ESTIMATING THE MARGINAL COST OF RAILWAY TRACK RENEWALS USING CORNER SOLUTION MODELS

Mats Andersson*, Andrew Smith**[1], Åsa Wikberg*, Phillip Wheat**

* Swedish National Road and Transport Research Institute, Department of Transport Economics, Box 920, 781 29 Borlänge, Sweden
** Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

Abstract

Economic theory advocates marginal cost pricing for efficient utilisation of transport infrastructure. A growing body of literature has emerged on the issue of rail marginal infrastructure wear and tear costs, but the majority of the work is focused on costs for infrastructure maintenance. Railway track renewals are a substantial part of an infrastructure manager's budget, but in disaggregated statistical analyses they cause problems for traditional regression models since there is a piling up of values of the dependent variable at zero. Previous econometric work has sought to circumvent the problem by aggregation in some way. In this paper we instead apply corner solution models to disaggregate (track-section) data, including the zero observations. We derive track renewal cost elasticities with respect to traffic volumes and in turn marginal renewal costs using Swedish railway renewal data over the period 1999 to 2009. This paper is the first attempt in the literature to apply corner solution models, and in particular the two-part model, to disaggregate renewal cost data in railways. It is also the first paper that we are aware of to report usage elasticities specifically for renewal costs and therefore adds important new evidence to the previous literature where there is a paucity of studies on renewals and considerable uncertainty over the effects of rail traffic on renewal costs. In the Swedish context, we find that the inclusion of marginal track renewal costs in the track access pricing regime, which currently only reflects marginal maintenance costs, would add substantially to the existing track access charge. This change would also increase the cost recovery ratio of the Swedish infrastructure manager, thus meeting a policy objective of government.

---

[1] Corresponding author: a.s.j.smith@its.leeds.ac.uk; +44 (0) 113 34 36654

1. INTRODUCTION

Marginal cost pricing of transport infrastructure wear and tear is of great importance from an economic efficiency standpoint. Over the last decade, research on the subject has gradually increased for all modes of transport (Nash and Sansom, 2001; Nash, 2003; Thomas et al., 2003; Nash and Matthews, 2005; Wheat et. al., 2009). One of the reasons for the renewed interest in the marginal cost of rail infrastructure costs has been the move in European railways towards vertical separation of rail infrastructure from train operations, driven by successive European legislation. This legislation requires countries to set rail infrastructure charges based on the direct cost of running different services, including: additional "wear and tear" costs of running more trains; scarcity charges; and environmental charges. Non-discriminatory mark-ups are also permitted. The changed model for organising rail transport in Europe has therefore created a key research need; namely to estimate the direct cost of running extra traffic on the network.

Sweden was the first country to undertake such a separation in 1988, with the rest of Europe following later, to a greater or lesser extent (see Nash and Matthews, 2009). The Swedish Railway Act stipulates two types of charges for the use of infrastructure (Banverket, 2009). Firstly, special charges, either covering the fixed costs of the infrastructure, or costs occurring when new infrastructure has been built as a special project. The second type is based on short run marginal costs. In turn, there are three different types of marginal cost based charges; the track charge, the accident charge and the emission charge. The first, and for our purposes most interesting, is the track charge, which mirrors the maintenance costs incurred by one additional tonne movement as a result of wear and tear on the network.

Importantly, to date, the wear and tear track charge has not taken into account incremental renewal costs. A renewal is an activity that will restore the infrastructure to its original standard. Renewals and maintenance are linked in such a way that lack of maintenance will force the infrastructure manager to renew at an earlier stage than if maintenance were undertaken properly and vice versa. An optimally managed railway track has a mix of maintenance and renewal in time over the life cycle and excluding renewals from the total picture of marginal infrastructure costs is therefore misleading.

More generally, most research on railway infrastructure wear and tear has focused on the relationship between maintenance costs and traffic, while controlling for infrastructure characteristics. The lack of empirical evidence concerning the size of the track renewal marginal cost has therefore recently drawn some attention in the literature and amongst policy makers (see Nash, 2005 and Wheat et. al., 2009).

Data on rail maintenance and renewal costs, outputs and network characteristics are typically recorded by rail infrastructure managers at the level of individual track sections. In

the rail marginal cost estimation literature, a track section represents the most disaggregate level at which cost data are recorded. In the case of this study, it is defined by the national track information system (BIS), administered by the Swedish Transport Administration (Trafikverket). However, as track renewals have long life cycles and therefore are rare events, disaggregate renewal cost data contain many zero observations – that is, no renewal is undertaken for a given track section in a given year.

In the small number of previous econometric studies on renewals marginal costs, this problem has been addressed by combining maintenance and renewal costs to create a measure of total costs (thus eliminating the zeros); see Andersson (2006; 2007a), and Marti et al. (2009). Alternatively, modelling has proceeded at a less disaggregate level (regional or even national, for a number of countries), thus eliminating zero renewal costs; see Wheat and Smith (2009), Smith (2008) and Smith et al. (2008), though again maintenance and renewals have been combined in the reported, preferred models. Both types of aggregation merely mask the problem of zero renewals. The result is that renewal cost elasticities have to be inferred from models based on maintenance and renewals combined, and there is therefore currently much uncertainty over the range of appropriate values to be used.

As an alternative way of circumventing the problem, Andersson and Björklund (forthcoming) applies survival analysis to estimate a deterioration elasticity with respect to traffic (tonnage) running on the network. Marginal costs are calculated as a change in present values of renewal costs from premature renewal following increased traffic volumes. One disadvantage of this approach is that it requires an assumption to be made about unit renewal costs in order to compute marginal costs. The latter is non-trivial given the considerable unit cost variation associated with different types and volumes (as unit costs vary with scale) of track replacement work.

Given the lack of previous evidence on track renewals marginal cost, and the associated methodological problems experienced in previous studies, new approaches to the problem and new evidence are called for. In this paper we utilise an alternative set of econometric models that are workable even for disaggregated data with a large proportion of zero renewals (Tobit, Heckit and the two-part model). These approaches derive marginal costs directly from the econometric cost model, based on actual cost data (avoiding the aforementioned problems associated with survival analysis), whilst ensuring that the zero data observations are utilised and modelled appropriately to ensure consistent estimates of the model parameters (a more satisfactory approach than simply aggregating the data). We explore the results of these alternative approaches using Swedish railway track renewal cost data.

