

**Adam Mantzos  
Brockley Consulting**

**Review of cost attribution and cost allocation  
approaches**

**A report for Network Rail**

**18 March 2015**

## **Contents**

Summary.....	3
Introduction.....	8
Cost allocation principles .....	10
Cost allocation in other UK regulated network industries.....	19
Railway legal framework.....	25
Variable costs .....	27
Fixed costs .....	40
Next steps .....	57
Annex A - Relationship between SRMC and total traffic related costs.....	59

### Summary

1. This report considers approaches to cost allocation for the GB railway network, and opportunities to refine those approaches. Objective, consistent and accurate cost allocation can help make the underlying reality of cost causation more transparent, and improve the quality of the information available to industry stakeholders, including Network Rail, funders and operators. This can help the effectiveness of decision making processes.
2. Refinements to cost allocation could in principle also have a role to play in making pricing mechanisms more effective through changes to the structure of charges. However, this would require many other factors to be considered, including:
  - a) practicability of implementation at a level which is sufficiently detailed, accurate and responsive to changing circumstances to reflect theoretical benefits, yet simple enough to have a meaningful influence on behaviour;
  - b) the timescale required for developing and fully consulting on changes to the structure of charges, and how that aligns with future charge control periods;
  - c) recognition of societal benefits;
  - d) the nature of the current franchising and competition regime; and
  - e) the interests of government.
3. These considerations are beyond the scope of this report, and therefore none of the observations in this report should be interpreted as a recommendation for any particular structure of charges.

### Variable costs

4. Within the current cost allocation framework in rail, a distinction is generally made between "variable costs" and "fixed costs":
  - a) the term "variable costs" tends to be used to refer to costs that vary in response to relatively small changes in traffic levels; and
  - b) the term "fixed costs" tends to be used to refer all other costs.
5. The current approach to identifying and allocating variable costs is mainly based on the concept of Short Run Marginal Cost (SRMC). It appears to broadly reflect the SRMC of costs that Network Rail incurs, assuming a network without capacity constraints, i.e. where demand is consistently below available capacity.
6. However, with the growth in traffic that the network has experienced over the last two decades, demand is not consistently below available capacity, and capacity constraints are widespread. In the presence of capacity constraints, the theoretical efficiency benefits of SRMC require it to reflect:
  - a) the short run costs that Network Rail incurs; *and in addition*
  - b) the short run costs that operators incur as a result of not being able to access a capacity constrained network as much as they would like, i.e. the value foregone by demand that cannot be accommodated by existing available capacity (sometimes referred to as "scarcity costs").

## Review of cost attribution and cost allocation approaches

7. The current cost allocation framework does not measure or allocate this additional set of short run costs, and so does not reflect the cost of capacity constraints. SRMC and variable costs are therefore understated in parts of the network that are capacity constrained.
8. One specific consequence of this partial reflection of SRMC is that the current cost allocation approach can lead to perverse results. In particular, a disaggregated analysis of the costs that Network Rail incurs suggests that variable costs are lower than average in capacity constrained parts of the network. However, this ignores the fact that variable costs resulting from capacity constraints are higher than average in those parts of the network. Therefore a move to disaggregation, using an approach which reflects the first set of costs but not the second, could serve to worsen rather than improve the cost reflectivity of cost allocations.
9. More generally, since costs relating to capacity constraints are not reflected, the management of capacity constraints must rely heavily on administrative mechanisms. Decisions on allocating scarce capacity and on investing in enhancement expenditure to relieve constraints, have to rely on mechanisms such as the specification of franchises, the approval process for track access contracts, and the Long Term Planning Process, in which Network Rail, funders, operators and other stakeholders consider a range of factors in the decision making process.
10. Identifying and allocating the costs associated with capacity constraints may help improve the effectiveness of these administrative mechanisms, by giving decision makers better quality information for consideration.
11. In principle, it may therefore be worth considering expanding the analysis of variable costs in the cost allocation framework to include the marginal costs associated with capacity constraints.
12. This consideration should take account of the likelihood that in practice, there will be two main difficulties in implementing such an expansion:
  - a) Measurement: Measuring the short run costs associated with capacity constraints is inherently very difficult, as it requires estimation of commercially confidential operator information on the value of unmet demand. However, it may be possible to gain similar benefits from an easier to implement long run approach, based on the concept of Long Run Marginal Cost (LRMC). LRMC considers the marginal cost to Network Rail of enhancing the network to alleviate capacity constraints.
  - b) Complexity: The costs associated with capacity constraints vary significantly by location and time of day. If the benefits of allocating these costs are to be realised, this suggests analysis at a very disaggregated level. Careful consideration would need to be given to the likely scale of benefits given the level of disaggregation that is likely to be achievable in practice.

### Fixed costs

13. The current cost allocation framework for what the industry describes as "fixed costs", i.e. all other costs not identified as variable, is based on allocating costs to franchised passenger operators in line with traffic metrics.
14. There does not appear to be any explicit linkage between the use of traffic metrics to allocate fixed costs between operators, and the costs that operators cause on the network. This is in contrast to a Long Run Incremental Cost (LRIC) or "avoidable cost" based approach. Such an approach aims at an objective and transparent allocation of fixed costs between operators, in a way which reflects long run patterns of cost causation as far as possible. Moving closer to a LRIC based approach, including a revised approach to the allocation of residual non-avoidable or "common" costs, may therefore bring benefits in terms of objectivity and transparency.

## Review of cost attribution and cost allocation approaches

15. It is important to stress that this remains a consideration of cost allocation, not of charging. Given the presence of government funding to finance the gap between fixed costs and charges, there need be no direct connection between the allocation of fixed costs, and the level and/or existence of fixed or variable charges for operators. However, a cost reflective LRIC based allocation of fixed costs could provide a clearer understanding of:
  - a) the allocation of government funding between operators; and
  - b) the pattern of support provided by the different providers of government funding.
16. The current use of traffic metrics as an allocation approach is not necessarily incompatible with some versions of a LRIC based approach, as long as those metrics are appropriately selected, to reflect the long term drivers of incremental capacity on the network. It may be worth considering a review of the traffic metrics applied with this in mind. For example, there may be a case for a greater emphasis on traffic during peak periods, which is likely to drive many incremental costs, rather than overall traffic.
17. The difference between the current traffic metric approach and other LRIC based approaches may be more significant. However, it is not clear that these other approaches would be clearly superior, and it may be appropriate to consider the benefits of moving to such approaches against the effort that would be involved.
18. The use and choice of traffic metrics is not the only likely source of differences between the current approach and a LRIC based approach. Differences of potentially comparable significance may be the result of way in which the current traffic based approach is implemented, particularly in respect of:
  - a) level of disaggregation;
  - b) consistency of operating route treatment; and
  - c) consistency of operator treatment.
19. Allocations under the current FTAC approach are generally performed at operating route or more aggregated levels. There may be considerable opportunities to improve the current approach by allocating fixed costs at a more disaggregated level where possible. This would bring the current approach closer to a LRIC based approach. Clearly this would involve significant effort, but increased disaggregation is something that could be implemented in a graduated evolutionary fashion.
20. There does not appear to be any objective cost allocation rationale for the current distinction between the treatment of the Scotland operating route from other operating routes in England and Wales. It would seem more appropriate to treat every operating route consistently, and allocate fixed costs based on operator activity on that route.
21. The current approach to allocating fixed costs appears to conflate the issue of cost allocation with the issue of charging. Fixed costs are allocated to franchised passenger operators only, and not to open access and freight operators.
22. The distinction between operator types does not appear to be based on any objective cost allocation considerations. Rather, the distinction appears to be based on the charging arrangements in place for different operators, and in particular on the fact that only franchised passenger operators are subject to fixed charges. However, given the existence of government funding, there is no necessary link between cost allocation issues, and charging issues.

