



Estimating Marginal Costs for the Austrian Railway System

*Gerhard Munduch, Alexander Pfister, Leopold Sögner and Alfred Stiasny**

Working Paper No. 78

February 2002

Abstract

This article presents an econometric analysis of the maintenance costs for the Austrian railway system. The data contain observations of track maintenance costs from 1998 to 2000. Our analysis identifies the cost driving factors in order to determine estimates of marginal costs, as required by the infrastructure provision principles of the European Union. The analysis identifies the variables “track length” and “transported gross-tons” as the principal cost determinants. Furthermore, we observe that total costs as well as marginal costs increase with (i) a high proportion of the track occupied by train stations, (ii) the number of switches within a track, (iii) narrow bends, and (iv) considerable slopes. Moreover average as well as marginal costs for secondary lines are significantly higher than for main lines.

Keywords: Railway maintenance costs.

JEL classification: C51, L92.

Acknowledgements: The authors thank the Österreichische Bundesbahnen (Austrian Railways) – especially Klaus Garstenauer from the department of infrastructure – for interesting discussions, the provision of the data and financial support. All remaining errors and shortcomings remain in our sole responsibility. A. Pfister and L. Sögner acknowledge financial support from the Austrian Science Foundation (FWF) under grant SFB#010.

***Address correspondence to:** Alfred Stiasny, Email: alfred.stiasny@wu-wien.ac.at, Tel.: +43 1 31 336 4541, Fax: + 43 1 31 336 755, Department of Economics, Vienna University of Economics and Business Administration, Augasse 2-6, A-1090 Vienna.

1. Introduction

One major goal of the European Union within the ongoing deregulation program is the liberalization of the European railway sector. Competition in the provision of railroad transportation services requires of course free market access of private companies. Presently, this market is dominated by national monopolies under public control. However due to a sub-additive cost structure (natural monopoly) with respect to the track infrastructure, a competitive system in this sector cannot be established easily.¹ Therefore the European Union has decided that monopoly in the provision of the track infrastructure will be maintained, yet under regulatory restrictions securing a high standard of efficiency. In the corresponding directive² Dir. 01/14 the regulation scheme is set out in full detail. The following analysis is motivated especially by article 7 of the directive, which imposes the requirement of marginal cost pricing. Allowances for external effects and “non-disturbing” fixed costs are permitted. However these peculiar exemptions from the principal rule will not be considered by our analysis since neither appropriate market prices nor reliable estimates are available.

It is worth mentioning that the regulation approach chosen by the European Union is not necessarily optimal. Economic theory provides a considerable number of alternative methods and mechanisms to implement a regulatory framework. For an overview the reader is referred to Train (1994); for the regulation of railroads in particular see Levin (1981b) as well as Kessides and Willing (1995). These methods differ in the objective they optimize and in their incentive structure. Demanding for prices corresponding to marginal costs results in a first best solution. Furthermore, a fast penetration of the markets can be expected in such a regime. However, a major problem arises if the cost function of the regulated business is sub-additive. Without cross-subsidies from other sectors the regulated firm will generate losses. This loss has to be compensated by a proper subsidy, which implies agency-problems e.g. due to false reporting. In practice therefore the outcome can be inefficient.

Another way to tackle the pricing problem is to impose average cost pricing. This scheme suffers from the disadvantage that a lower degree of competition is expected since the usage charges would be much higher compared to marginal cost pricing. In the case of Austria for instance our econometric analysis will show that marginal costs cover only between 20% and 30% of the average maintenance costs. This result is corroborated by Swedish data in the work of Johansson and Nilsson

¹ For a definition and the welfare-theoretic implications of a natural monopoly the reader is referred to standard microeconomic literature, e.g. Gravelle and Rees (1992) or Mas-Colell et al. (1995).

² This directive (EU(2001)) of the European Union will abbreviated by Dir. 01/14. Articles are denoted by parentheses.

(1999). Therefore, average cost charges would be far above marginal cost fees. Hence significantly weaker competition could be expected. Last but not least, agency problems due to false reporting activities do not disappear in this regime.

Since the European Union has chosen a regulation framework with the requirement of marginal cost pricing for the practical implementation it is necessary to determine empirically a reliable estimate of marginal costs. The following investigation tackles this problem for the case of Austria. A few similar studies can be found in the recent literature: For the US, Levin (1981a) investigated the railroad cost structure. A more detailed study on railway maintenance costs is given by the work of Johansson and Nilsson (1999), focussing on the Swedish railways. Given data from 1994 to 1996 (with some 170 observations per year), the authors use an augmented Cobb-Douglas specification for maintenance costs. In their specification the track length, gross-tons, the distinction between main lines and secondary lines, track quality, the number of switches, the number of bridges and the number of tunnels are used as explanatory variables. The authors identify the first four variables as the most significant ones. The estimated marginal costs are in the range of 12 to 28% of average costs.

This article is organized as follows. Section 2 describes the data used in our econometric approach, while section 3 illustrates the estimation methodology. Section 4 deals with the model specification and presents the estimation results. It turns out that the most significant cost factors are track length and gross-tons. Furthermore, we observe that total costs and marginal costs increase if (i) a higher percentage of the track length can be attributed to stations, (ii) more switches are within a track, (iii) the radiuses of the bends are small, and (iv) considerable slopes are present. Moreover, a significant impact can be attributed to the difference between main lines and secondary lines.

