Centre on Regulation in Europe

Improving network and digital industries regulation

Track access charges: reconciling conflicting objectives

Project Report

Prof. Chris Nash, CERRE & University of Leeds

With contributions from:

Prof. Yves Crozet, CERRE & University of Lyon (LAET)

Dr. Heike Link, German Institute for Economic Research (DIW)

Prof. Jan-Eric Nilsson, Swedish National Road and Transport Research Institute

Prof. Andrew Smith, 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

cerre Centre on Regulation in Europe

Improving network and digital industries regulation

Table of contents

Centre on Regulation in Europe

Improving network and digital industries regulation

About CERRE

Providing top quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network and digital industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach;
- the widely acknowledged academic credentials and policy experience of its team and associated staff members;
- its scientific independence and impartiality;
- the direct relevance and timeliness of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many Member States, regulators are part of a relatively recent profession.

The project, within the framework of which this report has been prepared, has received the financial support of a number of CERRE members. As provided for in the association's by-laws, it has, however, been prepared in complete academic independence.

The views expressed in this CERRE report are *attributable only to the authors in a personal capacity and not to any institution with which they are associated.* They do not necessarily correspond either to those of CERRE, to any sponsor or to any (other) member of CERRE.

About the authors

Chris Nash is a CERRE Research Fellow and Research Professor at the Institute for Transport Studies, University of Leeds (ITS); he was Director of ITS for seven years. He works primarily on road and rail transport economics and policy, including pricing, regulation and project appraisal. He has been coordinator of six EU Framework Programme research projects including UNITE, GRACE and Imprint-Europe, and contributed to many more, including the SUSTRAIL project. He has acted as specialist advisor to many organisations, including the European Commission High Level Group on Infrastructure Charging, the OECD and the Transport and European Union Select Committees of the British Parliament.

Yves Crozet is a CERRE Research Fellow, and a Professor of Economics at Lyon University (Institute of Political Studies) and a member of the Laboratoire Aménagement Économie Transports, which he headed from 1997 to 2007. In 2008, he became the head of the Energy, Environment and Transport Observatory set up by the French government to assess the carbon content of different transport modes. Prof. Crozet is also a former Secretary General of the World Conference of the Transport Research Society.

Heike Link is a Senior Research Manager in the Department of Energy, Transport and Environment at the German Institute for Economic Research in Berlin (DIW Berlin). She has led a number of national and international research and consultancy projects on infrastructure issues, transport pricing and railway economics focusing on areas such as access pricing, rail competition and efficiency analysis in the rail sector. The projects were commissioned by a range of national and European institutions, including the Transport Ministry, the Federal Environmental Board and the Federal Road Authority in Germany; the European Commission; the OECD; and national institutions in a number of other European countries. Her work has been published both in international refereed journals and as book contributions.

Jan-Eric Nilsson is Professor of Transport Economics at the Swedish National Road and Transport Research Institute (VTI) in Stockholm. Prior to this, he was the Research Director in the Transport Economics Department of Dalarna University from 2000 to 2008, having previously been a senior lecturer and researcher at the University from 1992. Prof. Nilsson also served for 10 years in the economics department of the Swedish National Rail Administration. He received his PhD in economics from Stockholm University**.**

Andrew Smith is Professor of Transport Performance and Economics at the Institute for Transport Studies, University of Leeds. His research is focused on: cost and efficiency modelling, with particular focus on rail transport, utilities (water and energy) and health; infrastructure access pricing and marginal cost research, rail regulation and the impact of rail reforms on economic efficiency; and modelling the economic impact of rail technology innovations. He has published important work in these areas and has led or is leading a wide range of research projects, funded by, for example, the British Office of Rail Regulation, Ofwat, the European

Commission, SNCF, UK Research Councils, National Institute for Health Research (NIHR), and the UK Department for Transport. Prof. Smith is also an academic advisor to the water regulator, Ofwat, on econometric benchmarking.

1. Introduction

Starting with some types of international freight in Directive 91/440, European rail policy has progressively opened the rail market to new entry. By 2007, all freight services, domestic and international were opened to new entry and in 2010 this was extended to international passenger services. Under the provisions of the fourth railway package, this process will be completed, with open entry to domestic commercial passenger services in 2020 and compulsory competitive tendering of domestic public service contracts in 2023, although with provision for exceptions.

Whilst the same company was responsible for both infrastructure and all train operations, there was no need for a system of track access charges, but as soon as the market was opened to new entry, this became essential. It was recognised from the beginning that the track access charging system should be non-discriminatory, and that required at the least that the incumbent implemented separate accounts for infrastructure and operations with transparent payments for use of the infrastructure. Current legislation requires that infrastructure and operations should each have their own management. Many European countries (including Sweden and Britain) have placed the infrastructure in a separate company from any train operator; elsewhere (as in Germany and France) they are separate subsidiaries of the same holding company.

Beyond this, legislation required that charges should be based on the direct cost of running the train service. Non-discriminatory mark-ups are permitted to contribute to funding, provided that these do not exclude any market segment able to pay direct cost. The mark-ups may be differentiated according to various market segments, including freight commodities and types of passenger train.

Charges are also permitted for scarcity and environmental costs, but environmental charges must not raise the overall cost of rail transport unless equivalent charges were levied on other modes.

