Optimal Pricing in Railway Passenger Transport. Theory and Practice in The Netherlands

van Vuuren, Daniël

Published in:
Transport Policy

DOI:
10.1016/S0967-070X

Publication date:
2002

Document Version
Publisher's PDF, also known as Version of record

PERGAMON

Transport Policy 9 (2002) 95–106

TRANSPORT POLICY

www.elsevier.com/locate/tranpol

Optimal pricing in railway passenger transport: theory and practice in The Netherlands

Daniel van Vuuren*

Department of Spatial Economics, Vrije Universiteit Amsterdam and Tinbergen Institute, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Abstract

This study tries to establish the link between the well-developed economic theory of optimal pricing, and recent empirical results concerning price elasticities of demand and marginal cost estimates for The Netherlands Railways. The *ex post* determination of Ramsey coefficients confirms that peak hour pricing is dominated by the aim of welfare maximization, while off-peak fares largely correspond to the profit-maximizing objective. It is argued that tariff differentiation according to track offers scope for a budget-neutral welfare improvement, and that a price-elastic off-peak market implies that lower off-peak fares result in a Pareto-improvement. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Optimal tariffs; Railway cost function; Public transport

1. Introduction

The optimal pricing of passenger transport by railway companies is a topic that has not received much attention in the empirical literature. Textbooks in microeconomics and industrial organization offer a rich theory, which could be applied in quite a direct way if all the required data were available. Lately, some empirical analyses have generated interesting results on the cost structure of The Netherlands Railways. In addition, a number of recent studies have investigated the sensitivity of travelers by train with respect to price changes. In this article, I try to establish the link between these empirical studies, and the well-developed economic theory of optimal pricing, particularly applied to the market for passenger transport by rail. Section 2 will first discuss a theoretical model, stressing the importance of high sunk and fixed costs in this market, after which empirical evidence for *increasing returns to scale* in the cost function of The Netherlands Railways will be discussed. The problem of X-inefficiency is also shortly addressed. Section 3 will review the relevant literature on optimal pricing. Two different kinds of *optimality* are considered: first, the *profit-maximizing* price schedule is characterized, and secondly, the theory of *second-best pricing* will be discussed. Second-best prices can be characterized as a compromise between the socially optimal price, which equals marginal cost, and the profit-maximizing price. The empirical results are discussed in Section 4. The tariff structure, recent marginal cost estimates and recent estimates of price elasticities will be discussed, and compared with the theoretical formulae of earlier sections. Two clear policy recommendations could be formulated as a result of this analysis.

2. The market structure of railway passenger transport

2.1. The general structure of railroad costs

The cost schedule of the passenger railway industry is crucial in order to determine the structure of the rail market. It will determine whether there are barriers to free entry and exit, both in the practical and the juridical sense, and which market form will produce services in the most efficient way. Principally, three types of cost can be distinguished: (i) *sunk costs*, (ii) *fixed costs*, and (iii) *variable costs*. Contrary to fixed costs, sunk costs cannot be eliminated, even at a zero production level. Examples in the railway industry are tracks and bridges.[1] On the other hand, fixed costs are the kind of costs that are necessary for production, but do not vary with the output level. An obvious example in the railroad industry is (the purchase of) locomotives. It is important to distinguish between sunk costs and fixed costs because it is the former that are the cause of entry

* Fax: +31-20-4446004.
E-mail address: dvuuren@tinbergen.nl (D. van Vuuren).

[1] It should be noted that costs for tracks and bridges are not totally sunk, because they can often be used for other purposes than transport by rail—however at considerable cost.

PII: S0967-070X(02)00005-7

barriers. In practice, it is not always clear whether costs should be categorized as sunk costs or fixed costs; the time horizon of the analysis is obviously crucial for this. As an illustration, consider the following cost schedule:

$$C(y) = c_0 + c_1\left\lceil \frac{y}{\alpha} \right\rceil + c_2 y,$$

where the three terms on the right-hand side correspond with sunk costs, fixed costs and variable costs, respectively, at given production level y, and for a certain, given, time horizon. One can think of y as the total amount of passenger train kilometers per year[2]. The function $\lceil x \rceil$ rounds x to the nearest integer, which is greater than or equal to x. If, for example, α is the capacity of one train-unit and c_1 is the cost associated with such a train-unit, then the fixed costs term corresponds to the costs that are associated with the use of the $\lceil y/\alpha \rceil$ train-units that are required for output y. The sunk costs for entering the market equal c_0, while the variable costs associated with the production of one unit of output are equal to c_2 in this example. Compare now the case that the industry is served by two firms[3] with respective outputs y_A and y_B, with the case of one single, monopolistic supplier with output $y \equiv y_A + y_B$. With two suppliers, the total cost of production at total output y equals

$$C(y_A) + C(y_B) = 2c_0 + c_1\left(\left\lceil \frac{y_A}{\alpha} \right\rceil + \left\lceil \frac{y_B}{\alpha} \right\rceil\right) + c_2(y_A + y_B),$$

while total costs in the case of one single supplier equal

$$C(y_A + y_B) = c_0 + c_1\left\lceil \frac{y_A}{\alpha} + \frac{y_B}{\alpha} \right\rceil + c_2(y_A + y_B).$$

As $\lceil y_A/\alpha + y_B/\alpha \rceil \leq \lceil y_A/\alpha \rceil + \lceil y_B/\alpha \rceil$, it turns out that two distinct firms can never produce more efficiently than the monopolist:

$$C(y_A + y_B) \leq C(y_A) + C(y_B). \tag{1}$$

At best, firms A and B can produce as efficiently as the monopolist if: (i) $c_0 \equiv 0$, i.e. there exist no sunk costs, and (ii) $\lceil y_A/\alpha + y_B/\alpha \rceil = \lceil y_A/\alpha \rceil + \lceil y_B/\alpha \rceil$, i.e. total fixed costs do not differ between the two different market structures that are considered in this example. The latter occurs for example if all train capacity is used, so no seats are left empty. For the general case, it is, however, likely that the two distinct firms will need more train-units than the single supplier will. The inequality in Eq. (1) between monopoly and duopoly production costs becomes bigger whenever (i) there are large sunk costs involved ($\alpha \to \infty$), or (ii) fixed costs are less flexible ($c_0 \to \infty$). The inequality in Eq. (1), termed sub-additivity, is the common definition for a single-product natural monopoly, when applied to all numbers of possible entrants of the industry:

