
Article:

Odolinski, K orcid.org/0000-0001-7852-403X and Boysen, HE (2019) Railway line capacity utilisation and its impact on maintenance costs. Journal of Rail Transport Planning & Management, 9. pp. 22-33. ISSN 2210-9706

https://doi.org/10.1016/j.jrtpm.2018.12.001

Railway line capacity utilisation and its impact on maintenance costs

Kristofer Odolinski[a], Hans E. Boysen[b]

[a] The Swedish National Road and Transport Research Institute (VTI), Department of Transport Economics, Box 55685, SE-102 15 Stockholm, Sweden (kristofer.odolinski@vti.se). Corresponding author.

[b] Royal Institute of Technology (KTH), Department of Transport Science, SE-100 44 Stockholm, Sweden (heboysen@kth.se)

Abstract

In this paper, we analyse how railway maintenance costs are affected by different levels of railway line capacity utilisation. Previous studies have focused on the wear and tear of the infrastructure, while this paper shows that it is important to also acknowledge the heterogeneity of the maintenance production environment. Specifically, we estimate marginal maintenance costs for traffic using econometric methods on a panel dataset from Sweden and show that these costs increase with line capacity utilisation. The results are significant considering that current EU regulation (2015/909) states that track access charges can be based on marginal costs, with the aim of achieving an efficient use of available infrastructure capacity.

Keywords: maintenance; marginal cost; rail infrastructure; capacity; track access charges.

1. Introduction

The use of track access charges has become a requirement within the European Union after the vertical separation between infrastructure management and train operations. It is established in the EU regulation 2015/909 that these charges should be based on the direct cost to the infrastructure manager of running a vehicle on the tracks. One part of these costs concerns the maintenance performed due to wear and tear of the rail infrastructure. The overall weight of rolling stock is an important cost driver in this aspect, and hence, gross tonne-km is a rather common charging unit among infrastructure managers in Europe.

There are other aspects that are also important for explaining the maintenance cost level. Different characteristics of the infrastructure such as the age and structure of the track, curvature, the number of switches and line speed are important cost drivers (see for example Öberg et al. (2007) and Odolinski and Nilsson (2017)), as well as vehicle and running gear characteristics, such as wheel slip, unsprung mass and curving performance (see Boysen and Andersson (1989)). These characteristics are often used as control variables in econometric studies that attempt to establish a relationship between traffic and costs (except the vehicle characteristics which can be used to differentiate the marginal costs; see Booz Allen Hamilton (2005), Öberg et al. (2007), and Smith et al. (2017)). Capacity utilisation is however a factor that has not been fully recognised in studies on marginal maintenance costs of rail infrastructure use.

The purpose of this paper is to estimate cost elasticities with respect to traffic that may capture potential differences in maintenance costs with respect to line capacity utilisation. These elasticities can be used to differentiate marginal maintenance costs. If these costs vary for different levels of capacity utilisation, then track access charges should be set accordingly in order to achieve a more efficient use of the infrastructure, according to the short-run marginal cost pricing principle.[1]

The literature on the marginal maintenance costs for rail infrastructure use has focused on the wear and tear caused by traffic (see for example Munduch et al. (2002), Johansson and Nilsson (2004), Öberg et al. (2007), Andersson (2008), Link et al. (2008), Wheat et al. (2009), Odolinski and Nilsson (2017)). From an engineering perspective, the wear and tear (need of repair) of the infrastructure may be non-linear with respect to traffic – that is, a proportional increase in traffic may result in disproportionate increases in wear and tear depending on the traffic level and the contributing damage mechanisms. For example, Öberg et al. (2007) find a non-linear relationship between axle load and track deterioration, while examples of studies that find a non-linear relationship between traffic and costs include Wheat and Smith (2008), Marti et al. (2009), Andersson (2011) and Odolinski (2016).

From a production perspective, different levels of traffic intensity will also result in different possibilities to maintain the assets. This effect is dependent on the line capacity utilisation. For example, if the available infrastructure capacity is heavily used, i.e. the line capacity utilisation is high, then the time slots for maintenance activities may be short and fragmented which creates more interruptions of the maintenance work, and/or maintenance activities need to be performed at night, which tends to be more costly. Indeed, according to Lidén and Joborn (2016), the planning regime for maintenance in Sweden lets the maintenance contractors apply for slots at a late stage in the planning process, which makes it difficult to find possessions that are cost efficient (with

---

[1] There are situations in which it is relevant to deviate from the marginal cost, see for example Rothengatter (2003). Still, as argued by Nash (2003), this does not change the fact that marginal cost should be the basis for an efficient pricing policy.

respect to maintenance production costs). In other words, traffic and infrastructure design with respect to capacity have an impact on scheduling track possessions. Moreover, considering that tracks with high capacity utilisation are more sensitive to delays (Lindfeldt (2015)), where disruptions can result in significant user costs, there is reason to carry out more (preventive) maintenance when capacity utilisation increases. The aim of this paper is therefore to study if and how capacity utilisation affects maintenance costs.

In general, capacity costs come in the form of congestion and scarcity costs. The former type of cost is the result of capacity related delays, which is considered by the track access charges in the United Kingdom. Specifically, a capacity charge is used to recover the delay costs incurred on the infrastructure manager (Network Rail) by increased traffic and is based on a relationship between line capacity utilisation and ‘congestion related reactionary delay’ (Rail Delivery Group (2014)). Scarcity costs, on the other hand, considers the cost of not meeting the demand for slots – that is, the opportunity cost of not allowing train services or maintenance personnel to receive the preferred slot. This part of capacity costs is particularly difficult to quantify within the railway industry as the train operators’ value of each time slot on the tracks is not known (Nash (2018)). Nilsson (2002) considers an auctioning procedure to reveal the opportunity cost and generate an efficient timetable and congestion charges. However, as pointed out by Nash (2018), this method is complex and is little used in practice, and it is only the United Kingdom that uses a specific method for calculating capacity (congestion) costs for railways, where scarcity costs are not included. For example, the Swedish infrastructure manager (Trafikverket, hereafter referred to as the IM) uses a capacity charge stating that the aim is to achieve more efficient use of railway capacity, but the charges are not based on empirical evidence on how capacity utilisation affects (maintenance) costs. Finding the capacity costs related to maintenance production is however not as complex as revealing the opportunity cost of train operators. The impact that capacity utilisation has on maintenance production is included in the maintenance cost, and empirical data on the variation in these two factors can be a way to establish a (possible) relationship.

This paper is organized as follows. Section 2 gives and overview of railway infrastructure capacity and its relationship with maintenance. This forms the basis for the infrastructure capacity variables that will be used in the estimation approach, which is presented in section 3. The model we estimate is presented in subsection 3.1, while the calculation of marginal costs for traffic is described in subsection 3.2. Descriptive statistics of the data used in the estimations are provided in section 4. Estimation results are presented in section 5, and section 6 concludes.

2. Railway infrastructure capacity and maintenance

For railways, there is a theoretical capacity that corresponds to a certain number of passengers or net cargo that can be transported past a point of the infrastructure (line or junction) during a certain time period. This measure is the product of train capacity (passengers or tonnes per train) and line capacity (trains per unit time), suggested by Boysen (2012). In this paper, we are interested in line capacity and its level of utilisation. When analysing the line capacity, the UIC (2013) states that one first and foremost needs a definition of the infrastructure and timetable boundaries (which should be interlocked). The next step is to calculate the capacity utilisation, which is defined as "…the utilisation of an infrastructure's physical attributes along a given section, measured over a defined time period." (UIC 2013, p. 13).[2] The Swedish IM bases its capacity calculations on the UIC leaflet and uses 6 hours per day as the additional time to secure quality of operation in the calculations, which include track possession for maintenance activities.

---

[2] Specifically, the percentage capacity consumption is defined as $\frac{Occupancy\ time + Additional\ times}{Defined\ time\ period} \cdot 100$, where "additional times" is set (by the infrastructure manager) in order to secure quality of operation.

The additional time used for maintenance is in reality heterogeneous. First of all, track possession times depend on the work to be performed, which may require possession times from one hour (or less) to several days. For example, signal repair, snow removal, and tamping of turnouts may take 1 to 4 hours, grinding and tamping of tracks may take 4 to 8 hours, whereas urgent repair may take several days (see Lidén (2014) for a list of maintenance activities with different time possessions and planning horizons). The required possession times, together with the planning horizon for maintenance and the planning process for obtaining possessions, will thus to a large extent determine the possession times given to maintenance production.