2. The Data

The data have been provided by the Austrian Railways and consist of observations of 220 tracks per year from the years 1998 to 2000. Two tracks had to be excluded from this data due to coding errors. In 1998 and 2000 the data for three of the remaining 218 tracks were missing. The cost data are denominated in Austrian Schilling (ATS). Although the econometric analysis is performed with the original data, we present the main results in EUR.

All (nominal) data have been deflated with respect to the base year 2000. The inflation factor for each year has been calculated from the construction cost index (1.11 and 1.08 for 1998 and 1999 respectively). After stacking the tracks there are up to 648 observations available. For each track, the following information is given:

Table 1: Information per track

Abbreviation	Description
<i>ID</i>	Track number
<i>HN</i>	0... for main lines (tracks of national or international importance) 1... for secondary lines (tracks of regional importance)
<i>K</i>	Track maintenance costs
<i>tkm</i>	Gross-ton-kilometers
<i>km</i>	Track length (for double-railed tracks, the double of the distance)
<i>t1</i>	Length of single-railed tunnels in meters
<i>t2</i>	Length of double-railed tunnels in meters
<i>r250</i>	Bends with radius < 250 meters (as percentage of track length)
<i>r500</i>	Bends with radius < 500 meters (as percentage of track length)
<i>s1</i>	Slopes of 10‰ to 20‰ (as percentage of track length)
<i>s2</i>	Slopes steeper than 20‰ (as percentage of track length)
<i>w_l</i>	Length of the switches in meters
<i>bh_proz</i>	Station rails (as percentage of track length)
<i>as5_15</i>	Rails with age from 5 to 15 years (as percentage of track length)
<i>as_15_25</i>	Rails with age from 15 to 25 years (as percentage of track length)
<i>as_25</i>	Rails older than 25 years (as percentage of track length)

Given the data base as described in Table 1, the following further variables were calculated: Gross-tons, length of the station rails in km, slopes steeper than 10‰ (as percentage of track length) and length of tunnels (as percentage of track length).

Table 2: Additionally calculated variables

<i>ton</i>	Gross-tons (tkm / t)
<i>bh_km</i>	Length of station rails in km ($bh_proz * km / 100$)
<i>steig10</i>	Slopes steeper than 10‰ ($s1 + s2$)
<i>tunnel</i>	Tunnels ($(t1 + 2*t2) / km$)

In order to get a first impression we present some descriptive statistics. Table 3 gives an overview of the average, the median, the maximum, the minimum, the total and the standard deviation of the most important variables used in the further analysis.

Table 3: Summary statistics of the most important variables

	<i>K</i> (Mill. EUR)	<i>tkm</i> (Mill.)	<i>km</i>	<i>ton</i> (Mill.)	<i>bh_km</i>	<i>w_l</i>	<i>r500</i>	<i>steig10</i>
Mean	9.46	311	32.79	9.23	14.54	1,766	16.50	17.87
Median	6.47	82.92	25.24	6.02	6.22	1,013	11.31	0.00
Maximum	57.02	3,620	185.00	179	417.38	15,821	100.00	100.00
Minimum	0.003	0.06	0.40	0.03	0.00	0.00	0.00	0.00
Std. Dev.	10.07	533	31.56	12.01	31.77	2,269	18.26	27.05
Total	6,130	202,000	21,246.99	5,980		1,144,565		

Before we proceed with our analysis it seems necessary to address a problem peculiar to any empirical analysis of cost data, namely the task of separating inefficiencies from technically efficient production. As long as markets are competitive, the firms are forced to minimize costs continuously. If only one supplier is present however, this incentive structure disappears. With regard to our empirical analysis the question arises what degree of inefficiency is within the available data for maintenance costs since these have been generated by a national monopoly. For Belgian railroads, De Borger (1993) investigated allocative and technical efficiency. The estimated degrees of inefficiency are in the range of 15-25% for allocative efficiency and 2% for technical efficiency, respectively. A major part of allocative inefficiency is caused by an inefficient use of electric energy. The author concludes that inefficiencies cause an increase in costs of 1-5%.

In our econometric analysis we proceed as follows: We use all the data available to estimate the cost function and identify outliers, by looking for tracks with costs far above or below the regression line. A strong indicator for inefficient production are tracks with costs far above the regression line (positive residuals). Surprisingly, all of the outliers in our analysis are tracks where the costs are far below the regression line (residuals are negative). Hence we conclude that no tracks, which are extremely less efficient than the mean, are in the data. However this procedure does not tell us whether the underlying production technology is efficient, since the whole track maintenance might be inefficient. Furthermore an international comparison of marginal costs is not helpful either because at present the number of railroad companies which provide information on marginal costs is still small and in addition the geographic differences between the railroads are large.

3. Estimation Methodology

The goal of our analysis is the estimation of the cost function and the determination of statistically significant cost driving variables. Therefore, let us start with a cost function $K = K(y)$, where $y = (y_1, \dots, y_n)$ is the vector of prediction variables; the response variable is the track maintenance costs K . Marginal costs MC , with respect to gross-ton-kilometers (tkm), are defined as:

$$MC \equiv \frac{\partial K}{\partial tkm}.$$

The choice of gross-ton-kilometers as the relevant output variable for the calculation of marginal costs has been motivated by the information that in Austria (as well as in other European countries) this variable will constitute the basis for the charges. However that does not necessarily mean that it represents the appropriate output variable in the cost function. In the case of a cost function which is linear in tkm this implies not only that gross tons ton and track length km have the same cost effect but also that the isolated effects interact multiplicatively. Similar

restrictions follow also for alternative functional forms, e.g. given a log-linear cost function equal elasticities for ton and km would be implied.