This paper is the first attempt in the literature to apply corner solution models, and in particular the two-part model, to disaggregate renewal cost data in railways (characterised by

a data structure comprising a large fraction of zero values for the dependent variable) and thus derive usage elasticities specifically for renewal costs[2]. We consider this to be an important addition to the literature, particularly given the paucity of studies of marginal track renewal costs in general and its importance in the context of setting track access charges in vertically separated rail systems. The methods used in this paper, whilst not previously applied in the context of track renewals marginal cost estimation, have been widely applied in transport applications more generally (see for example, Train, 1986 and Mannering and Hensher, 1987).

The paper is organised as follows. In section 2, we introduce the modelling approach followed by a description of the data set in section 3. Section 4 covers the econometric specifications and results, while we discuss the results and draw some conclusions in section 5.

2. MODELLING APPROACH

There exists an extensive literature on statistical modelling techniques for use when data are censored or truncated. When a relevant part of the population generating the data is unobserved, the data set is said to be truncated. In this case, data on both the dependent and independent variables are not observed. In the case of censored data, the dependent variable is not observable for some part of the population (though data on the independent variables are available). The corner solution model, sometimes described as being a type of censored data model, is relevant to a situation where a firm or household makes an (observable) choice for a variable, y, where y takes the value zero (the corner solution) with a positive probability, and otherwise is a continuous, strictly positive random variable (Wooldridge, 2002).

Examples of the corner solution model include household expenditure on life insurance or health services. In these cases researchers are analysing continuous variables (expenditure) containing a spike or probability mass at zero. The zeros are not censored versions of some underlying variable, they are "true" zeros, since they are the actual choices of the relevant decision maker. For this reason, the corner solution model is not actually a censored model, though it produces the same model specification and can thus be treated as the same in terms of estimation (Greene, 2007). The corner solution model is clearly appropriate to our case, where track renewal costs may be positive or zero, depending on the choice of the infrastructure manager.

---

[2] The term 'usage elasticity' was first used by Wheat et. al. (2009) to refer to the elasticity of cost (be it maintenance and/or renewal) with respect to traffic.

In this paper we apply three well established econometric techniques that are appropriate for application to our dataset: the standard Tobit model (Tobin, 1958; Amemiya, 1985), the two-part model (Cragg, 1971) and the sample selection or Heckit model first proposed by Heckman (1979). The standard Tobit model can be written as follows (see for example, Greene, 2007):

$$y_i^* = x_i' \beta + \varepsilon_i \tag{1}$$

$$y_i = y_i^* \qquad if \ y_i^* > 0 \tag{2}$$
$$y_i = 0 \qquad otherwise$$
$$\varepsilon_i \sim N[0, \sigma^2]$$

Note that in the context of the corner solution model, which is relevant here, y* is simply a construct to help us formulate the model. In the corner solution model, y is both the observed data and the variable that we are interested in understanding. The Tobit model corrects for the piling up of zeros, which violates the standard OLS assumption of the error term having a zero conditional mean, avoids negative predictions and also give more reasonable estimates of partial effects (Wooldridge, 2009). The Tobit model proceeds by applying maximum likelihood estimation to all of the data points (including the zeros).

Cragg (1971) proposes an alternative two-part model that nests the Tobit model as a special case. The first part of the two-part model can be written as a standard probit model:

$$z_i^* = x_{1i}' \beta_1 + u_{1i}$$
$$I_i = 1 \ if \ z_i^* > 0 \tag{3}$$
$$I_i = 0 \ otherwise$$
$$u_{1i} \sim N(0,1)$$

where I_i is a binary choice variable, taking the value 0 or 1, and z_i^* is an unobservable latent variable (underpinning the decision whether to renew or not). The second part is then a truncated regression model:

$$y_i \mid (I_i = 1) = x_{2i}' \beta_2 + u_{2i} \tag{4}$$

where $E(u_{2i}|I_i=1)=0$ and u_{2i} need not be normally distributed. The model here implies that the value of y (say expenditure), given that it is positive, and after controlling for the regressors (x), is independent of the decision whether to make any expenditure at all.

The two-part model considers that the data generating process (DGP) for the decision to participate – in this case, to renew or not – may be different from the DGP for the decision of how much to spend. This flexibility arises, firstly since the regressors can differ in each decision equation (x_{1i} need not equal x_{2i}) and second, even if the regressors are the same, the coefficients can be different (β_1 need not equal β_2). We may expect a priori that some of the candidate regressors for this study maybe statistically significant in one equation but not in the other or take different signs (see section 4.1 for further discussion of this point in respect of this study). If the same regressors appear in both parts, and $\beta_1 = \beta_2$ then the two-part model simplifies to the Tobit case (and this restriction is testable).

It should be noted that the two-part model imposes the assumption that the errors in each part are independently distributed. Thus, whilst correlation between the parts is allowed for via the included regressors, the residuals are not correlated. Correlation between residuals often arises if the reason for censoring is due to sample selection. However there is no sample selection issue in the corner solution interpretation; the zeros are true zeros. However correlation in residuals could result due to correlation between unobserved effects in each part of the model. Thus it is an empirical matter as to whether the assumption of independent errors is reasonable, though we further note that Dow and Norton (2003) show that the two-part model performs better in the context of a corner solution example. Importantly, therefore, although the problem of sample selection bias is not applicable to the corner solution model, the less restrictive Heckit model is considered as a candidate method even in this case (Dow and Norton, 2003 and Leung and Yu, 1996).

The Heckit model is set out below for completeness. It can be formalised as a selection equation (5) and an outcome equation (6).

$$z_i^* = x_{1i}'\beta_1 + u_{1i}$$
$$I_i = 1 \; if \; z_i^* > 0 \tag{5}$$
$$I_i = 0 \; otherwise$$

$$y_i \mid (I_i = 1) = x_{2i}'\beta_2 + \sigma_{12}\lambda(x_{1i}'\hat{\beta_1}) + u_{2i} \tag{6}$$

$$(u_{1i}, u_{2i}) \sim N(0, \Sigma)$$

$$\Sigma = \begin{bmatrix} 1 & \rho\sigma_2 \\ \rho\sigma_2 & \sigma_2^{\,2} \end{bmatrix}$$

and where $\lambda(x'_{1i}\hat{\beta}_1)$ is the estimated inverse Mills ratio $\phi(x'_{1i}\hat{\beta}_1)/\Phi(x'_{1i}\hat{\beta}_1)$ and $\phi(.)$ and $\Phi(.)$ are the standard normal pdf and cdf respectively. The correlation between the errors in the two stages is given as $\rho = \sigma_{12}/\sigma_2$. A simple *t* test of whether or not $\sigma_{12} = 0$ (or ρ = 0) can be used to test the null hypothesis that the two-part model is correct (the alternative hypothesis being that the Heckit model is the correct model)[3].