However, there remained a major divergence in terms of how the directives were interpreted. Some countries, for instance Germany, which saw the infrastructure manager as a commercial body, which should be charging train operators the full cost of providing and maintaining the infrastructure, essentially interpreted direct cost as average cost less those costs borne by the state budget. Others, including Sweden, saw the infrastructure manager as a public agency, providing the infrastructure out of general tax revenue and charging for its use on a marginal cost basis (ITF, 2008).

When defining rules for track access charges (TAC), the latest statement is to be found in the recast of the first railway package, Directive 2012/34/EU and the regulation on calculating infrastructure charges, Regulation (EU) 2015/909. Direct cost may include additional maintenance and renewal costs and operating costs where additional staff is required. It was

recognised that charges based on direct cost would only give the right incentives regarding choice of rolling stock if they were differentiated according to the characteristics (such as mass or axle weight), and also that direct cost may vary according to the characteristics of the track (such as curvature). Nevertheless, averaging is permitted to avoid excessive complication.

It is clear from the way direct costs are defined that these are intended to reflect short-run marginal social cost – that is to say, they exclude capital costs of expanding capacity. However, there is a long-standing debate in economic literature between the merits of short and long-run marginal cost pricing. Whilst short-run marginal cost pricing may give optimal incentives to train operators regarding the choice of services to operate over the existing infrastructure, it may give perverse incentives to train operators and infrastructure managers regarding long-run development of the infrastructure. For instance, if the infrastructure manager behaves commercially, the presence of scarcity charges provides infrastructure managers with an incentive to keep capacity scarce. This perverse incentive is countered in the legislation by a requirement for IMs to prepare and appraise capacity enhancement plans where capacity is declared scarce. In any event, all the main European rail infrastructure bodies are publicly owned and may be expected to use social cost-benefit analysis in assessing the case for infrastructure enhancement.

Some authors also argue that marginal cost pricing in general fails to promote efficiency by failing to give both infrastructure managers and train operators sufficient incentive to control total costs. Again, this is tackled by a separate measure requiring member states to ensure that infrastructure managers have an incentive to reduce costs either through the regulatory system or through a multi-annual contract with the state. The use of performance schemes is required to compensate train operators for poor performance on behalf of the infrastructure manager (and vice versa). Such schemes may also have the effect of avoiding incentivising the infrastructure manager to achieve cost savings at the expense of quality of service.

In short, then, the setting of track access charges involves conflicting objectives, particularly between efficient use of existing capacity and cost recovery. The outcome in any country is likely to vary according to the relative importance attached to each objective.

The Directive allows for this by saying that whilst charges must be at least at the level of direct cost, non-discriminatory mark-ups are permitted up to any desired degree of cost recovery (of course with a maximum of 100%) provided that they do not exclude market segments willing to pay direct cost. There is a clear distinction between Sweden and Britain, where the emphasis to date has been on promoting efficient use of the network, and France and Germany where the emphasis is on full cost recovery (after allowing for certain direct payments from government). However, in response to the recast of the first railway package, France and Germany have both substantially revised their charging system to make it clearly conform to the direct cost plus mark-ups rule. Britain also now looks set to make more use of mark-ups on direct cost.

The objective of this report is to examine research and practice in a sample of European countries (namely Great Britain, Sweden, France and Germany) and to reach recommendations to guide the European Commission, infrastructure managers and regulators, including dealing with the following key issues:

- How to measure and charge for the direct cost of wear and tear on the infrastructure?
- How to charge for congestion and scarcity?
- How to charge necessary mark-ups in a way which does as little damage as possible to efficiency of infrastructure use?

We first give a brief overview of the situation in our case study countries, and then review different approaches and evidence on the various elements of direct cost and also on mark-ups. We also briefly review the relevance of performance regimes and other incentive mechanisms for efficiency before reaching our conclusions. We do not consider charges for the use of stations or other facilities. This report is accompanied by detailed case studies of the four countries concerned.

2. Overview of track access charges in Britain, France, Germany and Sweden

2.1 Charges designed to promote efficient use of the network – Britain and Sweden

Both Britain and Sweden aim at recovering marginal maintenance and renewal cost through a usage charge. Econometric evidence suggests that the current charges are far too low. In the case of Britain, the usage charge is based on the results of an engineering model and is a charge per vehicle km. It is highly differentiated, distinguishing between vehicles by a variety of factors including weight, axle-weight, track friendliness of the bogies, and yaw stiffness. The Swedish charge is simpler, being levied per gross tonne km but differentiated according to three bands of axle weight. It is based on earlier econometric analysis of maintenance costs, but considers renewals costs to be entirely fixed. Only Sweden has an externalities charge for diesel traction, based on fuel consumption and type of engine.

Both countries also have charges related to congestion or scarcity. Britain has a capacity charge which is based on estimated congestion costs, derived from a regression model relating reactionary delay to the level of capacity utilisation on the line (Gibson et al, 2002). This charge is levied per train km and is differentiated according to the route and day of the week (weekends versus weekdays) but not by time of day, as Network Rail has stated that its billing system could not cope with that. It appears that as part of its current periodic review, the rail regulator may decide to abolish this charge. Sweden has a charge per path kilometre, which varies between three categories of route according to how busy they are. It also has surcharges for entering the three main cities in peak periods. Both countries regard costs of planning and operating the rail infrastructure as totally fixed.

