UDATING THE MARGINAL COST OF RAILWAY TRACK RENEWALS USING CORNER SOLUTION MODELS[1]

Jan-Eric Nilsson, Sherzod Yarmukhamedov, Kristofer Odolinski, Åsa Wikberg

1. INTRODUCTION

Much in the same way as in the road sector, there may be two different sources of marginal costs of railway wear. One concerns the impact of additional train services (cf. road vehicles) on the need for ongoing maintenance and the other the consequences of variations in train traffic for reinvestment spending, i.e. track renewal (resurfacing of roads).

The standard approach for addressing the first aspect, i.e. the impact of variations in track (road) use on day-to-day maintenance, is to regress costs against traffic and covariates in a cross section-time series model approach. If the regression coefficient(s) representing traffic is (are) significant, this signals that there indeed is an effect. Haraldsson (2007) makes use of this approach for roads and so does Swärdh & Jonsson (2014).

Based on Small et al (1987), there is since long an approach for addressing marginal reinvestment costs in the road sector. Using a large number of observations of road sections, Haraldsson (2007) and Nilsson et al (2014) base their papers on this approach. One reason for that this is not immediately transferrable to the marginal railway reinvestment costs is the far lower number of observations available for railway sections. Since renewal is a rare event over the life cycle of tracks-and-structures, this means that there is a large number of years with zero spending before a rather large amount of resources is spent on rehabilitation. This makes it inappropriate to use standard regression techniques for establishing the link between traffic and costs. Rather, an elaborate conceptual model for understanding what triggers track renewal as well as its costs has to be established before data processing can be initiated.

A track renewal is an activity that restores the infrastructure to its original standard. Renewals and maintenance are linked in so far as that lack of maintenance will force the infrastructure manager to renew at an earlier stage than if maintenance were undertaken properly and vice

---

[1] This paper is produced as an input for a government assignment to VTI to assess the social marginal costs for infrastructure use.

versa. An optimally managed railway track therefore has a mix of maintenance and renewal over its life cycle, and excluding renewals from the total picture of marginal infrastructure costs is therefore misleading.

The present paper is based on Anderson et al. (2012) which explores three different regression models to analyze track renewal data, the Tobit, the two-part and the Heckit models. Both on theoretical and empirical grounds they establish that the two-part model is the preferred approach. The first part of this model concerns the variables that have an impact on the decision to renew tracks; the second an estimate of marginal costs given that this decision has been made. This paper is subsequently referred to as ASWW after the surnames of the authors.

The purpose of the present paper is to provide a new estimate of the marginal cost of track renewals. Except for updating ASWW with a larger number of observations (more years), the previous paper did not have access to information about the extent of traffic at station areas. Since this information now has been made available, it is feasible to generalize results in this dimension. Still based on the same model, the paper also considers the possibility to add new, or replace previous variables in order to better understand what drives decisions to undertake track renewals and the costs for doing so.

The paper starts with a summary of the ASWW model in section 2. Section 3 describes the extended dataset and section 4 makes a stepwise analysis of the updating of data and the replacement and addition of more observations.

2. MODELLING APPROACH

ASWW base their approach on the corner solution model. This approach is relevant for situations where a decision-maker makes an (observable) choice for a variable, y, which takes the value zero (the corner solution) with a positive probability, and otherwise is a continuous, strictly positive random variable. The zeroes are not censored versions of some underlying variable, they are “true” zeroes, since they are the actual choices made by the decision maker.

The corner solution model is appropriate to use in the present context since spending on track renewal may be positive or zero any given year, depending on the choice of the infrastructure manager. The first step of the two-part model is a standard probit model:

$z_i^* = x'_{1i}\beta_1 + u_{1i}$

$I_i = 1$ if $z_i^* > 0$

$I_i = 0$ otherwise

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

I_i is a binary choice variable, taking the value 1 (or zero) when a decision is taken to implement a reinvestment (or not). z_i^* is a latent variable, i.e. it is unobservable in this context, underpinning the decision whether to renew or not. The second part is then a truncated regression model:

$$y_i|(I_1 = 1) = x'_{2i}\beta_2 + u_{2i}$$

Here, the random term E(u_{2i}|$\mathrm{I_i} = 1$)=0 but u_{2i} need not be normally distributed. The model implies that the value of *y*, which here is expenditure on new tracks, 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 decision to renew or not may therefore be based on other information than the decision of how much to spend. In this way, x_{1i} need not equal x_{2i}. Even if the regressors are the same, the coefficients can be different (β_1 need not equal β_2). Some of the candidate regressors may be statistically significant (may provide an explanation for the decision) in one equation but not in the other or take different signs.