In this paper we apply the above methods to the problem of obtaining marginal costs of railway track renewal. Our data set (see section 3) comprises data on 190 track sections where in any given year the observed track renewal costs are either positive or zero (and where almost 60% of the observations are zeros, since track has a long asset life). Further, as noted earlier, the zero observations are true zeros, and thus the corner solution interpretation is the relevant one in our case. Adopting the appropriate model from this model class ensures that consistent estimates are produced, which would not be the case if analysis proceeded by simply carrying out OLS on all the data, or OLS on the positive values[4].

3. THE DATA

There is no readily available, single database containing all data on costs, traffic and infrastructure required for our analysis. Therefore, our dataset been gathered from different sources within the Swedish Transport Administration (*Trafikverket*)[5]. The collection and assimilation of this data was therefore, in itself, a major undertaking.

---

[3] As discussed in Dow and Norton (2003), the *t* test can be used as a test of the null that the Two part model is true (against the alternative that the Heckit is true), and therefore the Two-part and Heckit models are partially nested (see also Leung and Yu, 1996).

[4] Though noting that this result is sensitive to the distributional assumptions (normality and homoscedasticity); see for example, Laitila (1989) and Greene (2007). It is also noted in the literature that whilst heteroscedasticity can impact on the parameter estimates, the marginal effects (that we are interested in here) are often similar to those resulting from a model assuming homoscedasticity (see Greene, 2003, page 768).

[5] The Swedish Rail Administration (Banverket) merged with the Swedish Road Administration (Vägverket) on April 1, 2010 and formed the Swedish Transport Administration (Trafikverket). All our data has been collected from Banverket, but we refer to Trafikverket as the provider of information as Banverket no longer exists.

The total data set contains 2093 observations and covers approximately 190 track sections for a period of eleven years, from 1999 to 2009. However, missing traffic data on some peripheral lines and station areas restricts us to consider a sample of 1663 observations. We also use a dummy variable based on whether a renewal occurs in the previous year, which means that data from 1999 is excluded, thus reducing our final sample size for estimation to 1507. Descriptive statistics are given in Table 1. The track sections are defined by the national track information system (BIS), administered by Trafikverket. The length of the track sections, including multiple tracks, varies from 2.6 kilometres to over 260 kilometres, with an average of about 78 kilometres. The number of annual observations varies between 145 and 159. One reason for this variation is that some track sections have been merged or abandoned, while some new sections have been formed during this period.

Table 1. Descriptive statistics

| Variable	Mean	Std.Dev.	Min	Max
Track renewal cost (SEK)[1]	3 342 144	16 100 000	0	243 000 000
Section length (meters)	77 801.3	52 394.8	2 642	261 561
Gross tonnes per track section (tonnage density)[2]	7 183 785	7 588 555	15.8	46 900 000
Number of trains	15 583.8	17 866.3	0.2193	132 501
Tunnels (meters)	383.1	1 487.0	0	13 802.4
Bridges (meters)	649.7	983.8	0	9 822
Number of joints	173.4	121.8	0	730
Number of switches	52.0	46.9	2	353
Switches (meters)	1575.8	1374.6	58.03	9 070
Switch age (years)	20.4	9.1	1	67.7
Rail weight (kg)	50.9	5.1	32	60
Rail age (years)	20.2	11.5	1	98
West region (dummy variable)	0.1606	0.3672	0	1
North region (dummy variable)	0.1294	0.3356	0	1
Central region (dummy variable)	0.1990	0.3994	0	1
South region (dummy variable)	0.2592	0.4383	0	1

[1] Annual cost in 2009 prices. [2] Defined as gross tonne-km divided by track-km.

The cost data originate from Trafikverket's accounting system, Agresso. The cost data cover track renewal costs at a track section level. The nominal cost data are converted into constant prices (2009 prices in Swedish Kronor (SEK))[6]. Track renewals make up roughly half of total rail infrastructure renewal costs. Out of the 1663 observations, 958 or almost 60 per cent of the track renewal cost observations equals zero. Further, approximately 2 per cent of the track sections have had track renewals in all of the studied years, while roughly 11 per cent of the track sections have not had track renewals in any of the years. At the overall network level, there has been a notable variation and increase in total track renewal costs during the period in question, as illustrated by figure 1. This reflects

---

[6] 1 SEK=0.11 EUR/0.15 USD, as of August 19, 2011.

a generally increased focus on track renewals, as well as an allocation of further resources to this area. Still, only 1.3 per cent of the total Swedish rail network track length was renewed in 2009.

Figure 1: Total Track Renewal Cost between 1999 and 2009 (2009 Prices).
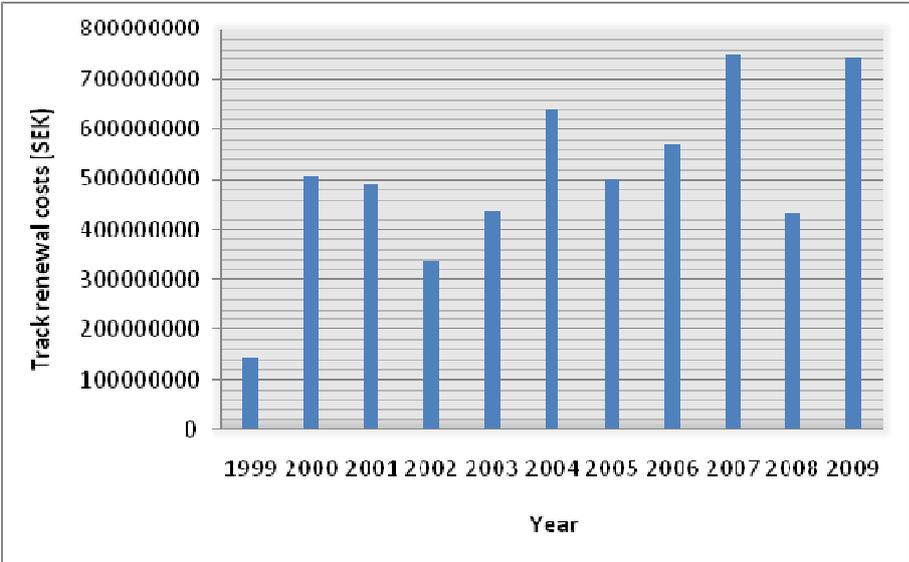
Since the separation of train operation from infrastructure management in Sweden in 1988, the supply of traffic data has become more problematic, particularly in view of the higher level of competition on the tracks. Traffic data were therefore retrieved from different sources such as train operators and published timetables, and for later years from Trafikverket. Generally, traffic has risen in the period in question from an average of 6 million gross tonnes per track section in 1999 to 7.7 million in 2009, peaking at 7.9 million in 2008.