Nilsson et al. (2015) describe the planning process in Sweden and the priority setting used: The maintenance (and renewal)[3] activities that have a long planning horizon and require exclusive and long consecutive track possessions in which the track is closed for traffic, are determined at an early stage in the timetabling process. In fact, these activities are planned before the train operators can make requests for train paths. Track possessions for the other maintenance activities, with shorter planning horizons, are determined simultaneously with the train operators' requests for train paths. When there is a conflict between requests for track possessions, the IM uses a set of priority criteria with the aim of finding the solution with the highest socio-economic benefit. The priority criteria are presented in the annual network statement by the Swedish IM (see for example Trafikverket (2015)). Requests for possession times for maintenance are in this case treated by calculating the alternative production costs for other possession times than those requested (the Swedish IM are however aware that the solution is complex, and that the model and priority criteria used are not optimal)[4]. When the train timetable has been set, there are (usually) free time slots still

---

[3] Note however that renewals are not considered in this study.

[4] See for example Brännlund et al. (1998), who presents an optimization approach for finding a profit maximizing timetable with respect to track capacity constraints, which now has resulted in attempts to develop an optimization tool

available. Train operators and maintenance contractors can apply for these available slots, at which the main principle is 'first come, first served'. The lengths of time slots vary depending on the capacity utilisation. Nilsson et al. (2015) provide an example from a maintenance contract, where four different sections of the track had different time slots available. One line section had 5 consecutive hours available, with one track open for traffic, while the other sections had between 2 and 6 consecutive hours with no other traffic running.

Clearly, the maintenance production environment is heterogeneous, and the track possession times available for maintenance can vary considerably between different track sections. Specifically, the timetabling process described above is interconnected to the infrastructure design and the traffic demand. As described in UIC (2013), Nelldal et al. (2009), and Boysen (2013), other important factors for the level of capacity available in railway systems are the number of tracks, the signalling system, the distances between passing sidings, interlockings (such as stations, nodes and junctions), train speeds and train speed heterogeneity. The interaction between these factors determines the production environment for maintenance work and its track possessions. In this study, we consider some of these factors in the assessment of whether and how line capacity utilisation has an impact on maintenance costs. In doing this, it is important to consider differences in railway asset types, where the IM is likely to have invested in high quality assets where the traffic volume is high, which can influence maintenance costs irrespective of the capacity. Variables for infrastructure characteristics that are included in the estimations are presented in section 4.

---

for timetabling (Nilsson et al. 2017). See also Lusby et al. (2011) for a survey of models and methods for railway track allocation, and Lidén (2016) for a treatment of the planning and scheduling problem for maintenance in coordination with traffic.

3. Estimation approach

The marginal cost pricing principle is the basis for the analysis in this paper, which means that the short-run marginal cost of infrastructure use is estimated. The marginal cost (MC) per train-km (TKM) is derived as (see Munduch et al. (2002) or Odolinski and Nilsson (2017)):

$$MC_{it} = \frac{\partial C_{it}}{\partial TKM_{it}} = \frac{TKM_{it}}{C_{it}} \frac{\partial C_{it}}{\partial TKM_{it}} \frac{C_{it}}{TKM_{it}} = \frac{\partial lnC_{it}}{\partial lnT_{it}} \frac{C_{it}}{TKM_{it}}, \qquad (1)$$

where C_{it} is maintenance costs on track section i in year t. Specifically, the cost elasticity with respect to trains ($\frac{\partial lnC_{it}}{\partial lnT_{it}}$) needs to be derived and multiplied by the average cost ($\frac{C_{it}}{TKM_{it}}$). From a line capacity usage perspective, we consider train-km to be a more relevant charging unit compared to gross tonne-km (however, we include a variable for the average tonnage density of the trains in the model estimation to capture the impact of heavier trains on the line).

The main approaches used in previous research to estimate the marginal cost of infrastructure use are the so-called bottom-up approaches (see Booz Allen Hamilton (2005) and Öberg et al. (2007)) and top-down approaches (see for example Munduch et al. (2002), Johansson and Nilsson (2004), Link et al. (2008), Gaudry and Quinet (2009) and Wheat et al. (2009)). The former approach uses engineering models to establish a relationship between traffic and wear and tear of the infrastructure, and then links the damage measures to costs, whereas the latter establishes a direct relationship between traffic and costs. The bottom-up approach is good at describing the infrastructure damage mechanisms caused by traffic (e.g. rolling contact fatigue, abrasive wear, track settlement and component fatigue), whereas the top-down approach is good at linking different cost drivers (such as traffic) to actual costs, allowing for various elasticities of production (depending on the cost function that is specified).

We use the econometric top-down approach, considering that the aim of this paper is to establish a relationship between maintenance costs and the traffic volume's interaction with line capacity. This implies that the cost impact of line capacity utilisation needs to be considered in this estimation, and the marginal cost charges need to be differentiated accordingly (the specification of our model in section 3.1 below reveals how this is achieved).

As previously noted, there are different factors that determine the level of capacity that is available in the railway system, such as the number of tracks, the signalling system, the distances between passing sidings, interlockings, train speeds and train speed heterogeneity and how the timetable is constructed. The factors considered in this study are infrastructure characteristics and traffic volume. Specifically, we use data from the Swedish IM's track information system 'BIS' and create two different variables for infrastructure capacity:

- Track length/Route length (average number of tracks), and
- Number of passing sidings per route-km

Note that track length only includes the main tracks, i.e. yard tracks are not included (which may be used for storage and thus do not have an impact on line capacity). The definition of passing sidings follows the definition provided in Lindfeldt (2009, pp. 13-14). For single track lines, there should be more than one track on a station in order to be defined as a passing siding. For double track lines, there should be more than two tracks, where at least one of the tracks is not classified as main track.

The traffic variables we use are the number of trains that have run on a track section during a given year and the average tonnage density of the trains, where the latter is used to separate the impact of heavier trains from the effect of increased capacity utilisation. That is, if more and heavier trains are running on a track section, the train density variable captures its impact on capacity utilisation while the average tonnage density variable captures the impact of higher axle loads. Regarding the impact of train speeds, we have information about the quality class number of a track section, which indicates the maximum speed allowed (higher speeds generally imply more trains per time period, yet this depends on the signalling system; see Nelldal et al. (2009)). However, its impact on capacity can be difficult to isolate from the effect line speed has on the wear and tear of the infrastructure, as well as from effects caused by differences in requirements on track geometry standard. Considering train speed heterogeneity, we do have information on whether the train is a passenger or a freight train. We can therefore (to some extent) capture the effect of traffic homogeneity with respect to speeds. We define this variable as $|\frac{\text{Passenger train-km}}{\text{Total train-km}} - 0.5|$, which thus can take a value on the interval [0, 0.5], where 0 implies a 50-50 mix between passenger and freight traffic, while 0.5 implies that either passenger or freight traffic is the only traffic type on the railway line (i.e. homogeneous traffic).

We consider the timetabling process to be relatively fixed, where any changes over time are due to changes in traffic demand and/or changes in infrastructure characteristics. If this is not the case, i.e. if the timetabling process changes due to factors not captured by our explanatory variables, we might have a problem with omitted variable bias. However, if these are general effects over the railway network, then they can be captured by year dummy variables (the specification of the model is presented below).

3.1 Model

To derive the cost elasticity with respect to traffic and capacity, we use a short run cost function

$$C_{it} = f(Q_{it}, \sum_{k=1}^{3} K_{kit}, \sum_{l=1}^{L} X_{lit}, \sum_{d=1}^{D} Z_{dit}), \tag{2}$$

where C_{it} is maintenance costs in track section i during year t. Q_{it} is the train density, and $\sum_{k=1}^{3} K_{kit}$ is our set of infrastructure capacity measures: track length/route length, number of passing sidings per route-km, and train speed homogeneity ($|\frac{\text{Passenger train-km}}{\text{Total train-km}} - 0.5|$). $\sum_{l=1}^{L} X_{lit}$ are other network characteristics such as track length and quality class (linked to line speed), including average tonnage density of the trains. $\sum_{d=1}^{D} Z_{dit}$ are dummy variables.

