COMPETITIVE AND SUSTAINABLE GROWTH
(GROWTH)
PROGRAMME

UNITE

UNIfication of accounts and

marginal costs for Transport Efficiency

Annex A4
Deliverable 10: Infrastructure Cost Case Studies
British Rail Infrastructure Case Study

Version 2.0
June 2002

Authors: Chris Nash and Bryan Matthews (ITS)

Contract: 1999-AM.11157
Project Coordinator: ITS, University of Leeds

Funded by the European Commission
5th Framework – Transport RTD

UNITE Partner Organisations

ITS/UNIVLEEDS (UK), DIW (De), NEI (Nl), CES/KUL (Be), TIS.PT (Pt), IWW/UNIKARL (De), VTI (Se), IER/USTUTT (De), CERAS/ENPC (Fr), HERRY (Au), EIET/ULPGC (Es), ISIS (It), STRATEC (Be), SYSTEMA (Gr), VATT (Fi), ECOPLAN (Ch), INFRAS (Ch), EKONO (Fi), EKI (Se)

UNITE

1999-AM.11157
<u>UNIfication of accounts and marginal costs for Transport Efficiency</u>

British Rail Infrastructure Case Study

This document should be referenced as:
Chris Nash and Bryan Mtthews (ITS) with contributions from partners – UNITE (UNIfication of accounts and marginal costs for Transport Efficiency) Working Funded by 5th Framework RTD Programme. ITS, University of Leeds, Leeds, August 2002.

June 2002
Version No: **2.0**
Authors: as above.

PROJECT INFORMATION

Contract no: 1999-AM.11157:
UNIfication of accounts and marginal costs for Transport Efficiency

Website: www.its.leeds.ac.uk/unite

Commissioned by: European Commission – DG TREN; Fifth Framework Programme
Lead Partner: Institute for Transport Studies, University of Leeds (UK)

Partners: ITS/UNIVLEEDS (UK), DIW (De), NEI (Nl), CES/KUL (Be), TIS.PT (Pt), IWW/UNIKARL (De), VTI (Se), IER/USTUTT (De), CERAS/ENPC (Fr), HERRY (Au), EIET/ULPGC (Es), ISIS (It), STRATEC (Be), SYSTEMA (Gr), JP-TRANSPLAN (Fi) VATT (Fi), ECOPLAN (Ch), INFRAS (Ch), EKONO (Fi), EKI (Se)

DOCUMENT CONTROL INFORMATION

Status: 1st submission to DG TREN
Distribution: UNITE partners
Availability: Public (only once status above is “Accepted”)
Quality assurance: C A Nash, University of Leeds
Co-ordinator’s review:

Signed: **Date:**

1. Introduction

The aim of this case study is to examine the evidence on the marginal cost of rail infrastructure use arising from the periodic review of rail infrastructure charges undertaken by the Rail Regulator in Britain in the period 1997-2000 (ORR, 2000a). It is particularly concerned with the estimation of marginal wear and tear costs, although the review also covered electric traction and congestion costs.

The next section discusses in principle the various elements of marginal social cost, and the following one describes the approaches to estimating marginal wear and tear costs that were developed during the periodic review. We then present the evidence on the size and variability of marginal wear and tear cost before drawing conclusions.

2. The Components of Short run Marginal Social Cost

The marginal costs of rail infrastructure use are the costs generated when an additional train uses that infrastructure. This case study focuses on short run marginal costs, assuming that capacity is held constant. It does not, therefore, consider costs which are fixed in the short run. The additional costs generated when an additional train uses the infrastructure can be divided into five main types: use-related wear and tear costs; congestion costs; scarcity costs; external accident costs and environmental costs. There may also be costs of signal operation, which some railways regard as part of infrastructure costs and in Britain certainly are a part of Railtrack’s costs recovered through access charges, as are the costs of providing and maintaining stations, freight terminals and depots where these are owned by the infrastructure manager.

Wear and tear of railway track is caused by a combination of usage-related and environmental-related damage. This results in the need to inspect, maintain and renew the track which gives rise to costs. Use-related wear and tear costs are that proportion of these inspection, maintenance and renewal costs, which result from trains using, and hence causing damage to, the track.

On roads, congestion usually manifests itself as volumes of traffic such that speeds are reduced below free-flow speed and/or queuing occurs at junctions. Since rail infrastructure managers control access to the network on a planned basis, rail congestion manifests itself in a different form. Indeed, it is useful to distinguish between two effects of shortages of capacity - congestion and scarcity.

Congestion represents the expected delays resulting from the transmission of delays from one train to another. The introduction of an additional rail service onto the network reduces the infrastructure manager’s ability to recover from an incident and increases the probability of delays. This becomes worse at high levels of capacity utilisation, since there is a lack of spare capacity to recover from any delays. Congestion costs are the costs associated with these expected delays. In this way, the consumption of additional capacity and the resulting congestion on the network imposes delay costs on train operators and, ultimately, rail passengers.