## Review of cost attribution and cost allocation approaches

23. Regardless of the cost allocation methodology ultimately adopted, there would be considerable merit in applying a single consistent approach to all operators. This would allow a clearer and more objective allocation of fixed costs, and a more transparent understanding of the distribution of government funding, without requiring any change to the level of charges paid by operators.

### Next steps

24. The current cost allocation framework does not appear to be always well understood. A better common understanding would help consideration of potential opportunities to refine the framework. It may be therefore be beneficial to develop a document which summarises the framework, including a description of the legal and regulatory context, details of current cost allocations and their inter-relationships, and explanation of the rationale for cost allocation approaches. Some of this is already summarised in the RDG's *Charges and Incentives User Guide*, which could form a good starting point.
25. The clearest opportunity for refining the current framework, in terms of likely benefits compared with implementation effort, would appear to be an improvement in the consistency and clarity of fixed cost allocation. This would be based on the current definitions of variable and fixed costs, i.e. including within the scope of variable costs, only Network Rail's SRMC.
26. As a first step, this could involve extending Network Rail's existing methodology for allocating fixed costs so that it is consistently applied to all operators and all operating routes. The impact of such a change could then be assessed and consulted on before considering potential implementation
27. Other steps might include:
  - a) clarifying the relationship between the freight avoidable cost analysis and the fixed cost allocation approach; and
  - b) improving the alignment between Network Rail's methodology and that applied in the ORR's *GB rail industry financial information* publication.
28. There also appear to be significant opportunities, again based on the current definition of fixed costs, to improve the accuracy of Network Rail's existing methodology for allocating fixed costs, and closing the gap with LRIC based approaches, by:
  - a) refining the choice of traffic metrics to better reflect drivers of incremental network capacity; and
  - b) applying traffic metrics at a more disaggregated level.
29. Realisation of these opportunities could require significant implementation effort in terms of data gathering and analysis; and questions remain as to alternative approaches to measuring LRIC. It may be best to explore these issues by conducting and consulting on a pilot study on a small section of the network before deciding whether to implement any changes more widely.
30. The remaining main area of opportunity relates to the possibility of expanding the analysis of variable costs to include the LLMCs associated with capacity constraints. The key issue here relates to the practicality of achieving the level of disaggregation necessary to deliver the potential theoretical benefits of such an expansion. Again, this may be best explored through a pilot study on a small section of the network, before deciding on potential implementation.

## Review of cost attribution and cost allocation approaches

31. Should all of these potential opportunities be implemented, the resulting cost allocation framework would comprise:
  - a) "variable costs", which would be allocated based on Network Rail's LRMC (i.e. the existing approach of Network Rail SRMC, expanded to include long run as well as short run marginal costs), resulting in a higher level of variable costs in aggregate; and
  - b) "fixed costs" (which based on (a) would be at a lower level than is currently assumed), which would be allocated based on a revised traffic metric approach more closely reflecting a LRIC or "avoidable cost" based approach, including a revised approach to the allocation of residual non-avoidable or "common" costs.

### Introduction

32. The broad principles of the current approach to cost allocation for Network Rail have been in place since the mid-1990s. The industry has developed significantly since then: traffic on the network has nearly doubled, major investment in network enhancement has been and continues to be made, details and properties of implemented and alternative charging approaches have been considered, and the industry governance structure and funding arrangements have evolved. These developments may create opportunities for cost allocation principles to be refined.
33. This report focuses on approaches to the allocation of costs. The allocation of costs to services should reflect the underlying factual reality of cost causation, i.e. by how variations in the provision of services affect costs incurred. This underlying reality exists, whether or not it is measured. Cost allocation simply makes that underlying reality transparent, helping the quality of information available.
34. The railway industry operates within a complex framework that includes public policy, social, legal, regulatory, environmental, and commercial factors. Moreover, the railway network itself is complex, with multiple uses and operators, and significant variations between different parts of the network and different times of the day. The management of the industry needs to consider and balance these complexities when making decisions.
35. Currently, this involves a significant administrative component: that is, pricing mechanisms which rely on interactions between charges and willingness to pay, are supplemented by administrative mechanisms in which Network Rail, its funders and operators consider such complexities in the decision making process.
36. The realisation of opportunities to refine the current approach to cost allocation should improve the effectiveness of these administrative mechanisms. Objective, consistent and accurate cost allocation can help make the underlying reality of cost causation more transparent, and improve the quality of the information available to industry stakeholders, including Network Rail, funders and operators. This can help the effectiveness of decision making processes.
37. In principle, a refined approach to cost allocation could also have a role to play in helping to make pricing mechanisms more effective, since cost reflective charges can help to incentivise the efficient use and development of the network. In addition, more effective pricing mechanisms may allow some reduction in the reliance placed on administrative mechanisms.
38. In practice, however, the ability to place greater reliance on cost reflective charges and pricing mechanisms depends on many other considerations.
39. Some considerations relate to the challenges of translating alternative cost allocation approaches into alternative charging structures, and in particular:
  - a) whether alternative charging structures can realistically be implemented, operated, kept up to date and responded to, at the level of complexity necessary to realise the theoretical benefits of greater cost reflectivity; and if not
  - b) whether simplified charging structures are capable of delivering net benefits, in view of the possibility that some charges will end up being non-cost reflective as a result of the simplification.
40. Other considerations which may limit the ability to place greater reliance on pricing mechanisms relate to the public policy, social, legal, regulatory, and other factors in the rail industry framework. These include:
  - a) the need for Network Rail to recover its total costs, which is achieved through a combination of charges to operators and government funding;



## Review of cost attribution and cost allocation approaches

- b) recognition of the benefits that passenger and freight railway services bring to society, above and beyond those that translate into operators' willingness to pay, as evidenced and supported by significant levels of government funding;
  - c) the need for the level and structure of charges to be mindful of the current passenger franchising regime, including the degree to which the regime insulates franchise operators from variations in charges, and the degree to which franchise specifications allow franchise operators the flexibility to adjust their patterns of operation in response to such variations;
  - d) the interests of government, as a major funder of the industry;
  - e) the level of on rail competition and the environment for freight; and
  - f) the framework within which the industry and in particular its funders make decisions in relation to investing in network enhancements.
41. The considerations which determine whether and how alternative approaches to cost allocation might go beyond providing better quality information to amending the structure of charges, and allowing a reduction in the reliance placed on administrative mechanisms, are complex, and beyond the scope of this report.
42. Where this report makes any observations on some of the potential consequences of alternative cost allocation approaches on the structure of charges, it should therefore be recognised that these are at this stage at a theoretical level only. None of the observations in should be interpreted as a recommendation for any particular structure of charges.

## Cost allocation principles

43. The allocation of costs to services should reflect the underlying factual reality of cost causation, i.e. by how variations in the provision of services affect costs incurred.
44. Cost causation approaches to cost allocation consider how costs change when a firm changes the pattern and level of its output. Different approaches adopt different perspectives to two principal dimensions of changes:
  - a) the scale of change, leading to the concepts of marginal and incremental cost; and
  - b) the time horizon over which the impact on costs is considered, leading to the concepts of short run and long run.
45. For the railway network, we are interested in how costs change when Network Rail changes its pattern and level of output, i.e. the nature and volume of train services running over the network.