Data on characteristics of the infrastructure have been retrieved from the national track information system, BIS. The list of variables capturing infrastructure characteristics includes, inter alia, rail age, switches, track length, bridges and tunnels. Further, dummy variables representing different track management regions of Sweden are included the data set. These variables will represent geographical differences, such as climate needs and potential differences in managerial skills.

Overall we have been able to collect a substantial high quality data set that enables us to explore the relationship between track renewal costs and traffic volume, controlling for a range of infrastructure characteristics and regional (dummy) variables. We now proceed to present and discuss the results.

4. ECONOMETRIC RESULTS

4.1 Model specification

As discussed in section 2, we apply three well established econometric techniques that are appropriate for application to our dataset. We carry out appropriate testing to arrive at a preferred model. Track renewal cost (in log form) is used as the dependent variable in the Tobit and the outcome equations of the two-part and Heckit models. A binary choice variable is used in the selection equations of the latter two models[7].

As independent variables, we use the logs of track section length, total gross tonnes per track section (or tonnage density), average switch age and rail weight, together with four regional dummy variables that will in part pick up remaining unobserved heterogeneity between the sections[8]. For the Tobit model and the outcome equation in the two-part and Heckit models we are thus using a log-linear (or double-log) functional form. As is standard in the rail cost function estimation literature, rail usage (and indeed the structure of the network) is assumed to be exogenously determined (see for example, Oum, Waters (II) and Yu, 1999). We also include nine dummy variables for the years 2001 – 2009, as well as a dummy variable that takes the value unity if there was a renewal on the section in the previous year (zero otherwise), in order to capture possible temporal correlation (as discussed further below). The inclusion of this dummy results in the sample size for estimation falling from 1663 to 1507. All other variables in Table 1 have not been found to improve our models, both in terms of fit and in providing plausible estimates.

Following the discussion in section 2, there is a question as to which of the three proposed models is, a priori, likely to be the most appropriate to our particular problem. At a conceptual level, it seems reasonable to think of track renewal costs being explained by a two part process: firstly there is a decision whether to renew, and secondly, there is a decision about the quality of the renewal to be undertaken, which will determine the unit cost and thus overall cost of the renewal.

The probability of a renewal occurring will depend on the state of the asset relative to relevant asset condition and safety standards. This in turn depends on the age and characteristics of the track and the volume and type of traffic that has run on the section. The quality of the renewal carried out, which naturally impacts on the observed renewal cost, will

---

[7] Given that our data set contains zero observations for the dependent variable, to estimate the Tobit model in log form, we first transform our data by finding the minimum log value of our positive observations, and setting the missing observations infinitesimally below the minimum value (see Cameron and Trivedi, 2009).

[8] Both switch age and rail weight (weight of the rail in kilograms per metre) are computed as averages over a track section. The average value is weighted by length of each segment that is included in the section.

depend on a range of factors, for example the type of track being replaced, and the loads that it is expected to bear. Other things equal, the age of the track would not be expected to impact on the cost of the renewal undertaken, since a renewal job is undertaken whether old or new track is being replaced. However, other things are not equal, as the age of the track could impact on whether the rail, sleepers, and ballast all need replacing together (or not), and indeed whether major work is required on the sub-structure. Thus track age impacts on cost in a potentially ambiguous way.

It is therefore possible to think of the variables included in the two-stages having different coefficients, possibly also with different signs. Whilst our primary concern in this paper is with the unconditional elasticity of cost with respect of tonnage density, it is worth briefly commenting on the expected signs of the coefficients in the selection and outcome equations for the two-part and Heckit models.

First, the length of a track section is included in both the selection and outcome parts of the model because the track sections are not of equal length. Thus, in the first stage, a longer track section is more likely to see part of the section renewed. Whilst some data exist at a further level of disaggregation (track segment), cost data are not available at that level. Our data set is therefore the most disaggregate level at which cost and cost driver information exists in Sweden. In the second stage, section length is a proxy for the size of the renewal undertaken. In both stages we expect the coefficient on section length to be positive.

Likewise, we expect total gross tonnes per track section (tonnage density) to increase both the probability of a renewal and the cost of renewal undertaken. At this point we note that in the first stage, (annual) tonnage density is acting as a proxy for cumulative tonnage. At present we do not have a robust measure of cumulative tonnage, but we hope to develop this and utilise it in future work. In the second stage, tonnage density is again likely to have a positive effect, but its effect may come through in respect of expected future tonnage, as that would affect the quality of a renewal to be done. To the extent that current tonnage is a good proxy for both past and future traffic, this distinction may be unimportant.

For age of track (which includes rail, sleepers, ballast and switches), we first tried to include rail age, which, surprisingly, proved to be insignificant in both stages of the model. However, switch age proved to be significant in both stages and is thus retained in the model. This variable can be partly thought of as a proxy for track age, but also reflecting the fact that in some cases switches may be replaced without any replacement of the remaining track infrastructure. We would expect older track assets (proxied here via switch age), other things equal, to have a higher probability of a renewal, as the asset reaches the end of its life, but its impact on the cost of renewal is driven by different factors as noted above. Increased rail weight would, other things equal, be expected to reduce the probability of a

track renewal, since the rail quality is higher. In the second stage rail weight would be expected to increase the cost of renewals as higher quality rail is being installed, but at the same time other factors, such as the age and type of sleepers and the condition of the sub-structure could come into play.