At the present time, the main mark-up in Britain is for services operated under a franchise, which is charged as a lump sum in accordance with allocated fixed costs. There is also a mark-up on freight traffic to recover avoidable fixed costs of freight services (e.g. freight only lines, passing loops which would not be needed in the absence of freight traffic). However, this is only levied on commodities considered able to pay it with little resulting loss of traffic, and currently these are coal, iron ore and nuclear waste. Whilst coal and iron ore both pay a mark-up of around £1 per 1000 tonne train mile (€0.8 per train km), nuclear waste pays around £20 per 1000 tonne train mile (€14 per train km).

There is a proposal to revise this charge. In the case of freight, the ceiling on the charge would become allocated total cost instead of just avoidable cost, but a study of ability to pay concluded that biomass was the only additional commodity on which it should be charged (CEPA, 2017a). In the case of passengers, it is proposed to introduce a mark-up on usage charges based on the current margin between train operating costs and revenue per train km (CEPA, 2017b). A study

has suggested that the only categories of traffic able to bear this charge are main intercity routes and densely-used, long-distance commuter services. The study suggested that these could bear a charge of £6-7 per train mile (€4-5 per train km).

One reason for introducing these mark-ups is to achieve equality between franchisees and open access operators; currently open access operation is very restricted in Britain and these operators only pay variable charges whereas franchisees pay a large fixed charge (and, in some cases, also a premium for the franchise). A second argument is that only if train operators pay the full cost of the infrastructure they use, will they have an incentive to work with the infrastructure manager to reduce costs. However other ways of sharing infrastructure cost risk have also been tried, including deep alliances in which differences in costs (and revenues) from budget are shared between the train operator and infrastructure manager and a sharing of changes in costs through the Route Level Efficiency Benefit Sharing scheme (Netirail, 2017).

Sweden only charges a mark-up for trains crossing the Oresund Bridge.

In Britain most passenger services, whether profitable or not, are run under franchises, and currently most franchises are let by central government. Regional services in Sweden, Germany and France are also franchised, albeit by regional government (although in France currently by law all franchises must be direct awards to SNCF). The interest of train operators in Britain to respond to incentives given by track access charges is limited, since the franchisee is compensated for any increases in track access charges during the life of a franchise. This removes any incentive for the train operator to work with the infrastructure manager to reduce costs, although they may in the long run influence franchise specifications.

2.2 Charges designed to recover total cost – Germany and France

Both Germany and France are moving to a charging system in which the only charges will be direct cost plus mark-ups. In Germany, direct costs are estimated using expert judgement as to what proportion of each category of costs are variable costs. Renewals are not included, but depreciation is. Our understanding is that this includes depreciation of enhancements, which should not form part of short-run marginal cost. For renewals, depreciation may provide a reasonable proxy for the long-run level of renewals depending on how it is calculated, but where data on actual renewals is available we regard it as preferable to use it, possibly smoothed in some way to reflect a steady-state level given the lumpy nature of renewals.

In France, charges for maintenance are based on econometric evidence, whilst those for renewals come from an engineering model of when renewals are necessary; the French estimates of direct cost are much higher than the German. In Germany, charges are per train km, but differentiated between long-distance and regional passenger trains, and freight, with a higher rate applying to particularly heavy freight trains. France is moving to charging per gross tonne km, with charges varying between conventional passenger train, high-speed passenger

trains and freight. Both countries regard a small proportion of timetabling and operations costs as variable, but the proportion of direct cost this accounts for is small.

Both countries used to have charges per path km, differentiated according to the quality of the path or the infrastructure, but both are now moving to charging mark-ups for commercial traffic based on Ramsey pricing principles, on the assumption that mark-ups lead to increased prices in the final market according to the proportion of costs or turnover the mark-up accounts for.

Mark-ups on high-speed trains in France are highly differentiated, with mark-ups for the most profitable high-speed trains approaching 1000%. In both countries, mark-ups for traffic carried under public service contracts are essentially politically determined. The only externalities charge is that Germany also has a bonus/malus scheme of charging for noisy wagons.

3. Charges to cover specific elements of costs

3.1 Wear and tear costs

Wear and tear costs are the costs of increased maintenance and renewals necessary when more traffic uses the infrastructure. They may vary both with the characteristics of the vehicle (gross weight, speed, axle load etc.) and the track (e.g. curvature) as well as the efficiency of the infrastructure manager concerned. Thus the absolute level of wear and tear costs may be expected to differ between countries.

Regulation (EU) 2015/909 permits three alternative ways of calculating the marginal cost for wear and tear, namely accounting, engineering or econometric methods. Nothing is said about how differences in cost estimates from using the respective approaches should be handled.

There is an ongoing puzzle regarding maintenance and renewal costs, as to why different approaches give very different values. 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 although absolute levels of wear and tear cost differed, the relationship between costs and gross tonne kilometres have traffic showed a reasonable degree of consistency. The range of (mean) cost elasticities with respect to gross tonne kilometres 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 complicated by the fact that the cost base included 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, indicating elasticities closer to 50% for this category of cost (see Andersson et. al. (2012); Odolinski and Nilsson (2017) and 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%** - but the lower part of the range of estimates could suggest a possible range of closer to **25-35%**.