$$C\left(\sum_{j=1}^{k} y^j\right) \leq \sum_{j=1}^{k} C(y^j), \tag{2}$$

for all $k = 2, 3, \ldots$. Despite its conceptual straightness, the practical use of the property of sub-additivity is rather limited. The reason for this is that it is a global property, which, in order to be tested empirically, needs a great amount of information which can hardly be available.[4] This is the main reason why the bulk of the empirical literature focuses on indicators for the property of sub-additivity, instead of focusing on the property itself. The most obvious property that is connected to the notion of sub-additivity is that of increasing returns to scale (IRS). This will occur if the elasticity of cost with respect to the scale of the output is smaller than one. Indeed, for a single product industry, satisfaction of this property is a sufficient condition for a sub-additive cost function. For a multiproduct industry, however, the property of IRS is neither a necessary nor a sufficient condition for a sub-additive cost function. Note that the multiproduct definition of sub-additivity is simply Eq. (2) with $y = (y_1 y_2 \ldots y_n)$. The IRS property can nevertheless be an important indicator of a sub-additive cost structure, in particular in combination with the property of *economies of scope*, which is to be discussed later. Keeler (1997) makes notice of three different kinds of economies of scale in the railroad industry, namely: economies of density, economies of length of haul, and economies of firm size. *Economies of density* occur due to fixed capital costs, such as maintenance of the infrastructure, but also due to the fact that higher traffic densities will allow for a more efficient use of the main production factors labor and equipment. In general, higher traffic densities will lead to a higher frequency of operation and also to longer trains, which implies lower material and labor costs per passenger-kilometer that is *produced*. Fixed expenses on railway stations may cause substantial *economies of length of haul*. Longer hauls then imply lower costs per kilometer, which makes it more attractive for the railway company to operate on an integrated nationwide system. *Economies of firm size*, i.e. lower costs for larger firms, will have this same effect. True, the just mentioned example of a cost function is simple and highly stylized, but it nevertheless underlines the importance of the large sunk and fixed costs in the railroad industry, which are the cause of the different kinds of economies of scale that have just been mentioned.

However, the existence of large sunk and fixed costs cannot be the only motivation for the conjecture of a sub-additive cost function in the railroad industry. The example

[2] Also, see the discussion later in this section on *revenue output* and *available output*.

[3] The case of three or more suppliers is essentially the same as the example with two firms that is being discussed here.

[4] Testing the opposite is less problematic: In order to show that an industry does *not* have a sub-additive cost function, one can make use of the fact that local rejection of this property implies global rejection, so that only local data are needed. Evans and Heckman (1984) have devised a statistical test for this.

that has just been considered concerns a single output industry, whereas the railroad sector is a typical *multiproduct sector* (Kessides and Willig, 1998). Transportation services are—amongst others—different at levels of quality (First Class, Second Class), times (night-train, morning peak hour train, etc.), origins and destinations, types of users (students, elderly, 'standard', etc.) and types of transportation services (intercity train, stopping train, etc.). In a multiproduct industry, the presence of economies of scope is a necessary condition for the property of sub-additivity; they will clearly increase the difference between total costs under monopolistic production and total costs in an industry with multiple firms. The appearance of economies of scope in the railroad industry is beyond doubt: most services are very similar from the production viewpoint, and share the same sunk and fixed costs, in particular those concerning infrastructure, locomotives, and even cars.

2.2. Cost (in-)efficiency

An important issue that has not been addressed yet in this article is that of *managerial efficiency*. In the above-mentioned example, it has been assumed that the cost schedule remains constant, regardless of the market structure. However, as has been argued in many theoretical articles, and shown in several empirical studies, the management of a monopolistic firm is usually less inclined to minimize the firms' costs, as is the management of a firm that faces competition from other firms in the industry. The management of a firm naturally has other interests (firm expansion, personal benefits) than the firm's shareholders (profit maximization). In a competitive market, shareholders are able to compare the management's performance with the performance of other companies in the sector, and if considered necessary, fire the management and hire a new one. This *yardstick competition*, which obviously does not exist in a monopolistic market, will be a clear incentive for the management to minimize the firm's costs. The efficiency of railway companies in 19 OECD countries during the 1978–1989 period has been the focus of study in Oum and Yu (1994). Because the considered railway companies in this article are subject to different levels of regulation, the authors have decided to use both the *available output* and the *revenue output* measures in order to determine managerial efficiency. Oum and Yu (1994) distinguish between two different concepts of output measures. Available output measures correspond to the level of capacity that is supplied, while revenue output measures indicate the level of output that is consumed by the firm's customers. For railway companies that are under tight regulatory control, revenue output will indicate the combined effect of managerial efficiency and regulatory efficiency, while available output will measure just managerial efficiency. For example, a railway company that produces many empty seats in order to match the regulators' requirements can still attain a high level of managerial efficiency. On the other hand, if government control is absent, it makes more sense to use revenue output as a measure for managerial efficiency, because the available output is under managerial control. The methodology that has been used, called *Data Envelopment Analysis*, derives a nonparametric production frontier by enveloping the observed input–output data. It turns out that The Netherlands Railways, together with British Rail and the Swedish and Finnish national railways, are the most efficient among those that have been considered in the survey.[5] Two general conclusions have been derived from the findings in this article: First, it appears that "railway systems with high dependence on public subsidies are significantly less efficient than similar railways with less dependence on subsidies", and secondly, "railways with a high degree of managerial autonomy from regulatory authority tend to achieve higher efficiency".[6] In order to reach to these conclusions, Oum and Yu (1994) have managed to adjust their efficiency measures for factors that are beyond managerial control, such as the geographical, demographic and institutional environments.

2.3. Returns to scale: empirical evidence for The Netherlands Railways

In the early 1990s, the old-style Netherlands Railways with integrated operations and infrastructure divisions was vertically split up into (i) *task organizations* for management and maintenance of the infrastructure[7] and (ii) the NS Group, which operates on the railway infrastructure. The NS Group has been horizontally split up into four different *business units* for passenger services, freight services, railway stations, and real estate (respectively called NS Reizigers, NS Cargo,[8] NS Stations, and NS Vastgoed). The task organizations are nonprofit oriented, and funded by the government. The four business units are allowed to act independently, as long as certain regulatory constraints are met. At present, no plans exist to either privatize NS Reizigers—in this article simply abbreviated as NS—or introduce any form of competition. In this section, I shall discuss recent empirical studies concerning the cost structure of NS Reizigers, mostly focusing on the possible existence of increasing returns to scale for this business unit. The focus on the business unit for passenger services implies that costs for infrastructure are not taken into account for this moment. In Section 2.4, I will come back to this.

[5] International comparative studies about the efficiency of railway operators remarkably often show that The Netherlands Railways are among the most efficient companies in Europe. E.g. see also Pestieau and Tulkens (1990).

[6] This second result is also found in Oum et al. (1999), who survey empirical findings on the productivity and efficiency of railway companies. Also see Gathon and Pestieau (1995), who provide statistical evidence for this conclusion. Also, see Dennis (2000).

[7] The three task organizations are NS Railinfrabeheer (maintenance and construction), Railned (capacity planning), and NS Verkeersleiding (traffic control).

[8] NS Cargo has now merged with its German counterpart, DB Cargo.