We claim that this approach is too restrictive a priori. The econometric analysis in section 4 will confirm this proposition. Therefore, the variable tkm has to be separated into km and ton to account for this problem. However to derive the marginal costs with respect to tkm in this setup, it is worth noting that the track length km represents a constant for a particular track, i.e. although tkm increases if an extra train passes the track between point A and point B, the distance between these points remains the same. Therefore $\partial(tkm) = \partial ton * km$ and we obtain

$$MC = \frac{\partial K}{\partial tkm} = \frac{\partial K}{\partial ton} \frac{1}{km}. \quad (1)$$

The empirical calculation of the marginal costs requires several steps with some methodological complications. In the following we will explain this procedure given a log-linear cost function since this specification will be used for the derivation of our main results and conclusions.

Let us consider the following cost function:

$$\log(K_i) = \mathbf{a}_0 + \mathbf{a}_1 \log(ton_i) + \dots + \mathbf{a}_k \log(x_{ki}) + \mathbf{a}_{k+1} \log(ton_i) z_{1i} + \dots + \mathbf{a}_{k+l} \log(ton_i) z_{li} + \mathbf{e}_i \quad (2)$$

where K_i are the costs of track i , $\mathbf{x}_i = (x_{1i}, \dots, x_{ki})$, $x_{1i} = ton_i$, $\mathbf{z}_i = (z_{1i}, \dots, z_{li}) \subset \mathbf{x}_i$ are the explanatory variables and \mathbf{e}_i is an iid normally distributed residual. The variables (z_{1i}, \dots, z_{li}) which form a subset of x_i may have an impact on the coefficient of $\log(ton)$. Estimates of model parameters are denoted by the symbol $\hat{\cdot}$, the tracks are indexed by $i = 1, \dots, n$.

To obtain a single estimate for marginal costs (“mean marginal costs”) we proceed as follows: In the first step we derive the marginal costs for the individual tracks. Secondly, we calculate average marginal costs by some proper averaging.

From the estimated cost function we obtain the expression for the elasticity of the maintenance costs with respect to gross-tons:

$$\frac{\partial \hat{K}_i}{\partial ton_i} \frac{ton_i}{\hat{K}_i} = \hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}. \quad (3)$$

This elasticity together with (1) leads to an estimate of the marginal costs with respect to tkm for a particular track i :

$$MC_i = \frac{\partial \hat{K}_i}{\partial tkm_i} = \frac{\partial \hat{K}_i}{\partial ton_i} \frac{1}{km_i} = \left(\frac{\partial \hat{K}_i}{\partial ton_i} \frac{ton_i}{\hat{K}_i} \right) \frac{\hat{K}_i}{tkm_i} = (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \frac{\hat{K}_i}{tkm_i}. \quad (4)$$

Examining (2) it is easy to see that the costs separate into a systematic component \hat{K}_i explained by the prediction variables x_i and z_i , and a non-systematic stochastic component e_i . In order to estimate \hat{K}_i , we recall that the distribution of $\log \hat{K}_i$ is normal with $\log \hat{K}_i \sim N[\log K_i - e_i, se^2]$, where se is the estimated standard deviation of the residuals. Consequently the distribution of \hat{K}_i is log-normal. Therefore, the estimate of \hat{K}_i (expected value of K_i) is given by

$$\hat{K}_i = \exp(\log(K_i) - \hat{e}_i + 0.5(se)^2),$$

which is simply the Laplace-transform of a normal random variable (see e.g. Mood et al, 1974, p. 117). This completes the calculation of the marginal costs for individual tracks.

Mean marginal costs can be now obtained by a weighted sum of marginal costs with the distribution of gross-ton-kilometers determining the weights:

$$MC = \sum MC_i \frac{tkm_i}{\sum tkm_i} = \frac{\sum (\hat{a}_1 + \hat{a}_{k+1} z_{1i} + \dots + \hat{a}_{k+l} z_{li}) \hat{K}_i}{\sum tkm_i}. \quad (5)$$

To account for the difference between main and secondary lines we define a dummy variable (indicator function) HN with $HN_i = 0$ if track i is on a main line and $HN_i = 1$ for tracks on secondary lines. The calculation of the respective mean marginal costs is straightforward. For the main lines it is given by:

$$MCm_i = MC_i (1 - HN_i)$$

$$MCm = \sum MCm_i \frac{tkm_i (1 - HN_i)}{\sum tkm_i (1 - HN_i)},$$

while for secondary lines it becomes

$$MCs_i = MC_i HN_i$$

$$MCs = \sum MCs_i \frac{tkm_i HN_i}{\sum tkm_i HN_i}.$$

Since the estimates of mean marginal costs are random variables in a finite sample, not only the expectation but also the standard deviation of the estimated marginal costs has to be estimated in order to investigate the statistical significance. Because the estimated mean marginal costs are a continuous function of the estimated parameters $MC = f(\hat{a})$ we can make use of the following theorem which states that $f(\hat{a})$ is normally distributed, such that estimates of the first two moments provide us with a sufficient statistic; i.e. the distribution of the marginal costs is fully described by these estimates.

Theorem [Greene (2000), p. 118]: Let $\hat{\mathbf{a}}$ be the OLS-estimate of the vector of parameters \mathbf{a} , and $f(\hat{\mathbf{a}})$ is a continuous transformation. Let the asymptotic distribution of $\hat{\mathbf{a}}$ be given by: $\hat{\mathbf{a}} \sim N[\mathbf{a}, \Sigma]$. Furthermore, define $\Gamma = \frac{\partial f(\mathbf{a})}{\partial \mathbf{a}'}$ and $C = \frac{\partial f(\hat{\mathbf{a}})}{\partial \hat{\mathbf{a}'}}$.