## Marginal and incremental cost

46. Marginal cost is the change in costs resulting from the addition or subtraction of a very small change to the existing level of output. Typically, this is interpreted as the change in costs associated with a single unit of output<sup>1</sup>.
47. For the railway network, marginal cost might be the change in costs resulting from adding or subtracting a single train over a section of track.
48. Incremental cost is the change in costs resulting from a change to the existing level of output equal to a defined "increment" of output, divided by the volume of the increment to generate a unit cost<sup>2</sup>.
49. The defined increment can be positive or negative. For the railway network, this would translate into considering the impact on costs of:
  - a) a specified increase in the number of trains running over a section of track; or
  - b) a specified reduction in the number of the trains running over a section of track.
50. Where the increment is negative (sometimes referred to as a "decrement"), incremental cost is sometimes referred to as "avoidable cost". That is, it considers the costs that would be avoided absent the increment.
51. The size of the defined increment can vary, depending on the nature of the issue which it is intended to inform<sup>3</sup>. The definition of the increment, in terms of size and sign (i.e. positive or negative) can have a very significant effect on the scale and nature of the resulting incremental cost. Because of this, any consideration of incremental cost should be made with a clear understanding of the definition of the increment. In particular, it cannot be assumed that an incremental cost measure from one context is similar to an incremental cost measure from a different context, unless the defined increments are known and compared.

---

<sup>1</sup> Technically, marginal cost is the derivative or gradient of the total cost curve, i.e. the change in cost associated with an infinitesimal change in output.

<sup>2</sup> Strictly speaking, "incremental cost" is the term for the change in costs, and "average incremental cost" is the term for the unit cost measure which results from dividing incremental cost by the volume of the increment. However, since it is usually the unit cost measure that is of interest, it is often described as "incremental cost".

<sup>3</sup> Marginal cost can be thought of as a special case of incremental cost, where the size of the defined increment is a single unit of output.

## Review of cost attribution and cost allocation approaches

52. In practice, the increment is typically defined so as to be relevant to the issue that incremental cost is intended to inform. For example:
- a) If a firm is considering supplying a new product, it may be interested in comparing the potential revenues from that product with the additional costs that would result from supplying that product. In such a case, the relevant increment might be the increase in output that would result from supplying the new product.
  - b) If a firm is considering the pricing of an existing product, it may be interested in comparing the actual revenues from that product with the proportion of its existing cost base caused by the supply of that product. In such a case, the relevant increment might be that product's contribution to the existing level of output, i.e. the reduction in output that would in principle result if that product were no longer supplied.
53. It is important to stress that while the increment should be relevant to the issue under consideration, that does not mean that the gain or loss of the increment must always be a realistic option under consideration. For example, in mobile telecoms Ofcom considers the costs attributable to the termination of voice calls, i.e. the costs that a network operator incurs when receiving an incoming call intended for one of its own customers from a different network. It does so by defining as an increment the loss of all incoming calls, even though it would be very hard to imagine a mobile telecoms industry where no incoming calls were accepted.

### Costs

54. In principle, the "costs" whose changes are considered in both the marginal cost and incremental cost concepts are defined widely. They comprise not only the costs incurred by the firm supplying the output, but also costs incurred by the firm's customers, and by society more widely.
55. The costs incurred by the firm's customers are considered further below. Societal costs can include costs associated with noise, pollution and environmental emissions. These are difficult to measure, and neither the current cost allocation framework, nor most alternatives considered, nor this report, attempts to do so. This limitation is one of the reasons why decisions need to take a range of factors other than cost allocations into account.

### Short run and long run

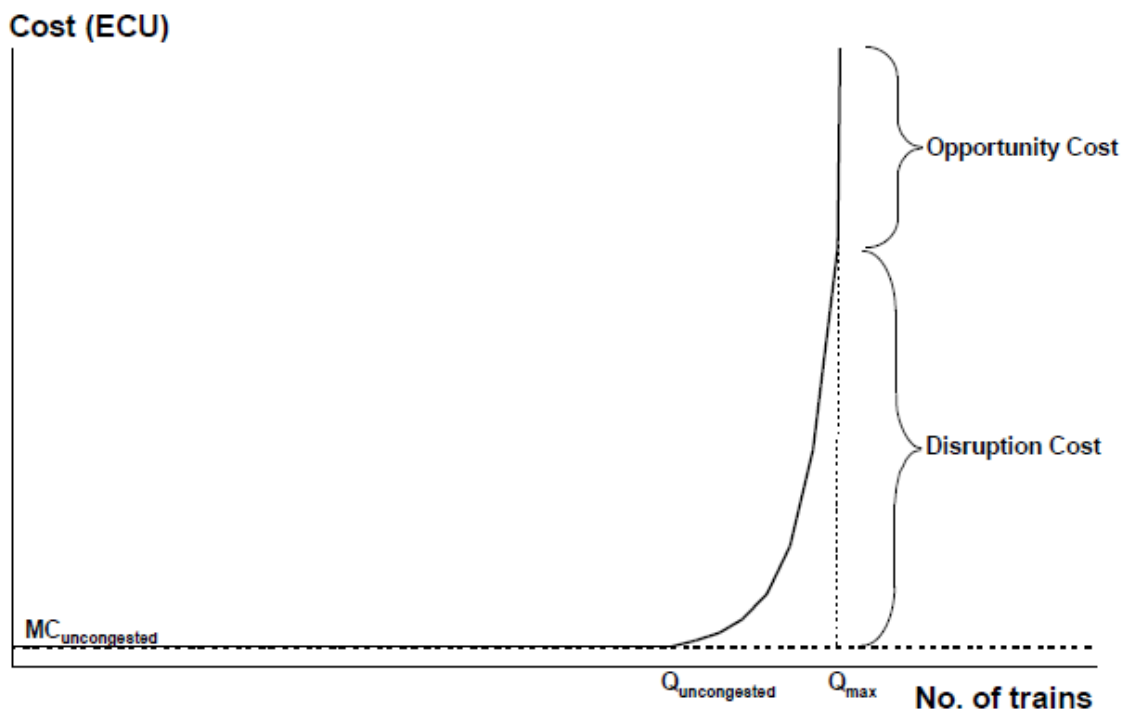
56. Both marginal cost and incremental cost require the estimation of the level of costs under a counterfactual scenario, after the addition or subtraction of a unit or increment of output, relative to the existing level of output. That estimation can adopt either a short run or long run perspective:
- a) Under a short run perspective, it is assumed that capacity is fixed at existing levels, regardless of how well suited that level of capacity is to the counterfactual level of output. Marginal cost and incremental cost considered under this perspective result in the measures termed Short Run Marginal Cost (SRMC) and Short Run Incremental Cost (SRIC)<sup>4</sup>.

---

<sup>4</sup> Note, "short run" refers only to the assumed inability of a firm to vary capacity in the short run. Apart from changes to the level of capacity, all other cost consequences of varying output are reflected in "short run" cost measures, even if they would take many years to realise. For example, the "short run" cost of track maintenance can reflect the fact that increased traffic levels may accelerate the future renewal of track through increased wear and tear, even if that renewal remains many years into the future.

## Review of cost attribution and cost allocation approaches

- b) Under a long run perspective, it is assumed that capacity is adjusted to the level best suited to the counterfactual level of output. Marginal cost and incremental cost considered under this perspective result in the measures termed Long Run Marginal Cost (LRMC) and Long Run Incremental Cost (LRIC).
57. The short run and long run measures feature different outcomes in situations where capacity is constrained, i.e. where output approaches available capacity.
58. In line with the assumption that capacity is fixed at existing levels, short run measures reflect the consequences of output approaching existing but rigid capacity limits. Where capacity becomes constrained in that sense, short run measures include the costs those capacity constraints impose not just on the firm supplying the output, but also on that firm's actual or potential customers. These include:
- the costs to other customers as a result of the increased disruption and delay resulting from accommodating additional marginal or incremental output; and/or
  - the value foregone by potential customers to whom output cannot be supplied as a result of accommodating additional marginal or incremental output (sometimes referred to as "opportunity cost")<sup>5</sup>.
59. These costs to customers can be very significant, and can cause short run cost measures to rise well above the levels suggested by looking at the supplying firm alone, and that prevail when capacity is not constrained. NERA illustrated this for SRMC with the following diagram in its 1998 study on rail infrastructure charges<sup>6</sup>:



Source: NERA (1998). Note, the dashed horizontal line indicates SRMC for an unconstrained (or uncongested) network.