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

A recent study using French data provides evidence that further supports the above ranges. An econometric study for SNCF-Réseau (see Smith et. al. 2016; 2017) reported elasticities for different maintenance asset classes. Whilst some models saw elasticities at the top end of the range noted above, the simpler models preferred by the regulator yielded estimates more towards the bottom end. Overall, taking all maintenance and renewals categories, the final results show variabilities of 17% for maintenance and 23% for renewals.

In terms of the methods actually used to estimate marginal wear and tear costs, the British engineering model suggests much lower marginal costs than those implied by these elasticities (below 10%). Similarly, the German estimates, based on expert judgement on the proportion of each category of costs that are variable, are relatively low, as are the Swedish figures, which are based on out-of-date econometric models that only looked at maintenance costs. Only the French values, which are based on recent econometric evidence for maintenance costs and a combination of an engineering and econometric model for renewals, are at much higher levels than those for all the other countries (Table 1).

The reasons for the differences are not entirely clear. The strong advantage of the econometric approach is that it is based on actual cost data – so 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. Here, unit cost measures are needed and 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. 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 data from cross-sectional analysis and panel 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).

Currently, data for such econometric exercises is limited to a small number of countries; we recommend that all infrastructure managers should institute a database which shows work done on each track section together with dates and costs, along with detailed information on traffic carried and characteristics of the track. However, where such a database does not exist it will take many years to accumulate sufficient data for econometric analysis, so it may be necessary to base charges on relationships established with data in other countries.

Table 1: Direct costs of track maintenance and renewals (2017, except France which is 2019) in euros (range for different train types)

Britain	0.38 per train km (average for passenger trains)
France	1.39 – 2.99 per 500 tonne train km
Germany	0.67-1.32 per train km (including operations)
Sweden	0.3 – 0.8 per 500 tonne train km

Source: Case study reports

Both engineering and econometric evidence suggest that wear and tear costs are heavily influenced by gross tonne km rather than just train km, and engineering research suggests that a number of other variables, including axle weight, are also important influences. A failure to take account of this in the charging system will have the potential to distort decisions on what timetable to run, by overcharging light trains and under charging heavy, and also to distort decisions about what type of rolling stock to use (Nash et al, 2014). Whilst the above study finds that the effectiveness of these incentives is limited by the short time horizons of train operating companies on short franchises, we still conclude that charges should be differentiated according to train weight and other characteristics of the train.

There are also strong engineering reasons and some econometric evidence for believing that marginal maintenance and renewals costs will vary according to the characteristics of the track, particularly curvature (Marschnig, 2016). Wheat et al (2009) finds that the marginal cost on high quality track is lower than on poorer quality track. Thus we believe that charges should also vary with characteristics of the track. However, care is needed here as to precisely which aspects of track quality / capability should be reflected in charges, to avoid providing perverse incentives if infrastructure managers were able to charge more for lower quality track.

3.2 Train planning and operations costs

Train planning and operation costs include the costs of preparing the timetable and signalling, and real-time train control. It seems plausible that with modern computerised systems such costs are fixed, to a large extent, in the short run, although there are still cases where system staffing will depend to some extent on workload. Thus we find the one case of econometric evidence, which finds a cost elasticity of 0.15, plausible (Wheat et al, 2009). This implies a small addition to direct cost of wear and tear to cover these costs, as is the case in the two countries

which include such a charge in their charges for direct costs (Table 2). Strictly, these costs should be charged per train km, as it is the number of train km run rather than the weight and other characteristics of the train that influences these costs.

Table2: Charges for Planning and operations (euros)

Britain	0
France (2019)	0.12-0.22 per train km
Germany (2018)	0.14 per train km
Sweden	0

Source: Case study reports

3.3 Capacity costs

Capacity costs refer to situations where existing capacity does not suffice to meet demand. Railways differ from roads in that the capacity problem is solved beforehand, i.e. queues never materialise. The annually updated time-table is the explicit realisation of how scarcity is handled, establishing the departure-arrival patterns that are permitted during the upcoming year.

However, the time-table does not make it explicit which departure-arrival patterns deviate from the ideal of the respective operators. Neither is it feasible to deduce which trains are not given slots because of scarcity, nor indeed which demand is never made explicit since the operators realise that they will not be accepted. The difference in the social value of the time-table that is operated and the time-table that would accept all explicit and implicit demand for slots is the cost of scarce capacity.

Simple as this definition is, it is today impossible to quantify. One reason is the mathematical complexity of the time-tabling problem, meaning that there does not exist an explicit optimisation mechanism to handle the trade-offs that have to be made. In addition, operators' value of each Departure-Arrival pattern is not known. For these reasons, time-tables are today established using ad hoc techniques and rely strongly on the skill of the responsible staff.

The only country with a specific methodology for calculating capacity costs is Britain. Here the costs are estimated in terms of the cost of increased unreliability as capacity utilisation rises, so this is not a scarcity charge. The methodology does not allow for what happens when demand for paths outstrips capacity and some requests cannot be met.