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 zeroes are true zeroes. Correlation in residuals could still result due to correlation between unobserved effects in each part of the model. It is therefore an empirical matter as to whether the assumption of independent errors is reasonable.

In order to calculate marginal costs, the first challenge is to establish an appropriate measure of traffic. Road wear may be caused by traffic at large but in particular by heavy vehicles (Nilsson et al 2014). In railways, a train may be short (railcar on secondary lines, it may be long and heavy or it may be long and light (loaded and empty freight train, respectively). The standard way to measure usage has therefore come to be the number of tons of trains using a railway line. More specifically, the number of gross tons per track kilometer, or tonnage density, provides our proxy for traffic. We also test for tonnage density between freight and passenger trains, respectively, in order account for the possibility that the impact of a certain tonnage density on tracks differs between the two categories.

The equation for cost elasticity for track section *i* with respect to tonnage density *(k)* is given:

$$\gamma_{ik} = \frac{\partial E[y]}{\partial x_k} \times \frac{x_k}{E[y]} = \beta_{2k} + \beta_{1k}\lambda\left(x_1'\hat{\beta}_1\right)$$

β_{1k} and β_{2k} are coefficient estimates for tonnage density *(k)* from the first and the second stages, respectively; $\lambda\left(x_1'\hat{\beta}_1\right)$ is inverse Mills ratio which equals $\frac{\phi(x_1'\hat{\beta}_1)}{\Phi(x_1'\hat{\beta}_1)}$, where ϕ(.) and Φ(.) are the probability density and cumulative distribution functions, respectively. This is derived from the estimation of a probit model in the first stage. Using the elasticity estimate, the equation for computing marginal cost is:

$$MC_{ik} = \gamma_{ik}\widehat{AC}_i$$

$$\widehat{AC}_i = \frac{\Phi\left(x_1'\hat{\beta}_1\right) exp\left(x_2'\hat{\beta}_2 + 0.5\sigma_2^2\right)}{GTKM_i}$$

$\widehat{AC}$ is the predicted average renewal cost; σ is the estimated standard error of second stage regression model; *GTKM* is a gross tonne km. A weighted marginal cost estimate is computed to take into account a traffic share per track section, so that marginal cost weighted by gross tonne km per track section is:

$$MC_{ik}^{W} = MC_{ik}\frac{GTKM_i}{\sum_i GTKM_i}$$

3. THE DATA

Sweden's railway network is 16 500 km long. Sweden's infrastructure manager *Trafikverket* (the Swedish National Traffic Administration), administers about 14 700 km and records information about costs for track maintenance and renewal as well as "output" (i.e. traffic) and network characteristics. The smallest level of observation of costs is the track section and the network comprises some 260 track sections.

Track sections vary greatly in structure. For example, they range in length from 1.5 to 290.6 kilometres, and in age from one to 65 years. The amount of traffic load also differs. Table 1 provides information both from the paper by ASWW and numbers used in the present paper. This difference is of importance for the analysis for several reasons. First, more years obviously generate more observations. Secondly, and as indicated above, data now includes additional observations of track sections that are railway stations. This has retroactively been added for all years. The third difference is that information about some track sections which were previously not available has become accessible for analysis. The new information in the first place means that the rather comprehensive picture of the core parts of the network that was previously available now is extended by older parts. Consequently, the average age of tracks and structures increases, irrespective of that substantial resources have been spent on reinvestment. In addition, and fourth, some sections that were included in the ASWW analysis have now been deleted due to obvious data inconsistencies.

Information about traffic per track section is still partly missing, mainly for sections with little traffic and sections used for industrial purposes. Marshalling yards have been excluded since the cost structure at these places can be expected to differ from track sections at large. Neither are privately owned sections, heritage railways, nor sidings and track sections that are closed for traffic included in the dataset.

