Centre on Regulation in Europe

Improving network and digital industries regulation

Track access charges: reconciling conflicting objectives

Case Study – Great Britain

Prof. Andrew Smith, University of Leeds

Prof. Chris Nash, CERRE & University of Leeds

9 May 2018

Centre on Regulation in Europe (CERRE) a.s.b.l.
Avenue Louise 475 Box 10|B-1050 Brussels|Belgium
+32 2 230 83 60|info@cerre.eu|www.cerre.eu

Table of contents

1. Introduction

In Britain, the approach taken to track access charges when infrastructure was separated from operations in 1994 was to charge a variable usage charge for wear and tear, whilst franchisees (and most passenger services were franchised) were also charged a fixed charge based on an allocation of remaining costs according to a formula based on train kilometres and revenue. Electricity was charged for separately, and there was also an electrification asset usage charge to cover wear and tear on the overhead line or third rail. Stations were leased to the main operator at that station, with a share of costs being recovered from other operators in proportion to train calls and numbers of passengers. Fixed charges were reduced in 2001 when Network Grant, a direct grant from the government to the infrastructure manager was introduced. Originally, a mark-up was applied to freight train kilometres on the basis of individual negotiations, but this was abandoned in 2001 as potentially discriminatory, and subsequently a mark-up applied to recover the avoidable fixed costs of freight operations (such as freight only lines) on the basis of ability to pay. Also, a capacity charge was introduced to reflect congestion costs. The latest proposal is to abandon the capacity charge, but to introduce mark-ups for passenger as well as freight services on the basis of ability to pay. Charges to cover external costs of accidents and environmental impacts have been consistently rejected on the basis that other modes do not pay such costs. Indeed, there is a grant scheme which contributes to the track access costs of certain freight operators on the basis of the reduced externalities resulting from attraction of traffic from road. Finally, stations and depots are mainly leased to a single operator, with other operators paying in proportion to the use they make of them.

The current level of Network Rail costs and the contribution made to them by each type of charge is shown in Table 1. It will be seen that charges in total cover a small proportion of the total cost of operation, maintenance and renewals, and that the variable usage charge designed to cover variable maintenance and renewals costs contributes only a small proportion of the total revenue from charges. This report will consider in turn: charges for wear and tear, charging for capacity and mark-ups, before reaching its final conclusions. In each case, the current charging system will be described, the evidence on which it is based outlined and other relevant research considered.

Table 1: Rail infrastructure cost coverage in Britain (2016/7) (£m)

Costs		**Revenue from track access charges**	
Operations	554	Variable usage charge	224.2
Maintenance	1319	Capacity charge	428.3
Renewals	2774	Fixed charges	410.8
		Use of electrification assets	16.1
		Stations and depots	353.0
		TOTAL	1432.4

Source: Network Rail (2017a) Statement 1 (expenditure), Statement 6a (Analysis of Income) and Statement 6c (Analysis of Income by Operator)

2. Charges for wear and tear

2.1 Introduction

The key distinguishing features of the charging system for wear and tear in Britain is the degree of differentiation that exists between charges for different vehicle types running on the network. This differentiation is based on detailed engineering models that assess the relative damage done by different types of vehicle and in principle should incentivise the use of more track friendly vehicles. The overall methodology used can therefore be seen as strongly quantitatively based, though as will be explained there remains an element of judgement in the determination of cost variabilities and, in turn, charges.

The approach to wear and tear charging has evolved since the creation of a separate rail infrastructure manager in 1994. More recently, one issue that has arisen is whether the overall level of charges is too low, particularly as compared to the top-down econometric evidence that exists on the variability of maintenance and renewals costs with respect to traffic. Even a highly differentiated charging system based on a very low level of variable charges may therefore be limited in its effectiveness in terms of incentivising efficient use of the infrastructure and encouraging whole system cost minimisation.

As noted elsewhere in this note, in Britain there is currently a consultation on whether to introduce mark-ups on variable charges. One motivation is to ensure that open-access operators are contributing to the fixed costs of the network. The arguments concerning this development are discussed elsewhere – however, it is worth noting that, to the extent that existing variable charges are too low currently, one way of ensuring that open access operators (and indeed all operators) contribute more towards the cost of running the rail network would be to ensure that variable charges are calibrated correctly.

The remainder of this section is structured as follows. In section 2.2, we briefly set the context, explaining how wear and tear related variable charges were set prior to a major change in methodology that was implemented into the charging framework in April 2009, following the 2008 Periodic Review of Network Rail's finances. In section 2.3, we then explain the current basis for determining wear and tear-related variable charges in Britain, implemented in April 2009 (there have been no fundamental changes to this approach since then). Finally, in section 2.4 we discuss how the level of usage-related variability within the British charging system compares with the top-down econometric evidence that informs charging in some other countries, most notably France.

2.2 Variable track access charges from privatisation to 2008

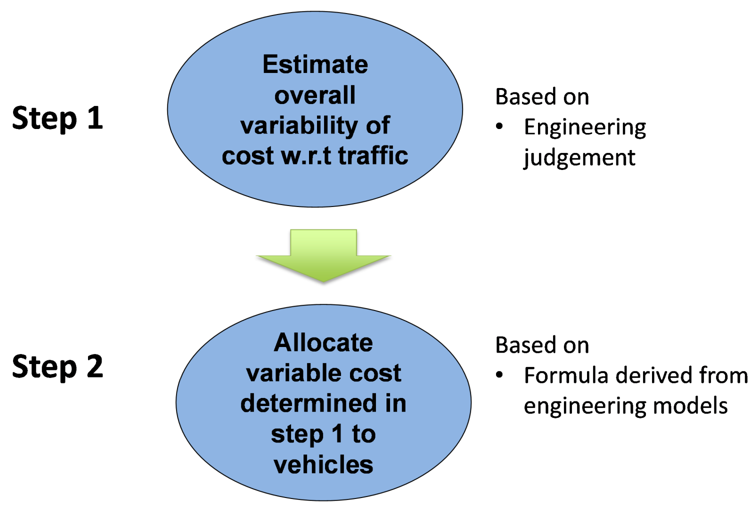
When Railtrack was privatised, the charging structure was set such that access charges were largely fixed (roughly 91% fixed, with the balance relating to wear and tear and electricity for traction (EC4T)). One feature of the planning process at privatisation (at least on the infrastructure side of the railway) was the assumption of a static or even declining railway; thus having a high level of fixed charges was seen as a way of guaranteeing a stable income for the infrastructure manager and thus aiding the privatisation process.