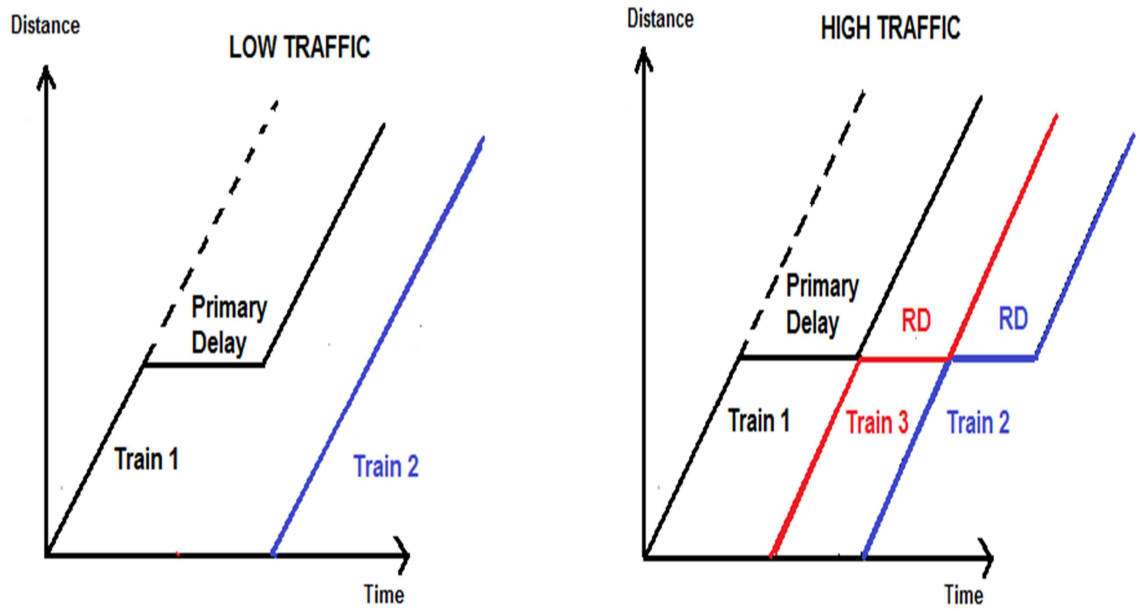
It should be stressed that what is being measured for the British capacity charge is not the cost of delays caused directly or indirectly by unreliability of the extra train itself; those costs are charged for as part of the performance regime. Rather it is the reduced ability of the system to recover from delays, brought about by the extra train, even if it is not in itself the cause of delays. Consider Figure 1. In the first diagram, there is a long gap between trains 1 and 2 so that the delay to train 1 will not impact upon train 2. But in the second part of the diagram, train 3

has been added between them. Now the delay to train 1 not only delays train 3, but also then has a knock-on effect on train 2 (Gibson et al., 2002)

Studies have been undertaken on the cost of being unable to accommodate all requested trains for Britain, estimating the relevant cost as the opportunity cost (in terms of foregone revenue, user benefits and externalities less saved costs) of excluding a desired service from the tracks (Wheat et al, 2009). In Britain, where most trains are franchised passenger services, it is usually assumed that the excluded train(s) would be franchised passenger services.

More generally, it has been suggested that this opportunity cost may be revealed by a process of auctioning paths (Nilsson, 2002). It is recognised however that this process would be complex, given the many different ways in which capacity may be turned into a timetable and the fact that the willingness to pay for one path depends on what other paths the train operator itself gets and its rivals get. Nilsson suggests dealing with this by requiring bids to include the change in the bid that getting a path earlier or later than desired would lead to, and by allowing bidders to change their bid in an iterative process when they see the overall allocations of paths. However, auctioning of paths remains little used in practice.

Figure 1: Reactionary delay



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

A further possible process is to set the initial scarcity charges on the basis of judgement at a low level and then to gradually raise them in accordance with the balance of supply and demand. This is, in practice, the approach used in Sweden.

Given the growth in rail traffic that is foreseen by the latest Transport White Paper (EC, 2011) and indeed the growth already taking place in some countries, it would seem that charging for

capacity should be an important element of rail track access charges. To the extent that reserving a path deprives another train operator of its use, whether it is actually used or not, capacity charges should be accompanied by reservation fees. On the other hand, congestion and scarcity costs may be assumed to be already internalised when both the train causing the problem and that affected by it belong to the same company. Thus charges for capacity may only be appropriate where there is no dominant operator. This does not apply for congestion charges where, as in Britain, there is a performance regime under which the infrastructure manager compensates operators for reactionary delays; in this case where the operator delayed is the same as the one causing the delay they are compensated for the delay through the performance regime, offsetting the fact that they have paid a congestion charge.

3.4 Environmental costs

In general it appears that the failure to charge rail operators for externalities is the result of the failure explicitly to make such charges for other modes. Of course, it would be better as a long term policy to levy such charges on all modes than to continue to ignore them on rail. Although uncertainties remain, the methodology to do this now exists (Ricardo-AEA, 2014). In general it would seem that noise costs should be charged per train km, but differentiated according to the characteristics of the train and its location. For air pollution, charges might more reasonably be per litre of diesel. Since electricity production is part of the Union-wide emissions trading scheme, it is not relevant to consider the consequences for the climate of using this category of traction.