Scarcity represents the inability for an operator to obtain the path they want, in terms of departure time, stopping pattern or speed. Therefore, in the presence of a capacity constraint, the value of any train which could not be run as a result of the path being allocated to the train in question would be added to the costs of track damage and of expected delays. The High Level Group on Transport Infrastructure Pricing identified scarcity, rather than congestion, as the dominant consequence of existing capacity constraints on the existing rail network (European Commission, 1999).

Environmental costs arise out of the impacts of local and regional air pollution, global warming and noise emitted by railways. Several methods have been developed for valuing these impacts and extensive national and international research has been conducted over the past decade to derive actual values (Friedrich et al, 1998).

Whilst rail is a relatively safe mode, there are costs arising out of railway accidents. When travellers use a rail service they expose themselves to the average accident risk on that service. At the same time, their use of that service may increase or decrease the accident risk for all other rail users and may increase or decrease accident risk for users of other transport modes. Moreover, part of the costs of accidents may be imposed on third parties (such as the National Health Service) and not recovered from the rail company or its insurers. The economic value of these consequences of additional rail use form the marginal accident cost. The users, in their decision to travel by rail, internalise the risk they expose themselves to, valued as the willingness-to-pay for safety on the part of the households to which they belong; they may, or may not, also take the willingness-to-pay for safety on the part of their relatives and friends into account in their decision. The marginal external cost then consists of the expected accident cost to the rest of society when the users expose themselves to risk by using the rail service (eg, medical and hospital costs) the willingness- to-pay of the household, relatives and friends and the rest of society related to the increase or decrease in the accident risk for all other rail users and the willingness- to-pay of the household, relatives and friends and the rest of society related to the change in accident risk for other transport users. Whilst the exact nature of the relationship between the use of rail infrastructure and the number and severity of accidents is not clearly understood, infrastructure charges may be viewed as a ‘convenient’ means of increasing the efficiency of the transport system by internalising marginal external accident costs.

With modern signalling systems signal operating costs are largely determined by the characteristics of the signaling system and the planned capacity rather than the actual volume moving over the system. There may be some scope to adjust staffing according to traffic volumes, and indeed to adjust the hours the route is open to traffic, but in general the latter will not form a part of the marginal cost of a specific train, and the former will be small. Signal operating costs are not considered further in this paper.

Station, terminal and depot provision and maintenance costs are generally determined by the capacity provided. Station operating costs may vary with both throughput of trains and number of passengers, but these costs are treated as part of supplier operating costs.

The main focus of this case study and, hence, of this paper is on infrastructure costs. Supplier operating, congestion, environmental and accident costs are the subject of case studies in other work packages, and are, therefore, not considered in detail in this paper. A note on railway congestion and scarcity, drawing also on the regulatory review, is provided as an annex to UNITE D7.

Much of the existing literature on railway costs and cost functions treats the costs of rail operations together with the costs of usage-related wear and tear to the infrastructure, whereas we are interested specifically in the latter. However, the recently completed periodic review of track access charges in Britain provides a rich source of information for this case study. The next section provides a summary of the different approaches to estimating the costs of infrastructure use put forward during this periodic review.

3. British Approaches

3.1 Background

Between 1997 and 2000 Railtrack, Britain's railway infrastructure manager, and its Regulator, the Office of the Rail Regulator (ORR), have reviewed the way in which charges for the use of rail infrastructure, levied by Railtrack on train operating companies, are calculated. This review has involved detailed examination of railway infrastructure costs and the ways in which they vary. The overall approach adopted in this review was to identify the incremental costs imposed on the network, and hence borne by Railtrack, by each train using the infrastructure.

The review considered three cost categories: usage costs; electricity costs; and congestion costs. Scarcity costs, along with environmental and accident costs, were, thus, not included.

Section 3.2 describes some of the mechanisms which give rise to usage costs, whilst section 3.3 seeks to briefly summarise Railtrack's Usage Cost Model and section 3.4 then describes the track usage sub-models which form the main element of the Usage Cost Model in more detail.

The overall approach to estimating usage costs put forward by Railtrack was a bottom up approach based on an understanding of detailed engineering relationships and the summation of individual elements of cost caused by additional trains. Somewhat by way of contrast, the Regulator put forward a top down approach which starts by identifying the total planned maintenance and renewal expenditure on different types of asset, then applies the percentage of these costs which vary according to number of trains run so as to derive a total variable cost for each asset type. It then uses detailed engineering relationships to allocate these total variable costs to particular vehicle types.

Congestion costs were included as the increased compensation paid by Railtrack for delays. This applies even where the delay is to a train of the same train operating company, and therefore already in a sense internalised, but in that case it is offset by the payment made by the TOC to Railtrack for the primary delay. (It may be argued whether in fact these costs are fully internalised given that ultimately they are borne by rail customers, and it is only if the resulting impact on train operators' revenue reflects the full cost to users that they are fully internalised). Further details of the approach to congestion costs is given in an annex to UNITE D7.

3.2 Usage Costs

For the purposes of analysing infrastructure usage costs, it is useful to divide railway infrastructure into four categories:

- Track;
- Signalling;
- Electrification equipment; and
- Structures.