Then the asymptotic distribution of $f(\hat{\mathbf{a}})$ is $N[f(\mathbf{a}), \Gamma \Sigma \Gamma']$.

In practical applications the covariance matrix is estimated by means of $\text{Var}[f(\hat{\mathbf{a}})] = C \hat{\Sigma} C'$, where $\hat{\Sigma}$ is the estimate of the covariance matrix of the regression coefficients. In our case we start from the continuous transformation:

$$f(\hat{\mathbf{a}}) = MC = \frac{\sum (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i}{\sum tkm_i}$$

and the components of the gradient vector

$$\mathbf{g}' = \left(\frac{\partial f}{\partial \hat{\mathbf{a}}_0}, \dots, \frac{\partial f}{\partial \hat{\mathbf{a}}_{k+l}} \right)$$

are derived from:

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_1} = \frac{\sum \left[\hat{K}_i + (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i \log(\text{ton}_i) \right]}{\sum tkm_i} \quad (6)$$

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_j} = \frac{\sum \left[(\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i \log(x_{ji}) \right]}{\sum tkm_i} \quad \text{for } j = 2, \dots, k \quad (7)$$

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_{k+j}} = \frac{\sum \left[\hat{K}_i z_{ji} + (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i \log(\text{ton}) z_{ji} \right]}{\sum tkm_i} \quad \text{for } j = 1, \dots, l. \quad (8)$$

By using the indicator function HN the gradient vector for the mean marginal costs of the main lines is given by:

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_0} = \frac{\sum (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i (1 - HN)}{\sum tkm(1 - HN)}$$

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_1} = \frac{\sum \left[\hat{K}_i (1 - HN) + (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i \log(\text{ton})(1 - HN) \right]}{\sum tkm(1 - HN)}$$

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_j} = \frac{\sum \left[(\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i \log(x_{ji})(1 - HN) \right]}{\sum tkm(1 - HN)} \quad j = 2, \dots, k$$

$$\frac{\partial f}{\partial \hat{\mathbf{a}}_{k+j}} = \frac{\sum \left[\hat{K}_i z_{ji} (1 - HN) + (\hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_{k+1} z_{1i} + \dots + \hat{\mathbf{a}}_{k+l} z_{li}) \hat{K}_i \log(\text{ton}) z_{ji} (1 - HN) \right]}{\sum tkm(1 - HN)} \quad j = 1, \dots, l$$

For the secondary lines the appropriate expressions can be calculated analogously. Last but not least we test for structural stability. In particular we address the question whether the parameters are stable over the three years considered. This will be accomplished by running a Chow-test (see e.g. Greene (2000), p. 287). The test statistic T is derived from:

$$T = \frac{(e_*'e_* - e_8'e_8 - e_9'e_9 - e_0'e_0)/(2k)}{(e_8'e_8 + e_9'e_9 + e_0'e_0)/(n - 3k)}, \quad (9)$$

where $e_*'e_*$ is the sum of squared residuals over all years, while $e_8'e_8$, $e_9'e_9$, $e_0'e_0$ are the corresponding sums from the years 1998, 1999 and 2000 respectively. k is the number of parameters to be estimated and n is the number of observations over all years. Under the null hypothesis that no structural break is present, the distribution of T is $F(2k, n - 3k)$.

4. Die Model-Selection and Estimation Results

In this section we proceed with the estimation of the cost function. A list of the explanatory variables $x_i = (x_{1i}, \dots, x_{ki})$ is given by:

<i>ton</i>	Gross-tons,
<i>km</i>	Track length,
<i>HN</i>	Main lines $HN = 0$ vs. secondary lines $HN = 1$,
<i>bh_km</i>	Length of station rails,
<i>r500</i>	Bends with radius < 500 meters (as percentage of track length),
<i>steig10</i>	Slopes steeper than 10‰ (as percentage of track length),
<i>tunnel</i>	Length of tunnels (as percentage of track length),
<i>w_l</i>	Length of switches in meters,
<i>alt</i>	Rails older than 15 years (as percentage of track length).

To begin with we have to decide which functional form is appropriate given our data. We restrict ourselves to the linear and the log-linear specification, since these functional forms are most commonly used in the empirical literature. One way to discriminate between these specifications is to estimate the corresponding average cost functions, keeping in mind, that average costs for log-linear total costs are again log-linear, whereas for linear total costs the average cost function is hyperbolic. We get the following OLS-estimates:

$$K_i / tkm_i = 0.073 + 156808.0 \cdot (1/ton_i)$$

(2.87) (13.7)

$$\log(K_i / tkm_i) = 0.26 - 0.86 \cdot \log(ton_i)$$

(29.6) (-37.9)

where the terms in parentheses are the corresponding t-values. In both cases the regression parameters are highly significant at a 5% significance level. However,

the coefficient of determination for the log-linear model is 0.69 compared to 0.23 for the linear model. Furthermore, a White-test on heteroscedasticity and a Ramsey-RESET-test on mis-specification are performed. Applying the White-test the squares residuals are regressed on ton and ton^2 (see Greene (2000), p. 503), while the Ramsey-Reset-test includes the squared forecasts \hat{K}^2 to the set of prediction variables (see Davidson and MacKinnon (1993), p. 195). For the log-linear model the F-statistics are 0.96 for the White test and 2.58 for the Ramsey test, i.e. the null hypotheses that the residuals are homoscedastic and that the model is properly specified cannot be rejected on a 5% significance level. With respect to the linear model these statistics are 11.8 and 61.2, indicating that both null hypotheses have to be rejected on a 5% level. We can therefore conclude that the log-linear specification outperforms the linear one. Consequently we restrict ourselves to the log-linear cost function.