Table 1. Descriptive statistics, average per track section.

	ASWW 1999-2009	Present study 1999-2012
No. of track units	162	194
Track renewal cost, m SEK*	3.5	3.3
Section length (m)	77 801	68 708
Tonnage density (gross tons per route)	7 183 785	7 908 212
No. of trains	15 584	17 329
Bridge length (m)	649	704
No. of joints	173	166
No. of switches	52	99
Switch length (m)	1 576	1 760
Switch age (years)	20	20
Rail weight (kg)	51	51
Rail age (years)	20	20
No. of stations	-	23

* ASWW report results in 2009 year prices. Here and subsequently this has been inflated to 2012 values by 4.9 percent representing change in CPI.

Most railway stations are part of a track section and a track section may include more than one station. In some cases, a train station is however defined to be a track section of its own. These sections have not been included in previous cost analyses due to missing information about traffic. Odolinski (2014) describes an algorithm which estimates traffic at stations based on information about traffic on adjacent track sections. This means that an additional 24 cross section observations are included each year.

While the network comprises 255 track sections, the elimination of some observations and the still missing information of other track sections results in 194 sections being available for analysis for the 1999 to 2012 period. While a comprehensive matrix would comprise (255 * 14 years =) 3570 observations, the following analysis is based on 2507 observations.

Table 1 demonstrates that average spending on each renewal is now slightly lower than before. Also section length is lower than before, one possible reason being the inclusion of track units that are railway stations. The same explanation may be relevant for the increase in the average number of switches.

4. ECONOMETRIC RESULTS

The presentation of results is done in two steps. First, section 4.1 basically recycles the ASWW model but includes the additional number of years as well as the stations. Section 4.2 then considers the pros and cons of adding or changing explanatory variables. All estimations are done in Stata 10 (StataCorp, 2007).

4.1 Model specification

As independent variables, we use the logs of track section length, total gross tons 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. A log-linear (or double-log) functional form is used. Dummy variables are used for all years between 1999 and 2012 except for year 2000, which is then the year of reference.[2] Other variables in Table 1 have not been found to improve our models.

---

[2] This slightly unorthodox way to choose year of comparison is motivated by that spending on track renewal was much lower in year 1999 than in subsequent years.

The purpose of table 2 is to compare results from the present analysis with ASWW. A first observation is that the difference between our new estimations with and without stations is very small. Since the overall purpose is to capture marginal costs for the network at large, it is obvious that the 24 station areas shall be included in the subsequent analyses.

Comparing the new estimation with results from ASWW with respect to the selection equation, differences are not large. In particular, the coefficient in focus here, tonnage density, diverges very little. Results deviate much more with respect to the outcome equation. Except for that the coefficient of tonnage density is substantially reduced, the significance of the new estimate is much lower. One possible reason for this is that ASWW has incorrectly used the entire sample in the outcome equation, while a subsample where actual reinvestment decisions were made (i.e. positive reinvestment costs are incurred) should have been used. Despite the imprecise estimation in ASWW, the table demonstrates that the marginal cost estimate is almost the same while elasticity is slightly lower compared to ASWW.