We include regional and year dummy variables to capture unobserved heterogeneity between sections, budget fluctuations and other time trends, but with no a priori expectation on signs. The sign of the coefficient on the prior year renewal dummy variable capturing whether a renewal takes place in the previous period is also ambiguous. Here it should be noted that a track section comprises several segments and the average length of a section is 78km. The renewal costs we observe at track section level are an aggregate of several renewals within a track section. Therefore, if a renewal on a track section takes place one year, this does not mean that a renewal cannot occur the following year, since only part of the section may be renewed. The sign on the coefficient could be negative if a large proportion of the section is renewed one year, thus reducing the possibility of future renewals. On the other hand it could be positive if segments within a section are at a similar stage in terms of degradation, so that a renewal on one segment in one year means that there are likely to be further renewals on other segments within the same section the following year. All estimations are done in Stata 10 (StataCorp, 2007).

Since we have good reasons to expect that the explanatory variables will have different impacts on the decision to participate and the decision of how much to spend on renewals, we would expect the two-part model to perform better than the more restrictive Tobit. However, this is an empirical matter, and we therefore also estimate the Tobit and perform the relevant tests. Likewise, given the absence of a sample selection issue in corner solution models, a priori we prefer the two-part model, though noting that the Heckit is still regarded in the literature as a candidate model for corner solution models. We therefore estimate and compare the results of all three of the candidate models noted above.

4.2 Estimation outputs

Concerning the choice between the Tobit, two-part and Heckit models, a likelihood ratio test shows that the Tobit restriction can be rejected at any reasonable levels of significance as compared to the more flexible two-part model. With regard to the choice between the two-part and Heckit models, we cannot reject the null hypothesis that the correlation between the errors in the two stages is zero even at the 10% level (based on the standard t test on the coefficient (σ_{12}) on the inverse Mills ratio in equation 6; p value = 0.125). This finding leads us to prefer the two-part model. However, since the power of this

test is affected by the multi-collinearity problems that often beset the Heckit model in empirical applications, we follow the approach recommended in Dow and Norton (2003), and utilise the empirical mean square error (EMSE) criterion (computed based on the main elasticity of interest, the estimated elasticity with respect to tonnage density). We find that the two-part model has the lower EMSE for this estimate, which again favours the two-part model according to this criterion.

The selection of our preferred model, based on empirical evidence, of the two-part model over the Tobit and Heckit models is in line with our a priori preference for the two-part model from a theoretical perspective. The estimation output for the two-part model is shown in Table 2 below. This is shown for completeness. Our key interest is in the elasticity of cost with respect tonnage density (and the associated marginal cost), which we discuss in section 4.3 below.

However, we first briefly comment on the coefficient estimates in the preferred, two-part model. As expected, section length and tonnage density have positive coefficients in both the probability (selection) and conditional regression (outcome) equations (and are statistically significant in both equations). Switch age also has the expected positive coefficient in the first equation (p value = 0.1240). As discussed in section 4.1, its sign in the second equation is ambiguous, and in this case found to be positive (p value = 0.0500). Rail weight also has the expected, negative sign in the first equation (p value = 0.1040), and as discussed earlier, its sign in the second equation is ambiguous, and in this case found to be negative also (p value = 0.2470). The coefficient on the prior year renewal dummy variable is positive and statistically significant in both equations of the two part model (p value=0.0000). It appears that this variable is therefore likely picking up the fact that segments within a section are at a similar stage in terms of degradation, meaning that if there are renewals on one segment in one year there are likely to be renewals on other segments within the same track section the following year.

4.3 Cost predictions and elasticities

Our main interest is in the cost elasticity with respect to tonnage density, as the marginal cost is the product of the estimated cost elasticity and the predicted average cost. Dow and Norton (2003) argue that where the two-part model is applied to corner solution data then it is the cost elasticities and marginal costs associated with the actual values of the dependent variable (cost) that are of interest rather than the elasticities and marginal costs of the latent variable.

Table 2. Two-Part Model Results

Variable / Equation	Coefficient	Robust standard error	z	P value	95% confidence interval	
Selection equation						
Ln (Section length)	0.3655	0.0555	6.5800	0.0000	0.2567	0.4743
Ln (Tonnage density)	0.2249	0.0381	5.9000	0.0000	0.1502	0.2996
Ln (Switch age)	0.1481	0.0962	1.5400	0.1240	-0.0405	0.3368
Ln (Rail weight)	-0.8857	0.5448	-1.6300	0.1040	-1.9535	0.1821
North region dummy	0.0536	0.1328	0.4000	0.6870	-0.2066	0.3138
Central region dummy	-0.0508	0.1192	-0.4300	0.6700	-0.2844	0.1828
South region dummy	0.1206	0.1198	1.0100	0.3140	-0.1142	0.3554
East region dummy	0.1482	0.1223	1.2100	0.2250	-0.0915	0.3879
Year 2001 dummy	0.3543	0.1629	2.1800	0.0300	0.0351	0.6735
Year 2002 dummy	0.7509	0.1559	4.8200	0.0000	0.4454	1.0564
Year 2003 dummy	0.5812	0.1577	3.6800	0.0000	0.2721	0.8903
Year 2004 dummy	0.6281	0.1611	3.9000	0.0000	0.3123	0.9439
Year 2005 dummy	0.5407	0.1636	3.3000	0.0010	0.2200	0.8614
Year 2006 dummy	0.8057	0.1632	4.9400	0.0000	0.4857	1.1256
Year 2007 dummy	0.7710	0.1609	4.7900	0.0000	0.4556	1.0864
Year 2008 dummy	0.8760	0.1635	5.3600	0.0000	0.5554	1.1965
Year 2009 dummy	1.2493	0.1709	7.3100	0.0000	0.9142	1.5843
Prior-year renewal dummy	0.8205	0.0768	10.6900	0.0000	0.6701	0.9710
Constant	-5.5999	1.9629	-2.8500	0.0040	-9.4471	-1.7527
Outcome equation						
Ln (Section length)	0.6668	0.1490	4.4800	0.0000	0.3748	0.9587
Ln (Tonnage density)	0.2120	0.0959	2.2100	0.0270	0.0240	0.4000
Ln (Switch age)	0.5481	0.2801	1.9600	0.0500	-0.0008	1.0971
Ln (Rail weight)	-1.5561	1.3438	-1.1600	0.2470	-4.1899	1.0776
North region dummy	0.1223	0.3168	0.3900	0.6990	-0.4985	0.7432
Central region dummy	-0.2050	0.3119	-0.6600	0.5110	-0.8162	0.4063
South region dummy	0.4196	0.3015	1.3900	0.1640	-0.1714	1.0105
East region dummy	-0.4685	0.2770	-1.6900	0.0910	-1.0113	0.0744
Year 2001 dummy	0.4649	0.6454	0.7200	0.4710	-0.8001	1.7300
Year 2002 dummy	-0.4338	0.6279	-0.6900	0.4900	-1.6644	0.7969
Year 2003 dummy	-0.1827	0.6212	-0.2900	0.7690	-1.4002	1.0348
Year 2004 dummy	-0.0332	0.6124	-0.0500	0.9570	-1.2334	1.1671
Year 2005 dummy	-0.2093	0.6322	-0.3300	0.7410	-1.4484	1.0298
Year 2006 dummy	-0.2956	0.6160	-0.4800	0.6310	-1.5029	0.9118
Year 2007 dummy	-0.4963	0.6075	-0.8200	0.4140	-1.6870	0.6945
Year 2008 dummy	-0.6691	0.5877	-1.1400	0.2550	-1.8209	0.4827
Year 2009 dummy	-0.3221	0.5885	-0.5500	0.5840	-1.4756	0.8314
Prior-year renewal dummy	0.8385	0.1890	4.4400	0.0000	0.4681	1.2089
Constant	7.3131	4.9614	1.4700	0.1400	-2.4110	17.0373