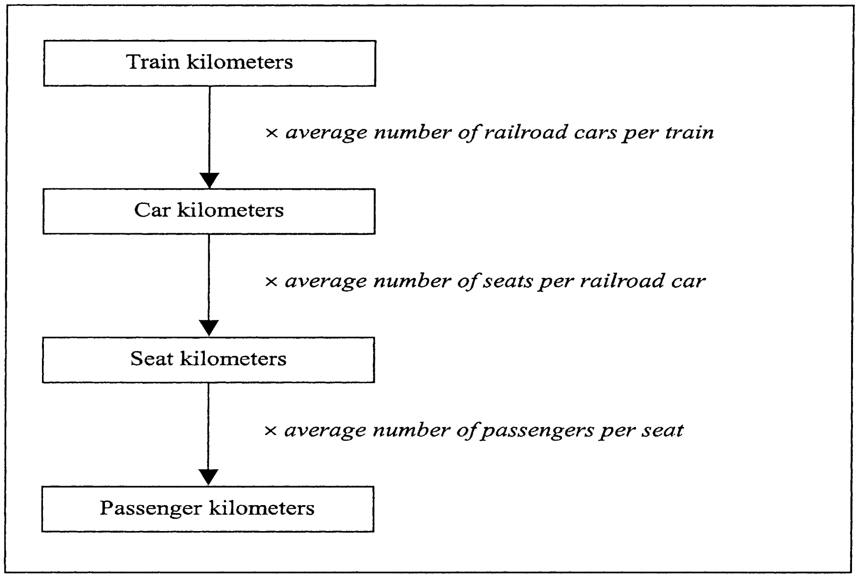

Fig. 1. Various output measures for passenger transport by rail.

As has been mentioned in Section 2.1, economies of scale are generally not a necessary nor a sufficient condition for natural monopoly to exist. However, the question whether scale economies are present is still an interesting one because it tells much about the cost structure of the firm. Moreover, the existence of economies of scope in the railway passenger transport industry is so apparent that the property of IRS will be an important indication for the sub-additivity property of the cost function.

In general, when empirically estimating a cost function, the two central variables are *cost* and *output*(*s*). However, the mere definition of these two variables forms a significant problem in the context of passenger transport by rail.[9] First, the attribution of fixed costs to the different outputs of the railway company will always be arbitrary to a certain extent. It is difficult, if not practically impossible, to compute the average cost of a passenger kilometer by train, because production factors are often common for different types of output such as peak and off-peak train kilometers.[10] Hence, the computation of average cost would require that such costs are attributed to the different categories according to a certain key, while according to Kessides and Willig (1998) "[t]here is simply no way to subdivide those costs in a mechanical fashion that is unique and has any foundation in economic logic". One also has to take into account that the available cost data have usually been reported by the railway company itself, so that the figures might be biased towards a direction that has been favorable for the company's own interests (or its management's interests)—also see Section 2.2. Secondly, there is the question which output measure(s) is the most appropriate for the aim of analysis. Fig. 1 shows four possible output measures. The first measure, train kilometers, is simply obtained by adding up all the distances that have been covered by all trains during a given time period. The three subsequent output measures can be obtained through respective multiplication with the factors in italics. It is debatable whether the last measure, passenger kilometers, can serve as a good indicator for the production side of the firm because it will be directly affected by the demand side. In terms of the distinction between output measures of Oum and Yu (1994), it is a revenue output measure. The other three output variables that are shown in Fig. 1—train, car, and seat kilometers—are much less sensitive to this point of critique. However, the choice for either one of these available output measures does matter, as will be discussed later.

Based on recent studies by Shires and Preston (1999), MuConsult (1999) and van Ooststroom (1999), Table 1 summarizes recent estimates of output elasticities for the mentioned outputs of The Netherlands Railways. MuConsult and Shires and Preston have estimated a *cost function*, based on data from roughly the same period (1971/1976–1993/1994). van Ooststroom estimates a *production function* for a longer period, starting in 1951. If *train kilometers* are taken as the output measure, the two first studies find constant returns to scale (CRS), while van Ooststroom finds increasing returns to scale (IRS). Two explanations can be given for this contradiction. Firstly, van Ooststroom uses a longer period, which contains a number of structural changes that are not (or hardly) contained in the periods used by the other two studies. To name two of such structural changes: (i) the 1950s and 1960s have shown a steady decrease in total employment

[9] For brevity, the discussion of other important issues with regard to the estimation of (transport) cost functions is omitted here. See Braeutigam (1999) for a recent overview.

[10] In fact, many more categories, such as international trips, student trips, etc. can be distinguished. This would only worsen the mentioned problem, as it will be even more difficult to allocate the proper shares to each different category.

Table 1
Recent empirical results about The Netherlands Railways cost function

Output	Shires and Preston (1971–1994)	MuConsult (1976–1993)[a]	van Ooststroom (1951–1993)
Train kilometers	CRS[b]	CRS[b]	IRS
Car kilometers	–	–	IRS
Seat kilometers	–	IRS	–
Passenger kilometers	–	IRS	–

[a] This study also discusses cost data of the period 1994–1998, but these have not been used in the statistical analysis, due to lack of consistency with the earlier data.
[b] Returns to scale are found to be slightly decreasing in these studies, though not as convincing as to reject the hypothesis of CRS.

at NS. This development has been halted in 1967, when NS and the Dutch government made an agreement about the compensation of losses of NS; (ii) in 1972, the total length of railroad track has been reduced by 10%. Before and after this year only relatively small changes have been made to the total size of the railway network. Thus, an important distinction between van Ooststroom's studies with the others is that it includes a longer time period with some significant structural changes, while MuConsult and Shires and Preston base their studies on more recent data sets with relatively small variations over time. A second explanation for the divergent results is that the production function approach taken by van Ooststroom has had to make several assumptions, of which one is particularly controversial for the current case, i.e. the supposition of *cost minimization*. As has been argued in Section 2.2, it is not so clear a priori that a publicly owned monopoly is inclined to minimize its production costs.

Van Ooststroom also takes into consideration *total car kilometers* as an output, thereby endogenizing the average number of railroad cars per train, cf. Fig. 1, and also finds scale economies for this case. However, the most suitable output measure for the aim of this article is *total seat kilometers*. Unlike train kilometers and car kilometers it is *customer-oriented*, because one seat kilometer precisely corresponds to the service that is offered to a potential customer. It is also better suitable than *passenger kilometers* because it is not directly affected by demand, as has just been mentioned. Thus, when it comes to the determination of optimal prices for customers, one should preferably take into consideration *seat kilometers* as the output measure. MuConsult estimate a cost function, and find convincing scale economies for this output. The main reasons for this are the possible use of longer trains and also making use of double deckers on busy tracks (cf. Fig. 1). Over the considered period (1976–1993), one can observe relatively little variation in total costs, while the total 'production' of seat kilometers shows a considerable increase, suggesting that the marginal price of additional seat kilometers is relatively low (compared to the marginal price of seat kilometers that were already produced).

2.4. Synthesis

As a result of law enforcement, the railway industry concerning passenger transport in The Netherlands is an industry that is not open to free entry and exit of firms. However, from an economical point of view, one may wonder whether the industry *should* be open to free entry and exit. The proper criterion for this is to examine whether the market is efficiently served if the industry consists of one single firm. In other words, the question is whether the industry is characterized by a sub-additive cost function. In Section 2.3, empirical results with respect to the operating costs of The Netherlands Railways have been discussed. The most relevant results show that

(A) train kilometers are produced under CRS;
(B) seat kilometers are produced under IRS.