In the absence of externality charging for all modes, EU legislation requires that externality charges do not increase overall revenue from track access charges, which implies a bonus/malus system (but this may push some charges below direct cost if there are no mark-ups).

4. Mark-ups

Given typical estimates of direct costs and the failure to charge either for externalities or for scarce capacity, then unless government is willing to pay most of the costs of rail infrastructure itself, substantial mark-ups are unavoidable. If the econometric evidence on direct cost and the need for scarcity and environmental charges were accepted, then the need for mark-ups would be less, but in many countries it would remain.

Generally, the current approach to estimating efficient mark-ups in Germany and France is to assume that the reaction of the operator to a mark-up is to pass it on in the form of higher prices to the final consumer. However, is the percentage price increase the percentage increase in TACs times the proportion of costs that are TACs, or times the proportion of turnover? The two would be the same if competition forced price to equal marginal cost. This may be a reasonable assumption in many countries in the case of freight, where increasingly competition between alternative train operators may be taken to ensure that price remains close to costs, although this does depend on there being effective competition in all parts of the rail freight market, which may not always be the case. It is less true of passenger services, where on-track competition remains the exception rather than the rule and where for many services decisions on what services to operate and what fares to charge are taken by public authorities through public service contracts.

For public service contracts, it is important that charges cover all the costs that would be avoided if the contract did not exist (including fixed costs of tracks including passing loops which would no longer be needed). This will ensure that these services are not cross subsidised by other services, and that the franchising authority is faced with the true costs of their decisions. The best way of charging for fixed costs would seem to be through a two-part tariff whereby the avoidable costs of the infrastructure they use and any required contribution to joint costs are charged as a fixed element of the tariff. Provided that whoever wins the contract pays the same fixed charge, this is not discriminatory. A problem does arise, however, in achieving a level playing field where services run on public service contracts are in effect in competition with purely commercial operations paying a mark-up.

For commercial passenger services, it may generally be assumed that the operator is already maximising revenue for the service they operate (subject to any price regulation of passenger fares in force, which is only the case in Britain). Thus they will not simply pass on the mark-up in the form of yet higher prices; rather they will cut services, as some services cease to be profitable, and only raise prices inasmuch as this is necessary to match supply and demand for the services they have left. Thus it seems that the best approach is that used by consultants to ORR in Britain who examined the difference between revenue and cost for individual passenger services in order to design efficient mark-ups for commercial passenger services (CEPA, 2017b). Of course this margin will vary greatly in time and space, suggesting that a complex system of

mark-ups varying by route and time of day/day of the week will be needed to produce efficient results.

One problem with the determination of efficient mark-ups is that it requires access to data on demand and costs which the train operator regards as confidential. Although there have been a number of independent studies of demand elasticities in the countries that have introduced mark-ups that have not depended on such confidential data, data on costs and revenues for individual market segments certainly are. It would appear that regulators need the power to access this data. This is the case in Britain, where all passenger train operators are required to use the same ticketing system and to make the data available in confidence to the regulator.

France and Germany distinguish between a small number of types of long-distance passenger, regional passenger and freight trains in applying mark-ups. In France, mark-ups on high-speed trains vary greatly by route. In Germany mark-ups are higher on higher speed services and services between the main metropolitan regions. The rationale for this is that these are the routes on which business travel is a significant proportion of total travel, and business travellers are less price sensitive than leisure. Only Britain distinguishes freight mark-ups by commodity (treating containers as a single commodity); this is straightforward in a country where most freight trains carry a single commodity but less so where wagonload services remain common. But the strength of road competition, and therefore the price elasticity of demand, does vary greatly by commodity, being much less strong for bulk commodities between private sidings than for manufactured goods and containers. From this point of view, being unable to distinguish between commodities in containers is not a serious problem since road haulage competitiveness does not vary much by commodity. Competitiveness does vary by length of haul, although if mark-ups were higher for long-distance traffic this might simply cause longer distance journeys to be split into two.

If cross subsidy between types of traffic are to be avoided then the revenue raised by mark-ups should at least cover the avoidable costs of the market segment in question, unless part of those costs is covered by government grant to the infrastructure manager. These avoidable costs will include the full costs of any tracks or passing loops required exclusively for these services. Provided that each market segment is at least covering its avoidable cost, there is no unique way of allocating the remaining costs which are truly joint. In Britain, in its last years, British Rail implemented a system of cost allocation which allocated these joint costs to the market segment considered to be the prime user of the assets in question, as part of a 'sector management' policy in which each sector (e.g. intercity passenger, regional etc.) had its own management, objectives and accounts. However, there is a degree of arbitrariness to this approach when several different sectors all make heavy use of the same assets.

None of the countries in our sample using mark-ups distinguishes between international and domestic traffic. Thus the mark-up for the end to end journey for international traffic is simply the weighted average of that in each country. It seems likely that the ability to pay for international traffic would be different from domestic, so that an efficient mark-up would

require direct negotiation between the different infrastructure managers and regulators regarding the level of mark-up and how it would be shared. This approach is encouraged by the European Commission. However, the failure to distinguish international traffic does imply that the risk of excessive charges for international traffic as a result of each infrastructure manager trying to earn as much as it can from the traffic (the so-called problem of double marginalisation) has not been a problem in practice in these countries.