However, it was seen during the first few years after privatisation that passenger numbers and train movements on the network were increasing very substantially. As the Office of Rail Regulation (ORR)[1] prepared for its first Periodic Review of Railtrack's finances in 2000 (PR2000), it was therefore apparent that the charging framework was not fit for purpose.

Amongst a number of changes, a new methodology was introduced for computing the variable usage charge. The variable usage charge is intended to ensure that the rail infrastructure manager can recover any short-run variations in operations, maintenance and renewals costs that are variable with traffic. The term short-run is used because it is assumed that the capability of the network remains unchanged, or if changes are made such changes would be funded through other means.

The new methodology adopted at that time is summarised in Figure 1 and operated as two distinct steps. In the first step, the overall variability proportions of different elements of cost with respect to traffic were estimated. This was based on engineering judgement and knowledge of operating, maintenance and renewal practices. The assumed variability proportions are shown in Table 1 (along with the changes made in the 2008 Periodic Review for comparison; the latter is discussed in section 2.3 below).

Figure 1: 2000 Periodic Review Methodology for Setting Variable Usage Charges



[1] Now re-named as the Office of Rail and Road (ORR).

Table 2: Cost variability proportions by activity / asset class

Activity / asset class	Variability Proportion: PR2000	Variability Proportion: PR2008*
Track - maintenance	30%	29%
Track – renewals (plain line)	36%	23%
Track – renewals (switches and crossings)	25%	17%
Signalling - maintenance	5%	5%
Civils – metallic underbridges	10%	8%
Civils – embankments	10%	5%

* The proportions introduced following PR2008.

Source: ORR (2008)

It should be noted that the categories included in Table 2 are those categories that are assumed to vary with traffic to some degree. All other cost categories, including, for example, signalling operations are assumed not to vary with traffic in the short-run. One category has been excluded, namely electrification assets. At PR2000 this charge was levied as a mark-up to the EC4T charge, meaning that it was based on the amount of electricity used rather than the volume of traffic. As noted below, this policy was changed at PR2008.

As a result of the changes, variable usage charges more than doubled (see Figure 2). At that time it was estimated that approximately 17% of maintenance and renewal activity was variable with traffic (see Pollitt and Smith, 2002).[2]

Once the overall level of variable costs was established in step 1 of the process (see Figure 1), the next step was to allocate these variable costs to different vehicles based on their differing characteristics. At PR2000 this allocation process was based on a measure referred to as Equivalent Gross Tonne Miles (EGTM). EGTM converts actual Gross Tonne Miles (GTM) into a measure that is more reflective of the damage done for each GTM, given the characteristics of the vehicle. There were two parts to this conversion, one for damage to track and one for damage to structures (bridges etc.):

[2] Note that the reference for the 17% figure in Pollitt and Smith 2002 is attributed to ORR incorrectly; this figure was obtained from a Network Rail internal presentation.

$$\mathbf{EGTM = K\ Ct\ A^{0.49}\ S^{0.64}\ USM^{0.19}\ GTM} \qquad \text{(for track)}$$

and $$\mathbf{EGTM = L\ Ct\ A^{3.83}\ S^{1.52}\ GTM} \qquad \text{(for structures)}$$

where: **K** is a constant

Ct is 0.89 for loco hauled passenger stock and multiple units and 1 for all other vehicles

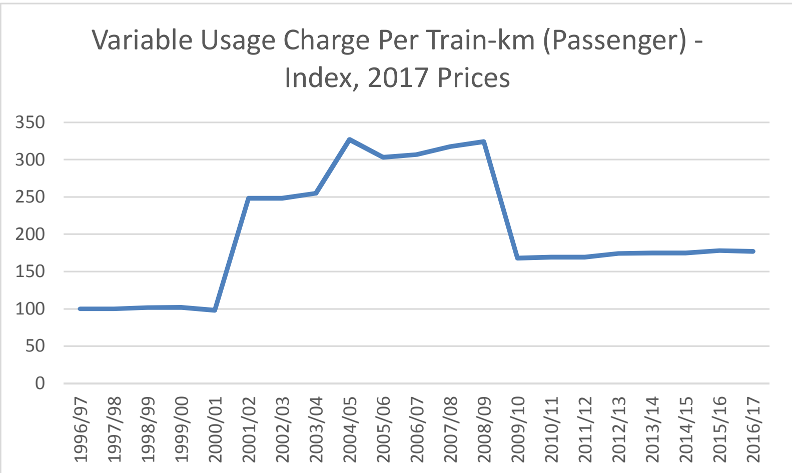
S is the operating speed [mph]

A is the axle load [tonnes]

USM is the unsprung mass [kg/axle]

GTM is gross tonne miles [Tonne-miles]

Figure 2: Variable Usage Charges Per Train-km (Passenger): Evolution over Time



Note: variable charges includes the Electrification asset usage charge introduced from 2009/10

Source: Own analysis based on Network Rail Regulatory Accounts. Traffic volume data sourced from Network Rail and from the ORR Portal

The parameters in this model were derived from models that regressed simulated damage measures derived from engineering-based damage models on a range of variables in order to identify the key drivers of damage and also their relative weights (based on a large volume of data).

Through this two-stage approach, PR2000 saw a substantial increase in the overall level of British wear and tear charges, and also resulted in a reasonably sophisticated differentiation in charging by vehicle type; meaning that each vehicle would pay a different charge depending on the amount of damage inflicted on the network (based on engineering models). The sharp increase in usage-related charges can be seen in Figure 2 (usage charges for passenger vehicles only are shown here for illustrative purposes and because other changes to freight charges were made during this period, which would otherwise distort the comparison).

A number of problems were identified with this method (see ORR, 2008). In particular, the per-vehicle calculation was based on only vertical forces and ignored wear and rolling contact fatigue. This could be seen as a serious weakness, particularly given events in Britain where the rolling-contact fatigue had proved such a problem. Further, the approach did not take into account variations in track quality or the condition of the vehicle wheelsets.

2.3 Variable track access charges from 2009 to date

At the 2008 Periodic Review (PR08) a significant change was made to the methodology for setting wear and tear related charges. These developments impacted on both Stage 1 and Stage 2 of the process (see Figure 1). In stage 1, the approach of relying on engineering judgement to derive variability proportions for different activities / asset classes was replaced by an engineering modelling approach in the form of Network Rail's Infrastructure Cost Model (ICM).