To capture the effect of capacity utilisation in the estimation of marginal costs, we need to consider the interaction between traffic and infrastructure capacity (K_{kit}), as well as non-linear effects of traffic. A flexible model that includes these types of effects is the Translog model, which was proposed by Christensen et al. (1971). It is a second order approximation of a cost (production) function (see Christensen and Greene (1976) for an application to cost functions). The cost model we estimate is

$$\begin{aligned}
lnC_{it} = {} & \alpha + \beta_0 lnC_{it-1} + \beta_Q lnQ_{it} + \tfrac{1}{2}\beta_{QQ}(lnQ_{it})^2 + \sum_{k=1}^{3} \beta_k K_{kit} + \\
& \tfrac{1}{2}\sum_{k=1}^{3}\sum_{k=1}^{3} \beta_{kk} lnK_{kit} lnK_{kit} + \sum_{k=1}^{K}\sum_{p=1}^{P} \beta_{kp} lnK_{kit} lnK_{pit} + \sum_{k=1}^{3} \beta_{kQ} lnK_{kit} lnQ_{it} + \\
& \sum_{l=1}^{L} \beta_l lnX_{lit} + \tfrac{1}{2}\sum_{l=1}^{L}\sum_{l=1}^{L} \beta_{ll} lnX_{lit} lnX_{lit} + \sum_{l=1}^{L}\sum_{r=1}^{R} \beta_{lr} lnX_{lit} lnX_{rit} + \sum_{l=1}^{L} \beta_{lQ} lnX_{lit} lnQ_{it} + \\
& \sum_{l=1}^{L}\sum_{k=1}^{3} \beta_{lk} lnX_{lit} lnK_{kit} + \sum_{d=1}^{D} \vartheta_d Z_{dit} + \mu_i + v_{it},
\end{aligned} \tag{3}$$

where α is a scalar, v_{it} the error term, and μ_i is the impact of unobserved track section specific effects. $\beta_Q, \beta_{QQ}, \beta_k, \beta_{kk}, \beta_{kp}, \beta_{kQ}, \beta_l, \beta_{ll}, \beta_{lr}, \beta_{lQ}, \beta_{lk}$, and ϑ_d are parameters to be estimated, and the symmetry restrictions $\beta_{kp} = \beta_{pk}$, $\beta_{lr} = \beta_{rl}$, and $\beta_{lk} = \beta_{kl}$ are used. The Cobb-Douglas constraint $\beta_{QQ} = \beta_{kk} = \beta_{kp} = \beta_{kQ} = \beta_{ll} = \beta_{lr} = \beta_{lQ} = \beta_{lk} = 0$ is tested using and F-test. We use a double-log specification as our functional form, which can reduce heteroscedasticity and skewness (Heij et al. (2004)). This functional form is common in the literature on rail infrastructure costs (see for example Munduch et al. (2002), Link et al. (2008), Wheat and Smith (2008), Odolinski and Nilsson (2017), Odolinski and Wheat (2018)).

We also include lagged maintenance costs (lnC_{it-1}) in the model to capture dynamic effects in the maintenance production; a change in a cost driver (such as traffic) during a year might also have an impact on costs in the subsequent year(s). This effect was for example found by Andersson (2008), Odolinski and Nilsson (2017), and Odolinski and Wheat (2018). The lagged maintenance costs lnC_{it-1} are however correlated with the (time-invariant) individual effects μ_i. We use the forward orthogonal deviation to remove these track section specific effects, a transformation proposed by Arellano and Bover (1995). Moreover, lagged maintenance costs are correlated with the error terms v_{it}. We therefore use instruments for the lagged variables. The best instruments available to us are further lags of the lagged variable(s) (which are not correlated with the error terms v_{it}), where a longer set of lags can improve estimation efficiency. To not lose observations when increasing the number of lags, we use the method by Holtz-Eakin et al. (1988) in which missing values are substituted by zeros when building our set of instruments for each time period. This generates the moment condition $\sum_{i,t} lnC_{i,t-2}\hat{v}_{it} = 0$ (we collapse the set of instruments to one column to restrict the number of instruments and not overfit the endogenous variables – see Roodman (2009) for details).

The estimates $\hat{\beta}_k$, $\hat{\beta}_{kk}$, $\hat{\beta}_{lk}$ and $\hat{\beta}_{kp}$ comprise the effects our infrastructure capacity measures have on costs, while $\hat{\beta}_Q$ and $\hat{\beta}_{QQ}$ capture the impact traffic has on costs. Moreover, the estimate $\hat{\beta}_{kQ}$ captures the cost impact of an increase in traffic when the level of infrastructure capacity increases – that is, it allows us to evaluate the cost elasticity for traffic with respect to different levels of infrastructure capacity, while holding the other variables constant. More specifically, the effect of a change in traffic is

$$\frac{\partial lnC_{it}}{\partial lnQ_{it}} = \hat{\beta}_Q + \hat{\beta}_{QQ} lnQ_{it} + \hat{\beta}_{kQ} lnK_{kit}, \qquad (4)$$

We test the inclusion of interaction terms between the squared capacity and traffic variables – that is, we include $\frac{1}{2}\beta_{kkQ}(lnK_{kit})^2 lnQ_{it}$, $\frac{1}{2}\beta_{QQk}(lnQ_{it})^2 lnK_{kit}$ and $\frac{1}{2}\frac{1}{2}\beta_{QQkk}(lnQ_{it})^2(lnK_{kit})^2$, which implies that we allow the interaction effect between traffic and the infrastructure capacity variables to be non-linear.

With a dynamic model, we can estimate so-called ‘equilibrium cost elasticities’ for traffic, where ‘equilibrium cost’ is used for a situation in which there is no tendency to change maintenance costs, *ceteris paribus* (Odolinski and Wheat (2018)). Hence, the equilibrium cost level is $lnC_{it} = lnC_{it-1} = lnC_{it}^{e}$. Note that this does not need to be an optimal level of maintenance costs, but it is rather the level chosen by the IM (we still consider that it has the objective of minimizing costs with respect to cost drivers such as traffic). Inserting the expression for equilibrium maintenance cost into equation (3), we get

$$lnC_{it}^{e} = \alpha + \beta_0 lnC_{it}^{e} + \beta_Q lnQ_{it} + \frac{1}{2}\beta_{QQ}(lnQ_{it})^2 + \sum_{k=1}^{3}\beta_k K_{kit} +$$

$$\frac{1}{2}\sum_{k=1}^{3}\sum_{k=1}^{3}\beta_{kk} lnK_{kit} lnK_{kit} + \sum_{k=1}^{K}\sum_{p=1}^{P}\beta_{kp} lnK_{kit} lnK_{pit} + \sum_{k=1}^{3}\beta_{kQ} lnK_{kit} lnQ_{it} +$$

$$\sum_{l=1}^{L}\beta_l lnX_{lit} + \frac{1}{2}\sum_{l=1}^{L}\sum_{l=1}^{L}\beta_{ll} lnX_{lit} lnX_{lit} + \sum_{l=1}^{L}\sum_{r=1}^{R}\beta_{lr} lnX_{lit} lnX_{rit} + \sum_{l=1}^{L}\beta_{lQ} lnX_{lit} lnQ_{it} +$$

$$\sum_{l=1}^{L}\sum_{k=1}^{3}\beta_{lk} lnX_{lit} lnK_{kit} + \sum_{d=1}^{D}\vartheta_d Z_{dit} + \mu_i + v_{it}, \qquad (5)$$

which can be expressed as

$$lnC_{it}^{e} = \frac{\alpha}{1-\beta_0} + \frac{\beta_0}{1-\beta_0} lnC_{it}^{e} + \frac{\beta_Q}{1-\beta_0} lnQ_{it} + \frac{1}{2}\frac{\beta_{QQ}}{1-\beta_0}(lnQ_{it})^2 + \sum_{k=1}^{3}\frac{\beta_k}{1-\beta_0} lnK_{kit} +$$

$$\frac{1}{2}\sum_{k=1}^{3}\sum_{k=1}^{3}\frac{\beta_{kk}}{1-\beta_0} lnK_{kit} lnK_{kit} + \sum_{k=1}^{K}\sum_{p=1}^{P}\frac{\beta_{kp}}{1-\beta_0} lnK_{kit} lnK_{pit} + \sum_{k=1}^{3}\frac{\beta_{kQ}}{1-\beta_0} lnK_{kit} lnQ_{it} +$$

$$\sum_{l=1}^{L}\frac{\beta_l}{1-\beta_0} lnX_{lit} +$$

$$\frac{1}{2}\sum_{l=1}^{L}\sum_{l=1}^{L}\frac{\beta_{ll}}{1-\beta_0} lnX_{lit} lnX_{lit} + \sum_{l=1}^{L}\sum_{r=1}^{R}\frac{\beta_{lr}}{1-\beta_0} lnX_{lit} lnX_{rit} + \sum_{l=1}^{L}\frac{\beta_{lQ}}{1-\beta_0} lnX_{lit} lnQ_{it} +$$

$$\sum_{l=1}^{L}\sum_{k=1}^{3}\frac{\beta_{lk}}{1-\beta_0} lnX_{lit} lnK_{kit} + \sum_{d=1}^{D}\frac{\vartheta_d}{1-\beta_0} Z_{dit} + \frac{\mu_i}{1-\beta_0} + \frac{v_{it}}{1-\beta_0}, \qquad (6)$$