Number of obs=1507. Log pseudolikelihood = -2321.5421. Wald Chi2(18) = 365.13. Prob > Chi2 = 0.0000.

Importantly, note that both the marginal costs and the elasticities for the two-part model depend on the coefficients from both stages of the models; the decision to renew and the cost of the renewal should it go ahead. Thus they represent the effect of increasing usage on cost taking into account the change in likelihood of undertaking a renewal and any change in the cost of a renewal should it be undertaken. It should be emphasised that both

marginal costs and elasticities are non-linear functions of multiple parameters. Dow and Norton (2003) derive the formula for the elasticity when the dependent variable is in log form (independent variables in non-logged form). In our case, when the dependent and independent variables are all in log-form, the formula for the two-part model is shown in equation (7) (see van de Ven and van Praag, 1981):

$$\frac{\partial E[y]}{\partial x_k} \times \frac{x_k}{E[y]} = \beta_{2k} + \beta_{1k}\lambda(x_1'\beta_1) \tag{7}$$

Table 3 shows the elasticity of cost with respect to tonnage, together with the standard error and confidence interval, for the preferred two-part model. The elasticity at the sample mean is around 0.41. That is, a one per cent increase in traffic will increase renewal costs by 0.41 per cent. The (average) marginal cost estimate for the two-part model is shown in Table 4, together with its standard error. The preferred model produces a weighted average marginal cost estimate of SEK 0.007 per gross tonne-km.

Table 3. Elasticity With Respect To Tonnage Density: Two part Model

Model	Elasticity*	Standard error	z	p value	95 % Conf. Interval	
Two-part	0.406	0.105	3.860	0.000	0.200	0.612

* Calculated at the sample mean

Table 4. Marginal Cost Estimates: Two-part Model

	Obs.	Weighted mean*	Standard error	95 % Conf. Interval	
Marginal cost	1507	0.0067	0.0002	0.0063	0.0071

* The marginal cost is weighted by gross tonne-km's per track section

Table 5. Variations in Marginal Costs and Elasticities by Rail Weight, Switch Age and Tonnage Density

	Marginal cost	Elasticity	Number of observations
Rail weight			
<45 kg	0.063	0.444	214
45-55 kg	0.016	0.404	925
>55 kg	0.006	0.388	368
Switch age			
<10 years	0.005	0.436	149
10-20 years	0.009	0.396	635
21-30 years	0.019	0.401	487
>30 years	0.061	0.422	236
Tonnage			
<5 mgt	0.032	0.426	830
5-10 mgt	0.006	0.394	253
11-20 mgt	0.005	0.380	286
>20 mgt	0.005	0.360	138

mgt=million gross tonnes. Note marginal costs and elasticities are averages of the observations within the respective ranges.

Table 5 shows how elasticities and marginal costs vary according to different dimensions. The finding of higher marginal cost for lower rail weight (lower quality rail) and older assets is intuitive. Previous studies have also found marginal cost falling with usage (with varying findings on the direction of change of the elasticity with traffic; see Wheat et. al., 2009). We see that classifying the network according to weight, age or density would make a substantial difference compared to the network average estimate in Table 4, thus suggesting a possible basis for differentiated track renewal charges. We return to this issue in Section 5.

We consider that our approach has produced robust estimates of the elasticity of track renewal costs with respect to usage that can be compared with the existing literature on maintenance and maintenance and renewals marginal cost estimation using disaggregate data (see Wheat et. al., 2009 and Table 6 and the associated discussion). The key innovation in the paper is the adoption of corner solution models to disaggregate data containing zero observations, as well the fact that this paper is the first to report separate marginal costs for renewals. Our model also controls for a range of infrastructure characteristics variables directly, as well as capturing unobserved heterogeneity at the

regional level via four regional dummy variables. We have further allowed for temporal correlation via the inclusion of a dummy variable denoting the implementation (or not) of a renewal in the previous year. We do not seek to model possible spatial correlation between sections, since our track section data is an aggregation of a number of segments with the average length of a section being 78km. Thus any spatial correlation would be expected to occur at the segment level (and indeed this effect appears to be being picked up via the prior-period renewal dummy variable)[9].

One possible area for future research would be considering further the dynamic nature of the track replacement decision (the previous literature has not attempted a dynamic estimation approach). However, given the very long asset life of rail infrastructure (on average more than 30 years), we would need a much longer time series than is available. We therefore consider that an alternative and more fruitful line of future research would be the incorporation of cumulative tonnage into the analysis. Cumulative tonnage is considered a key for renewal decisions in the rail technology literature. In this paper, we have attempted to model renewal costs using annual tonnage measures on the right hand side of the model, together with capability, condition and age measures. Whilst a robust measure of cumulative tonnage is not yet available in Sweden, this could be a realistic possibility with additional data collection and analysis in the future.