Reviews of the literature undertaken by Railtrack and the Regulator make it clear that maintenance and renewal of the track is the most significant component of usage costs. They are also the best understood component, in terms of their underlying engineering relationships. In addition a small part of the cost of structures, signals, and electrification equipment are considered to vary with usage. Further details of the percentages of costs which vary with usage for each of the four asset types are given in section 4.

Railway track comprises three main components, rails, sleepers and ballast. Damage to these components ultimately results in their replacement or renewal. Prior to their renewal, there is also a need to carry out maintenance work, which can be considered under three headings; the maintenance of track geometry, the inspection and patrolling of the track and a final category to cover the wide variety of other maintenance activities which are undertaken. Table 1 summarises the elements of track which vary, to a greater or lesser extent, with usage.

Table 1: Track costs variable with usage

Cost	Description
Track geometry	In the case of vertical geometry, deterioration is primarily due to differential ballast settlement under loading; this requires maintenance activity (tamping). The amount of maintenance will depend on the rate of deterioration, the standard required (line speed) and the effectiveness of maintenance.
Rail	One of the major causes of rail fatigue is loading (and cumulative loading); Maintenance is required to manage defects; cumulative loading will determine renewal. Rail wear, which takes place on the railhead and on the side of rail in curves, is a direct function of usage.
Sleepers	Affected by impact loads and (for concrete sleepers) abrasion due to contact with ballast.
Ballast	Accumulation of fine material generated from usage (including the maintenance process itself).
Switches & crossings	Subject to the same damage mechanisms as plain line track.

Maintenance	Inspection rates vary if total traffic passes threshold levels. Some minor maintenance activities, such as changing rail pads, are also usage-dependent.

Sources: Railtrack; BAH analysis

Structures are subject to deterioration in the long term which may be accelerated by the loads they carry. Signalling may be more likely to fail according to the number of signalling operations performed, and the overhead electric wire, or third rail, is subject to wear from contact with the pantograph or shoe. However, these costs appear to be much smaller in magnitude than that of wear and tear of the track.

Railtrack, and the consultants who worked for the Regulator, both take the view that engineering models are the most appropriate approach to the measurement of usage costs. Statistical analysis of past expenditure on maintenance and renewal costs is subject to a number of caveats. Firstly, the cyclical nature of much maintenance and renewals work makes it difficult to establish the traffic volume associated with the work carried out in any one year. Where there are constrained budgets, this may be compounded as a result of certain maintenance activities, which are often one of the first candidates for postponement when funds become tight, being deferred from one year to another. These deferred activities are then undertaken at some later stage when funding becomes less constrained but, clearly, in this kind of situation the relationship between cost causation and expenditure becomes strained. Therefore, “while statistical procedures can be helpful in understanding the behaviour of infrastructure maintenance expenses, they must be applied with an understanding of the inherent deficiencies of the input data and the validity of the results should be tested by logic and engineering judgement based on experience” (Booz Allen and Hamilton, 1999).

Nevertheless, statistical models are used in another UNITE workpackage, and our view is that, provided that the volume of data is sufficient for patterns to emerge despite the year to year fluctuations in attention to any specific section of track, they may be a valuable aid to understanding in this field. However, the necessary data to apply such models was not available in the British case study.

3.3 The Railtrack Usage Cost Model

Railtrack commissioned consultants to develop an infrastructure usage cost model to enable them to estimate the marginal costs of wear and tear to the infrastructure that can be directly attributable to a train service using that infrastructure. Railtrack wanted, for the purposes of being able to provide a predictable and steady profile of usage charges, to be able to estimate the costs as the wear and tear occurs, rather than as the expenditure on maintenance and renewal activities occurs. Thus the preferred approach is one in which the cost of bringing forward future maintenance and renewals is taken into account, rather than simply allocating current expenditure.

Track being the most significant and most well-understood element of rail infrastructure costs, Railtrack and their consultants focused their efforts on development of a detailed model to estimate track usage costs. The resultant track usage cost model has 7 track damage sub-models as follows:

Rail Maintenance;
- Rail Life;
- Sleeper Life;
- Track Geometry Maintenance;
- Ballast Life;
- Switch & crossing;
- Track inspection.

The models which emerged are built upon the earlier track damage models used in Mini-MARPAS, a computer model developed in the late 1980s based on considerable research into the mechanisms of track deterioration. The consultants also considered how the network should be segmented and how vehicles should be classified for the purposes of these models.

The approach firstly uses information on vehicles and track type to calculate the effect of an additional train on either maintenance requirements or the life of the asset and, secondly, uses data on the cost of the maintenance and renewal activities to derive an overall cost for the work required. That is, engineering relationships are used to estimate the damage caused and the effect on the life of each type of asset of the passage of a vehicle of a given type, and unit costs are then applied to these effects. This two-stage approach enables the usage models to be calibrated firstly against asset lives and actual quantities of maintenance work done and, secondly, against maintenance and renewal cost data.

The usage cost model uses a variety of data inputs. The different types of input data are: traffic data (eg Train services, Route of each service, Train consist of each service); infrastructure data (eg, Track type, sleeper type, line speed by network segment); vehicle data (eg, Specification of vehicle types and axle data); and cost data (eg, Unit costs of maintenance and renewals).