The ICM was developed as Network Rail's strategic planning tool to support the charging framework implemented in PR2008. The ICM was designed to predict operations, maintenance and renewal requirements over the long-term (up to 40 years), and operated at a geographically disaggregated level (with around 300 strategic route sections, but with more granular disaggregation operating beneath this level). The ICM brought together a number of models into a single modelling system.

The modelling framework implemented at PR2008 saw the ICM run under different scenarios: one at the current level of traffic and one at current levels + 5%. The resulting change in cost from moving to the higher traffic level was used to determine the variability proportions.

This general approach – to base the cost variability assessment on engineering models – has been retained since 2008, though the exact modelling framework has been through a number of developments. The majority of variable usage charges relate to track maintenance and renewals costs, and these are modelled using the vehicle track interaction strategic model (VTISM)[3] which is now regarded as the best available model for forecasting track related costs. The remaining variable costs, amounting to around 14% of the variable element, are still determined based on a top-down approach, determining variability proportions based on engineering judgement combined with some empirical evidence (see ORR, 2013).[4]

Importantly, as a result of this change in PR2008, the variability proportions underpinning variable usage charges were based on quantitative evidence from engineering models, as

[3] See Rail Delivery Group (2014) Charges and Incentives User Guide; https://cdn.networkrail.co.uk/wp-content/uploads/2017/01/RDG-charges-and-incentives-user-guide.pdf
[4] Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19 Page 569; http://orr.gov.uk/__data/assets/pdf_file/0011/452/pr13-final-determination.pdf

opposed to being based on engineering judgement. The biggest difference in terms of the variability proportions was a reduction in the assumed variability of track renewals (see Table 2).

The second major element of the changes was to develop step 2 of the process to incorporate lateral and longitudinal forces (which relate to rail wear and rolling contact fatigue) into the formula allocating variable costs to different vehicles. This element was added as an additional factor affecting the calculation, along with the EGTM formula shown above for vertical forces (the latter was also now based on new modelling undertaking using VTISM).[5] Vertical forces make up approximately 70% of the variable costs of track.[6]

As a result of the changes made in PR2008, changes were made to the variability proportions for certain assets as noted (see Table 2) and the overall level of variable charges fell substantially as shown by Figure 2 (the new charges were introduced in 2009/10). Since PR2008 the approach to setting variable usage charges – as set out here – has not fundamentally changed and this fact can also be seen from the data in Figure 2. However, the models have been developed and new evidence has emerged that has altered the allocation of costs for vehicles (for example, RDG note that heavy vehicles were found to be more damaging than previously thought, whilst higher speed vehicles less so).[7] It was also noted above that electrification asset charges were previously linked to electricity usage. This changed at PR2008, with a new electrification asset usage charge introduced based on the ICM, combined with engineering judgement – thus establishing an engineering relationship between traffic and asset damage and rectification cost for electricity assets for the first time.

As noted earlier, the variability proportion resulting from the charges introduced at PR2000 was 17% of maintenance and renewals costs. Further work was done between PR2000 and PR2008 as part of the ORR's Review of Structure of Charges in 2005 which resulted in a similar variability proportion being reported (19% of maintenance and renewal costs).[8] However, the changes introduced in PR2008 imply a much lower variability proportion. Based on the latest available data, the proportion of maintenance and renewal costs deemed variable with traffic is around 6% in 2016/17; (see Table 1).[9] This variability percentage is much lower than previous estimates

[5] See Rail Delivery Group (2014) Charges and Incentives User Guide; https://cdn.networkrail.co.uk/wp-content/uploads/2017/01/RDG-charges-and-incentives-user-guide.pdf and Serco (2013), VTISM Analysis to Inform the Allocation of Variable Usage Costs to Individual Vehicles https://www.networkrail.co.uk/wp-content/uploads/2017/02/Serco-VTISM-analysis-final-report.pdf

[6] Serco (2013), VTISM Analysis to Inform the Allocation of Variable Usage Costs to Individual Vehicles https://www.networkrail.co.uk/wp-content/uploads/2017/02/Serco-VTISM-analysis-final-report.pdf

[7] See Rail Delivery Group (2014) Charges and Incentives User Guide; https://cdn.networkrail.co.uk/wp-content/uploads/2017/01/RDG-charges-and-incentives-user-guide.pdf

[8] See Booz Allen Hamilton (2005) Review of Variable Usage and Electrification Asset Usage Charges: Final Report.

[9] It should be noted that care is needed in computing variability proportions as the denominator may differ in different comparisons. However, the 19% variability proportion reported for the 2005 Structure of Charges Review is based on total maintenance and renewal costs (even including stations; see also section 2.4 below) for forecast costs for a year that was designed to represent a steady state position (2013/14); see Booz Allen Hamilton (2005) Review of Variable Usage and Electrification Asset Usage Charges: Final Report. Further, the 6% variability proportion is robust to changes in the base year (e.g. using 2014/15 or 2015/16 in place of 2016/17).

at PR2000 or during the 2005 Review of Structure of Charges; and much lower than the econometric evidence, to which we now turn.

2.4 The econometric evidence on cost variability

Econometric evidence for Britain and several other European countries suggests that existing variable charges in Britain are too low (see, CATRIN (Cost Allocation of TRansport INfrastructure cost), Deliverable 8 - Rail Cost Allocation for Europe, Funded by Sixth Framework Programme: authors, Wheat, P., Smith, A. and C. Nash, 2009). The CATRIN research, which informed the European Commission in its drafting of relevant legislation, has been extended in subsequent work (in particular, the FP7 SUSTRAIL project; the Horizon 2020 NeTIRail-INFRA and a study for SNCF Réseau).

A major study of rail infrastructure marginal costs across Europe in 2009 – applying a common econometric approach across several case studies (including studies relating to data from Britain, France and Sweden) – concluded that the range of (mean) elasticities from the case studies was broadly 20% to 35% for maintenance. However, it was also found that the elasticity increased with traffic density (tonne-km per track-km). Thus a range was established of 20% to 45% for maintenance, depending on usage levels. For renewals, there were only two studies undertaken and in both cases the evidence was further compounded by the fact that the cost base included both maintenance and renewals (M&R) costs. Overall, it was concluded, albeit on relatively weak evidence, that an elasticity of 35% was reasonable for renewals (EU CATRIN project; see Wheat et. al., 2009).