However, it should be taken into account that the NS are essentially a multiproduct sector. Both train kilometers and seat kilometers are differentiated according to time and place, so that the property of IRS alone will not be sufficient for sub-additivity. Economies of scope are a necessary condition for this last property, but it has been argued in Section 2.1 that the presence of such economies is beyond doubt for the current case. These are strong indicators for a sub-additive cost function for the production of seat kilometers, so that NS can be characterized as a natural monopoly. Finally, note that the inclusion of infrastructure costs will most likely lead to higher IRS, and thus a stronger natural monopoly. This is very relevant for the Dutch case, because currently a debate is taking place in The Netherlands whether (and how) NS should be charged for making use of the railway infrastructure. It is not unthinkable that this will even lead to higher scale economies with respect to the production of seat kilometers than before due to an increased amount of fixed costs in the cost schedule of NS.

3. The economic theory of optimal pricing

3.1. Optimal pricing from the profit maximizing perspective

In this section, I shall concisely review the possibilities of a monopolist to maximize its profits through price discrimination. This section is largely based on the standard works of Varian (1989) and Tirole (1988).

According to Stigler (1987), price discrimination arises "when two or more similar goods are sold at prices that are in different ratios to marginal costs". There are several reasons for preferring this broad definition over the more 'classical' definition that price discrimination is present when the same commodity is sold at different prices to different customers. The main reason is that charging the same price to different customers may well be regarded as a form of price discrimination. For instance, people who travel by train during off-peak hours and pay the full tariff, are charged the same price as people who travel during peak hours, whereas it is likely that marginal cost is lower during the off-peak than during peak hours—which will be made explicit in Section 4.1—implying that the railway operator receives a higher relative profit from the off-peak traveler. This form of *third degree price discrimination* will be discussed later, but I will first briefly discuss the profit maximizing *nondiscriminatory* monopoly price, and price discrimination of the *second degree*. The theoretical concept of price discrimination of the *first degree*, also referred to as *perfect price discrimination*, occurs when each customer is charged exactly his willingness to pay (WTP), so that the entire consumer surplus is soaked up into the firm's profits. The information that is required for this—the WTP distribution of all potential customers—will not be available in practice, so that the second and third degree price discrimination rules remain more realistic alternatives.

3.1.1. Nondiscriminatory pricing

A monopolist which operates on a single market, and is not able to discriminate among its customers, maximizes its profits if mark-up[11] equals the inverse of the absolute price elasticity of demand:

$$\frac{p-c}{p} = \frac{1}{\epsilon}, \tag{3}$$

where p equals the (uniform) price of the product, c denotes marginal cost and $\epsilon \equiv \epsilon(p)$ is the absolute elasticity of demand (at price level p). Alternatively, this formula can be written as

$$p = c\left(1 - \frac{1}{\epsilon}\right)^{-1}. \tag{4}$$

This familiar result, found in any microeconomic textbook, is in fact nothing but a restatement of the first order condition for profit maximization, setting equal the monopolist's marginal revenue $p(1-(1/\epsilon))$ to marginal cost at the optimal price p. Comparing this monopoly price with the efficient *marginal cost price* that results from a competitive market ($p \equiv c$), it is seen that the total lack of competition inflates[12] the price by a factor $(1-(1/\epsilon))^{-1}$. This inflation factor becomes larger as the price elasticity gets smaller in magnitude. In other words, the monopolist will be able to achieve a higher mark-up, and hence profits, if its customers are not too sensitive to price changes. On the other hand, the higher the price elasticity, the closer the monopoly price will approach the competitive price. For instance, this may be the case if a railroad monopolist faces tough competition from other modes of transport, such as car, bicycle and bus. As the firm's potential customers have enough good alternatives in that case, they will be sensitive to price changes, and substitute away from the demand for transport by train if its price becomes too high. This will induce the monopolist not to charge too high a price, in order to maximize its profits. Thus it is seen that the welfare loss that is caused by the monopoly price in Eq. (4), usually referred to as the *deadweight loss*, tends to be larger if the monopolist's customers are less price sensitive. On the other hand, distortions from the social optimum can become negligibly small if customers are very sensitive to price changes ($\epsilon \rightarrow \infty$). Finally, note that the marginal revenue of the monopolist will be strictly positive whenever $\epsilon > 1$, and strictly negative if $\epsilon < 1$. Hence, fulfillment of the first order condition for profit maximization implies that the monopolist's optimal output is at a point on the demand curve where price elasticity exceeds unity: if the monopolist chose an output for which the price elasticity is smaller than one, then it would charge too low a price (too high an output), and it could simply increase profits either by charging a higher price or restricting its output.

3.1.2. Second degree price discrimination

Price discrimination of the second degree occurs when prices differ depending on the amount of the good that is consumed, but not across consumers. Therefore, this form of price discrimination is often referred to as *nonlinear pricing* or nonlinear price discrimination. An example in the context of railway transport is the seasonal ticket.

The design of profit-maximizing nonlinear tariffs is thoroughly discussed by Wilson (1992). It is hard to derive a general formula for the optimal nonlinear tariff and usually different sets of assumptions are postulated in order to derive an analytic expression for the optimal nonlinear price schedule. The most general result is that the customer with the highest preference for the produced service (the *marginal consumer*) is served efficiently, i.e. he faces a marginal price which is equal to marginal cost. It can be shown that in many cases, it turns out to be optimal to offer quantity discounts, implying that, apart from the marginal consumer, all consumers face a price that is above marginal cost.

3.1.3. Third degree price discrimination

A monopolist which serves different markets will in

[11] Sometimes this is also called the *Lerner index*.

[12] It is implicitly assumed that the good that is being offered by the monopolist is a *normal* good, i.e. its own price elasticity is strictly negative.

general be able to raise its profits by charging different prices in each market. Different markets can be thought of as markets for different products but also markets for the same product, but with different, *separable*, groups of customers. For the latter case, one can think of special fares for students and the elderly for essentially the same product, namely transportation by train under exactly the same conditions as for the standard customer. In this case, substitution effects between markets will not exist, and the optimal pricing rule in each market i will be given by

$$\frac{p_i - c_i}{p_i} = \frac{1}{\epsilon_i}, \tag{5}$$

which is nothing more than a restatement of Eq. (3) for each market separately. This is the familiar *Ramsey principle*, which states that a monopolist which serves separate markets should equate mark-up to the inverse elasticity of demand on each market *separately*, in order to maximize its profits. Stated differently, relative profits are higher in the market with low demand elasticity, and lower in the market with a relatively high demand elasticity. In some cases, marginal cost will be constant across different sub-markets.[13] Price discrimination is then justified by different price elasticities in each sub-market, resulting in optimal prices that are obtained by equating the marginal revenues of each sub-market:

$$p_i\left(1 + \frac{1}{\epsilon_i}\right) = p_j\left(1 + \frac{1}{\epsilon_j}\right) \qquad \text{for all } i,j.$$

This pricing rule is relevant when e.g. price discrimination according to age (student and senior fares) is applied.

In other cases, however, like e.g. the market for transportation during peak hours and the market for transportation during off-peak hours, substitution effects between markets may well exist, and optimal prices will be given by

$$\frac{p_1 - c_1}{p_1} = \frac{1}{\epsilon_1} - \frac{\partial p_2}{\partial q_1}\frac{q_2}{p_1} \tag{6}$$

By symmetry, the optimal price for market 2 follows by interchanging 1 and 2 in the subscripts in Eq. (6). It is seen that the higher the substitution effects between markets is, the greater will be the deviation from the Ramsey principle in Eq. (5).