<sup>5</sup> Section 3.2.1, *An examination of rail infrastructure charges*, NERA et al, May 1998; and Section 2.1, *Potential Generator Market Power in the NEM*, NERA, June 2011

<sup>6</sup> *An examination of rail infrastructure charges: Final report for the European Commission*, NERA et al., May 1998

## Review of cost attribution and cost allocation approaches

60. In principle, where capacity is constrained, SRMC/SRIC rises as high as is necessary for SRMC/SRIC based charges to reduce demand, i.e. price existing traffic off the network and deter potential traffic, until the level of remaining demand willing to pay those charges matches the level of existing capacity<sup>7</sup>.
61. For the railway network, where capacity is constrained, in principle SRMC includes, in addition to the marginal wear and tear associated with the running of an additional train:
  - a) the cost to other operators of the increased disruption resulting from running an additional train, in terms of the net revenues those operators lose as a result of that increased disruption<sup>8</sup>; and
  - b) the value foregone, in terms of operator net revenue not earned, as a result of removing an existing train and/or denying a potential train from running on the network<sup>9</sup>.
62. Long run measures adopt a different approach where capacity starts to become constrained. Whereas short run measures reflect the cost to customers that arise from the assumed inflexibility of capacity, long run measures reflect to the cost to the supplying form of alleviating the capacity constraint. LRMC/LRIC include the investment cost that the supplying firm would incur to expand capacity in order to accommodate the additional marginal or incremental output.
63. For the railway network, where capacity becomes constrained, LRMC includes, in addition to the marginal wear and tear associated with the running of an additional train, the cost to Network Rail of enhancing the network to accommodate that additional train.
64. This can make LRMC very volatile: if traffic is below available capacity on a particular section of track, LRMC incorporates no additional enhancement expenditure; but if capacity is constrained, LRMC could allocate, to a single additional train, all of the costs associated with enhancing that section of track.
65. Where the increment in LRIC is defined as being negative, i.e. a reduction in output, the long run perspective will incorporate as part of LRIC the costs that could be avoided in the long run as a result of reducing capacity in response to the reduced level of output (sometimes referred to as "avoidable costs").
66. For example, in the railway network, if traffic were reduced on a section of line with two track pairs (i.e. four tracks in total) it might in the long run be possible to reduce that section to one track pair. All of the costs of the second track pair would then be considered avoidable, and form part of LRIC.

---

<sup>7</sup> Section 3.2.1, *An examination of rail infrastructure charges*, NERA et al, May 1998; and Section 2.1, *Potential Generator Market Power in the NEM*, NERA, June 2011

<sup>8</sup> Net revenues lost comprise any losses in fare and other operator revenue, net of any savings in running costs.

<sup>9</sup> *Charging for Scarce Rail Capacity in Britain: A Case Study*, Johnson and Nash, March 2008

## Relationship between SRMC, LRMC, SRIC and LRIC

67. All four cost standards (SRMC, LRMC, SRIC and LRIC) give different perspectives on the same cost causation issue, i.e. what costs are objectively caused by, and therefore attributable to, different changes in output. The different perspectives depend on the treatment of two dimensions:

- a) the scale of change (marginal versus incremental); and
- b) the time horizon over which the impact on costs is considered (short run versus long run).

	Marginal	Incremental
Short run	Short Run Marginal Cost	Short Run Incremental Cost
Long run	Long Run Marginal Cost	Long Run Incremental Cost

68. Economic theory suggests that setting prices equal to SRMC (including the costs to customers that arise from capacity constraints) encourages efficient consumption. SRMC signals the cost to society of supplying a product at a particular time:

- a) if the product is priced below SRMC, some customers are incentivised to buy output at the price charged, even though the cost of providing that output exceeds their willingness to pay - resulting in an inefficiently high level of output; whereas
- b) if output is supplied at a price above SRMC, some customers are incentivised not to buy output at the price charged, even though their willingness to pay exceeds the cost of providing that output - resulting in an inefficiently low level of output<sup>10</sup>.

69. An important feature of SRMC based pricing is that all output, including existing output, is priced at SRMC, i.e. at the marginal cost associated with an increase over and above the existing level of output. No distinction is made between the pricing of existing output, or existing customers, and the pricing of new output or new customers. This is because the customer incentives described above which could lead to an efficient level of output at non-SRMC prices apply equally to existing and new customers.

70. For example, consider a situation where SRMC at the existing level of output of 100 units is £10 per unit, and an existing customer of one of those 100 units is charged £5, and is not willing to pay more than £5. That customer is incentivised to continue buying that output, preventing output from falling to 99 units. In effect this causes an increase in the level of output from 99 units to 100 units, even though the cost of providing that increase, equal to the SRMC of £10<sup>11</sup>, exceeds the existing customer's willingness to pay for that increase.

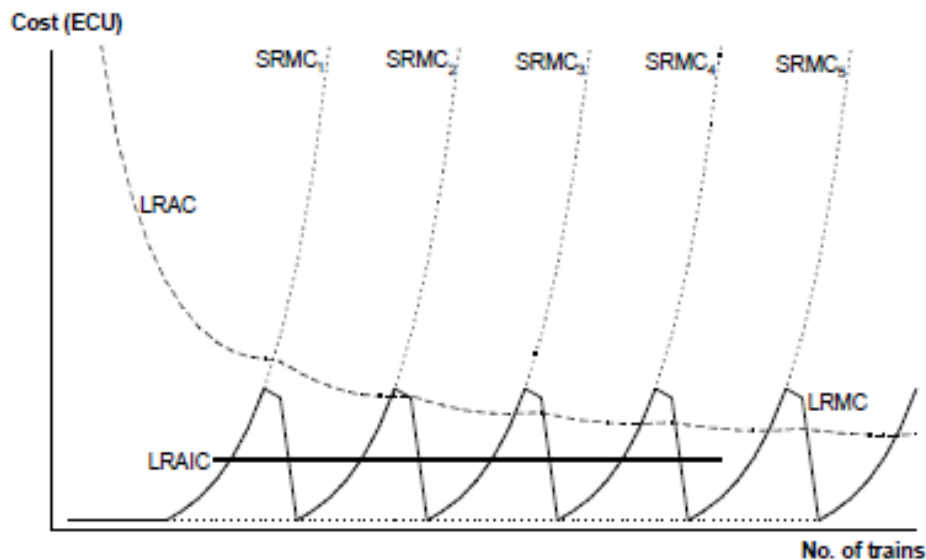
71. For the railway network, making no distinction between existing and new output in SRMC based charging effectively means that every train is treated as being "last on" to the network, regardless of the history of that train service.

<sup>10</sup> Section 3.3, *An examination of rail infrastructure charges*, NERA et al, May 1998

<sup>11</sup> Since SRMC considers a very small change in output, the cost of a small decrease is typically considered identical to the cost of a small increase.

## Review of cost attribution and cost allocation approaches

72. In theory, SRMC, LRMC, SRIC and LRIC are all linked:
- SRMC and LRMC are linked over the long term. As demand approaches and exceeds capacity, the opportunity cost of unmet demand rises. Eventually, SRMC increases to the point at which it exceeds the cost of expanding capacity to meet that unmet demand, i.e. LRMC. At that point capacity is in principle expanded and both SRMC and LRMC fall. Because of this link, SRMC and LRMC should not diverge materially over the long term<sup>12</sup>. However, divergences in the short to medium term can be material if the investment required to expand capacity is "lumpy", as in the railway network.
  - SRIC/LRIC represent smoothed versions of the SRMC/LRMC curves, since similar absolute cost changes are spread across larger increments.
73. NERA (1998) illustrated the relationship between SRMC, LRMC and LRIC (termed "LRAIC" by NERA) with the following diagram, which shows LRMC peaking and then falling with each capacity expansion<sup>13</sup>:



Source: NERA (1998)

74. Therefore, to the degree that LRMC, SRIC and LRIC can be considered linked to SRMC, all can be thought of sharing some of the efficiency benefits of SRMC.