Table 2: Comparing Results

	New estimation: without stations		New estimation: with stations		ASWW	
	Coef.	SE	Coef.	SE	Coef.	SE
Selection equation						
Ln (section length)	0.5010***	0.0612	0.4675***	0.0515	0.4541***	0.0513
Ln (tonnage density)	0.2777***	0.0480	0.2973***	0.0441	0.3023***	0.0359
Ln (switch age)	0.2243**	0.1059	0.2144**	0.0995	0.2196**	0.0930
Ln (rail weight)	-1.0640	0.7173	-1.2432*	0.6837	-1.0832**	0.5210
West region	Reference category					
North region	-0.0134	0.1672	0.0905	0.1575	0.2416**	0.1230
Central region	0.0392	0.1394	0.0347	0.1317	-0.0021	0.1097
South region	0.1712	0.1411	0.1775	0.1303	0.2085*	0.1101
East region	0.2719*	0.1444	0.2408*	0.1351	0.2841**	0.1130
Year 1999	0.1978	0.1345	0.1575	0.1297	Reference category	
Year 2000	Reference category				-0.1237	0.1630
Year 2001	0.3060**	0.1324	0.3173**	0.1270	0.1883	0.1630
Year 2002	0.7975***	0.1416	0.7522***	0.1393	0.6245***	0.1577
Year 2003	0.7889***	0.1453	0.7620***	0.1442	0.5999***	0.1609
Year 2004	0.7981***	0.1457	0.7951***	0.1440	0.6335***	0.1625
Year 2005	0.7742***	0.1579	0.7583***	0.1545	0.5581***	0.1654
Year 2006	1.0037***	0.1538	0.9648***	0.1503	0.7791***	0.1671
Year 2007	1.0336***	0.1544	1.0505***	0.1490	0.8149***	0.1654
Year 2008	1.0379***	0.1421	1.0208***	0.1414	0.9502***	0.1640
Year 2009	1.3624***	0.1677	1.3542***	0.1612	1.3071***	0.1697
Year 2010	0.7849***	0.1658	0.7539***	0.1558		
Year 2011	0.6755***	0.1599	0.6272***	0.1533		
Year 2012	0.2527	0.1588	0.2356	0.1551		
Constant	-7.2032***	2.4261	-6.3720***	2.3146	-6.9080***	1.8680
Outcome equation						
Ln (section length)	0.7648***	0.1756	0.6977***	0.1393	0.7878***	0.1446
Ln (tonnage density)	0.1476	0.1152	0.1692*	0.1013	0.2846***	0.0965
Ln (switch age)	0.2258	0.2771	0.2111	0.2553	0.5445**	0.2704
Ln (rail weight)	-1.5586	1.6881	-1.4726	1.6296	-2.4683*	1.3617
West region	Reference category					
North region	0.1589	0.4261	0.0592	0.3821	0.1063	0.3122
Central region	-0.1336	0.3766	-0.2307	0.3629	-0.1212	0.3094
South region	0.2596	0.3043	0.1702	0.2892	0.4938*	0.2995
East region	-0.1748	0.3043	-0.2987	0.2933	-0.3681	0.2771
Year 1999	0.3124	0.6332	0.4151	0.6027	Reference category	
Year 2000	Reference category				-0.1852	0.6579
Year 2001	0.4089	0.6253	0.5350	0.5843	0.2243	0.5047
Year 2002	-0.6126	0.5979	-0.4175	0.5703	-0.6515	0.4849
Year 2003	-0.2861	0.5930	-0.2155	0.5609	-0.1897	0.4801
Year 2004	0.1022	0.5976	0.1911	0.5673	-0.0660	0.4664
Year 2005	-0.0887	0.6061	0.1333	0.5768	-0.2310	0.4891
Year 2006	-0.1582	0.5873	-0.1570	0.5597	-0.3769	0.4766
Year 2007	-0.4979	0.6001	-0.3216	0.5636	-0.4841	0.4655
Year 2008	-0.4702	0.5935	-0.3784	0.5604	-0.6968	0.4416
Year 2009	-0.2083	0.5909	-0.1693	0.5581	-0.4201	0.4408
Year 2010	-0.4427	0.6469	-0.3004	0.6175		
Year 2011	0.3330	0.6113	0.5210	0.5816		
Year 2012	0.0245	0.6624	0.0443	0.6254		
Constant	8.5280	5.8574	8.6538	5.7432	8.9866*	4.8202
Elasticity	0.4275***	0.1406	0.4709***	0.1246	0.547***	0.105
Marginal cost	0.0094	0.0002	0.0098	0.0002	0.009	0.002

***, **, * Significant at 1%, 5% and 10%, respectively.

While table 2 only provides a comparison with previous results, some further tinkering with the model specification may be relevant, and the following specifications are therefore tested. Model I makes use of the entire sample and includes a station dummy. Model II excludes all year dummies from Model I, and compares the AIC values of Model I and Model II. The model with the lowest AIC value should be preferred. The third model specification (Model III) is built according to the AIC values of Model I and Model II and also includes the pertinent elasticity and marginal cost estimates.

By comparing AIC values in table 3, Model I is preferred for the selection equation, while in the outcome equation Model II should be chosen. This implies that our preferred version is Model III which comprises the first part of Model I and the second part of Model II: the selection of track sections is better understood by including year dummies, whereas the explanatory power of the outcome equation is not improved by adding period effects.[3] The preferred model generates an elasticity estimate 0.43 and marginal cost estimate 0.0089. These values are slightly lower compared to the table 2 results.