Since then, further work has been done particularly in the area of renewals. Using Swedish data, Andersson et. al. (2012) and Odolinski and Nilsson, 2017 found elasticities for track renewals of 55% (track renewals) and 50% (all renewals), respectively. Using Swiss data, an elasticity for maintenance and renewals was derived of 50% (see Walker et. al., 2015). Other evidence, using national-level railway data across European countries indicated elasticities for maintenance and renewals together in the range 45% to 51% (see Smith, 2012 and Smith and Wheat, 2012). Partly on the basis of this work, as part of the SUSTRAIL project, the recommendation for the renewals elasticity was changed to 45% (see Wheat et. al., 2015).

An overall view of the latest position, taking into account the previous findings and the updated evidence, can be summarised as follows (see Smith and Wheat, 2017) where it was noted that:

- Overall, the evidence seems to suggest variability for M&R could be as high as **40-45%**.
- The lower part of the range of estimates could suggest a possible range of closer to **25-35%**.
- Whilst there exists a range, and of course, some uncertainty in these estimates, the evidence is drawn from a considerable body of evidence from multiple countries – following similar methodologies and with a range of data aggregations.

- Certainly, the evidence does **not** seem to support variability **below 20-25%**, taking maintenance and renewals together.

One further study is relevant and worthy of further discussion; namely a recent econometric study for SNCF-Réseau, using French data. This work provides evidence that further supports the above ranges (see Smith et. al. 2016; 2017); also reporting maintenance elasticities for different asset classes for the first time.

Table 3: Maintenance Elasticities from the Study for SNCF Réseau

	Preferred models in Model Selection				**Regulator Choice**
	BT	BC	BCI	Average	TL
Track	51%	41%	69%	54%	29%
Switches and crossings (S&C)	33%	49%	61%	48%	23%
Signalling	43%	42%	39%	41%	33%
Catenary	15%	20%	25%	20%	22%

Source: Smith et. al. 2016 supplemented by data provided by SNCF Reseau.

BT, BC and BCI refers to the chosen Box-Tidwell (BT), Box-Cox (BC) and interacted Box Cox (BCI) models respectively. TL refers to the translog-model. Elasticities are the sum of the passenger and freight elasticities (where relevant).

A few points stand out. First of all, the results differ to some extent depending on the method and, in this case, a range of advanced approaches were applied. In some cases, there was reasonable agreement between methods; in other cases, there was greater variation. In the case of track, the translog model produced lower elasticities and this model was selected by the regulator as compared to utilising the more advanced approaches (and applying model averaging) which was the approach taken by SNCF Réseau. For the other assets, the results do not differ significantly between methods, though the translog marginal costs were generally lower across all asset classes than the other models (see Table 4).

Table 4: Maintenance Marginal Costs from the Study for SNCF Réseau[10]

MC in €/k tonne-km						
Track		BT4	BC4	BCI4	Average	TL2
	Pass	€1.50	€1.04	€1.25	€1.26	€0.86
	Freight	€0.55	€0.39	€0.81	€0.58	€0.40
	Overall	€1.20	€0.83	€1.11	€1.05	€0.71
S&C		BT4	BC4	BCI4	Average	TL2
	Pass	€0.30	€0.34	€0.36	€0.33	€0.27
	Freight	€0.17	€0.25	€0.36	€0.26	€0.22
	Overall	€0.26	€0.31	€0.36	€0.31	€0.26
Signalling		BT3	BC3	BCI4	Average	TL2
	Pass	€0.54	€0.50	€0.45	€0.50	€0.44
	Freight	€0.25	€0.22	€0.23	€0.23	€0.20
	Overall	€0.44	€0.41	€0.38	€0.41	€0.36
		BT4	BC2	BCI4	Average	TL2
Catenary	Overall	€0.29	€0.39	€0.48	€0.39	€0.19
(EALE)		€0.09	€0.09	€0.09	€0.09	€0.09
Catenary (tot)		€0.38	€0.48	€0.57	€0.48	€0.28
Marginal cost	Pass	€2.34	€1.88	€2.06	€2.09	€1.57
	Freight	€0.97	€0.86	€1.40	€1.07	€0.82
	Overall	€1.91	€1.55	€1.85	**€1.77**	**€1.33**

Source: Own Work based on Data from SNCF Réseau.

The numbers after the model types (e.g. BT4, BC4 and BCI4) refer to model types in the original work.

It should be noted that one rationale for using more advanced approaches was the observation in past work that the translog had not produced intuitive results concerning the direction of change in elasticities. However, simplicity, transparency and the ability to replicate results led the regulator to select the translog model in place of the alternatives (Frontier Economics, 2017).

More widely, it can be seen that whilst some models for track and switches and crossings saw elasticities at the top end of the range from the past literature noted above, the simpler models preferred by the regulator yielded estimates more towards the bottom end. Thus, at least for these categories, the elasticities seem broadly in line with the past literature. There may however be concerns about the size of the signalling elasticities as there does not seem to be an underlying engineering reason to support elasticities as high as those reported. To a lesser extent, the elasticities for catenary maintenance also appear high.

[10] I am indebted to Jean-Christophe Thiebaud for preparing this table.

The renewals elasticities were developed through a mix of an engineering and econometric approach. An engineering model was used to predict costs over a very long time period (300 years) and the simulated renewals series is then translated into an average measure which (for track) was then regressed on a single traffic variable (tonne-km) plus control variables (the type of tracks, how they were laid, the type of sleepers and a dummy variable denoting a track section in an area with high traffic density). It should be noted that the track sections referred to here are different to the sections used in the maintenance econometric analysis (for track, there were over 100,000 sections used in the renewals regression).

Overall, taking all maintenance and renewals categories, the final results show variabilities of 17% for maintenance and 23% for renewals (see Table 5); or around 20% for maintenance and renewals combined. This emphasises firstly that the translog models chosen in general had lower elasticities; but secondly that in arriving at a final figure for the proportion of maintenance and renewal costs that are variable with traffic, it is necessary to take account of those categories that have zero variability. The overall results for French maintenance and renewals are therefore in line with the lower end of the range from past econometric studies.

Table 5: Share of Variability for the Components of Maintenance and Renewal[11]

	Maintenance	Renewal
Tracks	20%	37%
S&C	26%	33%
Signalling	28%	0%
Catenary	19%	20%
Structures	0%	0%
Others	0%	0%
Overall	**17%**	**23%**

Source: Own Work based on Data from SNCF Réseau.