### Common costs

75. In general, there will be a difference between total costs and the aggregate of marginal costs or incremental costs. These differences, known as "common costs", tend to be very significant in network industries. The analysis of common costs can be complex, but it is important they are clearly understood and treated in any cost allocation framework. They are therefore explained below by way of an illustrative worked example.

<sup>12</sup> Section 2.3, *Potential Generator Market Power in the NEM*, NERA, June 2011.

<sup>13</sup> The diagram also shows a curve for Long Run Average Cost (LRAC). i.e. the total long run cost for a given level of output divided by the quantity of output.

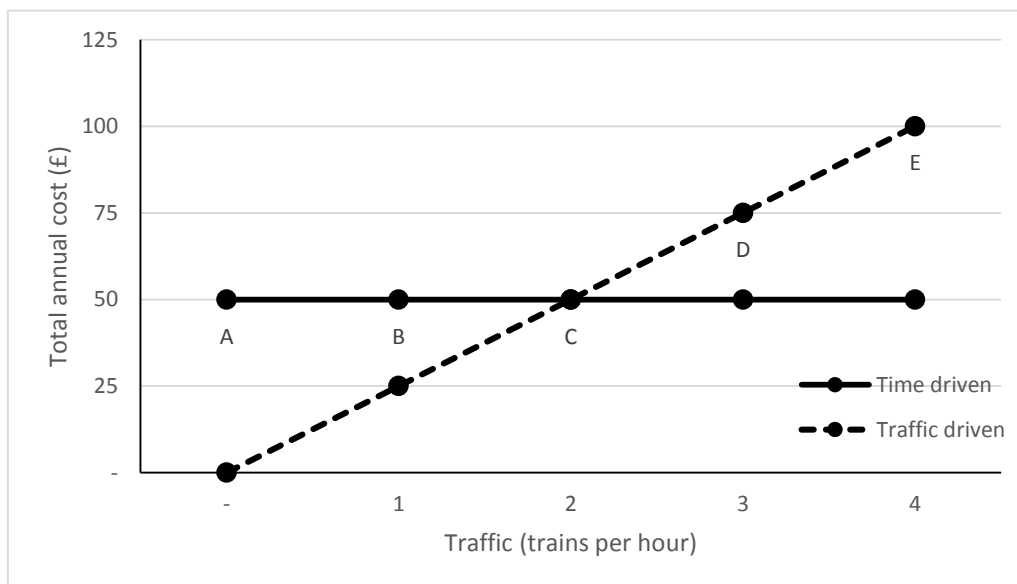
## Review of cost attribution and cost allocation approaches

76. Consider a bakery which sells bread and cake, and has costs comprising ingredients, staff, electricity, an oven and some premises.
- The incremental cost of bread, i.e. the costs that would be avoided by no longer supplying the increment of all bread, will include ingredients, staff costs and electricity for the oven. It will not however include the cost of the oven, since that is still necessary for the supply of cake, and will not therefore be avoided by no longer baking bread. Nor will it include the cost of the premises, which are still necessary for the supply of cake.
  - Similarly, the incremental cost of cake will exclude the cost of the oven and the premises.
77. Thus neither of the incremental costs of the two products supplied include the cost of the oven or the cost of the premises; and the aggregate of the two incremental costs will come to that much less than the total cost of the bakery.
78. Such costs are described in costing terminology as "common costs". Such costs are required for the supply of any single product, and at the same level, for the supply of all products.
79. Note that there is a distinction between costs that are shared between products, and common costs. For example, the cost of electricity is shared: there is a single supply of electricity, and a single bill at the end of each month. However:
- some electricity is avoided if bread is not baked;
  - some electricity is avoided if cake is not baked; and
  - some electricity is required for the premises to remain open, regardless of how many products are supplied.
80. The first two elements form part of the incremental costs of bread and cake respectively. It is only the third element, which is required at the same level for the supply of any or all products, that is a common cost.
81. Because of the presence of common costs, cost allocations based on marginal cost and incremental cost will often result in an aggregate allocation of costs that is less than total cost. This is particularly true in network industries, where common costs are often significant. In circumstances where it is important that all costs are allocated, it is therefore necessary to allocate common costs.
82. It is important, however, that common cost allocations recognise the fact that common costs change as the definition of the increment in incremental cost changes. Returning to the bakery example, assume now that in addition to bread and cake, the bakery now also sells coffee:
- the incremental cost of bread will as above be limited to ingredients, staff costs and electricity;
  - the incremental cost of cake will also be limited to ingredients, staff costs and electricity; and
  - the incremental cost of coffee will be limited to ingredients, staff costs and electricity.



## Review of cost attribution and cost allocation approaches

83. The oven and the premises remain common costs. However, there is an important difference between the two:
- the premises would be required for the supply of any combination of bread, cake and coffee; but
  - the oven would not be required if coffee alone were supplied.
84. So, in this example, the oven forms part of the incremental cost of the increment (bread + cake), or "baked goods". It is a common cost between bread and cake, but incremental to bread and cake collectively. Therefore, in considering how the common cost of the oven should be allocated between the three individual products bread, cake and coffee, there is a clear basis for allocation to bread and cake, but not to coffee.
85. It is therefore important, when allocating common costs within a cost allocation framework, that it is recognised where common costs are common to only a subset of products, but not to all products. This includes the cost allocation framework for the railway network, where common costs are widespread, but frequently only common to services using a specific part of the network.
86. As an example, some maintenance costs may be common costs. Suppose, for a section of track, maintenance costs are driven by two factors:
- the passage of time, which causes a certain level of maintenance expenditure regardless of traffic levels; and
  - traffic levels, which cause additional wear and tear.
87. The two factors are illustrated in the diagram below. Maintenance costs are driven by the time factor at traffic levels up to 2 trains per hour, above which the traffic factor dominates:



88. The total cost curve across all levels of traffic is defined by the kinked line ABCDE. It can be seen that £50 of maintenance costs (the part of the line defined by ABC) are required, regardless of the level of output. This element of maintenance costs cannot be avoided by the cessation of an subset of output, and is therefore a common cost between the output provided on that section of track. However, it should only be allocated to output on that section of track, and not to output on other sections of track.

## **Review of cost attribution and cost allocation approaches**

89. Common costs will often be encountered for cost categories where a proportion of costs is required, regardless of the level of output. For example, in addition to maintenance costs, this might include:
- a) signalling operations staff costs: if a minimum of say two people is required to operate a part of the network, but more people are required at higher levels of traffic, the minimum of two will be a common cost; and
  - b) central office costs: some functions, for example finance or IT, may have a minimum size, in which case that minimum size will be a common cost.

## Cost allocation in other UK regulated network industries

90. The railway network differs in many respects from other UK regulated network industries. For example:
- a) costs are funded through a combination of charges to network customers (i.e. operators) and government funding, reflecting wider societal benefits;
  - b) most customer demand can only be met by very specific elements of the network (for example, a service from Oxford to Didcot can only realistically use one section of track);
  - c) the costs associated with expanding capacity for specific elements of the network can often be very large relative to existing costs (i.e. investment is very “lumpy”);
  - d) much customer demand and revenue is reliant on a franchising regime which affects both customers’ exposure to charges and their flexibility to adjust patterns of demand; and
  - e) other customer demand (e.g. open access and freight) is subject to very individualised conditions.
91. The railway network does however share with other UK regulated network industries the underlying cost characteristics of:
- a) a high level of fixed costs required to serve a wide geographical coverage;
  - b) a relatively low proportion of costs which vary with small changes in traffic; and
  - c) geographically diverse constraints on capacity whose alleviation requires location specific network enhancements.
92. It may therefore be helpful to consider approaches taken to cost allocation in these other industries.