3.1.4. Simultaneous second and third degree price discrimination

As mentioned, it is difficult to derive general results for the case of second degree price discrimination, and likewise, the present case does not allow for general characteristic formulae. It should be realized that even when there are zero cross-price elasticities between the goods on the different sub-markets, then the optimal nonlinear tariffs could still not be determined for each market separately (Spence, 1980). Research on *optimal multiproduct nonlinear pricing* is still going on, with recent contributions by Armstrong (1996) and Sibley and Srinagesh (1997). Also, see chapters 13 and 14 in Wilson (1992).

Finally, it should be noted that the design of an optimal tariff is one thing but the practical application of it is another. Intricate tariff structures can lead to high administration costs, and are often not accepted by regulators. The design, maintenance, and marketing lead to high *menu costs* for the supplier. Empirical evidence from the United States shows that even when no regulatory restrictions exist, the cost argument often prohibits application of a theoretically better, more sophisticated tariff structure (Wilson, 1992). In addition, the individual customer's choice of the appropriate tariff may involve considerable *search costs*. The development of ICT is promising in this respect as it offers scope for bringing down both menu and search costs, e.g. by replacing different ticket types by one electronic chipcard which allows for a flexible variety of tariffs. This is scheduled to take place in The Netherlands in the near future.

3.2. Optimal pricing from the welfare maximizing perspective

3.2.1. Why regulate natural monopolies?

In contrast with the previous section, which discussed optimal prices in the light of profit maximization of the monopolist, this section will postulate *welfare maximization* as the desired optimum. However, the maximization of total welfare, defined here as the sum of producer and consumer surplus, is an objective that contrasts with the profitability of the firm. As has been discussed in Section 2.1, the cost structure of a natural monopoly is characterized by high sunk and/or common costs. The welfare maximizing price schedule, with prices set equal to marginal cost, will therefore lead to a loss that is equal to such costs. This *cost recovery problem* is the main problem of natural monopolies. Losses could in theory be recovered by lump-sum taxation of the whole population, but this might be regarded as being 'unfair' to the ones who are not making use of the monopolist's services. Alternatively, only customers who make use of the service that is provided by the monopolist may pay a lump-sum amount in order to make use of the service ('two-part tariffs'). However, this will cause a certain number of customers not to make use of the service, in particular those ones with a relatively low willingness-to-pay.

Braeutigam (1989) gives an excellent account on how to deal with this problem. The first crucial piece of information which determines whether to impose regulatory constraints or not concerns the deadweight loss. This welfare loss equals the difference in welfare between the *second-best* case, i.e. the constrained welfare maximizing allocation, and the socially optimal case of *marginal cost pricing*. This constraint concerns the profitability of the firm: either breaking even, not exceeding a certain amount of losses, or

[13] This case is extensively discussed by Trotter (1985), with special reference to British Rail.

exceeding a certain amount of profits. If this deadweight loss is high, then the government should intervene and impose regulatory constraints in order to achieve a more efficient allocation. However, natural monopoly itself is not always a sufficient condition for regulation. As will be discussed later in this section, competition *for* the market may in some circumstances lead to a quite efficient allocation of resources, and therefore in some cases it may be optimal not to intervene at all in the market. If the deadweight loss between first-best and second-best allocation is not too high, then this latter possibility may be a serious alternative to second-best regulation. In practice, this will not always be feasible however, so that regulation in order to achieve a second-best allocation remains a serious option.

The second major problem concerning natural monopolies is the problem of *information*. Monopolistic firms are often able to make use of informational advantages, in their own interest or in the interest of their managers, but to the detriment of public interest. This has also been discussed in Section 2.1. Especially when it comes to designing optimal policies by the regulator, being properly informed is of crucial interest.

3.2.2. Second-best pricing

The general aim of the literature on second-best price schedules is to find a way in which distortions from the first-best price schedule are minimized. As has been noticed in Section 3.2.1, the first best allocation that results from marginal cost pricing will result in losses for the monopolistic firm. One of the ways to deal with this problem is the introduction of *Ramsey prices*. The basic principle of Ramsey pricing is that the mark-up is set equal to a fraction between 0 and 1 of the inverse elasticity:

$$\frac{p-c}{p} = \frac{\alpha}{\epsilon}, \tag{7}$$

The number α can then be chosen so as to exactly let the monopolist meet the imposed profitability constraint. Note that for $\alpha = 1$, one obtains monopolistic pricing without price discrimination as in Eq. (3), and that $\alpha = 0$ corresponds to marginal cost pricing. Therefore any value of α between these extremes corresponds to a compromise between welfare maximization ($\alpha \downarrow 0$) and profit maximization of the monopolist under uniform pricing ($\alpha \uparrow 1$). The present formulation does not allow for nonlinear pricing but does so for third degree price discrimination. For that case, the above rule can be applied for each separate sub-market, with varying elasticity values for each sub-market. (This, of course is the mere justification for third degree price discrimination, see Section 3.2.1.)

Matters become even more interesting if nonlinear pricing (or second degree price discrimination) is allowed. Citing Roberts (1979), "[i]n essence, nonlinear [price] schedules offer maximum scope for the minimization of distortions". It can be shown that it is always possible to raise total welfare by moving away from nondiscriminatory monopolistic pricing towards a nonlinear price schedule. Willig (1978) has even shown that the most simple nonlinear price schedule, the *two-part tariff*, Pareto dominates a flat rate monopoly price, implying that not only total welfare increases, but also that all customers and the firm are at least as well off as under the flat rate schedule. However, the nonlinear tariff structure that maximizes profits is in general not maximizing total welfare (Spence, 1977; Roberts, 1979), and it may even lead to a decrease in total welfare (Katz, 1983). Both Roberts, (1979) and Goldman et al. (1984) consider welfare maximization subject to a break-even constraint for the firm. Goldman et al. (1984) prove a very interesting result, stating that the optimal nonlinear tariff is directly related to the optimal tariff under third degree price discrimination, i.e. the Ramsey pricing formula in Eq. (7):

$$\frac{p(q)-c}{p(q)} = \frac{\alpha}{\epsilon(p(q),q)}. \tag{8}$$

The intuition behind this result is that each quantity q can be regarded as an infinitesimally small 'market', and hence for each given quantity there exists an optimal price which conforms to the Ramsey pricing rule. This formula also holds for the case of simultaneous second and third degree price discrimination, in which case a subscript has to be added to the price elasticity, indicating each separate sub-market.