There are 4 stages in the process:
1. Network categorisation;
2. Population of the vehicle damage matrix;
3. Estimation of the life of each type of asset on each part of the network and determination of the amount of 'damage' done by a service to each part of its route; and
4. The application of unit costs to each of the 'damage records' generated in Process 3 (taking into account the 'amount' of the network category within each segment).

Network categorisation identifies the amount of each network category (track miles of plain line or numbers of S&C units) in each segment of constant traffic. Essentially, this is done by comparing the contents of two Railtrack databases. It is a ‘one-off’ process and only need be up-dated when there is a change to the infrastructure or the routing of train services.

The Usage Cost Model contains 121 'network categories', each one being a combination of: Track type (CWR, Jointed or S&C); sleeper type (Concrete, Hardwood, Softwood or Metal); and track quality band (11 bands defined by line speed).

The 'vehicle damage matrix' stores the usage-related damage attributable to an individual vehicle for each network category. 'Damage' is defined here as: “The proportion of the life of an asset that is consumed when a single vehicle of the given type passes over a piece of track of the given network category (for the 'life' models); and the proportion of a maintenance operation that is required when a single vehicle of the given type passes over a section of track of the given network category" (1).

The vehicle damage matrix is populated by combining results from individual degradation models to produce an estimate of total asset damage, and hence a proportion of life used. Each cell in this vehicle damage matrix represents a specific combination of track, vehicle and operation characteristics so the calculation is repeated for every relevant combination of vehicle, track and operation characteristics. Again, population of the vehicle damage matrix is a 'one-off' exercise, which only need be up-dated when the vehicle data changes, or if the parameters of the track models are changed.

The Mini-MARPAS damage models did not separate usage and environmental-related damage and expressed asset life in numbers of years. However, rather than completely rewrite them for use in the Usage Cost Model, usage and environmental damage were separated where feasible and asset life was converted from years to number of vehicles. This enabled the structure of the existing Mini-MARPAS equations to be utilised. "The comparisons of the asset lives produced by the Usage Cost Model and those experienced in service indicates that the damage values stored in the vehicle damage matrix are reasonable" (1).

The life of an asset in a given track section is calculated by summing the total annual usage damage from all the traffic that uses it and the environmental damage, and dividing the result into the total amount of damage before the asset must be replaced. The total annual usage damage is found by summing the appropriate values from the 'vehicle damage matrix' for every vehicle using the section in a year. This calculation is undertaken for each type of asset (rail, ballast and sleeper), and for each network category within each segment of constant traffic.

The amount of 'damage' imposed on the track by a service is calculated for each combination of: vehicle in the service's train consist; segment in the service's route; and network category within each segment of the service's route. A typical service will have several hundred combinations. The relevant 'damage' information is looked up in the 'vehicle damage matrix' (for the vehicle type and network category concerned).

The final step is for the service to be costed. This is done by applying unit costs to each of the 'damage records' generated in the previous step (step 3), taking into account the 'amount' of the network category within each segment, and then summing over all costs for that service.

In addition to being able to separate usage-related damage from environmental damage, the model improves on the earlier Mini-MARPAS model in a number of ways. Firstly, it uses more detailed data on vehicle characteristics, with regard to speed, number of axels, axel load etc. Secondly, it incorporates 3 recent developments in technology: Stone blowing and design overlift; improvements in rail steel metallurgy; and cold bolt hole expansion. Thirdly, a number of improvements were made to the track geometry model, the earlier version of which was recognised to be poor.

Railtrack's estimates of non-track asset cost variability with usage are based on more general, top down, engineering assessments. For structures, Railtrack analysed 'total expenditure on heavy maintenance and renewal of structures and in February 1999 estimated that the usage related costs of structures was equivalent to 10% of track usage costs and represented £30-40 million pa over the network. Initial work to develop a bottom-up approach to these costs, focusing firstly on costs associated with underbridges, was being undertaken by Railtrack but the current status of this work is not known. For signalling, a top-down assessment, based on the judgement of Railtrack's engineers, concluded that the impact of usage on signalling renewals is not significant. For signalling maintenance, the total costs assessed as being related to usage amounts to 5-6% of total maintenance and renewals spend – equivalent to about 10% of signalling maintenance spend pa. BAH view this as being an over-estimate of variability as it relies on an assumption that additional usage will lead to a proportionate increase in the number of points movements, though points are not moved for every train movement. For electrification assets, a top-down study in October 1999 concluded that most maintenance costs did not vary with usage but that usage-related degradation would be a primary driver of renewals requirements. Overall, Railtrack estimated that usage-related costs amount to 15% of total electrification spend, or some 38% of renewals spend pa.

3.4 The Track Damage sub-Models

Rail life and rail maintenance requirements are determined by rail failure rates. Rail failure, and ultimately rail replacement, usually results from one of two different mechanisms:

rail fatigue - for a number of different reasons, all of which are strongly related to the ‘loading environment’, cracks form in the rails. Most types of rail fatigue can be detected through inspection and then corrected by some form of maintenance activity, but if not detected, rail fatigue may lead to rails fracturing. Because the number of such cracks forming on any particular rail increases with the cumulative loading on that rail to date, the maintenance input increases over time. At a certain point, maintenance activity reaches a level whereby it becomes more economic to replace the rail.