## **Telecoms**

93. LRIC is widely used in the UK telecoms industry to allocate network costs to products. For example:
- a) Ofcom uses LRIC to allocate costs to the termination of voice calls on both mobile and fixed telecoms networks<sup>14</sup>; and
  - b) BT uses LRIC to allocate costs to the different products it offers to wholesale and retail customers<sup>15</sup>.
94. For reasons that are specific to the nature of cost recovery and competitive dynamics in the telecoms industry, charge controls set by Ofcom for voice call termination are not intended to cover fixed costs, but only variable costs. In that respect there are parallels with the allocation of variable costs in the rail industry.

---

<sup>14</sup> *Wholesale mobile voice call termination: Statement*, Ofcom, March 2011; *Review of the fixed narrowband services markets Statement on the proposed markets, market power determinations and remedies*, Ofcom, September 2013

<sup>15</sup> *Long Run Incremental Cost Model: Relationships & Parameters*, BT, August 2014

## Review of cost attribution and cost allocation approaches

95. In setting charge controls for voice call termination, Ofcom has recognised the efficiency benefits of SRMC, but has suggested that LRIC has advantages due to its lower volatility. As noted above, LRIC can be thought of as a smoothed version of SRMC over the long run. This makes it related to, but less volatile than, SRMC. As explained by Ofcom:

*"Economic theory suggests that prices set at marginal cost lead to efficient outcomes, and are closer to the prices you might expect to find in a competitive market (assuming no fixed costs or externalities). Following this logic we should seek to set regulated MTRs as close to marginal cost as possible..."*

*In network industries (such as mobiles) the short run marginal cost of a service may be very low or very high depending on whether usage is a long way from, or effectively at, installed capacity. This leads to very low (or zero) marginal cost most of the time, with small increments over which marginal cost is very high. In regulatory practice, long-run incremental cost has, therefore, been applied as a proxy, avoiding the volatility implied in setting prices on the basis of marginal cost which can be very variable in response to small changes in output."*

<sup>16</sup>

96. In telecoms, LRIC is estimated for a specific product by defining, as an increment, the loss of all of the traffic volumes associated with the product under consideration, assuming the continued supply of all other products<sup>17</sup>. The long run reduction in costs, or avoidable costs, that would result from this loss is then divided by those traffic volumes to generate a unit cost. That same unit cost is then applied equally to all customers.
97. An important feature of this LRIC modelling is that a distinction is made between the costs associated with a network's coverage, and the costs associated with enhancing a network with a given coverage to create additional capacity in order to accommodate traffic volumes.
98. This distinction is described in the European Commission's 2009 Recommendation of the regulatory treatment of voice call termination:

*"Coverage can be best described as the capability or option to make a single call from any point in the network at a point in time, and capacity represents the additional network costs which are necessary to carry increasing levels of traffic. The need to provide such coverage to subscribers will cause non-traffic-related costs to be incurred which should not be attributed to the wholesale call termination increment. Investments in mature mobile markets are more driven by capacity increases and by the development of new services and this should be reflected in the cost model. The incremental cost of wholesale voice call termination services should therefore exclude coverage costs but should include additional capacity costs to the extent that they are caused by the provision of wholesale voice call termination services."*<sup>18</sup>

99. The distinction between coverage and capacity can be interpreted as a constraint over the degree to which the network would be reconfigured absent traffic volumes, even in the long run. Such a constraint, referred to as a "scorched node" assumption, is intended to reflect the fact that the current network configuration is likely to feature legacy inefficiencies arising from the historical development of the network, which cannot realistically be avoided. BT assumes, in its modelling of the costs of a network with lower traffic volumes, that:

*"BT maintains its existing geographical coverage in terms of customer access and connectivity between customers, and provides the infrastructure to do this from existing network nodes."*<sup>19</sup>

---

<sup>16</sup> Paragraph 8.34 (including Footnote 573), *Wholesale mobile voice call termination: Statement*, Ofcom, March 2011

<sup>17</sup> This is a simplification of the approach used by BT, which decomposes products into component elements, and considers the loss of all traffic volumes associated with those component elements.

<sup>18</sup> *Commission Recommendation of 7 May 2009 on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU (2009/396/EC)*

<sup>19</sup> Section 4.3.1, *Primary Accounting Documents*, BT, August 2014

## Review of cost attribution and cost allocation approaches

100. One instance of potential relevance for rail is the way in which BT implements this assumption its modelling of the long run avoidable costs of its duct network, which is a significant component of its cost base. BT's underground duct network comprises the buried tubes which through which copper and fibre cables are routed, the tubes themselves being passed through underground tunnels or "bores". Many sections of duct use only a single bore, but sections which carry more traffic may require additional bores to accommodate additional tubes and cables.
101. In its LRIC modelling, BT calculates the avoidable costs of its duct network by estimating the reduction in the number of bores that could be achieved at lower traffic volumes, subject to a minimum of a single bore, so as to hold constant the existing geographical pattern and total distance of duct, and maintain customer access and connectivity. BT describes this as network "thinning"<sup>20</sup>.
102. Historically, the term "LRIC" has also been used in the telecoms industry in quite a different sense, in order to estimate, for charge control purposes, the efficient total cost base of a network on a forward-looking basis, as an alternative to relying on efficiency adjusted accounting or forecast data from the incumbent network operator(s) (including data relevant for the calculation of a Regulatory Asset Base).
103. Under this approach, only a single increment is ever considered, being all of the services provided by the network. This "all service LRIC" equates to the incremental costs associated with increasing output:
  - a) from the provision of no services at all;
  - b) to the provision of all services.
104. Since it is assumed that the long run costs of providing no services at all are zero, the "all service LRIC" simply equates to the aggregate cost of the network providing all services, i.e. total network costs.
105. Clearly that single figure of total network costs can by itself be of no assistance in allocating those costs between products. In practice, all service LRIC models have been used to allocate costs to products, but such allocation tends to rely on the making of assumptions on the relative degree to which different products consume capacity<sup>21</sup>.
106. A key difficulty with such an approach is that perhaps the best way to understand the relative capacity consumption of different products is to estimate the LRIC of products individually. Yet this is precisely what an all service approach precludes. As a result, all service LRIC models are significantly less helpful than product LRIC models in informing the allocation of costs between products.

## Electricity

107. Transmission network use of system charges recover the costs of installing, operating and maintaining the electricity transmission network<sup>22</sup>. The underlying rationale for the charging methodology is that:

*"efficient economic signals are provided to Users when services are priced to reflect the incremental costs of supplying them".*

---

<sup>20</sup> Pages 126 to 127, *Long Run Incremental Cost Model: Relationships & Parameters*, BT, August 2014

<sup>21</sup> For an example of an all service LRIC model, see *The LRIC model of UK mobile network costs, developed for Oftel by Analysys, September 2001: A Manual for the Oftel model*, Analysys, January 2001

<sup>22</sup> Sections 14.14 and 14.15, *Connection and Use of System Code*, National Grid, October 2014

## Review of cost attribution and cost allocation approaches

108. Transmission charges are therefore based on allocated costs which:

*"reflect the impact that Users of the transmission system at different locations would have on the Transmission Owner's costs, if they were to increase or decrease their use of the respective systems"*.

109. Transmission costs are defined as including maintenance, investment and enhancement costs; and increases or decreases in use are estimated on a marginal basis, changing injection into each network node by 1MW during the network peak period. This appears to equate to LRMC.

110. Distribution network use of system charges for extra high voltage customers are set using one of two methodologies: Forward Cost Pricing (FCP) and an approach termed LRIC. Both methodologies are intended to reflect a zonal analysis of the incremental long run cost of future reinforcements resulting from an incremental increase in demand. The FCP approach considers an increment equivalent to the expected future increase in demand over a 10 year forecast period (i.e. a version of what is described in the principles section above as LRIC); the LRIC approach considers a small 0.1MW increment in demand (i.e. what is described in the principles section above as LRMC)<sup>23</sup>.