3.2.3. Regulation

The practical implementation of Ramsey prices can be achieved through a number of ways. In this paragraph, I will mention the most important, although a detailed discussion is beyond the scope of this article. One of the most prominent forms of regulation in order to induce Ramsey prices is the imposition of a *price cap*. A price cap may be imposed on either the average price or the maximum price for one single good, or possibly both at the same time. Another possibility is to impose a price cap on a price index concerning the various outputs of the monopolist. In that case, an (average) price increase for one good will have to be offset by a decline in price for the other good that is offered by the monopolist. A concise overview can be found in the appendix of Train (1997). If the price cap is set at the optimal level, then this will induce the firm to set Ramsey prices, and thus attain a second-best allocation. Many recent theoretical articles deal with the regulation of monopolistic industries through price caps. For instance, Armstrong et al. (1995) consider a specific form of price cap regulation, called *average revenue regulation*. In general, if price cap regulation is applied in a correct fashion, it will: (i) induce firms to be cost efficient, both in the short and in the long term, and (ii) lead to an overall welfare increase. A possible drawback is that it may lead to *underinvestment* (Helm and Thompson, 1991). Other forms of regulation are rate of return regulation (Train, 1997; Srinagesh, 1986) and output floor regulation (Katz, 1983; Armstrong and Vickers, 1991). For natural monopolies, output floor regulation is equivalent to price cap regulation. A different way to arrive at Ramsey

Table 2
Estimates of marginal cost per train kilometer (e-cents), according to MuConsult

Marginal cost definition	Peak	Off-peak
No additional rolling-stock or seat	–	0.3
One additional seat	3.4	0.7
Additional rolling-stock	9.3	3.4

prices is the introduction of a form of competition into the market. An up-to-date overview of different competition models for the market for railway passenger transport is given by van de Velde (1999).

4. Optimal pricing in The Netherlands: an empirical analysis

4.1. Marginal costs

As has been discussed in the two previous sections, marginal costs play a crucial role in the determination of the optimal tariff structure of a natural monopolist. However, it is often difficult to find the adequate data for reasons mentioned in Section 2, i.e. the definitional problem of cost (Section 2.3), and the agency problem (Section 2.2), which may result in a bias in the reported cost structure. Nevertheless, the earlier mentioned recent study by MuConsult (1999) does provide interesting information with respect to this matter, as it tries to compute marginal costs from different definitional viewpoints. The following practical definitions have been assigned to the term 'marginal' in this study:

1. The cost of transportation of one extra traveler over 1 km, without requiring additional rolling-stock or seat.
2. The cost of transportation of one extra traveler over 1 km, requiring an additional seat but no additional rolling-stock.
3. The cost of transportation of one extra traveler over 1 km, requiring additional rolling-stock.

Estimates for these three definitions of marginal cost are found in Table 2. During peak hours, transportation of an extra passenger without making use of additional rolling-stock or seat is in general not possible, because during that period supply is close to the maximum capacity. Therefore, peak hour marginal cost estimates are only reported in terms of an additional seat or additional rolling-stock. Estimation of the aggregated cost function has yielded estimates of 3.4 and 9.3 e-cents, respectively.[14] At a less aggregate level, one is able to observe congestion problems in western and central parts of The Netherlands, while tracks outside these regions face these problems to a lesser extent. Despite being apparent on at least some tracks, and possibly even on the network as a whole (Shires and Preston, 1999), congestion problems are not explicitly taken into account in the report of MuConsult (1999). This implies that the estimated marginal cost of 9.3 e-cents per kilometer is possibly on the conservative side, especially for tracks in the western and central parts of The Netherlands. The marginal cost estimates for the off-peak period turn out to be much lower, being just 20–35% of the peak hour estimates. In absolute terms, the off-peak marginal cost where no additional rolling-stock is required turns out to be very low, especially when compared to revenues. This comparison will be done in the next sections.

4.2. The tariff structure

Table 3 gives some information on the various types of cards that are offered by The Netherlands Railways, and their contribution to total returns. Considering Second Class passengers, it is found that the largest share in passenger kilometers is for the student cards (30%). It is followed by holders of reduced fare cards (22%), which allow one to travel with a 40% reduction in the fare, but only after the morning peak. Not more than 19% of the passenger kilometers yields the full fare revenues to The Netherlands Railways. Railway pass holders, who pay a temporal fixed fee in order to travel at zero marginal price,[15] account for 15% of total distance traveled.

The shares in the total returns are rather different: the largest share (27%) comes from full fare passengers. It is followed by student card holders (22%), reduced fare passengers (20%) and railway pass holders (16%). These passes, once they have been purchased, allow free use of the entire network or a certain trajectory, respectively.

The average receipts per kilometer are high for full fare tickets (one-way and return), and particularly low for the 'other tickets' category (children), and also for the student card holders. The variation in the receipts per traveler kilometer indicates that The Netherlands Railways have been able to arrive at a substantial degree of price discrimination among its customers. Note that in some cases the average receipts may be very different from the marginal prices paid by travelers. For example, in the case of the railway pass the marginal receipts of an extra passenger kilometer equal zero. With the reduced fare card, the marginal receipts will also be lower than the average receipts, since the latter include the price of the card giving the right to buy tickets at a reduced fare.

[14] Note that Rietveld and Roson (2001) make use of a not very distinct value of 8.8 e-cents for peak hour marginal cost requiring additional rolling-stock. The off-peak marginal cost where no additional rolling stock or seat is required is set to 0.1 e-cent per kilometer in their article. These figures are based on direct computation from raw NS data.

[15] This can be either on a certain given track or on the entire network. It is also possible to include other means of public transport (bus, tram, metro) with a pass for the entire railway network.

Table 3
Average revenues of The Netherlands Railways for passenger transport (1996/1997), source: NSR

Type of card	Share in passenger km (%)	Share in returns (%)	Return per passenger km (e-cents/km)
First class	7	10	10.2
Second class	93	90	6.7
of which			
full fare	19	27	9.8
of which			
single	6	11	11.6
return	9	12	10.1
other	4	5	8.4
reduced fare	22	20	6.1
other tickets (children)	4	1	2.2
railway pass	15	16	7.6
student card	30	22	5.1
international	4	5	8.6

4.3. Price elasticities

Numerous authors have gauged price elasticities for the demand for transport by train. Oum et al. (1992) give an overview of price elasticities for various means of transport, mainly based on studies from the 1980s. There appears to be quite some dispersion between estimates, but this can for a large part be attributed to the type of data that has been used (cross-section, panel, aggregate time series) as well as to the market segment (business, leisure, etc.) for which the elasticity has been estimated. As has been mentioned by Oum (1989) comparison of different estimates is often difficult, especially when the models to be compared are not nested in a more general model. Nevertheless, if the data source and the model assumptions are taken into account well, comparison of different elasticity estimates can be fruitful. Table 4 summarizes recent estimates of price elasticities for The Netherlands. Peak hour is defined as the period before 9 a.m. during weekdays. The periods after 9 a.m. and the weekends are defined as the off-peak. Based on an aggregate time series data set, Oum (1992) reports a price elasticity of −1.16. However, this study does not distinguish between peak and off-peak hour demand, nor between First and Second Class travelers. A recent study by Vuuren and Rietveld (2002) reports a long run estimate for the price elasticity during the off-peak hour at −1.37. This figure is based on a repeated cross-section dataset of The Netherlands Railways during the period 1992/1993. In his overview concerning transport elasticities in The Netherlands, Waard (1990) reports a morning peak hour elasticity value of −0.68. Although these figures are not directly comparable to one another, they are in accordance with what is to be expected a priori: the demand for train kilometers during peak hours is inelastic, while the demand during the off-peak is elastic. Motivation plays an important role for this, as commuters usually do not have many alternatives and therefore cannot easily switch to other modes or another departure time or another track as is often possible for recreational trips, resulting in a relatively low price elasticity for the demand during peak hours.