As for other results, we can note, first, that the length of a track section is included in both the selection and outcome parts of the model. In the first stage, a longer track section is more likely to see part of the section renewed. In the second stage, section length is a proxy for the size of the renewal undertaken. This possibility is, however, tested in another way in the next section. In both stages we expect the coefficient on section length to be positive.

A second observation is that in the first stage, (annual) tonnage density is acting as a proxy for cumulative tonnage; a subsequent version of the paper will seek to use instead the accumulated weight of trains. The expected positive effect of tonnage density in the second stage may rather derive from expected future tonnage; higher expected usage would trigger higher standard and cost of a renewal. To the extent that current tonnage is a good proxy for both past and future traffic, this distinction may be unimportant. Regional and year dummy variables are included to capture unobserved heterogeneity between sections, budget fluctuations and other time trends, but with no a priori expectation on signs

---

[3] The same conclusion holds for the base model results in Table 2, however, complete results with corresponding AIC values are not presented to save a text space. Besides, in Table 2, the choice of covariates is the same as in ASWW for comparison purposes.

Table 3. Estimation results: different model specifications

Row		Model I		Model II		Model III	
		Coef.	SE	Coef.	SE	Coef.	SE
1	*Selection equation*						
2	Ln (section length)	0.4948***	0.0565	0.4524***	0.0517	0.4948***	0.0565
3	Ln (tonnage density)	0.2839***	0.0466	0.2693***	0.0446	0.2839***	0.0466
4	Ln (switch age)	0.2110**	0.0988	0.3223***	0.0926	0.2110**	0.0988
5	Ln (rail weight)	-1.1481*	0.6890	-0.6707	0.6265	-1.1481*	0.6890
6	West region	Reference category					
7	North region	0.0852	0.1565	0.0695	0.1480	0.0852	0.1565
8	Central region	0.0421	0.1325	0.0193	0.1243	0.0421	0.1325
9	South region	0.1805	0.1300	0.1855	0.1210	0.1805	0.1300
10	East region	0.2510*	0.1360	0.2367*	0.1247	0.2510*	0.1360
11	Year 1999	0.1614	0.1293			0.1614	0.1293
12	Year 2000	Reference category					
13	Year 2001	0.3183**	0.1271			0.3183**	0.1271
14	Year 2002	0.7553***	0.1390			0.7553***	0.1390
15	Year 2003	0.7645***	0.1438			0.7645***	0.1438
16	Year 2004	0.7978***	0.1435			0.7978***	0.1435
17	Year 2005	0.7603***	0.1543			0.7603***	0.1543
18	Year 2006	0.9661***	0.1500			0.9661***	0.1500
19	Year 2007	1.0513***	0.1488			1.0513***	0.1488
20	Year 2008	1.0219***	0.1412			1.0219***	0.1412
21	Year 2009	1.3566***	0.1608			1.3566***	0.1608
22	Year 2010	0.7548***	0.1557			0.7548***	0.1557
23	Year 2011	0.6273***	0.1532			0.6273***	0.1532
24	Year 2012	0.2342	0.1553			0.2342	0.1553
25	Stations (dummy)	0.1574	0.1207	0.1460	0.1120	0.1574	0.1207
27	Constant	-6.8519***	2.3446	-7.6521***	2.1261	-6.8519***	2.3446
28	AIC	2900.5539		3046.4248			
29	*Outcome equation*						
30	Ln (section length)	0.7580***	0.1555	0.7764***	0.1579	0.7764***	0.1579
31	Ln (tonnage density)	0.1365	0.1110	0.1422	0.1107	0.1422	0.1107
32	Ln (switch age)	0.2047	0.2564	0.1400	0.2498	0.1400	0.2498
33	Ln (rail weight)	-1.2617	1.6404	-1.4726	1.6086	-1.4726	1.6086
34	West region	Reference category					
35	North region	0.0414	0.3856	0.0296	0.3893	0.0296	0.3893
36	Central region	-0.2118	0.3626	-0.1963	0.3659	-0.1963	0.3659
37	South region	0.1712	0.2837	0.1390	0.2883	0.1390	0.2883
38	East region	-0.2823	0.2873	-0.2792	0.2931	-0.2792	0.2931
39	Year 1999	0.4439	0.6027				
40	Year 2000	Reference category					
41	Year 2001	0.5317	0.5853				
42	Year 2002	-0.4077	0.5698				
43	Year 2003	-0.2036	0.5607				
44	Year 2004	0.1910	0.5683				
45	Year 2005	0.1313	0.5771				
46	Year 2006	-0.1534	0.5606				
47	Year 2007	-0.3268	0.5649				
48	Year 2008	-0.3780	0.5605				
49	Year 2009	-0.1709	0.5582				
50	Year 2010	-0.2991	0.6181				
51	Year 2011	0.5209	0.5813				
52	Year 2012	0.0353	0.6258				