5. DISCUSSION AND CONCLUSIONS

In this paper, we have analysed railway track renewal costs using Swedish track section data from 1999-2009. We have estimated three different regression models; the Tobit, the two-part and the Heckit. All of these models have properties to make them suitable for estimation when the data contain a large fraction of true zeros in the dependent variable. Our preferred model is the two-part model.

We find that the cost elasticity with respect to output (gross tonne-km) is around 0.41. This is higher than previously found for analyses of maintenance costs, which suggests a range of 0.20-0.35 (Wheat et. al., 2009); and is in line with a priori expectations, since engineering evidence suggests that renewals are more variable with traffic than maintenance (see Abrantes et. al., 2008)[10].

---

[9] If the model was implementing using more disaggregate, segment level data, where spatial correlation may be present, the Probit estimator (and thus in turn the elasticity estimates from the two-part model) would be inconsistent (see Pinkse and Slade, 1998).

[10] Though recent work done based on British data indicates that the cost variability of track renewal costs may be lower than for track maintenance, though still higher than for other types of maintenance (for example, electrification and signalling asset maintenance) and other types of renewals; see ORR (2008) page 232.

We now turn to consider how our results fit into the previous literature in respect of renewals costs (see Table 6). A few points need to be borne in mind at this stage. First, as compared to studies of maintenance marginal costs, there is a relative shortage of studies involving renewals costs. Second, and perhaps more importantly, all of the previous studies have modelled maintenance and renewals together, and these studies have produced a wide range of estimates for the total maintenance and renewals cost elasticities. The result is that renewal cost elasticities have to be inferred from models based on maintenance and renewals combined, and there is therefore currently much uncertainty over the range of appropriate values that should be used. Our paper is therefore the first paper in the literature that we are aware of to report usage elasticities specifically for renewals costs (in our case, track renewal costs).

Table 6. Studies on railway infrastructure renewal costs

Study	Data	Cost category	Average elasticity*
This paper	**Track section level Sweden 1999 – 2009**	**Track renewals**	**0.41**
Andersson (2006)	Track section level Sweden 1999 – 2002	Maintenance and Renewals	0.26
Marti et al. (2009)	Track section level Switzerland 2003 – 2007	Maintenance and Renewals	0.28
Wheat and Smith (2009)	Maintenance delivery unit level Great Britain 2006	Maintenance and Track renewals	0.49
Smith et al. (2008)	Regional level 5 European countries 2002-2006	Maintenance and Track renewals	0.43-0.44
Smith (2008)	National level 13 European countries 1996 – 2006	Maintenance and Renewals	0.48-0.51
Wheat et al. (2009)	A range of country case studies	Maintenance only	0.20-0.35
Andersson and Björklund (forthcoming)	Track segment level Sweden 1999 – 2005	Renewals only	-0.3**

* Elasticity of cost w.r.t. traffic volume; ** Elasticity of expected life time w.r.t. traffic volume

Nevertheless, it is interesting to consider how our results compare against previous work. Given difference in network quality between countries resulting in substantial differences in average cost, Wheat et al. (2009) recommend that generalisation from one study to other networks should be based on elasticities. As noted, the majority of the studies reported in Table 6 cover both maintenance and renewals (M&R) cost, and we would thus

expect our results to have a higher elasticity than in those studies based on engineering evidence. The reported elasticity of 0.41 from our preferred model does indeed lie towards the top of the range of previous estimates for maintenance and renewals. It should further be noted that the higher M&R elasticities in Table 6 derive from the results of more aggregated data (national, regional or maintenance delivery unit). The more relevant comparators in this case are the studies by Andersson (2006) and Marti et al. (2009) – which utilise disaggregate (track section) more similar in nature to the data set used in our study – and they report much lower elasticities (0.26-0.28).

Overall, we conclude that our findings on the elasticity of cost with respect to tonnage make sense in the context of previous work (econometric and engineering), though the different cost categories used make a more in-depth comparison problematic. Importantly, by presenting the first renewals-only study, we consider that we have added new clarity to the literature, and indeed increased certainty regarding the elasticity of track renewal costs with respect to traffic.

The average marginal cost per gross tonne-km is estimated to be approximately SEK 0.007 or €0.07c. Marginal cost estimates are either not reported in the other previous studies shown in Table 6, or are non-comparable since they are based different cost bases (i.e. they include maintenance); and as noted above, generalisation from one country to another should proceed based on elasticities rather than marginal costs in any case. Turning to the evidence in respect of Sweden, our estimates are higher than found in Andersson and Björklund (forthcoming), using survival analysis and a unit cost for track renewal (estimated marginal cost of 0.002 SEK). We expect the present estimates to be higher as they cover a larger track renewal cost share. Since the current pricing scheme in Sweden only covers the marginal infrastructure cost for maintenance activities, the inclusion of our estimate of marginal infrastructure renewal costs would add substantially to the current track charge, which is only SEK 0.0036 per gross tonne-km.

The results derived in this paper are relevant in the context of two important policy debates which are occurring both at the national (Swedish) and wider European levels. The first concerns the choice between a network-wide charge for all track sections and a more differentiated solution. We have shown in Table 5 how marginal costs and elasticities vary with switch age, rail weight and tonnage density. The network weighted average, by traffic volume, has the benefit of being revenue neutral, i.e. it gives the same amount of revenue as a pricing scheme based on track section specific marginal costs and is easy to apply in a pricing scheme. However, it is less efficient as it does not take into account the underlying variation in marginal costs over the network. With a constant network charge, based on the weighted average, we see that traffic on high quality tracks (low switch age and high rail weight), where marginal cost is relatively low, is effectively subsidising traffic on low quality

tracks. However, a more differentiated wear and tear charge could lead to access charges that would make rail too expensive on peripheral lines, but at the same time would potentially enable funds to be prioritised towards improving the main network. Of course, the notion of higher charges for lower quality infrastructure, and prioritisation of one part of the network over the other, may be hard to sustain from a political perspective.