4.4. Optimal prices

As has been argued earlier in this article, The Netherlands Railways (NS) is a monopolistic firm, which, as has been shown in recent empirical work, faces a cost function that exhibits scale economies. In this section, I will compare the collected figures of the previous sections on marginal costs, average revenues, and price elasticities, and conclude to what extent the current pricing schedule is optimal, both from the viewpoint of welfare maximization and profit maximization.

First, Table 4 has shown that the price elasticity of travelers during the morning peak hour is quite low at a value of about 0.7. Clearly, this implies that the railways are not allowed to maximize their profits in this segment of the market: At the current given capacity, profit maximization would have implied that prices should be higher. Definitely, some travelers will be lost, but overall the profits will rise due to the inelastic nature of this sub-market. However, government regulation has not allowed this to happen, and as can be read from Table 3, the average revenue per passenger kilometer for full tariff travelers is restricted at 9.8 e-cents, and 7.6 e-cents for railway pass holders. Given that during peak hour nearly half of the travelers have a railway pass and about 20% use full tariff tickets,[16] the average revenue will be around 8–9 e-cents per kilometer. Comparing these figures with the marginal cost estimates in Table 2, it can be seen that a significant portion of peak hour travelers are paying a price that is lower than marginal cost. In this respect, it is not strange to find that over the past years the prices of railway passes have been considerably raised. In addition, suggestions have been made to disallow students to travel at zero marginal prices before 9 a.m. On less dense tracks, where no additional rolling-stock would be required in order to *produce* one extra seat kilometer, NS are able to achieve a fairly large mark-up for peak hour travels. In order to determine to what extent NS are able (i.e. *allowed*) to

[16] Source: Basisonderzoek NS 1992/1993.

Table 4
Reported price elasticities for passenger kilometers by train in The Netherlands

Study	Reported price elasticity	Data set	Period
Oum (1992)	−1.16	Aggregate data from different sources, 1977–1991	–
van der Waard (1990)	−0.68	Landelijk Model, disaggregate data, 1995 (predicted value)	Morning peak
van Vuuren and Rietveld (2002)	−1.37	Baissonderzoek NS, 1992/1993 (repeated cross-section)	Off-peak

make profits, one can compute the Ramsey coefficient in Eqs. (7) and (8) *ex post*:

$$\alpha = M\epsilon, \tag{9}$$

where mark-up equals $M = (p - c)/p$, and ϵ is the price elasticity of demand. The value of α will then be an indicator for the extent to which NS are able (allowed) to make profits. On less dense tracks, the thus computed indicator equals 0.4, which shows that the compromise between the welfare maximizing price ($\alpha = 0$) and the profit maximizing price ($\alpha = 1$) is slightly in favor of the former. However, NS does make profits on these tracks, but this cannot be said of dense tracks. As has just been shown, prices on dense tracks hardly reach the level of marginal cost. If infrastructure capacity is the bottleneck, then the negative gap between price and marginal cost will most likely become significant, implying that peak hour seat kilometers on dense tracks are *underpriced*, both from the profit-maximizing perspective and from the welfare-maximizing perspective. With these considerations, it appears to be odd that the price of one seat kilometer is equal on the entire railway network, while marginal costs may substantially differ between different tracks, as was seen in Table 2. As Rietveld and Roson (2002) have shown in recent work, the constraint that prices are equal on different tracks may easily lead to welfare losses. Given current empirical evidence, it seems reasonable to study the possibility of lowering tariffs in less dense areas, while raising peak hour prices on dense tracks. Such tariff differentiation—a new form of third-degree price discrimination—offers scope for a budget-neutral improvement of total welfare.

During off-peak hours, a substantial percentage of customers travels at reduced fare, which equals 60% of the full fare. Thus, apart from the relatively small category other tickets, average revenues during the off-peak are in the range 6.1–11.6 e-cents (Table 3). Comparison with the marginal cost estimates in Table 2 then shows that NS are able to earn a considerable mark-up in this sub-market. The production of additional off-peak seat kilometers will mostly not require additional investment in infrastructure or rolling-stock. This implies that mark-up practically equals unity, so that Eq. (9) rewrites as

$$\alpha \approx \epsilon. \tag{10}$$

This means that a perfectly elastic market ($\epsilon = 1$) implies the profit-maximizing price, while an inelastic market would correspond with Ramsey prices (Eqs. (7) and (8)). An elastic market ($\epsilon > 1$) implies sub-optimal prices, both from the profit- and the welfare-point of view. In other words, lowering prices will lead to an increase in both profits and welfare. In addition, such a lower price will not harm one single (potential) customer, so that it can be termed a *Pareto-improvement*. Results in Section 4.3 suggested price-elastic off-peak travels, implying that lowering the off-peak fares should be seriously considered. In this case, the lower receipts due to the lower price of train tickets will be offset by the higher receipts from additional sales of seat kilometers.

The preceding analysis strongly hinges on reported estimates of marginal cost and demand elasticities. Tables 5 and 6 report mark-up and ex post Ramsey coefficients (cf. Eq. (9)) for varying levels of price elasticities and marginal costs. In Table 5, a peak kilometer price of 8.5 e-cents has been used, and for the off-peak kilometers (Table 6) the used price is 7 e-cents. In Table 5, Ramsey coefficients corresponding to welfare maximizing prices are in bold, while Table 6 indicates profit maximizing price with bold Ramsey coefficients. The computed mark-ups and Ramsey coefficients should be regarded as being indicative. Peak hour Ramsey coefficients show that, regardless of the elasticity value, marginal costs between 7 and 9 e-cents practically correspond to marginal cost pricing.

Table 5
Peak hour Ramsey coefficients ($M_p\epsilon_p$) at varying cost and elasticity levels

Marginal cost	2	3	4	5	6	7	8	9
Implied mark-up	0.8	0.6	0.5	0.4	0.3	0.2	0.1	< 0
$\epsilon = -0.2$	0.2	0.1	0.1	0.1	0.1	0.0	**0.0**	< 0
$\epsilon = -0.4$	0.3	0.3	0.2	0.2	0.1	0.1	**0.0**	< 0
$\epsilon = -0.6$	0.5	0.4	0.3	0.2	0.2	0.1	**0.0**	< 0
$\epsilon = -0.8$	0.6	0.5	0.4	0.3	0.2	0.1	**0.0**	< 0
$\epsilon = -1.0$	0.8	0.6	0.5	0.4	0.3	0.2	**0.1**	< 0

Table 6
Off-peak hour Ramsey coefficients ($M_o\epsilon_o$) at varying cost and elasticity levels

Marginal cost	0.1	0.3	0.5	0.7	1	2	3	4	5
Implied mark-up	1.0	1.0	0.9	0.9	0.9	0.7	0.6	0.5	0.4
$\epsilon = -0.6$	0.6	0.6	0.6	0.5	0.5	0.4	0.4	0.3	0.2
$\epsilon = -0.8$	0.8	0.8	0.7	0.7	0.7	0.6	0.5	0.4	0.3
$\epsilon = -1.0$	**1.0**	**1.0**	**0.9**	**0.9**	0.9	0.7	0.6	0.5	0.4
$\epsilon = -1.2$	1.2	1.2	1.1	1.1	**1.0**	0.9	0.7	0.6	0.4
$\epsilon = -1.4$	1.4	1.3	1.3	1.3	1.2	**1.0**	0.9	0.7	0.5
$\epsilon = -1.6$	1.6	1.5	1.5	1.5	1.4	1.2	**1.0**	0.8	0.6

This implies that the earlier finding that peak hour prices 'hardly reach the level of marginal cost on dense tracks' seems quite robust. It is also confirmed that tracks with lower marginal costs will in general be able to generate some profits. As was already seen in Eq. (10), low marginal costs—say lower than 1 e-cent—imply that profit maximization coincides with a perfectly elastic market. Marginal costs exceeding 1 e-cent mostly imply Ramsey pricing.