Ultimately, as noted above, there seems to be a very strong evidence base in support of an overall elasticity for maintenance and renewal costs of not less than 20%; and potentially higher than that. The evidence from past top-down, cost allocation exercises in Great Britain indicates a range of 17-19% which is not too far away from the econometric evidence. However, the engineering approach in Britain puts cost variability much lower (around 6%).

[11] I am indebted to Jean-Christophe Thiebaud for preparing this table. Note that the percentage variability proportions do not match exactly with the elasticities reported for the translog model in Table 3. The reason is that the elasticities in Table 3 are weighted averages (weighted by tonnage) across all track sections. The marginal costs (MC) used for the variability proportions in in Table 5 are also weighted averages, based on section level MCs, but this computation does not give the same overall marginal cost number as using the weighted average elasticity multiplied by overall network average cost (AC). In other words, the averages are taken at a different point in the calculation.

2.5 Reconciling the econometric and engineering evidence?

The reasons for the differences between the econometric and engineering-based approaches are not entirely clear. The strong advantage of the econometric approach is that it is a transparent method based on actual cost data – so that it reflects what actually happens on the ground, rather than on assumptions as in the engineering approach. There could be questions as to which is optimal of course, and whether the actual decisions on the ground are inefficient compared to those based on engineering understanding – though it is not clear that this would be the case in all countries and all time periods. Nevertheless, it is hard to think that such a large discrepancy could be solely for this reason. One weakness of the engineering method is that whilst it may be relatively good at predicting damage, it is much harder to translate damage into remedial activity and importantly cost, where unit cost measures are needed which may be hard to obtain (for example, see Smith et. al., 2017).

It is also worth noting that the body of evidence from econometric studies is extensive, and considerable attention has been paid to ensuring common data definitions (where possible) and common methodological approaches. Annex 1 provides further details on the costs included within previous studies; but broadly speaking, we can say that except where we know that certain cost categories are excluded, the cost definitions used in the various studies show a fairly extensive coverage of costs, with perhaps only station costs excluded in general.

It is also the case that a wide variety of different levels of data disaggregation has been used, but with similar results (that is, data at the level of track sections, maintenance delivery units, regions and country level data); and studies have utilised both cross-sectional and panel data analyses. Thus, at least at the high level, it does not seem to be the case that data aggregation greatly influences the results, although this may be an area worth further investigation in future since it could be the case that the cost allocation issues that are present in respect of very disaggregate data structures (e.g. track section models) have the potential to distort the results depending on how the allocations are done.

For example, if a percentage mark-up on top of section-level direct costs is used to allocate indirect costs to track sections (as in the French data), whilst this will not distort the elasticity estimated (in log models at least), it will give the impression that the variability proportion obtained from the model applies to a wider body of the costs (whereas in fact the indirect costs may vary little with traffic). Distortions could also occur if indirect / overhead costs are allocated based on traffic volumes. Of course such issues become less of a problem with more aggregated data.

Overall, given the extensive nature of the econometric evidence, it seems hard to ignore the conclusion that marginal wear and tear costs for maintenance and renewals should be above 20% - and that further work is needed to understand why engineering methods produce values so much lower than econometric techniques (given that the EU legislation permits both approaches).

One final point should be made. Any modelling exercise, whether econometric or engineering is data and effort (method) intensive. However, in respect of the econometric method, there exists a wide body of literature that is available such that even in the absence of a country-specific study, as long as average cost data is available, a range for marginal costs can be obtained by applying elasticities from the literature to the average cost data for the country concerned. One new approach developed as part of the Horizon 2020 NeTIRail-INFRA project introduces the possibility of a mixed approach that could be used to assist in situations where a new country study is implemented using country-level data, but also drawing on prior information about elasticities from previous studies to help improve the estimation process (see Dekker et. al., 2017). This approach is expected to have the greatest benefits where sample sizes are small, for example where a country is attempting an econometric study for the first time. It is also possible to determine whether the data being analysed is compatible with the past literature. This approach is still in the pilot stage, however.

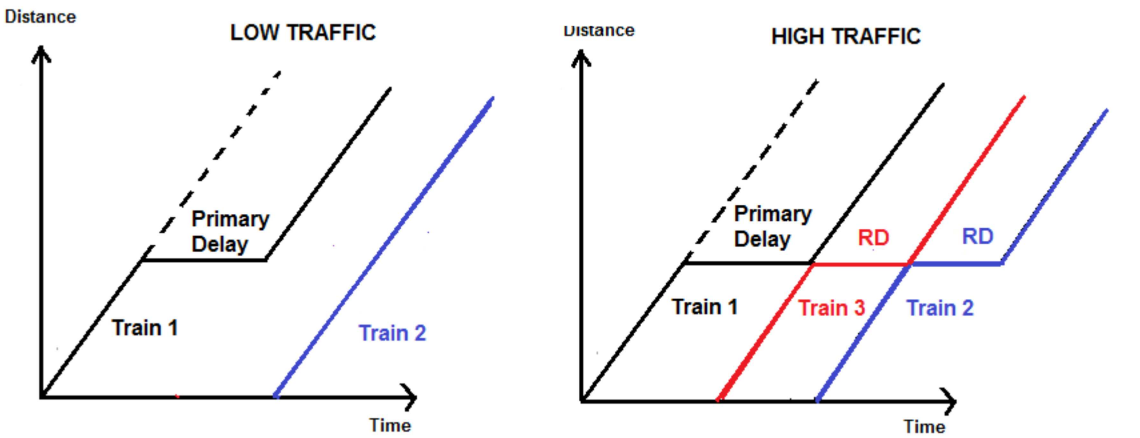
3. Charging for capacity

The current capacity charge in Britain is based on congestion costs rather than pure scarcity (congestion costs occur as a result of increased delays arising as infrastructure use becomes close to capacity, whereas scarcity costs relate to the opportunity cost of squeezing other trains off the network when all demands cannot be met). It is estimated by examining the relationship between reactionary delays and capacity utilisation, using a capacity utilisation index derived by looking at the proportion of capacity which would be used if the trains on the existing timetable were squeezed as close together as is possible without reducing scheduled speeds (Gibson, Cooper and Ball, 2002). The capacity utilisation does not take account of congestion at stations and junctions, or of the desire by passenger operators for a particular spacing of trains, such as a regular interval service.