The equilibrium cost elasticity for traffic is then

$$\gamma_{it} = \frac{\partial lnC_{it}^{e}}{\partial lnQ_{it}} = \frac{\beta_Q}{1-\beta_0} + \frac{\beta_{QQ}}{1-\beta_0} lnQ_{it} + \sum_{k=1}^{3}\frac{\beta_{kQ}}{1-\beta_0} lnK_{kit}, \qquad (7)$$

3.2 Marginal costs

To calculate marginal costs, we use a fitted cost

$$\hat{C}_{it} = \exp\left(\ln(C_{it}) - \hat{v}_{it} + 0.5\hat{\sigma}^2\right) \qquad (8)$$

which derives from the double-log specification of our model that assumes normally distributed residuals (see Munduch et al. (2002) and Wheat and Smith (2008)). The average cost for train-km is calculated as

$$\widehat{AC}_{it} = \hat{C}_{it}/TKM_{it} \qquad (9)$$

The marginal cost is calculated by multiplying the average cost by the estimated cost elasticities.

$$MC_{it} = \widehat{AC}_{it} \cdot \hat{\gamma}_{it} \tag{10}$$

A weighted marginal cost is calculated for the entire railway network included in this study:

$$MC_{it}^{W} = MC_{it} \cdot \frac{TKM_{it}}{(\sum_{it} TKM_{it})/N} \tag{11}$$

where N is the number of observations in the sample.[5] The weighted marginal cost will generate the same income to the IM as if it would use each observation's marginal cost (eq. 10) for the different track sections.

4. Data

The data has been provided by the Swedish IM and covers a large part of the Swedish railway network during the period 1999 to 2014. Five regional units and a central planning unit within the IM administers the state-owned 14 100 track-km network. Information about the infrastructure is available at different levels of detail. Technical aspects of the tracks, such as rail weight, type of sleeper and quality class are provided for segments of the track that can be shorter than 100 meters, whereas information on costs is available for track sections of the network that comprise 3 to 290 track-km. In total, there are about 250 track sections during the period 1999-2014 (there are changes where sections merge, as well as splitting into new sections). Our dataset does however not include

---

[5] Munduch et al. (2002) and Andersson (2008) use a different expression for weighted marginal costs ($MC^{W} = \sum_{it} MC_{it} \cdot \frac{TKM_{it}}{(\sum_{it} TKM_{it})}$), which generates the same value as the average value of equation (11). Using equation (11), we can provide average values for different parts of the railway network with respect to capacity utilisation.

all sections, partly due to missing information, and partly due to the exclusion of marshalling yards, sections closed for traffic, and heritage railways. Moreover, we exclude so called stations sections in our analysis, i.e. sections that have a short route length but many parallel tracks. The reason is that the traffic structure is different compared to most other track sections as these station sections are not only used for overtaking or crossing, but can also be used for shunting, changing locomotives, as well as starting or terminating train services (UIC (2013)). In total, we observe on average 164 track sections per year during 1999-2014, comprising on average 11 936 km, which is the majority of the state-owned railway network. Descriptive statistics of our dataset are presented in Table 1.

Information on the technical characteristics of the infrastructure has been collected from the track information system 'BIS' administered by the IM. As mentioned above, this information is available at a more disaggregate level than the cost data, which means that we use weighted averages (with track lengths as weighting factors) of variables such as rail weight and quality class in the model estimations made at the track section level. Traffic data has been collected from the IM and comprise information on train-km and the gross tonnage of the trains reported by the train operators. We use a density measure for trains that is calculated as train-km/route-km and can be described as the average number of trains that have run on the entire route length of the section. We also use an average tonnage density measure, calculated as tonne-km per train-km on a section and year.

The maintenance cost data include costs for all activities conducted to maintain the rail infrastructure, including snow removal, inspections and minor replacements. Specifically, it includes maintenance of all the infrastructure assets, i.e. tracks (sub- and superstructure), electrification, signalling, and telecommunications. Major replacements are defined as renewals and are not included in this analysis as it has a different data generating process and requires a different model approach; see for example Andersson et al. (2012), Andersson et al. (2016) and Odolinski and Wheat (2018) who use corner solution models, survival analysis and vector autoregressive models, respectively.

Table 1. Descriptive statistics, track sections, 1999-2014 (2619 observations)

	Median	Mean	St. dev.	Min	Max
Maintenance cost, million SEK in 2014 prices	8.30	11.71	11.37	0.01	110.75
Traffic and line capacity variables					
Train-km, thousand	474	763	875	0	4 778
Tonne-km, thousand	169 144	390 953	536 092	1	4 176 261
Train density (Train-km/Route length), thousand	11	16	19	0	146
Average tonnage density (Tonne-km/Train-km)	406	523	493	52	6 011
Dev. from 50-50 mix in traffic (Deviation from 50-50 mix between passenger train and freight trains)	0.36	0.32	0.16	0.00	0.50
Track length/Route length (Average number of tracks)	1.006	1.260	0.463	1.000	3.517
No. of passing sidings	5.00	6.65	5.81	1.00	40.00
No. of passing sidings per route length	0.11	0.14	0.12	0.01	1.11
Infrastructure characteristics and weather					
Route length, km	47.21	59.26	43.41	0.97	258.10
Track length, km	59.65	72.92	51.69	3.18	290.65
Switch length, km	1.18	1.53	1.34	0.06	9.07
Rail weight, average kg of one meter rail	49.98	51.14	5.03	39.86	60.00
Share of track length with concrete sleepers	0.84	0.62	0.39	0.00	1.00
Share of track length with wooden sleepers)	0.16	0.38	0.39	0.00	1.00
Share of track length with slab track	0.00	0.00	0.00	0.00	0.00
Max axle load, tonnes	22.50	23.09	1.75	16.00	30.00
Average quality class number, 1-6	3.00	2.99	1.23	1.00	6.00
Snow (mm precipitation when temperature <0° Celsius)	98	111	64	2	344
Organisational dummy variables					
Tendered in competition	0.00	0.48	0.50	0.00	1.00
Mix tendered and not tendered in competition	0.00	0.06	0.24	0.00	1.00
Sections in region West	0.00	0.18	0.39	0.00	1.00
Sections in region North	0.00	0.13	0.33	0.00	1.00
Sections in region Central	0.00	0.19	0.39	0.00	1.00
Sections in region South	0.00	0.26	0.44	0.00	1.00
Sections in region East	0.00	0.24	0.43	0.00	1.00

Starting in 2002, maintenance was gradually exposed to competitive tendering. Odolinski and Smith (2016) found that this reduced costs by about 11 per cent. To control for the impact tendering had on maintenance costs, we include dummy variables indicating when a track section belongs to an area tendered in competition.

Sweden is a large country with climate differences, especially between the northern and southern parts. This can have an impact on the maintenance production, especially since snow removal is included in this study. We have therefore collected weather data from the Swedish Meteorological and Hydrological Institute (SMHI), comprising information on daily mean temperatures and mm of precipitation. We define a variable for snow as mm of precipitation when the daily mean temperature is below 0 degrees Celsius.

5. Results

The dynamic model is estimated with the generalized method of moments (GMM), where we use the two-step System GMM, an approach proposed by Arellano and Bover (1995) and Blundell and Bond (1998). The variables in our model have been divided by their sample median prior to taking a logarithmic transformation. In that way the first order coefficients can be interpreted as elasticities at the sample median. However, the dummy variables and the variables for sleeper type have not been log-transformed. The percentage change in costs from a change in these variables is therefore calculated as $\frac{\Delta C}{C} = 100 \cdot [\exp(\hat{\beta}_l \Delta X_l) - 1]$, where $\hat{\beta}_l$ is the estimated coefficient for variable X_l.

The estimation results are presented in section 5.1 below. The Windmeijer (2005) correction of the variance-covariance matrix is used to avoid downward biased standard errors. All estimations are carried out using Stata 12 (StataCorp, 2011).

5.1 Estimation results

The estimation results are presented in Table 2. First, we can note that the coefficient for lagged maintenance costs is positive and statistically significant, which is in line with the results in previous studies on long panel data sets (see Wheat (2015), Odolinski and Nilsson (2017), and Odolinski and Wheat (2018)).[6] Hence, an increase in a cost driver in year $t-1$ will have an impact on maintenance costs in year t; the IM is not able to adjust its maintenance cost level within the current year after a sudden change in a cost driver. Furthermore, we note that the first order coefficients for track length, switch length, rail weight, snow, and sleeper type (concrete sleepers, with wooden sleepers as baseline), have the expected signs.[7] However, the estimates for snow and concrete sleepers are not statistically significant.