To account for further cost driving factors, the following augmented log-linear specification will be investigated:

$$\begin{aligned} \log(K_i) = & \mathbf{a}_0 + \mathbf{a}_1 \log(ton_i) + \mathbf{a}_2 \log(km_i) + \mathbf{a}_3 \log(bh_km_i) + \mathbf{a}_4 r500_i + \\ & \mathbf{a}_5 steig10_i + \mathbf{a}_6 tunnel_i + \mathbf{a}_7 \log(w_l_i) + \mathbf{a}_8 alter_i + \\ & \mathbf{a}_9 \log(ton_i) \log(bh_km_i) + \mathbf{a}_{10} \log(ton_i) r500_i + \mathbf{a}_{11} \log(ton_i) steig10_i + (10) \\ & \mathbf{a}_{12} \log(ton_i) tunnel_i + \mathbf{a}_{13} \log(ton_i) \log(w_l_i) + \mathbf{a}_{14} \log(ton_i) alter_i + \\ & \mathbf{a}_{15} HN_i + \mathbf{a}_{16} \log(ton_i) HN_i + \mathbf{a}_{17} \log(km_i) HN_i + \mathbf{e}_i. \end{aligned}$$

Starting with this general specification we first address the question whether the variable tkm should be included instead of ton and km – an issue already discussed in section 3. For this reason we perform an F-test for the restrictions $\mathbf{a}_1 = \mathbf{a}_2$ and $\mathbf{a}_{16} = \mathbf{a}_{17}$. The corresponding F-statistic is 23.72, which implies that the null hypothesis can be rejected at a 5% level. This result exhibits an interesting implication from an economic point of view. Since the data support a separate treatment of ton and km , a growth of one percent in the gross-tons has a significantly different effect on costs than an increase of the track length. The estimated coefficients given in Table 4 show that the effect of km is approximately six times higher than the effect of ton .

Next we proceed with model selection (exclusion of variables) by applying the Akaike and Schwartz information criteria. We finally arrive at the following parsimonious specification:

$$\begin{aligned} \log(K_i) = & \mathbf{a}_0 + \mathbf{a}_1 \log(ton_i) + \mathbf{a}_2 \log(km_i) + \mathbf{a}_3 \log(bh_km_i) + \mathbf{a}_4 r500_i + \\ & \mathbf{a}_5 \log(w_l_i) + \mathbf{a}_6 \log(ton_i) \log(bh_km_i) + \mathbf{a}_7 \log(ton_i) steig10_i + (11) \\ & \mathbf{a}_8 \log(ton_i) tunnel_i + \mathbf{a}_9 \log(km_i) HN_i + \mathbf{e}_i. \end{aligned}$$

Compared to equation (10) either the direct effects on total costs of the variables *alter*, *HN*, *tunnel*, *r500*, *steig* and *w_1* or the cross-effects with *ton* have been eliminated.

The estimation results are presented in Table 4. This table reads as follows: The bold numbers are statistically significant parameter estimates at a 10% significance level, terms in parentheses are the corresponding t-values. In addition for the implied estimates of mean marginal costs the (95%) confidence intervals are given in square brackets.

Table 4: Estimation results, equation (11)

Variable	All years	2000	1999	1998
<i>constant</i>	10.7620 (17.47)	10.8052 (11.20)	10.6772 (10.56)	10.4857 (9.84)
$\log(\text{ton})$	0.1049 (2.88)	0.0883 (1.47)	0.1037 (1.84)	0.1456 (2.49)
$\log(\text{km})$	0.6170 (17.33)	0.5919 (9.00)	0.6479 (12.20)	0.6175 (9.53)
$\log(\text{bh_km})$	-0.3049 (-1.42)	-0.6695 (-1.94)	-0.2984 (-0.75)	0.2373 (0.78)
<i>r500</i>	0.0083 (5.20)	0.0091 (3.01)	0.0082 (3.12)	0.0079 (2.56)
$\log(w_l)$	0.0803 (2.27)	0.1284 (2.44)	0.0777 (1.21)	0.0291 (0.49)
$\log(\text{ton})\log(\text{bh_km})$	0.0358 (2.58)	0.0555 (2.60)	0.0349 (1.38)	0.0050 (0.24)
$\log(\text{ton})\text{steig10}$	0.0003 (3.40)	0.0003 (2.07)	0.0003 (2.01)	0.0003 (1.90)
$\log(\text{ton})\text{tunnel}$	-0.0004 (-1.80)	-0.0005 (-1.73)	-0.0002 (-0.90)	-0.0004 (-0.84)
$\log(\text{km})\text{HN}$	0.073 (2.54)	0.0459 (0.86)	0.0915 (2.09)	0.0802 (1.71)
R2 (adjusted)	0.80	0.81	0.79	0.81
Akaike	2.038	-	-	-
S.E.	0.67	0.65	0.71	0.63
<i>MC</i>	0.0076 (6.06)	0.0081 (4.18)	0.0076 (3.50)	0.0065 (2.81)
[conf. interval]	[0.0051, 0.0101]	[0.0043, 0.0119]	[0.0033, 0.0119]	[0.0019, 0.0111]
<i>MCm</i>	0.0066 (5.90)	0.0072 (4.20)	0.0065 (3.32)	0.0053 (2.66)
[conf. interval]	[0.0044, 0.0088]	[0.0038, 0.0106]	[0.0027, 0.0104]	[0.0014, 0.0093]
<i>MCs</i>	0.0425 (5.28)	0.0378 (3.12)	0.0458 (3.44)	0.0450 (3.04)
[conf. interval]	[0.0267, 0.0583]	[0.0139, 0.0617]	[0.0196, 0.0721]	[0.0158, 0.0741]