Rail wear - rails become worn, both on the ‘head’ of the rail and, in the case of curves, along the sides of the rail. Lubrication can be used to reduce wear but rails will require replacement when satisfactory wheel/rail contact angle conditions cannot be guaranteed, or the rail gauge becomes excessive

Whilst the mechanisms of both rail fatigue and wear have been well researched and models are available, only rail fatigue models were incorporated in Railtrack’s Usage Cost Model. The omission of rail wear has been criticised as a major simplification but Railtrack argue that the realistic representation of rail wear requires extremely complex modelling and do not consider it to be a sufficiently major cost driver. Also, rail grinding has not been incorporated as, at the time of the review, it was not used extensively in the UK.

Rail maintenance costs are derived from the average rail failure rate throughout the whole life cycle of the rail. “The Usage Cost Model determines this from the annual rail failure rate which is determined by summing the individual rail failure rates throughout the life of the rail assuming an annual traffic level of 10 MGTPA (Million Gross Tonnes Per Annum) of the vehicle being assessed (AEAT, 1999). It is necessary to make an assumption about the annual traffic level because the rail damage models used in Mini-MARPAS had been formulated using annual traffic level and age as a measure of cumulative damage to date. The relevant equations are as follows:

For Continuously Welded Rail

annual failure rate = (beta) * tache ovale failure rate + squat failure rate + thermic failure rate + flashbutt factor * flash failure rate

For Jointed track
annual failure rate = (beta) * tache ovale failure rate + squat failure rate + (alpha) * bolthole failure rate.

Rail life is set by calculating when the rail reaches a predetermined rail failure limit at which it is judged to be more economic to replace the rail. For CWR track this limit is set as 1.38 failures/mile/year and for jointed track this limit is set as 1.76 failure/mile/year. These limits were determined historically to ensure that the Mini-MARPAS rail life predictions agreed with actual rail lives experienced on BR.

Railtrack state that it is very difficult to determine the life of sleepers. Failure can arise from a number of mechanisms, none of which appear to be very well-understood. Hence, the sleeper life sub-model which forms part of the track usage model is empirical in nature and based closely on the original Mini-MARPAS model. The model is expressed as:

Sleeper Life = 1/(Annual Traffic Damage + Annual Environmental Damage)

For empirical purposes, the Annual Environmental Damage may either be a constant or vary as a function of the annual traffic level. Annual traffic damage is calculated from a Sleeper Damage parameter which is a function of axle load, P2 forces and rideforces, as follows:

Sleeper Damage alpha (Axle load + P2 Force + Dynamic Ride Force) squared

The track geometry maintenance costs sub-model and ballast life sub-model are both based on predictions of the rate of track geometry deterioration with traffic. the rate of track geometry deterioration is used to calculate how frequently track geometry maintenance work is required and as this maintenance work contributes to damaging the ballast it then forms an input to the ballast life model.

A review of track geometry deterioration, focusing on the deterioration of vertical geometry (Shenton, 1984) suggested that the basic ballast settlement (delta) was proportional to the maximum load due to a passing axle, multiplied by the number of axles raised to the power of 0.2.

the incorporation of track roughness, resulting from differential settlement, into the Railtrack track usage model was a significant enhancement to previous geometry deterioration models. The contribution to track roughness from three components - Dipped Joints, the Unsprung mass and Ride Forces - were incorporated by estimating the total standard deviation (SD) of the track as follows:

SDTotal (in mm) = square root of((SDDipJoints) squared * (SDTotUnsprungMass) squared * (SDTotRideForce) squared))

Thus, the track geometry deterioration model can estimate the settlement which will result after the passage of different quantities of traffic. From this, the rate of track quality deterioration can be calculated. This can then be combined with track standards in order to determine how frequently track geometry maintenance is required. However, in actuality the amount of track geometry maintenance work undertaken at any particular location is dependent on Railtrack Track Quality Group targets and the local condition of the ballast so, in order to account for these factors, a calibration value of 0.8 is applied to the model.

In addition, alternative maintenance treatments such as Stoneblowing and Design OverLiftTM, which reduce the rate of track geometry deterioration and extend ballast life, have been incorporated into the track geometry maintenance model. This has been done by again applying factors to account for the greater durability resulting from these maintenance treatments and then applying appropriate unit costs and varying the proportions of work undertaken by each method to reflect the mix of maintenance treatments.

The cost of manual maintenance work by track gangs is also modelled using the track geometry maintenance model, as the fundamental deterioration mechanisms are thought to be the same. However, a scaling factor is then used to relate the track geometry maintenance damage stored in the vehicle damage matrix to actual cost.