The quality classification of the railway line (linked to line speed) can be an important factor for maintenance costs (higher speeds may increase wear and tear and are also linked to stricter requirements on track geometry etc.). However, its coefficient in the estimations is small and not statistically significant. Here we can note that the correlation coefficients between this variable and the number of tracks and train density are -0.56, and -0.42, respectively. Dropping the quality classification variable does not change the estimations results significantly.

Turning to the first order coefficients for the line capacity variables, we can see that the coefficient for the average number of tracks (no. of tracks) has a positive sign, yet its second order effect is negative. The average cost elasticity with respect to the number of tracks – evaluated at the sample median of the other variables – is -0.3828 (standard error 0.1602 and p-value 0.018).

---

[6] Andersson (2008) found a negative impact using a much shorter panel (years 1999 to 2002).

[7] Note that newer rails are usually heavier. Rail weight therefore picks up the impact track age has on costs to some extent (a variable for the age of the tracks was also considered in the estimations, but it did not have an impact on the results when rail weight was also included).

That is, increasing the number of tracks on a line lowers maintenance costs at the sample median, *ceteris paribus,* which is in line with the hypothesis in this paper as more tracks imply higher line capacity available. The coefficient for the number of passing sidings per route-km is positive, yet the estimate is small and not statistically significant. Moreover, the interaction terms between passing sidings per route-km and traffic were not statistically significant and were therefore dropped from the model. The parameter estimate for the level of traffic mix is is positive and not statistically significant.

Turning to the impact of traffic, the first order coefficient for train density is 0.2601 and the second order coefficient is 0.0342 (both statistically significant), which shows that the effect of train density is increasing. Specifically, this indicates that running one more train is costlier on tracks with higher traffic volume, i.e. with higher capacity utilisation. The overall equilibrium cost elasticity (see equations 5-7) with respect to train density is 0.25 (including the impacts from the interactions with average number of tracks) and statistically significant at the 1 per cent level.[8] The estimate shows that we have considerable economies of density, where a 10 per cent increase in train density generates a 2.5 per cent increase in maintenance costs. Importantly, this implies that track sections with a higher traffic density have lower average costs, i.e. cost per train-km. The cost elasticity with respect to traffic is in line with estimates in the literature on rail infrastructure costs (see Link et al. (2008) and Wheat et al. (2009)). Moreover, the average tonnage density estimate is

---

[8] The overall cost elasticity with respect to train density without the impact from lagged maintenance costs is 0.1967. Considering that the estimate for lagged maintenance costs is 0.2144, we have a rather fast adjustment period: the cost impact from train density left in the next year is (0.1967*0.2144=) 0.0422, and in the subsequent year it is (0.1967*0.2144^2=) 0.0090. Note that the equilibrium cost elasticity includes the effects from the entire adjustment period, and is 0.1967/(1-0.2144)=0.2504.

0.0492 (p-value 0.100), indicating that increasing the average weight of the trains will increase maintenance costs.

Table 2. Econometric results

	Coef.	Corr. Std. Err.	[95% Conf.	Interval]
Constant	12.4495***	0.8936	10.6869	14.2122
ln(maintenance cost t-1)	0.2144***	0.0560	0.1040	0.3248
ln(track length)	0.6214***	0.0600	0.5030	0.7399
ln(track length)^2	0.0851	0.0949	-0.1022	0.2724
ln(switch length)	0.1226***	0.0445	0.0348	0.2104
ln(switch length)^2	0.0835	0.0530	-0.0211	0.1881
ln(rail weight)	-0.5084*	0.2611	-1.0235	0.0067
ln(average quality class)	0.0357	0.0439	-0.0509	0.1222
ln(max axle load)	-0.2200	0.3061	-0.8238	0.3838
ln(max axle load)^2	7.4073***	2.7776	1.9283	12.8863
Share of concrete sleepers	-0.0860	0.0607	-0.2057	0.0336
Share of slab sleepers	191.6174**	80.0568	33.6977	349.5371
ln(passing sidings per route length)	0.0061	0.0342	-0.0614	0.0737
Deviation from 50-50 mix in traffic	0.0098	0.0097	-0.0093	0.0288
ln(no. of tracks)	0.1151	0.2676	-0.4128	0.6430
ln(no. of tracks)^2	-1.1746**	0.5359	-2.2317	-0.1175
ln(average tonnage density]	0.0492*	0.0297	-0.0095	0.1078
ln(train density)	0.2601***	0.0336	0.1938	0.3263
ln(train density)^2	0.0342***	0.0118	0.0108	0.0575
ln(train density)ln(no. of tracks)	-0.4479**	0.1826	-0.8082	-0.0877
ln(train density)[ln(no. of tracks)^2]	1.4072***	0.4617	0.4964	2.3180
[ln(train density)^2]ln(no. of tracks)	0.0379	0.1016	-0.1624	0.2383
[ln(train density)^2][ln(no. of tracks)^2]	-0.4478*	0.2313	-0.9041	0.0085
ln(track length)ln(no. of tracks)	-0.2093*	0.1196	-0.4453	0.0267
ln(track length)ln(switch length)	-0.0682	0.0688	-0.2040	0.0677
ln(track length)ln(max axle load)	0.0716	0.2976	-0.5155	0.6586
ln(no. of tracks)ln(switch length)	0.1703	0.1122	-0.0510	0.3916
ln(no. of tracks)ln(max. axle load)	0.5585	0.8069	-1.0332	2.1501
ln(switch length)ln(max axle load)	-0.3029	0.1943	-0.6861	0.0803
ln(snow)	0.0380	0.0253	-0.0119	0.0878
Mix tendered in competition, dummy	-0.0306	0.0378	-0.1052	0.0441
Tendered in competition, dummy	-0.1178***	0.0358	-0.1884	-0.0473
Year dummies	Yes[a]	-	-	-
Region dummies	Yes[b]	-	-	-

***, **, *: Significance at 1%, 5%, 10% level, [a] Jointly significant (F(14, 189)=11.85, Prob>F=0.000), [b] Jointly significant (F(4, 189)=4.98, Prob>F=0.001). Test of Cobb-Douglas constraint: F(15, 189)=4.67, Prob>F=0.000. No. of instruments: 64.

To evaluate how the cost elasticities with respect to traffic vary with capacity utilisation, we turn to the coefficients for the interaction terms between the traffic and the infrastructure capacity variables. The parameter estimate for the interaction between train density and number of tracks (ln(train density)ln(no. of tracks)) is -0.4479 and statistically significant at the 5 per cent level. The interpretation of the coefficients is that the cost elasticity with respect to traffic is decreasing with the degree of infrastructure capacity, as measured by the average number of tracks, which is in line with the hypothesis in this paper. Note that we also have an interaction between the squared variable for the number of tracks and the traffic variable (ln(train density)[ln(no. of tracks)^2]), which is positive – thus, the negative impact this capacity measure has on the cost elasticity for traffic diminishes. We also have interaction terms between squared traffic and the number of tracks ([ln(train density)^2]ln(no. of tracks)) and between squared traffic and the squared number of tracks ([ln(train density)^2][ln(no. of tracks)^2]), where the former is positive (0.0379) and not statistically significant and the latter is negative (-0.4478) and statistically significant. These estimates imply that the positive second order effect of traffic diminishes (and turns negative) when the number of tracks increases.

The impact of these estimates can be seen in Figure 1 below, where track sections with an average number of tracks in the interval [1.00, 1.75) have cost elasticities that increase with the traffic volume (capacity utilisation), whereas track sections with more tracks (interval at [1.75, 3.52]), have decreasing cost elasticities with respect to traffic. In general, cost elasticities are higher when there are fewer tracks available for a certain traffic volume (comparing the elasticities between different intervals of average number of tracks at a certain point on the x-axis in Figure 1) – that is, when capacity utilisation is higher. One exception is the comparison between the highest intervals [1.13, 1.75) and [1.75, 3.52] at the lower levels of traffic volume, which indicates that there may be differences in maintenance production strategies (activities) that are not captured by our explanatory variables. Moreover, the decreasing cost elasticities for sections with an average number of tracks above 1.75 suggests that these sections have a relatively low capacity utilisation, making a traffic increase less costly compared to the other track sections. In other words, receiving cost efficient time slots for maintenance does not seem to be a problem for the current traffic levels on these sections. And indeed, for higher levels of traffic, these sections have lower cost elasticities than sections with fewer tracks (i.e. with higher capacity utilisation).