5. Incentives for efficiency

EU legislation requires that infrastructure managers are given incentives for efficiency either through the regulatory regime or the multi-annual contract with the state. One way of doing this is through setting track access charges and other sources of income for the infrastructure manager on the basis of efficient costs rather than any other way to measure costs. This is designed to give infrastructure managers an incentive to produce efficiently, since otherwise they will not be able to cover their costs. However, given that infrastructure managers are government owned and there are no shareholders to bear the costs, there are doubts as to how effective this incentive is.

This requires a way of estimating efficient costs. The British regulator has in the past used international benchmarking for this, and with the decentralisation of Network Rail into separate routes it intends to use internal benchmarking in the future. The British regulator also has the specific role of checking that the requirements the state places on the infrastructure manager in terms of capacity and performance are consistent with the finance it provides, again assuming the infrastructure manager delivers them efficiently.

In Germany, the regulator sets charges on an RPI-x basis, where x is the estimated growth in productivity of industry as a whole. The state provides a grant to cover renewals costs and new investment, and incentivises efficiency through setting quality standards to be met within a budget constraint, and with penalties for non-performance. In France, the Regulator has disallowed a recent proposed increase in track access charges partly because of doubts as to whether the infrastructure manager's level of costs is efficient; the regulator also concluded that the current multi-annual contract between the state and the infrastructure manager is unrealistic, but unlike in Britain has no specific powers to intervene on this issue.

Whatever combination of regulatory action and multi-annual contracts is chosen as a way of incentivising efficiency, there is a problem in that all European infrastructure managers are publicly owned, so there are no private shareholders to bear the costs of financial penalties or pressures. Most countries do have bonus schemes for senior managers related to performance, and in Britain these do come within the scope of the regulator's powers, whilst reputational incentives may also be important (again the British regulator plans to make more use of these by benchmarking individual Network Rail routes against each other – i.e. the use of yardstick competition).

EU legislation requires the use of performance regimes to ensure that neither infrastructure managers nor train operators minimise their own costs at the expense of imposing costs on other parties through poor performance. Britain has the most sophisticated performance regime, calibrated to compensate train operators for the loss of revenue and any extra costs imposed by delays or by track closures for maintenance work. It may be argued that these incentives are inadequate, however, as they do not take account of costs of delays imposed on customers. Particularly where demand is inelastic, these may not be adequately reflected in loss

of revenue to the train operator. If penalties including these costs were paid to the train operator then it would over-compensate them, so any element of the performance regime reflecting costs other than those imposed on the train operator should be paid to the state.

Where there is separation between infrastructure and operations, no one body will seek to optimise systems costs as a whole unless ways are found to incentivise them to do so. Whilst track access charges and performance schemes help, these will not incentivise train operators to help the infrastructure manager reduce any elements of costs they do not bear. Moreover, a performance scheme designed purely to compensate train operators for the impact on their finance of performance failures will not encourage the infrastructure manager to take account of the full social cost of poor performance (such as delays to passengers who do not change their behaviour). Thus train operators will have an incentive to lobby for the highest possible levels of quality and capacity, and to exploit the rights regarding track occupation times for maintenance and renewals to the full. Britain has experience of ways of sharing infrastructure cost and revenue risk, through deep alliances or other risk sharing schemes, to give such an incentive.

In short, we recommend use of benchmarking to establish efficient costs and the setting of track access charges and other sources of revenue for the infrastructure manager on the basis of these costs (with allowance for the fact that it may take time to reach these cost levels, so reductions in charges may need to be phased). Careful thought also has to be given to ways of incentivising train operators to help optimise elements of infrastructure costs they do not bear.

6. Conclusions

The most important costs that need to be reflected in track access charges are wear and tear and congestion or scarcity costs; mark-ups also are becoming a very substantial element of track access charges.

Probably the most reliable evidence on the absolute level of wear and tear costs comes from econometric evidence, given the extensive range of studies conducted through co-ordinated research programmes. This evidence suggests that charges for these costs are generally far too low, except in France. Data for such exercises is limited to a small number of countries; we recommend that all infrastructure managers should institute a database which shows work done on each track section together with dates and costs, along with detailed information on traffic carried and characteristics of the track. However, where such a database does not exist it will take many years to accumulate sufficient data for econometric analysis, so it may be necessary to base charges on relationships established with data in other countries.

There remains a puzzle that engineering models and judgement seem to suggest a far lower level of variability of maintenance and renewals costs than does the econometric approach, and this does require further investigation.

It appears that maintenance and renewals costs also vary with characteristics of the vehicle and of the track on which they are running. The degree to which econometric evidence provides enough detail for such differentiation is limited (this being more possible with larger datasets as in France), although there is clear evidence for charging according to gross tonne km rather than simply train or vehicle km. Further differentiation may need to be based on engineering models. A failure to differentiate appropriately will distort choice of rolling stock and choice of route, as well as the basic question of what level of output to offer. At the least, charging per gross tonne km with some differentiation according to axle load, as in Sweden but at levels above the Swedish, appears desirable, as does varying charges with the characteristics of the track.