The elasticity with respect to the track length is 0.62 while it is only 0.10 with respect to the gross-tons. This corroborates our presumption that *km* and *ton* do not exert equal effects. The estimate of mean marginal costs for all tracks amounts to ATS 0.0076 with the 95% confidence interval given by [0.0051, 0.0101]. For the main lines we derive an estimate of 0.0066, while for secondary lines a value of 0.0425 has been estimated. Our estimates imply e.g. that a train with 1300 tons passing a track of a length of 300 km has to be charged 2500 ATS (\approx 182 EUR) if marginal cost pricing (with equal charges for all lines) is applied.

Furthermore, to check for structural stability we apply a Chow-Test. The F-statistic takes the value 1.32, i.e. the null hypothesis that no breaks are present cannot be rejected at a 5% level. With respect to mis-specification the Ramsey-Reset test yields a F-statistic of 0.12 with a prob-value of 0.73, such that no mis-specification can be detected.

With respect to the assumption of homoscedasticity we encountered a problem since the White test indicated possible heteroscedasticity. Therefore we applied as an alternative estimation method WLS (weighted least squares (see Greene (2000), pp. 514-515). This method has the disadvantage that one has to choose some prediction variables for the estimation of the scedastic function. Since the final results turned out to be very sensitive with respect to this choice we dismissed this method. Instead we adhered to the original estimation method but calculated the t-values by the Newey-West method (see Greene (2000), p. 464), which is robust with respect to a rather general class of error processes.

Next we investigated the distribution of the residuals. A Jarque-Bera test (see Greene (2000), p. 398) strongly rejected the null hypothesis that the residuals are normally distributed. As an illustration Figure 1 presents a QQ-plot³ of the residuals against the normal distribution. The convex deviation from the ideal straight line is clearly visible. A closer inspection leads to the conjecture that only a few observations might be responsible for the apparent non-normality. We tried to identify those points by defining all tracks as outliers whose estimated residuals are larger than three times the standard deviation, i.e. $|e_i| > 3se$. Thereby we isolated 10 tracks as outliers and reestimated the cost equation without these observations. The distribution of the resulting residuals were now significantly closer to normality as shown by the QQ-plot in Figure 2. The implied estimates of marginal costs however were not significantly different from the original results. Consequently we concluded that non-normality is not an essential problem in our analysis and we again decided to adhere to the results summarized in Table 4.

³ A QQ-plot compares two distributions by plotting the quantiles of the one series against the quantiles of another series or a theoretical distribution. If the two distributions are the same, the plot should lie on a straight line.

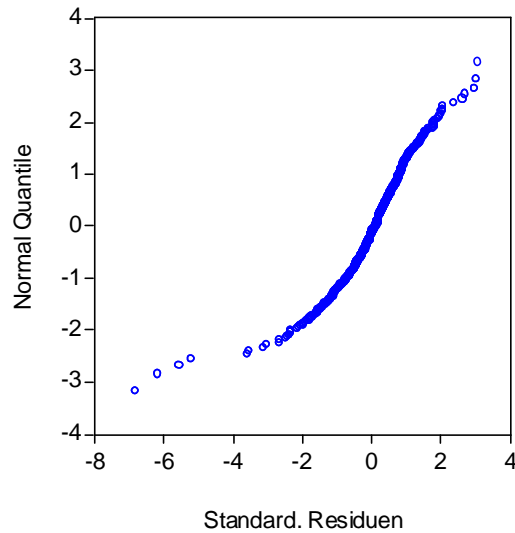


Figure 1 Q-Q-plot of standardized residuals of (11)

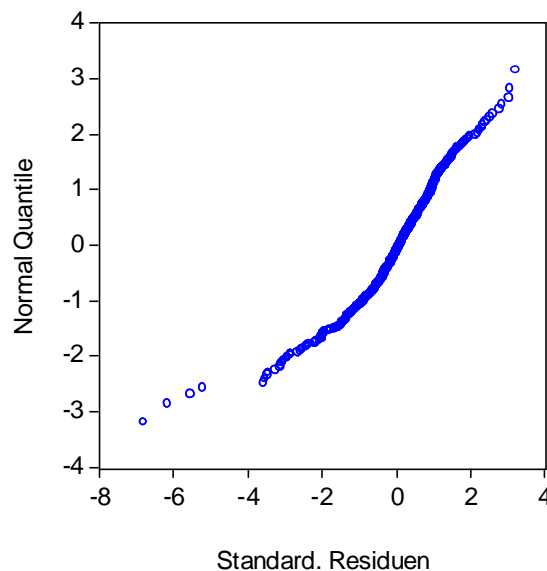


Figure 2. Q-Q-plot of standardized residuals, outlier adjusted

Finally a few graphical representations should provide the reader with some intuition, how the regression model is able to identify the variables which mainly influence the cost function. First we focus on the difference between tracks on main lines versus tracks on secondary lines. In Figure 3 the estimated marginal costs (for individual tracks) and in Figure 4 the (observed) average costs have been plotted against gross-ton-kilometers. Tracks on main lines are denoted by the symbol \circ , while for those on secondary lines the symbol \bullet is used. The falling pattern in both diagrams is clearly visible, which indicates strong economies of scale. In accordance with this first diagnosis we see in both cases a concentration of scatter points for secondary lines in the upper left region, whereas points for main lines

are concentrated in the lower right part. Of course this is a reflection of the higher utilization of main lines.