### Gas

111. Transmission network charges comprise capacity and commodity charges, each of which is levied on both entry to and exit from the network<sup>24</sup>.

112. Capacity charges are principally designed to recover network maintenance, renewals and enhancement costs. Entry capacity charges are determined through auctions but are subject to reserve prices. Exit capacity charges are applied on an administered basis. Both entry capacity auction reserve prices and exit capacity administered charges are based on LRMC. A model estimates the long run costs of reinforcing the network to transport a small increase in gas, over existing peak levels, between entry and exit points. Marginal costs are calculated at a disaggregated level for each entry point to each relevant exit point.

113. Commodity charges are principally designed to recover network operations costs. Such charges are allocated equally to all units of gas entering and exiting the network.

### Water

114. The water industry does not appear to have a common approach to cost allocation. Ofwat has put in place certain "Charging Guidelines" but these are at a high level only, stating:

*"Companies should set their charges to reflect the broad balance of costs imposed by different customer classes"* <sup>25</sup>.

115. Ofwat has suggested that there is likely to be a focus on cost allocation in the coming years, in light of increased prospects for competition in the industry following The Water Act 2014. In particular, access pricing rules will need to be developed to support entrant access to monopoly assets and services<sup>26</sup>. However, it is not yet clear what direction this will take.

---

<sup>23</sup> Schedules 17 and 18, *Distribution and Connection Use of System Agreement*, November 2014

<sup>24</sup> Section Y, *Uniform Network Code*, Joint Office of Gas Transporters, December 2014

<sup>25</sup> Ofwat website

<sup>26</sup> *Consultation on wholesale and retail charges for 2015-16 and charges scheme rules*, Ofwat, May 2014

### Post

116. Post differs from most other regulated network industries in that it is labour intensive rather than capital intensive. Royal Mail allocates costs to products using an Activity Based Costing approach. Total costs are split between different activities, and the costs are allocated to products on the basis of the relative workload that different products impose on each activity<sup>27</sup>.
117. Activity Based Costing approaches tend not to identify marginal or incremental costs explicitly. However Royal Mail started to consider the development of product LRIC estimates in 2010, using a broadly similar methodology to that used by BT<sup>28</sup>, and began to create a LRIC model shortly afterwards. Postcomm, its regulator at that time, encouraged the development of LRIC estimates, stating that:

*"In competitive markets prices should reflect long-run marginal costs (LRMC), but LRMC can be very difficult to calculate in practice. LRIC, as a proxy for LRMC, is used as a practical regulatory benchmark for efficient pricing..."*

*LRMC is often considered a better measure than SRMC for regulatory purposes. The reason is because if a company is producing at capacity, increasing output by one unit could mean significant levels of SRMC, whereas when it is not producing at capacity SRMC can be much lower. SRMC can thus be viewed as being quite 'lumpy' depending on when the demand change is assumed to occur. In contrast, because LRMC reflects the costs that occur when all inputs can be varied in response to a sustained demand change, there is an averaging out of this 'lumpiness' and so it can provide better signals to consumers and the market*

*In the postal services sector it is difficult to calculate LRMC in practice because assessing the long run cost impact of a small increment (e.g. an extra letter) is never realistic. In reality, inputs (such as staff, vehicles and machines) can only be sensibly varied in larger discrete amounts (e.g. as a result of network, rostering or fleet changes), which reflect the changes that can be made in response to variations in much larger volumes of outputs.*

*LRIC can be used as a proxy for LRMC to reflect this reality; it has the benefit of being able to be more readily estimated from business planning information and/or accounting records than LRMC."*<sup>29</sup>

### Airports

118. The CAA does not appear to have a specific approach for allocating airport costs between different products and customers. For example, in 2014 it ruled on a complaint brought against Heathrow Airport Limited (HAL) alleging that the structure of its passenger and landing charges were unreasonably discriminatory<sup>30</sup>.
119. HAL submitted evidence in respect of its passenger charges which included an analysis of the "asset costs" of handling different categories of passenger, i.e. on the "theoretical terminals [required] for different passenger types, the terminals having different space requirements for the different passenger groups". By considering the long run infrastructure requirements of different passenger categories, this appears to be similar to a LRIC approach.

<sup>27</sup> Consolidated Regulatory Accounting Guidelines, Ofcom

<sup>28</sup> The Development of Long Run Incremental Cost Estimates in the Postal Sector by Royal Mail: A discussion document, Royal Mail, December 2010

<sup>29</sup> Chapter 6, The building blocks for a sustainable postal service - Annex B: Cost transparency and accounting separation, Postcomm, April 2011

<sup>30</sup> Appendix I, Investigation under Section 41 of the Airports Act 1986 of the structure of airport charges levied by Heathrow Airport Limited - CAA decision, CAA, March 2014

## Review of cost attribution and cost allocation approaches

120. The CAA concluded that this aspect of HAL's approach was not unreasonable, but indicated that other approaches could also have been acceptable, stating that:

*"there is not one 'right' way to allocate sunk and common costs and there are a number of legitimate approaches HAL could adopt. It has a margin of discretion about how it recovers the sunk and common costs associated with providing airport infrastructure."*<sup>31</sup>

121. The CAA considered but rejected an "all service" LRIC approach to estimate a total cost base for Gatwick Airport Limited's most recent charge control<sup>32</sup>. However as noted above, this is of limited relevance for cost allocation issues.

---

<sup>31</sup> Paragraph 3.27, *Investigation under Section 41 of the Airports Act 1986 of a complaint made by bmi against Heathrow Airport Limited: A Consultation*, CAA, July 2011

<sup>32</sup> Pages 266 to 276, *Economic regulation at Gatwick from April 2014: notice of the proposed licence*, CAA, January 2014



## Railway legal framework

122. Under the charging principles set out in EU legislation, transposed into the Access & Management Regulations, access charges levied on operators should at least cover:
- "the cost that is directly incurred as a result of operating the train service"*<sup>33</sup>.
123. Mark-ups above the costs *"directly incurred"* are permitted in certain circumstances, but only where such mark-ups do not price market segments off the network<sup>34</sup>.
124. From a pure cost allocation perspective, SRMC, SRIC, LRMC, LRIC are all *"directly incurred"* by the relevant increment, in that they are an increase in cost that is directly caused by the addition of the increment (or avoidable by the loss of the increment). So depending on what increments are consistent with the term *"the train service"*, it is arguable that any of these cost allocation approaches could, from a non-legal perspective, be considered compliant with the phrase:
- "the cost that is directly incurred as a result of operating the train service"*.
125. However, the standard industry interpretation of the phrase appears to have been to constrain it to the SRMC of Network Rail's costs only. For example, in 2012 the Independent Regulators' Group - Rail (IRG-Rail) issued a position paper on the issue, stating:
- "The IRG-Rail charging working group supports the view that the "cost that is directly incurred" should be interpreted as "short-run marginal cost" (SRMC), and that short-run marginal cost should be taken to include:*
- *Operating costs (e.g. signalling);*
  - *Maintenance costs (e.g. wear and tear repairs);*
  - *Renewal costs."*<sup>35</sup>
126. Moreover, it is possible that the interpretation of the phrase may become further constrained. In 2014, the EC's Single European Railway Area Committee issued a discussion paper which goes beyond the principle of SRMC into specific detail as to how SRMC should be calculated<sup>36</sup>.
127. Based on the current interpretation of the phrase *"directly incurred as a result of operating the train service"*, it therefore appears that any charge in excess of the SRMC of Network Rail's costs is subject to the requirement that the excess does not price market segments off the network. That is, mark-ups above the SRMC of Network Rail's costs are constrained by assessments of operators' ability to pay without being priced off the network.
128. It is not always entirely clear how these considerations are taken into account. For example, charges for High Speed 1 (HS1) appear to be set above SRMC of infrastructure costs, and include elements such as long run costs, without any clear reference to ability to pay<sup>37</sup>. It seems clear however that the potential implications of any mark-ups would need to be considered.