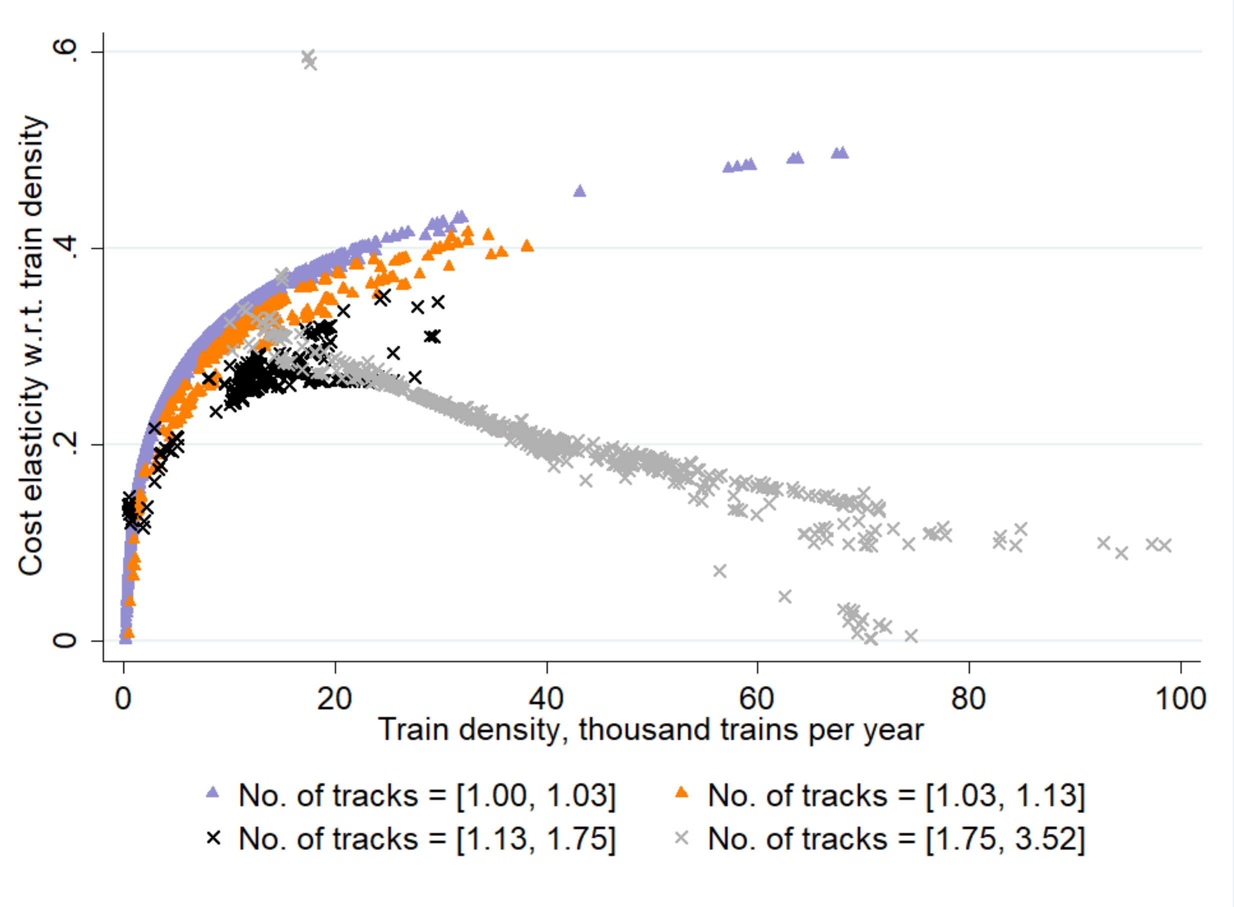

Figure 1. Cost elasticities with respect to train density[9]

---

[9] This figure excludes 55 negative cost elasticities with respect to traffic.

The cost elasticities are summarized in Tables 3 and 4, in which we have grouped the observations based on traffic volume and the average number of tracks. A comparison of average cost elasticities going from the top left to the bottom right of Tables 3 and 4 (excluding sections with tracks in the interval [1.75, 3.53]) shows that elasticities are increasing with capacity utilisation as shown by Figure 1. Moreover, even though the results for sections with the highest number of tracks ([1.75, 3,52]) stand out, for high traffic volumes they are still in line with the general pattern of increasing cost elasticities at higher levels of capacity utilisation as measured by the number of tracks (see Table 4).

Table 3. Average cost elasticities with respect to different levels of capacity utilisation, train density intervals [0K, 5K), [5K, 10K) and [10K, 20K).

	Elasticity	**Obs.**	**Elasticity**	**Obs.**	**Elasticity**	**Obs.**
No. of tracks	*(Train density [0K, 5K))*		*(Train density [5K, 10K))*		*(Train density [10K, 20K))*	
[1.75, 3.52]	0.72	1	-	0	0.38	56
[1.13, 1.75)	0.16	24	0.24	7	0.28	131
[1.03, 1.13)	0.16	36	0.27	72	0.33	92
[1.00, 1.03)	0.18	576	0.30	369	0.35	452

Table 4. Average cost elasticities with respect to different levels of capacity utilisation, train density intervals [20K, 30K), [30K, 40K), and [40K, 146K].

	Elasticity	**Obs.**	**Elasticity**	**Obs.**	**Elasticity**	**Obs.**
No. of tracks	*(Train density [20K, 30K))*		*(Train density [30K, 40K))*		*(Train density [40K, 146K])*	
[1.75, 3.52)	0.26	86	0.22	108	0.15	217
[1.13, 1.75)	0.28	31	-	0	-	0
[1.03, 1.13)	0.38	30	0.40	11	-	0
[1.00, 1.03)	0.40	41	0.43	5	0.48	11

5.2 Marginal costs

We calculate the marginal costs by multiplying the estimated equilibrium cost elasticities with the average costs, as described in section 3.2. The average cost, marginal cost and the weighted marginal cost for the entire sample are presented in Table 5. We can note that the charge used by the Swedish IM in 2018 is about SEK 0.0107 per gross tonne-km (in 2014 prices), which covers both maintenance and renewal costs. The charge for maintenance is about SEK 0.0051 per gross tonne-km, considering that maintenance covers 47 per cent of the cost according to the marginal cost estimates that the charge is based on. Using the average tonnage density in our data (407 tonnes), the maintenance cost charge corresponds to SEK 2.07 per train-km, which is significantly smaller than our estimated weighted marginal cost at SEK 3.96 per train-km. This difference is bigger for higher levels of capacity utilisation (see Tables 6 and 7 below).

Table 5. Average costs, marginal costs and weighted marginal costs, SEK per train-km

Variable	Mean	Std. Err.	[95 % Conf. Interval]	
Average cost	44.2151	2.3482	39.6104	48.8198
Marginal cost	8.1151	0.3123	7.5026	8.7275
Weighted marginal cost	3.9647	0.0728	3.8220	4.1074

To evaluate the impact that capacity utilisation has on marginal costs, we plot these costs against traffic volume and differentiate with respect to the average number of tracks on a section. See Figure 2 below, where the observations in the figure correspond to a marginal cost for each track section (i) in each year (t). Specifically, Figure 2 shows that marginal costs fall sharply with train density, which is similar to the shapes presented in for example Wheat et al. (2009). This is expected, considering that the cost elasticities with respect to train density are below 1, indicating economies of density with decreasing average costs. Still, the marginal cost per train-km are generally higher for track sections with a lower average number of tracks – that is, when comparing the marginal costs at certain point on the x-axis. However, there are observations that are not in line with this general pattern.