It is important to understand that the additional reactionary delays as capacity is reached are distinct from the delays directly imposed by an additional train as a result of unreliability. These are paid for directly under the performance regime. The congestion costs are the additional delays that arise simply because of the presence of an additional train on the network even if it does not directly cause any delays.

The principle behind the calculation of the charge was that the marginal cost of congestion was calculated as being equal to the additional calculated cost in reactionary delay arising from the addition of another train to the compressed timetable. The way in which adding trains may add to reactionary delay is illustrated in Figure3. When there is low traffic, only trains 1 and 2 run. It will take a very large delay to train 1 to delay train 2; the level of primary delay shown in the diagram will not do this. But if train 3 is added, then the delay shown to train 1 will delay train 3 and that in turn will delay train 2 and potentially subsequent trains.

Figure 3: Reactionary delay



Source: Adapted from diagram in: Network Rail (2012) Periodic Review 2013 – Consultation on the Capacity Charge

In order to estimate this cost, the British rail network was divided into a large number of geographic sections which formed 24 different strategic routes (e.g. the West Coast Main Line) and into different times of day /days of the week. Reactionary Delay per Train Mile (the dependent variable) was then regressed on the capacity utilisation index (the explanatory variable). The exponential form was chosen as providing the 'best' relationship between capacity utilisation and reactionary delay and is written as shown in Equation (1):

$$\mathbf{D_{it} = A_i * \exp(\beta C_{it}).} \qquad \mathbf{(1)}$$

Where $\mathbf{D_{it}}$ is the **reactionary delay per train mile on section *i* in time period *t*; $\mathbf{A_i}$ is a section specific constant; β** is the coefficient of the capacity utilisation which varies by route and $\mathbf{C_{it}}$ is the capacity utilisation calculated for section ***i*** and time period ***t***.

A recent recalibration exercise confirmed that the exponential form is the most appropriate choice to use, to calculate the costs of congestion (Arup, 2013).

Whilst the regression analysis permitted derivation of congestion charges for a large number of different combinations of location and time of day, in practice the charging structure was much simplified. Network Rail argued that its billing system did not permit charges to be differentiated by time of day, or at a finer degree of disaggregation than by service code (which essentially differentiated services by their origin, destination and route). So charges vary by service code, with the only temporal differentiation being between weekdays and weekends.

As noted above, scarcity, as opposed to congestion, arises when some trains simply cannot be accommodated on the network. Its cost may be regarded as the social value of the marginal train which cannot be accommodated. One way of obtaining this value would be to auction paths. However, the number of possible paths available is large, as many different timetables incorporating different combinations of paths could be operated. A way of taking account of this would be to invite train operators to bid for their ideal path and to indicate which discount they would require per minute earlier or later their allocated path turned out to be (Nilsson, 2002). Also, because the value of one path to an operator depends on what other paths they obtain, and on what paths their competitors obtain, an iterative procedure would be needed, and the complexity of this has put infrastructure managers off this approach. In any event, the bid made by an operator would only reflect the value of the path to itself, not to its customers or anyone affected by external costs or benefits, unless a set of subsidies and taxes could be implemented to take these factors into account.

In Britain, currently a large proportion of capacity is taken up by franchised passenger services, and in practice decisions about the use of capacity are typically choices between allowing these franchised operators or open access operators to use any spare paths. A number of studies have sought to estimate the marginal social value of allocating additional paths to the franchisee, or the marginal social cost of removing paths at the margin from the franchisee (Johnson and Nash, 2008).

Relevant items are operator profits, marginal infrastructure costs, user benefits and the benefits of diverting traffic from or to road (tax revenues, congestion and environmental impacts including noise, local air quality, greenhouse gases and safety).

Obviously the results vary very much by time of day and location. Although ORR has considered scarcity charges on several occasions, it has faced opposition from the industry (particularly freight operators, who argue they would have to bid for work without knowing what track access charge they would pay, since they would not know what path they would obtain) and has never been implemented (SDG, 2014).

4. Mark-ups

As noted above, for some years, Britain has had mark-ups on freight track access charges designed to recover the avoidable fixed costs of freight services. These charges are subject to an ability to pay test, and for that reason were only applied to three commodities – coal, iron ore and nuclear waste – where it was concluded that the charge would have very little or no impact on the volume of rail traffic. The test was applied by assuming the percentage increase in freight charges would equal the percentage increase in track access charges multiplied by the proportion of costs accounted for by track access charges, and estimated demand elasticities in the final market used to assess the impact of this increase on traffic.

ORR is currently consulting on a proposal to apply mark-ups more widely to recover allocated fixed costs, subject to an ability to pay test. These charges would replace capacity charges for all services and part or all fixed charges for passenger franchisees. They will be capped at allocated total cost, either using the existing methodology or the new methodology proposed by Network Rail (Network Rail, 2017b). It is argued that such charges will improve the incentives to train operators to help reduce total cost, and also make for fair competition between franchised train operators, who currently pay fixed charges and open access operators who do not, thus protecting the contribution to the costs of the rail system currently made by the *premia* paid by franchisees as a result of the bidding process. Currently, open access operators are only allowed to enter the market if they attract new business to rail rather than simply abstracting revenue from franchisees, but there is a proposal to remove this constraint.

For freight, ORR will continue to regard commodities (with inter modal as a single commodity) as appropriate market segments. It will continue to charge mark-ups on coal, iron ore and nuclear waste, but also introduce mark-ups on biomass (the argument being that rail is much cheaper than road and access charges are a small part of delivered price, so mark-ups unlikely to reduce biomass consumption). It will not introduce mark-ups on any other freight commodities as demand is too sensitive to price.