53	Stations (dummy)	0.3454	0.2782	0.3435	0.2814	0.3435	0.2814
54	Constant	7.6369	5.7745	8.3055	5.6392	8.3055	5.6392
55	AIC	4841.3912		4831.9967			
56							
57	Elasticity	0.4246***	0.1348	0.4078***	0.1335	0.4304***	0.1351
58	Marginal cost	0.0087	0.0002	0.0088	0.0002	0.0089	0.0002

***, **, * Significant at 1%, 5% and 10%, respectively.

4.2 Changing explanatory variables

As researchers, we are stuck with information that happens to be available for empirical analysis; this is not necessarily the data that would be used if data collection could start from scratch. Since some information that previously was not available now has become accessible, it is reason to examine if this could be used for further improving the quality of the estimations.

The first additional piece of information refers to reported errors. *Trafikverket* collects information about manually reported errors on a track section level. Examples include fire, animal accidents, material wear and tear/ageing, faulty maneuvering, abnormal temperatures, loose details etc.; cf. further Odolinski 20xx. Observations are, however, only available from year 2003, slightly reducing the size of our panel.

The total number of errors per year and track unit (*NRERR*) is used as an additional explanatory variable to explain reinvestment decisions in the first step of the model. The expectation is that more errors will increase the stage probability of renewal. The following causes are included when registering the number of *NRERR* observations per track unit and year: insufficient maintenance, geotechnical problems, loose details, electronic malfunctioning, materials fatigue or ageing, improper construction, improper assembly, unexpected electric or mechanic strain, rust/verdigris, rail displacement and improper insulation.

NRERR is obviously a count variable that includes zeroes and that is right-skewed. For this reason, we transform this variable by decomposing into two parts: First, a dummy which indicates zero values of *NRERR*, i.e. *NONRERR*=1 if *NRERR*=0, 0 otherwise; Second, we log transform *NRERR*, i.e. ln(*NRERR*), replacing resulting missing values (due to zeros) after transformation with the lowest value of ln(*NRERR*). At the end, we include *NONRERR* and ln(*NRERR*) variables into the model instead of *NRERR*[4].

---

[4] We perform another type of transformation as well to check for the sensitivity of the results to the transformation method. A variable NRERR is log transformed using a constant (C), i.e. ln(NRERR)= ln (NRERR + C) in order to avoid missing values after log transformation in cases where NRERR is zero. As a new transformed variable ln(NRERR) is

Table 4 indicates that it is justified to include this additional variable since the AIC value is lower than if it is not included. The LN(NRERR) coefficient is positive and significant: The larger the number of errors, the higher the probability that tracks will be upgraded. Interestingly, adding the number of errors to the estimation results in lower estimates of elasticity and the marginal cost is almost halved, 0.26 and 0.0048, respectively, compared to the results in Table 3 (0.43 and 0.0089).

Trafikverket also measures track geometry on a regular basis using a track measurement car. Errors are categorized into three classes: A-, B-, and C-class errors, with the latter being most severe that require remedial action short after being registered (see Odolinski 2013 for details). This information is, however, only available for 131 track sections over the period 2002-2012. Including the number of C-class errors as an additional explanatory variable for those parts of the network where information is available however turned out to be insignificant in explaining reinvestment decisions. Results are suppressed to save space. Since observations of a large number of severe problems with respect to track geometry can be expected to increase the probability for track renewal, it is reason to test the explanatory power of the variable once data is more complete.