Similar scatter plots which are not presented for the sake of brevity show that significant “length of station rails” is concentrated in the lower cloud of the scatter points and that the cost driving characteristics “narrow bends” and “steep slopes” are chiefly restricted to tracks on secondary lines.

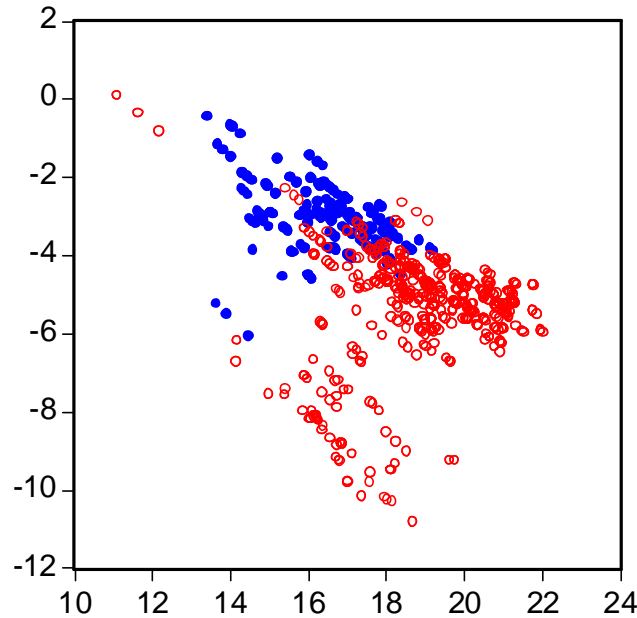


Figure 3: Estimated marginal costs as a function of tkm on a logarithmic scale;
○ denotes main lines ($HN = 0$), the symbol ● secondary lines ($HN = 1$).

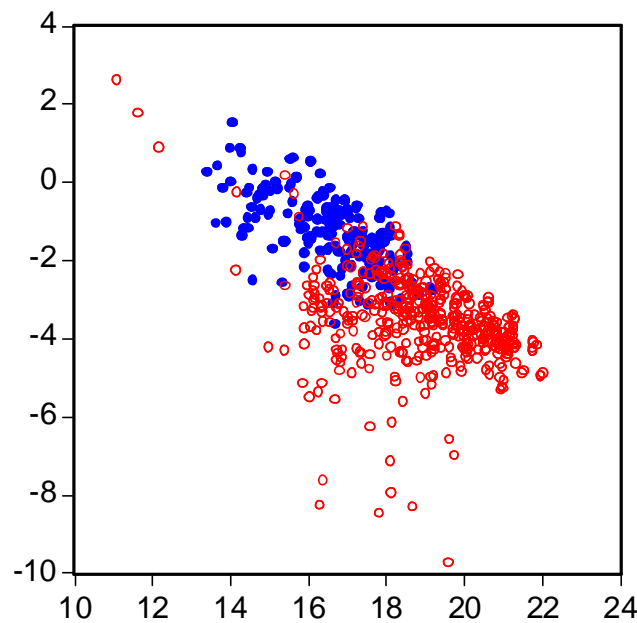


Figure 4: Average costs as a function of tkm on a logarithmic scale;
○ denotes main lines ($HN = 0$), the symbol ● secondary lines ($HN = 1$).

Finally we derive from the estimation results and expression (5) a function with mean marginal costs dependent on tkm and all other variables fixed at their mean values (“mean marginal cost function”):

$$MC(tkm) = f(tkm, \bar{x}_2, \dots, \bar{x}_k),$$

using the reformulated cost function:

$$\log(K) = a_0 + a_1 \log(tkm) + (a_2 - a_1) \log(km) + \dots,$$

and analogously we calculate an “average cost function”. Due to our specification (11) these functions are linear on a logarithmic scale. Figure 5 confronts the mean marginal cost function to the estimated mean marginal costs for individual tracks.

Figure 6 on the other hand compares both functions. The difference between these functions reflects once again the presence of economies of scale.

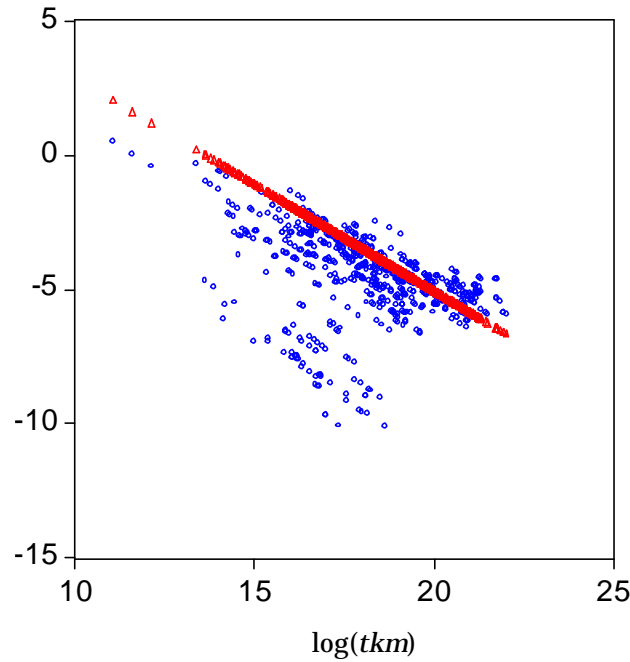


Figure 5: Mean marginal cost function (Δ) and estimated marginal costs (\circ) (logarithmic scale).