Where routes are operating at high levels of capacity utilisation, and there is a variety of operators, some combination of congestion and scarcity charging makes sense. It raises additional revenue in a way which gives train operators an appropriate incentive to economise on the use of scarce capacity, whereas mark-ups distort decisions on prices and service levels. Modelling of congestion costs, as in Britain, appears appropriate to deal with the costs of increased unreliability as track utilisation approaches capacity, whilst scarcity charges may rely on a combination of modelling and judgement, adjusting from a small initial value according to whether that charge has the desired impact on demand.

Mark-ups for freight traffic may reasonably make use of the assumption that final prices rise in accordance with the proportion of costs accounted for by track access charges, and apply final demand elasticities scaled by this proportion to determine optimal mark-ups. Ideally these would vary by commodity, as the evidence is that there are substantial differences in price

elasticities between commodities. But it is recognised that this may complicate billing systems particularly where individual trains carry a variety of commodities. In addition, for some traffic such as containers, the commodity carried is not known, so containers themselves must be regarded as the market segment. To the extent that road haulage competition for containers varies with length of haul rather than commodity carried, this is not a problem.

For mark-ups for commercial passenger services, it is not clear that the assumption that these are simply passed on as higher prices is appropriate, except in the limited number of cases where there is strong on-track competition. In the general absence of on-track competition, it may be assumed that commercial passenger operators are already maximising revenue, and the issue therefore is more one of the extent to which increased track access charges lead to the operation of fewer services. Minimising distortion here will require sophisticated differentiation by route and time of day/day of week, basing mark-ups on the estimated margin between revenue and train operating and direct infrastructure costs.

The issue of the mark-up for services operated under public service contracts is, and inevitably must remain, largely political, but such services should at least cover their avoidable costs, including the total cost of any routes only used by such services and the costs of additional tracks or passing loops on shared routes.(Indeed, this is a rule for all categories of service unless the relevant costs are covered by a grant to the infrastructure manager if cross subsidisation is to be avoided). These costs are best charged as a fixed charge, which is non-discriminatory as long as whichever company wins the franchise is charged the same. To the extent that charges with mark-ups to cover avoidable costs leave an element of joint costs uncovered, there is no unique way of allocating this cost, although the prime user approach formerly used by British Rail has some rationale.

Promoting efficiency for a publicly owned monopoly infrastructure manager is not straightforward as financial penalties will tend to impact on the state or the companies' customers rather than shareholders. Nevertheless, we recommend use of benchmarking to establish efficient costs and the setting of track access charges and other sources of revenue for the infrastructure manager on the basis of these costs, relying on reputational incentives and the impact on management bonuses to make these incentives effective. We also recommend a performance regime which incentivises both the infrastructure manager and train operators on the basis of the full social cost of delays or interruptions to services due to maintenance work. There remains an issue as to how to incentivise train operators to work with the infrastructure manager to reduce total infrastructure costs. This requires some degree of sharing of infrastructure cost risk between the infrastructure manager and train operators.

One final comment is in order. The charging of marginal cost plus mark-ups determined by Ramsey pricing for rail infrastructure is only economically optimal when other modes are appropriately charged. To the extent that this is not the case, there is no guarantee that implementing this approach to rail track access charges will improve the efficiency of the transport system; it may actually make matters worse. There is provision in the legislation for

time-limited subsidies for rail track access charges to compensate for under charging on other modes, and some countries, including Britain, make some use of this. However, efficient transport pricing is not an issue that can be addressed for one mode in isolation; progress needs to be made on all modes simultaneously. European legislation does permit kilometre-based charges for heavy goods vehicles to reflect marginal social cost but few member states have implemented such a charge.

References

CEPA (2017a) Market-can-bear analysis: Freight services, report by CEPA for ORR

CEPA (2017b) Market-can-bear analysis: Passenger services, report by CEPA and Systra for ORR, revised version

European Commission (2011), 'Road map to a single European transport area', Transport White Paper, COM(2011) 144, Brussels: European Commission

Gibson, S., G. Cooper, and B. Ball (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

ITF (2008) Charges for the use of rail infrastructure OECD, Paris.

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

Marschnig (2016) Innovative track access charges. Transportation Research Procedia 14 1884 – 1893

Nash CA, Smith ASJ, Goodall R, Kudla N and Merkert R (2014). *Economic Incentives for Innovation: A comparative study of the Rail and Aviation Industries*. Report, Institute for Transport Studies, University of Leeds

Netirail (2017) Deliverable D1.7 Incentives Final Report Annex 1 Research on Incentives for Innovation

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

Odolinski, K. & 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

Ricardo-AEA (2014) Update of the Handbook on External Costs of Transport. Final Report

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

Smith ASJ, Walker P, Wheat PE, Guiraud L and Silavong C (2017) Estimating the marginal maintenance cost for the French railway network: a comparison of models, ITEA Conference, Barcelona 2017

Smith ASJ, Wheat, PE, 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., Marti, M. and A.S.J. Smith (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, PE, Smith, ASJ and Matthews, B (2015) Deliverable 5.3: Access Charge Final Report, SUSTRAIL.

Wheat, PE, Smith, ASJ and Nash, CA (2009) CATRIN D8 Rail cost allocation for Europe. University of Leeds, Leeds.