As ballast deteriorates with age and the ability to retain a satisfactory track geometry is consequently lost, the track usage cost model includes a ballast life sub-model to determine ballast renewal rates. The model is based on the size of reservoir within the structure which is available to collect fine material, and the rate of generation of the ballast fines; the ballast life is expired when the density of the fines reaches a specific failure criterion. The main sources of deposition of fine material in the ballast are damage caused by tamping, spillage from wagons, breakdown under traffic loading and those fines deposited from the atmosphere. The basic structure of the model is as follows:

Ballast Life = Reservoir Density/Density of fines generated per year

Fines generated per year = (Fines deposited from the atmosphere + Wagon spillage fines + Fines due to the breakdown of ballast under traffic + Fines generated by tamping)

Fines deposited from the atmosphere were excluded from the track usage model because these are not usage related. Further modifications to the model were made in the form of factors to reflect the increased durability resulting from the new maintenance treatments and parameters to model the fines generated. The ballast life is then calculated for an assumed 10 MGTPA of traffic for each of the relevant different vehicle types. This done, the proportion of the life attributable to a single vehicle of a particular type is calculated and stored in the vehicle damage matrix. Lastly, the cost of wet bed eradication is incorporated via the application of a scaling factor, determined by calibration against actual cost values, to the ballast life damage stored in the vehicle damage matrix.

Switch and crossing life is estimated in the track usage model by relating it to sleeper life. This is argued as being not unreasonable “as S&C components are likely to be replaced individually until the point is reached where the sleeper and fastening system have deteriorated to a poor condition” (AEAT, 1999). Sleeper usage damage is thus multiplied by a factor of 1.111 so as to account for the higher dynamic forces exerted on S&C as opposed to on sleepers.

Switch and crossing maintenance costs are modelled using the original Mini-MARPAS models for Switch Failures and Crossing Failures, derived using data from the mid 1980s. The models for each of the main types of failure relate the rate of failure to speed and traffic volume. However, few details of these models are available.

The inspection costs sub-model considers the costs of routine visual inspection, ultrasonic rail inspection and track geometry recording.

The total inspection costs for a network segment are calculated based on the EMGTPA of the total traffic using the track section and the relevant frequency of inspection set out in Railtrack's Group Standards. In conjunction with a unit cost to inspect a unit length of track, the frequency of inspection and length of network segment are used to calculate the total annual cost of inspection.

The Regulator's consultants, Booz Allen Hamilton (BAH) undertook a critical review of the Railtrack model and identified what they considered to be a series of potential and actual short-comings. Firstly, their review of the individual track damage sub-models highlighted a number of concerns. For instance, they argue that the exclusion of the impacts of rail wear within the rail damage sub-model is a significant over-simplification. Furthermore, they criticised the sleeper life model on the grounds that it is simply a 'curve-fitting' exercise. BAH argue that by taking regular gauge restraint and rail cant angle measurements, it would be possible to determine degradation rates which are both site specific and traffic-related but highlight that this kind of data does not appear to figure within the model. They conclude that "overall, this sub-model is not robust" (BAH, 1999) and indicate that AEAT themselves have advocated it's further development. They further conclude that including the inspection cost sub-model within the track usage model is likely to overstate usage costs. They argue that the impacts of increasing usage on inspection costs are not continuous and that inspection frequency depends more on line speed than simply on usage. Hence, there would only be an impact on costs if change in usage were to be substantial. Secondly, they identify a number of difficulties within the calibration process, including the use of a number of non-verifiable assumptions relating to costs and expenditures [Railtrack later refuted this criticism, at least in part]. Thirdly, they raise a number of concerns regarding the incorporation of different standards of track quality across the network.

Having said all this, BAH acknowledge that the Railtrack model does produce "usage-related costs for the various vehicle types that are internally consistent and generally conform to prior expectations" (BAH, 1999). Nevertheless, while acknowledging the value of Railtrack's detailed analysis, concern is expressed that complexity should not be mistaken for 'spurious accuracy'. That is, given all of the potential difficulties they identify with the detail of the Railtrack approach, a simpler approach may provide an adequate means of calculating costs.

4. Cost variability and estimates of marginal wear and tear costs

The BAH approach relied on a traditional cost accounting distinction between fixed and variable costs. However, the categorisation they used relied on an extensive review of empirical evidence.

Research carried out for Railtrack using the model described above indicated that approximately 20% of its total maintenance and renewals costs vary with asset usage. This variable component is dominated by track maintenance and renewal, and may be broken down as follows:

Track maintenance and renewal - 78%;
of which Geometry maintenance: 23%
    Rail maintenance: 2%
    Manual maintenance: 14%
    Rail renewal: 5%
    Sleeper maintenance: 9%
    Ballast renewal: 19%
    S&C renewal: 6%
Structures maintenance: 12%
Signalling maintenance: 5%
Electrification renewal: 5%

Certain assets are excluded from the analysis:

those whose costs do not show much variability with traffic volume - eg parking and housing;
those which relate to supplier operating costs - ticket-selling facilities;
non-transport-related assets - eg shops, restaurants etc;
rail stations, urban public transport infrastructure and inter-modal freight terminals.

Thus, the major component of the variable element of railway maintenance and renewal costs are associated with the track. Therefore, most attention has been given to these costs in cost research. Table 2 shows Railtrack's estimates of how different cost components vary with usage for the British rail network.