For passenger, ORR regards the difference between revenue and cost of running a particular train as indicating scope for mark-ups, rather than using the approach based on elasticities as for freight. Presumably, this is due to the difference in the structure of the market and the fact that passenger operator response may take the form of reductions in frequency as well as increase in price. In the passenger market in Britain there is little on-track competition. Almost all passenger services are run under franchises awarded through competitive tendering. Some fares are regulated but otherwise train operators are free to raise fares to profit maximising levels. Thus, in practice the direct response to higher track access charges is likely to be reductions in train services rather than increases in price (of course reducing capacity may itself bring about an increase in price to restrict demand to capacity offered). On this basis, ORR concludes that mark-ups are possible for intercity passenger services and long-distance commuter trains to London. But Network Rail still indicates that charges cannot differentiate below service code level and cannot differentiate by time of day because they do not capture train km below this level. The latter seems a very serious constraint on the ability to differentiate according to what the market can bear; profitability varies greatly by time of day so that charges which the market can bear at some times of day will make services unprofitable at others. Mark-ups will be levied per train km and for franchisees will be revised annually in line with actual train km (at present fixed for entire franchise period)

It is argued that charging mark-ups to cover some or all fixed costs will give train operators an incentive to consider the long-run costs their decisions to run services cause, but the charges are not based on long-run marginal capacity costs, which are only positive where there is a shortage of capacity. In practice, mark-ups will encourage withdrawal of some services which are covering their (short-run and long-run) marginal costs. Of course, for passenger franchisees any change to the fixed charge will lead to an equal and opposite change to the premium/subsidy, and will only affect train operator decisions where service level changes are not prevented by the franchise agreement. It is for open access operators that the impact may be greater.

The Network Rail (2017b) consultation relates to the new approach to allocating fixed costs developed as a pilot project by Brockley Consulting, which has now been applied to all train operators. This does improve on the current approach by working at track section level, and only allocating the costs of each track section to operators actually using that track section. However, it is assumed that no track section can be completely closed. Under this assumption, 27% of fixed costs are avoidable if all operators withdrew. But the method then proceeds arbitrarily to allocate all joint costs for the track section on the basis of train miles. The result is an allocation which bears no relation to the amount of cost that the train operator in question can influence by its decisions. For this purpose, an assessment of avoidable fixed costs (those costs which could be avoided if the services of that operator no longer ran) would be of more value.

5. Conclusions

Britain has gone further than most European countries to try to develop track access charges that give the right incentives to train operators regarding the services they run and the rolling stock they use. The high degree of differentiation of variable user charges between vehicle types, based on engineering models, goes a long way to doing this. However, the econometric evidence from multiple case studies covering several European countries suggests that the overall level of variable charges in Britain considerably understates true wear and tear costs and may therefore encourage excessive provision of services (although since most services in Britain are required by franchise agreements, this may not be a big issue).

Given that EU legislation permits different approaches to be used as a basis for charging, including cost allocation (accounting) methods, econometric approaches and engineering methods, we consider that new research is needed to better understand the reasons for the differences coming from the various approaches. In respect of Great Britain, cost allocation methods would indicate cost variability for maintenance and renewals of around 20%; this being also the bottom end of the range coming from the econometric evidence (with many studies indicating higher variability proportions – up to 45%). However the current variable charges, based on engineering models, only recover around 6% of maintenance and renewal costs.

The capacity charge in Britain is really a congestion charge, reflecting the increased unreliability as capacity utilisation rises. It does not deal with the cost of scarcity, where some trains simply cannot be accommodated at all, and although some research on this has taken place, such a charge has been rejected by the regulator. The capacity charge is currently too crude to play a major role in promoting more efficient use of capacity, in particular given the lack of differentiation by time of day, and only deals with congestion as opposed to true scarcity, but is at least giving some incentive in the right direction, which its proposed replacement by a system of mark-ups will remove.

Basing mark-ups for freight on demand elasticities by commodity, assuming that increases in track access charges will be passed on to the final consumer, seems appropriate. For commercial passenger services, it is likely that the response of operators will be to withdraw marginally profitable trains rather than to raise prices, although this will only be possible where these trains are not required as part of a franchise agreement. Avoiding reductions in service levels below the efficient level, where they are permitted (for instance for open access operators), will require mark-ups that are highly differentiated in time and space.

References

Andersson, M., Smith, A.S.J., Wikberg, T. and Wheat, P.E. (2012), Estimating the marginal cost of railway track renewals using corner solution models, Transportation Research Part A: Policy and Practice, 46, pp.954-964.

Arup (2013), Recalibrating the Capacity Charge for CP5. Final Report. ORR, London.

Dekker, T., Wheat, P.E. and Smith, A.S.J. (2017), NeTIRail-INFRA Deliverable D1.7 Incentives Final Report Annex 4: Marginal Wear and Tear Costs in France Using Bayesian Analysis.

Frontier Economics (2017), Estimation des couts marginaux d'entretien du reseau ferre national Expertise des travaux économétriques. Report committed for ARAFER.

Gibson, S., Cooper, G. and Ball, B. (2002), Developments in Transport Policy. The evolution of capacity charges on the UK rail network, Journal of Transport Economics and Policy Vol.36 Part 2 pp. 341-354.

Johnson, D.H. and Nash, C.A. (2008), Charging for Scarce Capacity: A Case Study of Britain's East Coast Main Line, Review of Network Economics, Vol.7.

Network Rail (2012c), Periodic Review 2013 – Consultation on the Capacity Charge. July 2012

Network Rail (2017a), Network Rail Infrastructure Limited Regulatory Financial Statements, Year ended 31 March 2017. Network Rail, London.

Network Rail (2017b), Consultation on methodology for allocating fixed costs to train operators in Control Period 6 (CP6), Network Rail, London.

Nilsson, J-E. (2002), Towards a Welfare Enhancing Process to Manage Railway Infrastructure Access, Transportation Research, Part A.

Odolinski, K. and Nilsson, J-E. (2017), Estimating the marginal maintenance cost of rail infrastructure usage in Sweden; does more data make a difference? Economics of Transportation, 10, 8-17.

ORR (2008), Periodic Review 2008 Determination of Network Rail's outputs and funding for 2009-14. ORR, London.

ORR (2013), Periodic Review 2013 Final Determination of Network Rail's outputs and funding for 2014-19. ORR, London.

ORR (2017), Periodic Review 2018 consultation on charges recovering fixed network costs. ORR, London.

Pollitt M.G. and Smith, A.S.J. (2002), The Restructuring and Privatisation of British Rail: Was it Really that Bad?, Fiscal Studies, 23, pp.463-502.

SDG (2014), The practicalities of scarcity charging: Final Report, ORR London

Smith, A.S.J., Kaushal, A., Odolinski, K., Iwnicki, S. and Wheat, P.E. (2017), Estimating the damage and marginal cost of different vehicle types on rail infrastructure: combining economic and engineering approaches, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, accepted February 2017.

Smith, A.S.J. and Wheat, P.E. (2012), Estimation of cost inefficiency in panel data models with firm specific and sub-company specific effects, Journal of Productivity Analysis, 37, pp.27-40.