The second issue concerns the overall funding of the Swedish rail network. The Swedish Transport Administration has been given a task by the Swedish government to double the access charge revenues by 2014 and increase the cost recovery rate. Sweden currently has amongst the lowest rail infrastructure charges in Europe (see Nash and Matthews, 2009). There are a number of means by which access charge revenues could be increased, but extending the wear and tear charge to also include marginal track renewals costs would contribute to both the financial target and increased economic efficiency (and of course European legislation requires charges to be cost reflective). Given the evidence provided by this paper, the inclusion of marginal track renewal cost would lead to an increase in charges of approximately 200 per cent.

ACKNOWLEDGMENTS

REFERENCES

Abrantes, P., P. Wheat, S. Iwnicki, C. Nash and A. Smith (2008). Review of Rail Track Cost Allocation Studies. Deliverable 1, EU Sixth Framework Project CATRIN (Cost Allocation of Transport Infrastructure Cost).

Amemiya, T. (1985). Advanced Econometrics. Harvard University Press.

Andersson, M. (2006). Marginal Cost Pricing of Railway Infrastructure Operation, Maintenance and Renewal in Sweden: From Policy to Practice Through Existing Data. Transportation Research Record: Journal of the Transportation Research Board, 1943, 1-11.

Andersson, M. (2007a). Fixed Effects Estimation of Marginal Railway Infrastructure Costs in Sweden. Scandinavian Working papers in Economics, VTI Series 2007:10, Swedish National Road and Transport Research Institute, Stockholm, Sweden.

Andersson, M. and G. Björklund (forthcoming). Marginal Railway Renewal Costs: A Survival Data Approach. VTI Working Paper.

Banverket (2009). Network Statement 2011. Borlänge, Sweden.

Burke, W. J. (2009). Fitting and interpreting Cragg's Tobit alternative using Stata. Stata Journal, 9, 584-592.

Cameron, C. and P. Trivedi (2009). Microeconometrics using Stata. Stata Press.

Cragg, J. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica, 39, 829-844.

Dow, W. and E. Norton (2003). Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions. Health Services and Outcomes Research Methodology, 4, 5-18.

Fin, T. and P. Schmidt (1984). A Test of the Tobit Specification Against an Alternative Suggested by Cragg. Review of Economics and Statistics, 66, 174-177.

Greene, W. H. (2003). Econometric Analysis. 5th Edition. Prentice Hall.

Greene, W. H. (2007). Censored Data and Truncated Distributions. In Mills, T. C. and Patterson, K. (eds), Palgrave Handbook of Econometrics: Volume 1 Econometric Theory. Palgrave Macmillan.

Heckman, J. (1979). Sample Selection Bias as a specification Error. Econometrica, 47, 153-161.

Laitila, T. (1989). Asymptotic Misspecification Biases for Heckman's Two Step Estimator. Communications in Statistics – Theory and Methods, 18, 2, 743-753.

Leung, S. F. and S. Yu (1996). On the Choice Between Sample Selection and Two-part Models. Journal of Econometrics, 72, 197-229.

Mannering, F. L. and D. A. Hensher (1987). Discrete / Continuous Econometric Models and their Application to Transport Analysis. Transport Reviews, 7, 3, 227-244.

Marti, M., R. Neuenschwander and P. Walker (2009). Annex 1B – Track maintenance and renewals costs in Switzerland. Deliverable 8, EU Sixth Framework Project CATRIN (Cost Allocation of Transport Infrastructure Cost).

Nash, C. (2003). UNITE (UNIfication of accounts and marginal costs for Transport Efficiency). Final Report for Publication. Funded by 5th Framework RTD Programme, Institute for Transport Studies, University of Leeds, Leeds, UK.

Nash, C. (2005). Rail Infrastructure Charges in Europe. Journal of Transport Economics and Policy, 39, 259-278.

Nash, C. and B. Matthews (2005). Measuring the Marginal Social Cost of Transport. Research in Transportation Economics, Vol. 14. Elsevier, Oxford, UK.

Nash, C. and B. Matthews (2009), European Transport Policy: Progress and Prospects, Community of European Railway and Infrastructure Companies (CER).

Nash, C. and T. Sansom (2001). Pricing European Transport Systems: Recent Developments and Evidence from Case Studies. Journal of Transport Economics and Policy, 35, 363-380.

Office of Rail Regulation (2008). Periodic Review 2008: Draft Determinations, June 2008.

Oum, T.H., Waters, W.G. (II) and Yu, C. (1999). A Survey of Productivity and Efficiency Measurement in Rail Transport. Journal of Transport Economics and Policy, 33 (I), 9-42.

Pinkse, J. and M.E. Slade (1998). Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models. Journal of Econometrics, 85, 125-154.

Smith, A. S. J. (2008). International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB Dataset (1996-2006). Report for the Office of Rail Regulation, Report written as part of PR2008, October 2008.

Smith, A. S. J., P. E. Wheat and H. Nixon (2008). International Benchmarking of Network Rail's Maintenance and Renewal Costs. Report written as part of PR2008, June 2008.

StataCorp. (2007). Stata Statistical Software: Release 10. Stata Press.

Thomas, J., F. Dionori and A. Foster (2003). EU Task Force on Rail Infrastructure Charging. Summary Findings on Best Practice in Marginal Cost Pricing. European Journal of Transport and Infrastructure Research, 3, 415-431.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. Econometrica, 26, 24-36.

Train, K. (1986). Qualitative Choice Analysis: Theory, Econometrics, and an Application to

Automobile Demand. MIT Press, Cambridge, Massachusetts.

Van de Ven, W. P. and B. M. van Praag (1981). Risk Aversion of Deductibles in Private Health Insurance: Application of an Adjusted Tobit model to Family Health Care Expenditures. In Health, Economics and Health Economics (J. van der Gaag and M. Perlman, eds), North Holland, Amsterdam, 125-148.

Wheat, P. and A. Smith (2009). Annex 1E – Renewals costs in Great Britain. Deliverable 8, EU Sixth Framework Project CATRIN (Cost Allocation of Transport Infrastructure Cost).

Wheat, P., Smith, A. and C. Nash (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), *Deliverable 8 - Rail Cost Allocation for Europe*. Funded by Sixth Framework Programme. VTI, Stockholm.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Mass.

Wooldridge, J. M. (2009). Introductory Econometrics. Fourth Edition. South-Western.