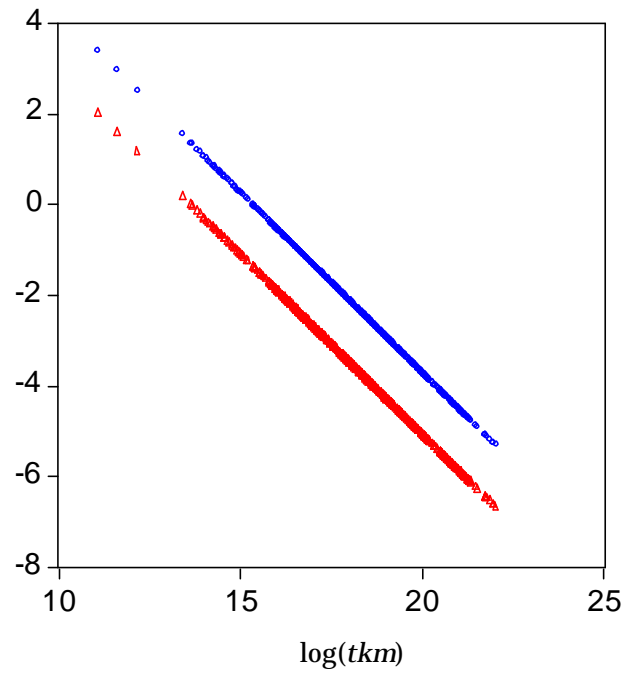


Figure 6: Average cost function (\circ) and mean marginal cost function (Δ) (logarithmic scale).

5. Conclusions

This paper dealt with an econometric investigation of the maintenance costs of the Austrian railways with the primary objective to obtain estimates of marginal costs which could form the basis of infrastructure charges. The analysis is based on data from 1998 to 2000. In particular observations with respect to gross-ton-kilometers, track length, main versus secondary lines, stations, bends, slopes, tunnels, age of rails and switches for about 200 tracks have been used.

By applying standard econometric tools, we found that a log-linear specification conforms best to the data. We identified track length and gross-tons as the dominant cost factors. Furthermore we derived the result that total costs as well as marginal costs increase with (i) the length of station rails, (ii) the length of switches, (iii) narrow bends, and (iv) considerable slopes. A significant influence can be attributed to the difference between main lines and secondary lines. In addition we argued that the inclusion of the variable gross-ton-kilometers in the specified cost function is not appropriate, since the effects of the track length and of gross-tons are clearly asymmetric. Moreover, the negative slopes of average costs and of estimated marginal costs support the conjecture that the Austrian railway infrastructure has to be regarded as a natural monopoly with pronounced economies of scale.

Summing up, the following table compares our estimates of (mean) marginal costs per gross-ton-kilometers with those presented by Johansson and Nilsson (1999) for the Swedish railways. All numbers are in EUR and in constant prices with respect to the base year 2000.

	Marginal Costs	95%-confidence interval	Johansson and Nilsson (1999)
All lines	0.00055	[0.00033 – 0.00074]	0.00043
Main lines	0.00048	[0.00028 – 0.00065]	0.00027
Secondary lines	0.00309	[0.00182 – 0.00413]	0.00207

We see that for Austria marginal costs per gross-ton-kilometer (EUR 0.00055), are about 25% higher than the Swedish estimates. However, the Swedish results are still in the 95%-confidence interval of the Austrian result. The available data do not permit to analyze whether this discrepancy is caused by differences in the tracks, higher labor costs or by inefficiencies in the maintenance.

Since mean average costs amount to EUR 0.0020 per gross-ton-kilometer mean marginal costs cover approximately 27 percents of the mean average costs. Therefore marginal cost pricing would imply drastically lower charges compared to fees based on average costs. Insofar, marginal cost pricing seems to be more appropri-

ate to meet the goal of the European Union to enhance efficiency and competition in the railway sector. Last but not least the obvious difference between main and secondary lines should be accounted for in any practical implementation of infrastructure charges.

References:

Davidson, R. and MacKinnon, J. G., 1993. *Estimation and Inference in Econometrics*. Oxford University Press, New York.

De Borger, B., 1993. The Economic Environment and Public Enterprise Behaviour: Belgian Railroads, 1950–86. *Economica* 60, 443–463.

European Union, 2001. Directive 2001/14/EC of the European Parliament and of the Council of 26 February 2001 on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification. Brussels.

Gravelle, H. and Rees, R., 1992. *Microeconomics*. Longman, Essex.

Greene, W. H., 2000. *Econometric Analysis*. Prentice Hall, Upper Saddle River.

Johansson, P. and Nilsson, J.-E., 1999. *An Economic Analysis of Track Maintenance Costs*. Mimeo, Uppsala.

Kessides, I. N. and Willing, R. D., 1995. *Restructuring Regulation of the Rail Industry for the Public Interest*. Public Policy for the Private Sector, Note No. 58. The World Bank.

Levin, R., 1981a. Railroad Regulation, Deregulation, and Workable Competition. *American Economic Review* 71, 394–398.

Levin, R., 1981b. Railroad Rates, Profitability, and Welfare under Deregulation. *Bell Journal of Economics* 12, 1–26.

Mas-Colell, A., Whinston, M. D. and Green, J. R., 1995. *Microeconomic Theory*. Oxford University Press, New York.

Mood, A. M., Graybill, F. A. and Boes, D. C., 1974. *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

Train, K., 1994. *Optimal Regulation*. MIT Press, Cambridge.