Smith, A.S.J. (2012), The application of stochastic frontier panel models in economic regulation: Experience from the European rail sector, Transportation Research Part E: Logistics and Transportation Review, 48, pp.503-515.

Smith, A.S.J. and Wheat, P.E., (2017), The Econometric Approach to Estimating Rail Marginal Wear and Tear Costs, Presentation to the Rail Delivery Group, January 2017, London.

Smith, A.S.J., Walker P., Wheat P.E., Guiraud, L. and Silavong, C. (2017), Estimating the marginal maintenance cost for the French railway network: a comparison of models, ITEA Conference, Barcelona 20.

Smith, A.S.J., Wheat, P.E., Walker P., and Marti, M. (2016), Modelling railway infrastructure maintenance cost in France: Overview of estimates, Final Report for SNCF Réseau.

Walker, P., Wheat, P.E, Marti, M. and Smith, A.S.J. (2015), Swiss Case Study, Annex 3 to Deliverable 5.3 Access charge, The sustainable freight railway: Designing the freight vehicle – track system for higher delivered tonnage with improved availability at reduced cost, SUSTRAIL.

Wheat, P.E., Smith, A.S.J. and Matthews, B. (2015), Deliverable 5.3: Access Charge Final Report, SUSTRAIL.

Wheat, P.E., Smith, A.S.J. and Nash, C.A. (2009), CATRIN D8 Rail cost allocation for Europe. University of Leeds, Leeds.

Annex 1: Data Definitions in Rail Infrastructure Marginal Cost Econometric Cost Studies

Maintenance – cost base to apply elasticities

During the CATRIN project considerable effort was put into trying to understand what cost categories were included in each case study in order to ensure their comparability (given that different cost categories are more or less variable with traffic). The CATRIN project defined four categories of maintenance cost:

- Permanent way maintenance – maintenance of the rails, sleepers and ballast;
- Signalling and telecoms maintenance – maintenance of signalling and any line side telecommunications;
- Electrification and plant maintenance – maintenance of contact electrification apparatus, plant (such as pumping stations, signal power supplies and point machines) and distribution equipment (such as cables and transformers);
- Other maintenance – including inspections and overheads.

The CATRIN project also reported the cost coverage of the different studies in tabular form. Most studies considered the whole of maintenance cost while the study in Great Britain used a more limited definition. However, the main cost category included in the British study (permanent way) is likely to be the most variable with traffic. In the CATRIN project the elasticity from the British study was therefore "scaled" on the assumption that all costs other than permanent way costs do not vary with traffic.

Cost considered in each country

Country	Maintenance category				Proportion of total maintenance cost considered in the study
	Permanent Way	Signalling and Telecoms	Electrification and plant	Other	
Sweden	✓	✓	✓	✓	100%
Switzerland (All of Contracting A)	✓	✓	✓	✓	100%
Great Britain	✓			(Part)	67%
France	✓	✓	✓	✓	100%
Pooled International	✓	✓	✓	(Part)	90%

Source: CATRIN Deliverable D8 (Wheat et. al., 2009)

For Britain, all stations, depots and lineside buildings expenditure is treated as renewals so these are excluded. Though CATRIN does not specifically refer to these costs, we are almost certain that stations are also excluded from the maintenance definitions in other countries as well (see detail below). Structures maintenance, though not specifically referred to in the CATRIN documentation, are almost certainly excluded in the case of Britain (see Network Rail Network Rail Infrastructure Limited Regulatory Financial Statements Year Ended March 2014), though this is in any case a very small cost category (most costs in Britain relating to structures are renewals). For Sweden and Switzerland, structures maintenance is included. It is not totally clear from the CATRIN definitions whether this category is included or not in the remaining studies, but in general it appears that the definition is broad and inclusive.

Based on further discussions with partners in Sweden and Switzerland responsible for the econometric marginal cost studies (post-project), we can say the following. For Sweden, the categories included in maintenance costs and in the renewal costs are:

- Track
- Signalling systems
- Electrification (overhead contact line, auxiliary power line)
- Telecommunication systems
- Power supply equipment (to some extent)
- Structures (bridges and tunnels)
- Rolling stock heating equipment (but not train depots)
- Buildings connected to the railway
- Crossings
- Platforms
- Fences and sound berms/noise barriers
- Snow removal

For Switzerland, again for both maintenance and renewal, the following costs, *inter alia*, are included:

- Track
- Structures (tunnels; bridges)
- Electrification
- Signalling
- Telecoms
- "Nature" (e.g. prevention of avalanches in mountain regions)
- Some station costs (platforms and shelter)

In general, it is clear that the Swiss study contains a broad range of costs. Depots (and marshalling yards) are excluded however. At present, the best assumption seems to be that platforms are included but not, in general, station buildings.

One difference that we have identified with respect to the Swiss and Swedish maintenance cost studies is that the former includes snow removal, but this is excluded from the Swedish study. Some minor renewals are also included in the maintenance definition for Sweden, but this is also true of Britain. It may be that different countries have different thresholds for what is maintenance and what is renewal. For example, for Britain, track replacement jobs less than 200 metres or signalling refurbishment costing less than £5000 were classed as maintenance. However, this may not be a major issue in general.

Renewals – cost base to apply elasticities

As noted, the British study referred to above includes only track maintenance. In the Swiss study, the definition was said to be “all lineside renewals”. This is then clearly a broader definition. As noted above, our conversation with the Swiss partners has confirmed the broad definition of renewals in this case. A further point that was noted is that the more up-to-date Swiss study done in SUSTRAIL, which has a higher elasticity for M&R, uses data at a more aggregate level than the earlier CATRIN Swiss study. It appears that this data then also means that the total M&R cost base included in the econometric model is very close to the totals reported by the company. With the previous, more disaggregated analysis done during CATRIN, it appears that a smaller amount of costs was included. There is also concern over whether the allocation of costs to track sections was accurate in that earlier study.

Summary

In general, it seems that there is a reasonably good understanding of the definitions of the costs included in studies in this area which typically show quite extensive coverage of maintenance and renewals. Perhaps the key aspects that typically seem to be excluded are station related costs. Of course in some cases, such as Britain and the study for SNCF Réseau (Smith et. al. 2017), we know that certain categories are excluded (Britain only included track) and the study for SNCF covered track, S&C, signalling and catenary only.