The overall conclusion from Tables 5 and 6 is that peak hour pricing is dominated by the welfare objective, while off-peak hour prices are dominated by the profit objective. Secondly, it has been confirmed that lowering tariffs in a price elastic off-peak market will lead to an increase in profits, and even to a Pareto-improvement. Finally, tariff differentiation according to track will offer scope for improvement of total welfare, and also railway profits.

Acknowledgements

I wish to thank Piet Rietveld, Erik Verhoef, Harry van Ooststroom, Jan Rouwendal, editor John Preston, and an anonymous referee for discussions and comments on earlier drafts of this article.

References

Armstrong, M., 1996. Multiproduct nonlinear pricing. Econometrica 64 (1), 51–75.

Armstrong, M., Vickers, J., 1991. Welfare effects of price discrimination by a regulated monopolist. Rand Journal of Economics 22 (4), 571–580.

Armstrong, M., Cowan, S., Vickers, J., 1995. Nonlinear pricing and price cap regulation. Journal of Public Economics 58, 33–55.

Braeutigam, R., 1989. Optimal policies for natural monopolies. In: Schmalensee, R., Willig, R. (Eds.). Handbook of Industrial Organization. Handbooks in Economics, vol. II, no. 10. North-Holland, Amsterdam, pp. 1289–1348.

Braeutigam, R., 1999. Learning about transport costs. In: Gómez-Ibáñez, J., Tye, W., Winston, C. (Eds.). Essays in Transportation Economics and Policy. Brookings, Washington, DC, pp. 57–98.

Dennis, S., 2000. Changes in railroad rates since the Staggers Act. Transportation Research 37E, 55–69.

Evans, D., Heckman, J., 1984. A test for sub-additivity of the cost function with an application to the Bell system. American Economic Review 74 (4), 615–623.

Gathon, H., Pestieau, P., 1995. Decomposing efficiency into its managerial and its regulatory components: the case of European railways. European Journal of Operational Research 80, 500–507.

Goldman, M., Leland, H., Sibley, D., 1984. Optimal nonuniform prices. Review of Economic Studies 51, 305–319.

Helm, D., Thompson, D., 1991. Privatised transport infrastructure and incentives to invest. Journal of Transport Economics and Policy 25, 231–246.

Katz, M., 1983. Non-uniform pricing, output and welfare under monopoly. Review of Economic Studies 50, 37–56.

Keeler, T., 1997. Competition, natural monopoly, and scale economies. In: Oum, T., Dodgson, J., Hensher, D., Morrison, S., Nash, C., Small, K., Waters II, W. (Eds.). Transport Economics. Harwood, New York, pp. 123–143.

Kessides, I., Willig, R., 1998. Restructuring regulation of the rail industry for the public interest. Railways: Structure, Regulation and Competition Policy, DAFFE/CLP(98)1, OECD, pp. 147–181.

MuConsult, 1999. Voorbereiding Prestatiecontract NS. Concept eindrapport.

van Ooststroom, H., 1999. Marktwerking en regulering bij de spoorwegen. PhD Thesis. Vrije Universiteit Amsterdam.

Oum, T., 1989. Alternative models and their elasticity estimates. Journal of Transport Economics and Policy 23, 163–187.

Oum, T., 1992. The structure of travel demands in The Netherlands, report. Faculty of Commerce and Business Administration, University of British Columbia, Netherlands Ministry of Transport and Public Works, Project Bureau of Integrated Transport Studies.

Oum, T., Yu, C., 1994. Economic efficiency of railways and implications for public policy. Journal of Transport Economics and Policy 28, 121–138.

Oum, T., Waters II, W., Yong, J.-S., 1992. Concepts of price elasticities of transport demand and recent empirical estimates. Journal of Transport Economics and Policy 26, 139–154.

Oum, T., Waters II, W., Yu, C., 1999. A survey of productivity and efficiency measurement in rail transport. Journal of Transport Economics and Policy 33 (1), 9–42.

Pestieau, P., Tulkens, H., 1990. Assessing the Performance of Public Sector Activities: Some Recent Evidence from the Productive Efficiency Viewpoint, CORE. Université de Liège/Université Catholique de Louvain.

Rietveld, P., Roson, R., 2002. Direction-dependent prices in public transport: a good idea? The back haul pricing problem for a public transport firm with market power. Transportation, in press.

Roberts, K., 1979. Welfare considerations of nonlinear pricing. Economic Journal 89, 66–83.

Shires, J., Preston, J., 1999. Getting back on-track or going off the rails? An assessment of ownership and organisational reform of railways in Western Europe. Paper presented at the Sixth International Conference on Competition and Ownership in Land Passenger Transport, Cape Town.

Sibley, D., Srinagesh, P., 1997. Multiproduct nonlinear pricing with multiple taste characteristics. Rand Journal of Economics 28 (4), 684–707.

Spence, M., 1977. Nonlinear prices and welfare. Journal of Public Economics 8, 1–35.

Spence, M., 1980. Multi-product quantity-dependent prices and profitability constraints. Review of Economic Studies 47, 821–841.

Srinagesh, P., 1986. Nonlinear prices and the regulated firm. Quarterly Journal of Economics 101, 51–68.

Stigler, G., 1987. Theory of Price. MacMillan, New York.

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press, Cambridge, MA.

Train, K., 1997. Optimal Regulation. MIT Press, Cambridge, MA.

Trotter, S., 1985. The price-discriminating public enterprise, with special reference to British Rail. Journal of Transport Economics and Policy 19, 41–64.

Varian, H., 1989. In: Schmalensee, R., Willig, R. (Eds.). Price Discrimination. Handbook of Industrial Organization, vol. 1. Elsevier, Amsterdam, pp. 597–654.

van de Velde, D. (Ed.), 1999. Changing Trains, Ashgate.

van Vuuren, D., Rietveld, P., 2002. The off-peak demand for train kilometres and train tickets: a microeconometric analysis. Journal of Transport Economics and Policy, in press.

van der Waard, J., 1990. Koncept Elasticiteiten Handboek, Rijkswaterstaat, Directorate General of Public Works and Water Management. Dienst Verkeerskunde.

Willig, R., 1978. Pareto-superior nonlinear outlay schedule. Bell Journal of Economics 9, 56–69.

Wilson, R., 1992. Nonlinear Pricing. Oxford University Press, Oxford.