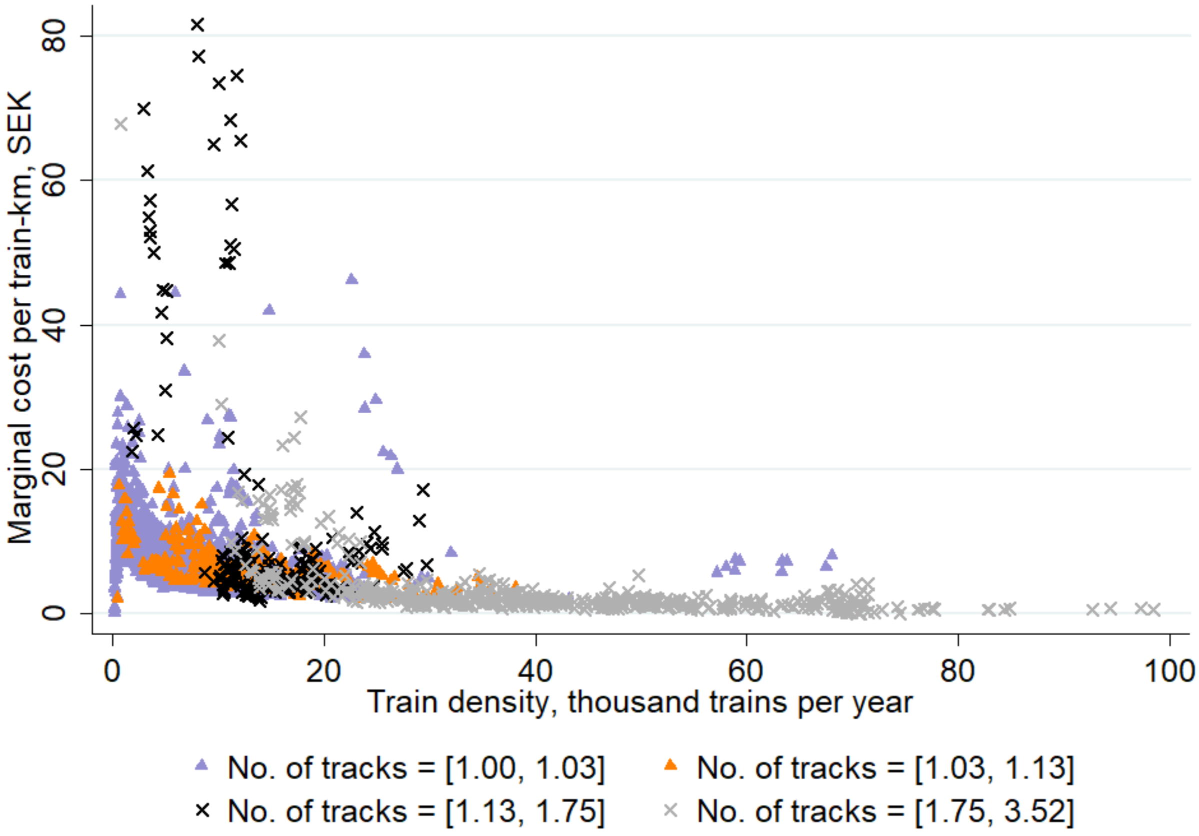

Figure 2. Marginal cost per train-km, SEK

The differences in costs with respect to capacity utilisation are slightly more apparent in Tables 6 and 7, in which we present weighted marginal costs that have been grouped based on traffic volume and the average number of tracks. The weighted marginal costs are mostly increasing with capacity utilisation (going from the top to the bottom of the table, i.e. comparing sections with different number of tracks), except when comparing costs with the highest interval [1.75, 3.52]. As the marginal costs are weighted with traffic, the sections with highest traffic volumes have lower marginal costs (due to economies of density). Specifically, going from left to right in Table 6 shows an increasing weighted marginal cost, whereas Table 7 (with the highest traffic volumes) shows a decreasing weighted marginal cost with traffic volume.

To sum up: holding train density constant, a comparison of sections based on the number of tracks shows that weighted marginal costs are generally increasing with capacity utilisation. Holding the number of tracks constant, a comparison between traffic volumes shows that weighted marginal costs are increasing with capacity utilisation, up to a train density at about twenty thousand per year. For higher levels of traffic, the weighted marginal cost is decreasing.

Table 6. Weighted marginal costs per train-km (SEK) with respect to capacity utilisation, train density intervals [0K, 5K), [5K, 10K) and [10K, 20K).

No. of tracks	**WMC**[a] *(Train density [0K, 5K))*	Obs.	**WMC**[a] *(Train density [5K, 10K))*	Obs.	**WMC**[a] *(Train density [10K, 20K))*	Obs.
[1.75, 3.52)	**0.70** (-) [-, -]	1	-	0	**4.54** (0.30) [3.93, 5.15]	56
[1.13, 1.75)	**1.28** (0.20) [0.88, 1.69]	24	**2.86** (0.53) [1.58, 4.14]	7	**4.16** (0.21) [3.74, 4.58]	131
[1.03, 1.13)	**1.42** (0.13) [1.14, 1.69]	36	**2.98** (0.29) [2.41, 3.55]	72	**5.82** (0.35) [5.12, 6.53]	92
[1.00, 1.03)	**1.87** (0.07) [1.74, 2.01]	576	**4.40** (0.19) [4.02, 4.78]	369	**5.84** (0.22) [5.40, 6.28]	452

[a] *Standard errors are in parentheses and 95 per cent confidence intervals in brackets.*

Table 7. Weighted marginal costs per train-km (SEK) with respect to capacity utilisation, train density intervals [20K, 30K), [30K, 40K), and [40K, 146K].

No. of tracks	**WMC**[a] *(Train density [20K, 30K))*	Obs.	**WMC**[a] *(Train density [30K, 40K))*	Obs.	**WMC**[a] *(Train density [40K, 146K])*	Obs.
[1.75, 3.52)	**5.80** (0.36) [5.09, 6.52]	86	**5.72** (0.31) [5.11, 6.33]	108	**3.21** (0.16) [2.89, 3.52]	217
[1.13, 1.75)	**4.47** (0.35) [3.75, 5.19]	31	-	0	-	0
[1.03, 1.13)	**4.47** (0.49) [3.47, 5.48]	30	**3.61** (0.27) [3.00, 4.22]	11	-	0
[1.00, 1.03)	**4.76** (0.37) [4.00, 5.51]	41	**2.79** (0.98) [0.06, 5.52]	5	**1.13** (0.17) [0.75, 1.52]	11

[a] *Standard errors are in parentheses and 95 per cent confidence intervals in brackets.*

6. Conclusions

Capacity utilisation can have an impact on the possibilities to maintain the railway, where for example the time slots for maintenance activities may be short and fragmented when capacity utilisation is high, which can increase production costs. This paper shows how differences in line capacity utilisation influence marginal maintenance costs for rail infrastructure usage, which is a contribution to the literature as previous studies have focused on the wear and tear caused by traffic. Specifically, the estimation results show that cost elasticities with respect to traffic are increasing with capacity utilisation, where fewer tracks on a line and/or a higher traffic level imply higher maintenance costs. One exception is the sections with the highest number of tracks, which have cost elasticities that decrease with higher traffic levels, suggesting that their capacity utilisation is relatively low and that one more train will not affect the maintenance production in this aspect (for example, cost efficient time slots for maintenance are still available). The weighted marginal costs per train-km – calculated as the product between average costs and the cost elasticities – have a similar pattern as the cost elasticities. However, due to economies of density, these marginal costs eventually decrease with higher traffic volumes.

The results are significant for future studies on marginal maintenance costs of rail infrastructure usage – that is, these studies need to recognize that high capacity utilisation may have an impact on possession times for maintenance, and that highly utilised tracks are more sensitive to delays, which can require more (preventive) maintenance. In this, we acknowledge that our results indicate that sections with the highest number of tracks have a cost structure that differs from the rest of the network, which needs to be investigated further in future research. Still, including the impact of capacity utilisation in the marginal cost estimation is important, especially since track access charges can be based on marginal costs, and that several countries set their charges based on econometric studies (examples are France, Sweden and Switzerland). Setting charges based on marginal costs that are differentiated with respect to capacity utilisation may well change the behaviour of the operators, and thus lead to a more efficient use of the infrastructure.

Acknowledgements

The authors are grateful to the Swedish Transport Administration for funding this research. Special thanks to Vivianne Karlsson, Anders Nilsson and Anne-Lie Mathisen for providing data. We also wish to thank Jan-Eric Nilsson, Per Köhler and Pär-Erik Westin for providing helpful comments. All remaining errors are the responsibility of the authors.

References

Andersson, M., 2008. Marginal railway infrastructure costs in a dynamic context. EJTIR, 8, 268-286.

Andersson, M., 2011. Marginal cost of railway infrastructure wear and tear for freight and passenger trains in Sweden. Eur. Transp./Transp. Eur. 48, 3–23.

Andersson, M., Smith, A., Wikberg, Å., Wheat, P., 2012. Estimating the marginal cost of railway track renewals using corner solution models. Transportation Research Part A, 46, 954-964. DOI: https://doi.org/10.1016/j.tra.2012.02.016

Andersson, M., Björklund, G., Haraldsson, M., 2016. Marginal railway track renewal costs: A survival data approach. Transportation Research Part A: Policy and Practice, 87, 68-77. DOI: https://doi.org/10.1016/j.tra.2016.02.009

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics. 68, 29-51. DOI: https://doi.org/10.1016/0304-4076(94)01642-D

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87, 115-143. DOI: https://doi.org/10.1016/S0304-4076(98)00009-8

Booz, Allen & Hamilton, TTCI, 2005. Review of Variable Usage and Electrification Asset Usage Charges: Final Report. Report R00736a15, London, June 2005.

Boysen, H., Andersson E., 1989. Track-friendly bogies for high speeds and heavy payloads. 4th International Heavy Haul Railway Conference, Brisbane.

Boysen, H., 2012. General model of railway transportation capacity. 13th International Conference on Design and Operation in Railway Engineering (COMPRAIL), New Forest.