Table 2: Railtrack's estimates of asset usage costs and cost variability

Asset type	Overall variable cost per year (£)	Percentage variability	Percentage of estimated total usage cost
Track	300m	50	77
Underbridges	40-50m	20	13
Signalling	15-20m	5	5
Electrification	15-20m	5	5

(Source: Railtrack)

Thus Railtrack's estimates show that, of an overall level of expenditure on maintenance and renewals of £1175-1350m p.a., some £370-390m (or 29-32%) may be regarded as variable.

Whilst Railtrack believe their estimates of the extent of cost variability is broadly in line with experience from elsewhere, two caveats should be noted. Firstly, the British rail network, particularly in comparison with the US, has a much greater share of its total costs in signalling and electrification, which are inherently less variable with volume than is track and, secondly, traffic densities on the British rail network are relatively low in international terms. It would therefore be expected that the proportion of variable costs would be much lower for Britain than, for example, for the US.

In their review undertaken for the Regulator, Booz Allen Hamilton (1999) suggest that the percentage of track maintenance and renewal costs that is variable will vary by track standard and by traffic volume [and perhaps also by the type and speed of trains using the infrastructure] . For example, for lines with low traffic volumes the fixed costs will dominate the variable costs such that they will experience proportionately little change in total cost (even if volume doubles), and vice-versa for lines with high traffic volumes. It is suggested by Booz Allen Hamilton that lines with tonnages in the region of 50 million gross tonnes per year “will have a very high proportion of variable cost and particularly renewals costs, which will be almost entirely driven by usage, and for such lines the variability will be close to 100%. Other things being equal, the incremental cost will generally be greater the higher the track quality and the higher the average line speeds, although this will also be a function of the standard of initial construction. Furthermore, work carried out by the Office for Research and Experiments (ORE) of the UIC during the 1980s linked track maintenance and renewal costs to speed, axle loads and the condition of the track. They found that costs vary at approximately 60-65% of the rate of change in both speed and axle-load. The rate of change was also sensitive to track condition, with the increase generally being greater the poorer the quality of the track.

In a review of international research, BAH found that 30-60% of track maintenance and renewal costs vary with the level of use. The research they analysed varied in its coverage and nature. Some studies used engineering estimates, whilst others used statistical analysis of past expenditure; some studies were of very high density railways (eg in Russia and China), whilst others were of lower density railways (eg Australia); some studies were of predominantly freight railways (eg USA), where as others were of predominantly passenger railways (eg Europe); and some studies included costs of structures as well as track. Very little research into the variability of maintenance and renewals costs of structures, of signalling and of electrification equipment exists.

Table 3 shows estimates by Booz Allen Hamilton of how different cost components vary with usage for the British rail network. These figures seek to account for the above caveats but, in doing so, rely heavily on judgement.

Table 3: Variable costs of infrastructure (per cent)

	% variable	% by asset category
Track		38
Maintenance	30	
Renewals		
Rail	95	
Sleepers	25	
Ballast	30	
S&C	80	
Structures	10	10
Signals		2
Maintenance	5 0	

Renewals		
Electrification		24
Maintenance		
AC	10	
DC	10	
Renewals		
AC	35	
DC	41	

(Source: BAH, 1999, TABLE 6)

The Booz Allen Hamilton work therefore suggests a somewhat lower level of variability than Railtrack; applying the above percent variability estimates to the Railtrack figures of cost by cost category gives an overall per cent variability in the range of 21-23%.

An advantage of the Railtrack approach is that it produces estimates at a level of fine detail for different types of vehicle. The figures finally derived by the Regulator used a 'top down' approach to estimate the overall variable costs (i.e. by splitting Railtrack's total maintenance and renewal costs into fixed and variable using the above percentages) but the Railtrack model to apportion these between vehicle types. Some examples of the resulting figures are given in Table 4.

Table 4
Typical examples of usage charges (p/vehicle km 1999/2000)

Diesel shunter (class 08)	2.6
Diesel loco (class 47)	63.9
Electric loco (class 90)	59.7
Passenger car (mk 3)	10.4
Diesel multiple unit (class 158)	10.4
Electric multiple unit (class 333)	
Powered car	15.4
Trailer car	11.9
Freight wagon	2.7 - 3.3 *

* p per gross tonne km

Source ORR (2000a, 2000b)

Variation in the mean charge per train km for different types of train is given in Table 5.

Table 5
Usage cost per train km (£1998)

Inter city passenger	1.116
London commuter passenger	0.406
Regional passenger	0.149
Bulk freight	1.790
Other freight	0.880

Source Sansom et al (2001)

These figures reflect the fact that inter city trains are heavier and higher speed than trains in other sectors, London commuter trains are predominantly lower weight and lower speed multiple units whilst regional trains are predominantly multiple unit trains of two or three cars only.

5. Marginality and the links with average variable costs

What the approach outlines in the previous section does is to estimate average variable cost and to assume that this may be taken to approximate marginal cost. It is worth considering in what circumstances this is a reasonable assumption.