---

an explanatory variable, the choice of C is not crucial for the analysis. Here, we let C=1. The results (suppressed) provide similar coefficient estimates in both transformation methods.

Table 4: Two-part model results with no. of errors in the selection equation, 2003-2012

Row		Without errors		With errors	
		Coef.	SE	Coef.	SE
1	*Selection equation*				
2	Ln (section length)	0.5188***	0.0653	0.4736***	0.0712
3	Ln (tonnage density)	0.2458***	0.0485	0.2205***	0.0506
4	Ln (switch age)	0.2360**	0.1199	0.2492**	0.1241
5	Ln (rail weight)	-1.3539*	0.7118	-1.1339	0.7471
6	West region	Reference category			
7	North region	-0.2000	0.1729	-0.2118	0.1736
8	Central region	-0.0762	0.1487	-0.1175	0.1483
9	South region	0.1734	0.1490	0.2028	0.1496
10	East region	0.3218**	0.1544	0.3130**	0.1558
11	Year 2003	Reference category			
12	Year 2004	0.0329	0.1076	0.0480	0.1085
13	Year 2005	-0.0084	0.1262	0.0007	0.1252
14	Year 2006	0.1945	0.1390	0.1944	0.1390
15	Year 2007	0.2840*	0.1446	0.2669*	0.1456
16	Year 2008	0.2546*	0.1374	0.2489*	0.1377
17	Year 2009	0.5878***	0.1461	0.5815***	0.1477
18	Year 2010	-0.0091	0.1454	-0.0210	0.1495
19	Year 2011	-0.1326	0.1468	-0.1870	0.1517
20	Year 2012	-0.5220***	0.1491	-0.5613***	0.1535
21	Stations (dummy)	0.2050	0.1410	0.1356	0.1473
22	Constant	-4.9955**	2.5033	-5.1468**	2.5679
23	Ln (number of errors)			0.0832*	0.0467
24	No error (dummy)			0.0298	0.1263
25	AIC	2190.8272		2159.0404	
26					
27	*Outcome equation*				
28	Ln (section length)	0.8048***	0.1639	0.8048***	0.1639
29	Ln (tonnage density)	0.0645	0.1131	0.0645	0.1131
30	Ln (switch age)	-0.0104	0.2536	-0.0104	0.2536
31	Ln (rail weight)	-1.4715	1.6378	-1.4715	1.6378
32	West region	Reference category			
33	North region	-0.0550	0.4257	-0.0550	0.4257
34	Central region	-0.2388	0.3820	-0.2388	0.3820
35	South region	0.1596	0.2893	0.1596	0.2893
36	East region	-0.2292	0.3017	-0.2292	0.3017
37	Stations (dummy)	0.3939	0.2927	0.3939	0.2927
38	Constant	9.6088*	5.6975	9.6088*	5.6975
39	AIC	3902.395		3902.395	
40					
41	Elasticity	0.2848**	0.1335	0.2600**	0.1311
42	Marginal cost	0.0052	0.0001	0.0048	0.0001

***, **, * Significant at 1%, 5% and 10%, respectively.

A third set of data that now is available is the length-of-tracks that are given a treatment, i.e. the extent of renewal activities actually implemented. This is used for replacing the section length variable now used in the outcome equation. Section length is significant in all previous models, signaling that it is more costly to undertake renewal activities the larger (longer) the track section. This is now substituted by the actual number of meters new track that is being replaced. Bearing in mind that the complete track section rarely is replaced but rather a fraction of the full length, the intuition is to check if this is not a better way to model the stage 2 estimation of costs. The length-of-tracks being renewed is captured as (*NEWRAIL*) in table 5. This variable also includes many zeroes, and is therefore transformed in the same way as *NRERR* variable is done.