Boysen, H., 2013. Quicker meets, heavier loads and faster empties – Effects on transportation capacity and cycle time. 10th International Heavy Haul Association Conference, New Delhi.

Brännlund, U., Lindberg, P.O., Nou, A., Nilsson, J-E., 1998. Railway Timetabling using Lagrangian Relaxation. Transportation Science, 32(4), 358-369. DOI: https://doi.org/10.1287/trsc.32.4.358

Christensen, L.R., Greene, W.H., 1976. Economies of scale in U.S. electric power generation. Journal of Political Economy, 84(4), 655-676. DOI: https://doi.org/10.1086/260470

Christensen, L. R., D. W. Jorgenson, och L. J. Lau (1971): 'Conjugate Duality and the Transcendental Logarithmic Production Function', *Econometrica*, 39(4), 225-256.

Gaudry, M., Quinet, E., 2009. CATRIN (Cost Allocation of TRansport INfrastructure cost), Deliverable 8, Rail Cost Allocation for Europe – Annex 1Di – Track Maintenance Costs in France. Funded by Sixth Framework Programme. VTI, Stockholm.

Heij, C., de Boer, P., Franses, P.H., Kloek, T., van Dijk, H.K., 2004. Econometric Methods with Applications in Business and Economics. Oxford University Press Inc., New York.

Holtz-Eakin, D., Newey, W., Rosen, H. S., 1988. Estimating vector autoregressions with panel data. Econometrica. 56(6), 1371-1395. DOI: https://doi.org/10.2307/1913103

Johansson, P., Nilsson, J-E., 2004. An economic analysis of track maintenance costs. Transport policy, 11, 277-286. DOI: https://doi.org/10.1016/j.tranpol.2003.12.002

Lidén, T., 2014. Survey of railway maintenance activities from a planning perspective and literature review concerning the use of mathematical algorithms for solving such planning and scheduling problems. Report, October 13, 2014. DiVA, id: diva2:754942

Lidén, T., 2015. Railway infrastructure maintenance – a survey of planning problems and conducted research. *Transportation Research Procedia* 10, 574-583. DOI: https://doi.org/10.1016/j.trpro.2015.09.011

Lidén, T., 2016. Towards concurrent planning of railway maintenance and train services. Linköping Studies in Science and Technology. Thesis No. 1746, Licentiate Thesis. Department of Science and Technology, Linköping University.

Lidén, T., and Joborn, M., 2016. Dimensioning windows for railway infrastructure maintenance: Cost efficiency versus traffic impact. *Journal of Rail Transport Planning & Management*, 6, 32-47. DOI: https://doi.org/10.1016/j.jrtpm.2016.03.002

Lindfeldt, A. 2014. Kapacitetsutnyttjande i det svenska järnvägsnätet: Uppdatering och analys av utvecklingen 2008 – 2012. Rapport, KTH, Stockholm. (In Swedish)

Lindfeldt, A. 2015. Railway capacity analysis – Methods for simulation and evaluation of timetables, delays and infrastructure. Doctoral Thesis in Infrastructure. KTH Royal Institute of Technology, School of Architecture and the Built Environment, Department of Transport Science, Stockholm 2015.

Link, H., Stuhlehemmer, A., Haraldsson, M., Abrantes, P., Wheat, P., Iwnicki, S., Nash, C., Smith, A.S.J., 2008. CATRIN (Cost Allocation of TRansport INfrastructure cost).

Deliverable D 1, Cost allocation Practices in European Transport Sector. VTI, Stockholm (March 2008).

Lusby, M.R., Larsen, J., Ehrgott, M., and Ryan, D., 2011. Railway track allocation: models and methods. *OR Spectrum*, 33, 843-883. DOI: https://doi.org/10.1007/s00291-009-0189-0

Marti, M., Neuenschwander, R., Walker, P., 2009. CATRIN (Cost Allocation of TRansport INfrastructure cost), Deliverable 8 – Rail Cost Allocation for Europe – Annex 1B – Track maintenance and renewal costs in Switzerland. VTI, Stockholm.

Munduch, G., Pfister, A., Sögner, L., Siassny, A., 2002. Estimating Marginal Costs for the Austrian Railway System. Vienna University of Economics & B.A., Working Paper No. 78, February 2002.

Nash, C., 2003. Marginal cost and other pricing principles for user charging in transport: a comment. Transport Policy, 10, 345-348. DOI: https://doi.org/10.1016/j.tranpol.2003.09.004

Nash, C., 2018. Track access charges: reconciling conflicting objectives. Project Report. CERRE, Centre on Regulation in Europe, Brussels, 9 May 2018.

Nelldal, B-L., A. Lindfeldt, and O. Lindfeldt (2009): 'Kapacitetsanalys av järnvägsnätet i Sverige. Delrapport 1. Hur många tåg kan man köra? En analys av teoretisk och praktisk kapacitet'. Rapport, KTH Järnvägsgruppen. (In Swedish)

Nilsson, J-E., 2002. Towards a welfare enhancing process to manage railway infrastructure access. Transportation Research Part A, 36, 419-436. DOI: https://doi.org/10.1016/S0965-8564(01)00011-8

Nilsson, J-E., Lindberg, P.O., and Aronsson, M., 2017. Optimering och tidtabelläggning. Slutrapportering av förstudie. VTI notat 24-2017. (In Swedish).

Nilsson, J-E., Pyddoke, R., Karlsson, R., and Johansson, A., 2015. Tidtabelläggning. Principer, tumregler och utfall. VTI rapport 880. (In Swedish).

Odolinski, K. 2016. The impact of axle loads on rail infrastructure maintenance costs. CTS Working Paper 2016:21. Centre for Transport Studies, Stockholm.

Odolinski, K. and Nilsson, J-E. 2017. Estimating the marginal maintenance cost of rail infrastructure usage in Sweden; does more data make a difference? Economics of Transportation, 10, 8-17. DOI: https://doi.org/10.1016/j.ecotra.2017.05.001

Odolinski, K. and Smith, A.S.J., 2016. Assessing the cost impact of competitive tendering in rail infrastructure maintenance services: evidence from the Swedish Reforms (1999 to 2011). Journal of Transport Economics and Policy, 50(1), 93-112.

Odolinski, K., and Wheat, P. 2018. Dynamics in rail infrastructure provision: Maintenance and renewal costs in Sweden. Economics of Transportation (In Press). DOI: https://doi.org/10.1016/j.ecotra.2018.01.001

Rail Delivery Group, 2014. Charges and Incentives User Guide. June 2014.

Roodman, D., 2009. A note on the theme of too many instruments. Oxford Bulletin of Economics and Statistics, 71(1), 135-158. DOI: https://doi.org/10.1111/j.1468-0084.2008.00542.x

Rothengatter, W., 2003. How good is first best? Marginal cost and other pricing principles for user charging in transport. Transport Policy, 10, 121-130. DOI: https://doi.org/10.1016/S0967-070X(02)00063-X

Smith, A., Iwnicki, S., Kaushal, A., Odolinski, K., Wheat, P., 2017. Estimating the relative cost of track damage mechanisms: combining economic and engineering approaches. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231(5), 620-636. DOI: https://doi.org/10.1177/0954409717698850

StataCorp, 2011. Stata Statistical Software: Release. StataCorp LP, College Station, TX.

Trafikverket, 2015. Network Statement 2017. Annex 4 B – Priority criteria. Edition 2015-12-10.

UIC, 2013. UIC Code 406, Capacity. 2nd edition, June 2013, Original, UIC, International Union of Railways.

Wheat, P., Smith, A.S.J., 2008. Assessing the marginal infrastructure maintenance wear and tear costs for Britain's railway network. J. Transp. Econ. Policy 42 (2), 189–224.

Wheat, P., Smith, A.S.J., Nash, C., 2009. CATRIN (Cost Allocation of Transport Infrastructure cost). Deliverable 8 – Rail Cost Allocation for Europe. VTI, Stockholm.

Wheat, P., 2015. The Sustainable Freight Railway: Designing the Freight Vehicle-track System for Higher Delivered Tonnage with Improved Availability at Reduced Cost SUSTRAIL, Deliverable 5.3: Access Charge Final Report Annex 4, British Case Study.

Windmeijer, F., 2005. A finite sample correction for the variance of the linear efficient two-step GMM estimators. Journal of Econometrics, 126(1), 25-51. DOI: https://doi.org/10.1016/j.jeconom.2004.02.005

Öberg, J., Andersson, E., Gunnarsson, J., 2007. Track Access Charging with Respect to Vehicle Characteristics, Second edition. Rapport LA-BAN 2007/31, Banverket.