Average variable costs equate with marginal costs when average variable costs are constant. It seems reasonable to assume that the damage done by a particular run of a vehicle over a stretch of track will not depend on the density of traffic on that track, in terms of trains per week, although it may depend on the cumulative volume of traffic since the track was last repaired or replaced (however, it is unlikely that one would ever wish to vary charges by this variable given the fact that each train will run over a variety of segments of track each at a different stage in their life cycle). It is normal to define the life of a stretch of track between repair and renewal in terms a measure of output such as gross tonne kilometres of traffic; the effect of an additional gross tonne kilometre is to bring repair or renewal forward in time. A doubling of traffic will therefore double the frequency with which output related tasks need to be undertaken.

Thus it appears from a priori reasoning that the assumption of constant average variable cost may be reasonable. To be more confident in this requires empirical evidence, of which not much is available. However, in a companion UNITE case study, Jan-Eric Nilsson has analysed track maintenance and renewal costs for Swedish and Finnish railways. He finds that marginal cost is typically x % of average cost, and that marginal cost is relatively constant (BRYAN – please check details). This suggest that the Booz Allen model may be a reasonable way forward when the detailed data for statistical analysis is not available or time or cost do not permit this approach.

.

6. Generalisation

Our view is that the methodological approaches to rail infrastructure costing proposed in Britain, described in this paper have the potential to be used for railways in other countries. Whilst the approach proposed by Railtrack would appear to offer the potential for deriving detailed and accurate costs, a potentially major inherent drawback of the approach is that it is dependent upon a great deal of detailed data and modelling specific to the railway network concerned, which may not be readily available in other countries. It may be possible to use some of the British data as a proxy for actual data for other railways, though no attempt has been made to test this.

It may be possible to use a relatively simple and transferable approach for usage costs. Clearly both levels of cost and cost conditions vary between railways so it will not be easy simply to modify values and transfer them from one country to another in a reliable way. It may be slightly more reliable to transfer cost elasticities (the cost elasticity is equal to the ratio of marginal cost to average cost, which also equals the average degree of variability if average variable cost equals marginal cost). he evidence cited above suggests that in European conditions the typical rail infrastructure cost elasticity may be in the range of 0.2-0.3. However, we have already seen that the cost elasticity varies substantially with the detailed structure of costs, for instance the ratio of signalling and electrification costs to track costs, the quality of the track, speeds, axle loads etc.

It may be possible to do at least a little better than assuming an overall cost elasticity by using the BAH approach. BAH suggest that "There is little argument over which infrastructure components are affected by usage, and probably general agreement over the broad order of magnitude of the level of variability" (BAH, 1999). For example, we have highlighted in section three, through a review of costing studies, that usage costs as a proportion of total maintenance and renewal costs tend to lie within a range of 20-30% in European conditions. It may be reasonable, in the absence of better information, to transfer the proportionate cost variabilities in Tables 1 or 2 between railway networks in different countries. Similarly the relativities between different train types presented in Tables 3 and 4 may be reasonably transferable. Thus, the approach proposed for usage costs by BAH, which is relatively simple and is consistent with the overall level of expenditure is probably more susceptible to being transferred for use in other rail networks.

For instance suppose we had two railways, A and B, whose cost breakdowns were as follows:

	Railway A	Railway B	% variability
Track	70	50	38
Structures	10	10	10
Signals	10	20	2
Electrification	10	20	24
Overall Elasticity	0.3	0.25	

By treating the overall percentage variability as an estimate of the cost elasticity for infrastructure costs, we may estimate that the cost elasticity of railway A is .3 and of railway B is .25. Given that the cost elasticity is the ratio of marginal to average costs, if we have data on total infrastructure costs for each railway we could readily estimate marginal costs. We might then use the relativities from the British study to estimate the marginal costs for different types of track or rolling stock, if that is needed.

However, more statistical research on the percentage of total costs which vary with usage, and into what determines the percentage variability, would be desirable to improve accuracy. Care needs to be taken in transferring results between different railways because of differences in track quality, tonnage, traffic densities and service frequencies. However, the range of variability does appear relatively consistent across the different studies and we can use our understanding of the ways in which variability depends on these different characteristics to judge where particular railways lie along that range of variability.

References

Booz, Allen Hamilton (1999) Railway infrastructure Cost Causation, Report to Office of the Rail Regulator, London.

European Commission (1999): Calculating Transport Infrastructure Costs, Final Report of the Expert Advisors to the High Level Group on Infrastructure Charging (Working group 1)

Friedrich R, Bickel P and Krewitt W, eds. (1998), External Costs of Transport. Stuttgart: Institute of Energy Economics and the Rational Use of Energy (IER).

Gibson, S (2000) Charging for the use of railway capacity, Seminar on infrastructure charging on railways, Helsinki.

NERA (1998) An Examination of Rail Infrastructure Charges, NERA, London

Office of the Rail Regulator (1999) The periodic review of Railtrack's access charges. Usage charges - a technical consultation document. ORR, London.

Office of the Rail Regulator (2000a) The periodic review of Railtrack's access charges. Final Conclusions. ORR, London.

Office of the Rail Regulator (2000b) Review of freight charging policy - consultation on costs. ORR, London.

Sansom, T et al (2001) Surface Transport Costs and Charges. ITS, University of Leeds.

Shenton, M.J. Ballast Deformation and Track Deterioration ICE Conference on Track Technology for the Next Decade, Nottingham, July 1984]