and-Tear Costs by Traffic Class in France”, *Working Paper* n° 04-02 CERAS, Ecole Nationale des Ponts et Chaussées, Paris

Laffont J.-J. and Tirole J. [1993], *A Theory of Incentives in Procurement and Regulation.* Cambridge, MA: MIT Press.

Australian Productivity Commission [1999], *Progress in Rail Reform,* Inquiry Report.

Vickrey,W.S., [1963]. Pricing in urban and suburban transport. *American Economic Review Papers and Proceedings* **53**, 452–465.