Table 5. Two-part model results with no. of errors in the selection equation and a renewed section length in the outcome equation, 2003-2012

Row		Coef.	SE
1	*Selection equation*		
2	Ln (section length)	0.4736***	0.0712
3	Ln (tonnage density)	0.2205***	0.0506
4	Ln (switch age)	0.2492**	0.1241
5	Ln (rail weight)	-1.1339	0.7471
6	West region	Reference category	
7	North region	-0.2118	0.1736
8	Central region	-0.1175	0.1483
9	South region	0.2028	0.1496
10	East region	0.3130**	0.1558
11	Year 2003	Reference category	
12	Year 2004	0.0480	0.1085
13	Year 2005	0.0007	0.1252
14	Year 2006	0.1944	0.1390
15	Year 2007	0.2669*	0.1456
16	Year 2008	0.2489*	0.1377
17	Year 2009	0.5815***	0.1477
18	Year 2010	-0.0210	0.1495
19	Year 2011	-0.1870	0.1517
20	Year 2012	-0.5613***	0.1535
21	Stations (dummy)	0.1356	0.1473
22	Constant	-5.1468**	2.5679
23	Ln (number of errors)	0.0832*	0.0467
24	No error (dummy)	0.0298	0.1263
25	AIC	2159.0404	
26			
27	*Outcome equation*		
28	Ln (renewed section length)	0.3400***	0.0850
29	No renewal (dummy)	0.4080	0.4038
30	Ln (tonnage density)	0.0387	0.1103
31	Ln (switch age)	0.0998	0.2172
32	Ln (rail weight)	0.0129	1.5421
33	West region	Reference category	
34	North region	-0.1536	0.3976
35	Central region	-0.2074	0.3659
36	South region	0.0648	0.3125
37	East region	-0.3289	0.3313
38	Stations (dummy)	-0.3959	0.2606
39	Constant	11.4909**	5.5426
40	AIC	3873.953	
41			
42	Elasticity	0.2342*	0.1284
43	Marginal cost	0.0046	0.0002

***, **, * Significant at 1%, 5% and 10%, respectively.

AIC values in the outcome equation in Table 4 (row 39, AIC=3902.395) and Table 5 (row 40, AIC=3873.953), suggest that a model with renewed track length better explains renewal costs in the outcome equation. Therefore, the results in Table 5 show that a renewed section length variable is positive and significant which implies that an increase in track section renewal length by 1 percent leads to a 0.34 percent increase in reinvestment costs.

Table XXX. Summary of the results for railway track renewals

	AC	Elasticity	Marginal cost
ASWW	3.5	0.547***	0.009
New estimation: Period 1999-2012	3.3	0.4304***	0.0089
New estimation: Period 2009-2012	4.5	0.2323	0.0052

Note: AC stands for average renewal cost, million SEK; AC value for ASWW has been inflated to 2012 values by 4.9 percent representing change in CPI.

***, **, * Significant at 1%, 5% and 10%, respectively.

5. Summary

This paper provides new estimates of the how variations in train traffic affects costs for track renewal. The best results for a panel of data covering the 1999-2012 period is that the elasticity is 0.43 and that the marginal cost is SEK 0.0089 per gross ton km. This is very close to previous estimates, primarily comparing our benchmark reference ASWW.

We have, however, also sought to extend the platform for the analysis by amending and changing explanatory variables. As a result elasticity is now almost halved to 0.23 and so is also marginal cost with a new estimate of SEK 0.0046 per gross ton km.

The reason for this drastic change does, however, not primarily originate in the change of explanatory variables. It is rather due to that the new variables only date back to 2003, making it necessary to reduce the size of the panel. This, in turn, indicates the possibility in substantial changes in the way that maintenance was implemented during the first years of the period compared to the situation in the latter years. In view of uncertainties with respect to data quality, it is not feasible to base any conclusions on the new results. There is, therefore, strong reason to address this and other question-marks observed in the above review in further research.

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. Bjorklund (forthcoming). Marginal Railway Renewal Costs: A Survival Data Approach. VTI Working Paper.

Andersson, M., A. Smith, Å. Wikberg, P. Wheat (2012). ESTIMATING THE MARGINAL COST OF RAILWAY TRACK RENEWALS USING CORNER SOLUTION MODELS. Transportation Research Part A, 46, 954-964.

Banverket (2009). Network Statement 2011. Borlange, 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.

Odolinski, K. (2013). Track geometry quality 2002-2012. Memorandum, VTI

Odolinski, K. (2014). PM:Trafikdata 1999-2012, järnväg. Memorandum, VTI (In Swedish)

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
22
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.