---

<sup>33</sup> Railways Infrastructure (Access and Management) Regulations 2005, as amended by Railways Infrastructure (Access and Management) (Amendment) Regulations 2009

<sup>34</sup> Paragraph 16.33, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013. This is a simplified characterisation of the charging principles. For example, they also allow for charges to reflect the scarcity of capacity of identifiable segments of the infrastructure during periods of congestion, but only where plans to enhance capacity are prepared and proceeded, unless a cost-benefit analysis shows they are not worthwhile (*Scoping study for scarcity charges*, Nash and Johnson, March 2006).

identifiable segment of the infrastructure during periods of congestion.

<sup>35</sup> *Position paper on the concept of "cost that is directly incurred"*, IRG-Rail, October 2012

<sup>36</sup> *Discussion paper: On the modalities for the calculation of the cost that is directly incurred as a result of operating the train service*, SERAC, July 2014

<sup>37</sup> *Regulation of High Speed 1: Statement by the Office of Rail Regulation*, ORR, October 2009

## **Review of cost attribution and cost allocation approaches**

129. As set out above, such considerations are beyond the scope of this report, which focuses on only on cost allocation, and does not extend to the merits of different charging structures. From that perspective, there are benefits to be gained from greater transparency in relation to the underlying reality of cost causation in the cost allocation framework, whether or not the resulting cost allocations are reflected in charges.

### Variable costs

130. The railway network's current cost allocation framework makes a distinction between what are described as "variable costs" and "fixed costs":
- a) the term "variable costs" tends to be used to refer to costs that vary in response to relatively small changes in traffic levels; and
  - b) the term "fixed costs" tends to be used to refer to all other costs.
131. Under this interpretation, the term "variable costs" broadly aligns with the concept of marginal cost as set out above. The different components of and perspectives on variable costs / marginal costs are considered in this section.

### Short run marginal costs

132. Short run marginal costs are those costs that vary with a small change in traffic, assuming that capacity is fixed. The assumption that capacity is fixed is significant, and clearly not always accurate. However, as noted above in principle short run and long run cost allocation approaches are linked.
133. As explained above, there are two main categories of SRMC<sup>38</sup>:
- a) Network Rail short run marginal costs, i.e. costs that Network Rail incurs that vary with traffic, but excluding enhancement costs, since capacity is assumed to be fixed; and
  - b) operator short run marginal costs, i.e. the cost of delays and opportunity costs of not being able to access the network, which result from assumed fixed capacity constraints.

### Network Rail short run marginal costs

134. Network Rail short run marginal costs are generally reflected in the current cost allocation framework and described as "variable costs".
135. The original charging regime put in place at privatisation focused on two categories of short run marginal costs:
- a) operating, maintenance and renewal costs caused by the marginal wear and tear from running an additional train; and
  - b) the cost of traction electricity<sup>39</sup>.
136. These categories remain, but others have been added.

### Operating, maintenance and renewal costs

137. The SRMC of operating, maintenance and renewal costs is estimated<sup>40</sup> and reflected in the Variable Usage Charge (VUC), on the basis of vehicle miles for passenger operators or 1000 gross tonne miles (kgtm) for freight operators. The VUC is not intended to reflect the costs of providing or enhancing the network.

---

<sup>38</sup> Excluding societal costs

<sup>39</sup> *Access pricing in rail - Principles and structures*, Thomas and McMahon, 2005

<sup>40</sup> In practice, although the SRMC of operating costs is considered, it is assumed to be zero.

## Review of cost attribution and cost allocation approaches

138. SRMCs are estimated by Network Rail using a model which considers the forward-looking cost impact, over a 35 year period, of a 20% increase in traffic over existing levels. Strictly speaking, SRMC is derived from considering much smaller increases. However, sensitivity analysis using smaller increases (e.g. 5% and 10%) indicates very similar unit costs, suggesting that the difference between what is estimated and strict SRMC is unlikely to be significant.

### Geographical disaggregation

139. Track wear and tear costs form a significant proportion of operating, maintenance and renewal costs. The SRMC of track wear and tear is currently calculated on a highly geographically disaggregated basis, but is then averaged nationally for the purpose of charging.
140. Disaggregated results suggest that marginal track wear and tear costs are lowest on busy parts of the network where capacity is most constrained. This is principally because the most efficient engineering and operational solution on busy parts of the network is to lay particularly high quality track. Such track has a relatively high up-front cost, but a relatively low marginal rate of wear and tear. This means that the SRMC of operating, maintenance and renewal costs overall, i.e. the basis of the VUC, tends to be lowest on those parts of the network that are busiest.
141. Network Rail has expressed concern that a move to disaggregate the existing nationally averaged charges to reflect these disaggregated results would risk:
- “potentially introducing perverse incentives where it is cheaper to operate on busier routes.”*<sup>41</sup>
142. The difficulty that Network Rail refers to is largely a result of the current framework’s incomplete approach to measuring and allocating SRMC. The fact that the SRMC of operating, maintenance and renewal costs overall are lowest where capacity is most constrained implies that:
- a) one element of SRMC, Network Rail short run traffic related costs, is lowest; in the same parts of the network where
  - b) another element of SRMC, operator short run traffic related costs relating to opportunity costs, is highest.
143. In principle, a complete analysis of SRMC would recognise, through the second element, the cost of operating on busy parts of the network, resulting in a cost allocation which reflected the balance between the two elements of SRMC. However, as discussed further below, operator short run traffic related costs are not fully reflected in the current cost allocation framework. As a result, a disaggregated application of the existing approach would result in SRMC being significantly understated on busy parts of the network, since lower Network Rail costs are not counterbalanced by higher opportunity costs.
144. It is this incomplete reflection of SRMC that means that cost allocations from a disaggregated analysis of the SRMC of operating, maintenance and renewals costs alone would be distorted, leading to perverse and inefficient signals to the industry.

### Vehicle disaggregation

145. SRMCs are differentiated by vehicle class, reflecting the variation in infrastructure wear and tear costs associated with different vehicle characteristics, for example vehicle operating speed and axle load. In the case of freight, costs are also differentiated by the commodity being transported, which is a key determinant of axle load<sup>42</sup>.

---

<sup>41</sup> Section 6.3, *Periodic Review 2013 consultation on incentives - Network Rail’s response*, Network Rail, February 2013

<sup>42</sup> Paragraph 16.89, *Periodic Review 2013: Final determination of Network Rail’s outputs and funding for 2014-19*, ORR, October 2013

## Review of cost attribution and cost allocation approaches

146. VUCs do not fully reflect calculated SRMCs. In particular, Network Rail's latest SRMC calculations for PR13 incorporated revised approaches for allocating track, structure and signalling costs. These suggested significant changes to VUCs for some vehicle types. However, VUCs were not revised fully in line with these changes, due to concerns over the potential impact of higher charges on levels of freight traffic<sup>43</sup>.

### Relationship between SRMC and total traffic related costs

147. As is typical with SRMC based cost allocation, even though SRMC is estimated on the basis of a marginal increase over and above all existing traffic, the resulting estimate is allocated to, and in this case charged to, all traffic.
148. One of the features of allocating costs in this way is that there can be a difference between:
- the aggregate of SRMC allocations over all traffic; and
  - total traffic related costs, i.e. the costs that are incremental to all traffic (or, short run incremental cost (SRIC), based on an increment of all traffic).
149. The nature of this difference depends on the way in which marginal cost varies as output varies. As set out in greater detail in Annex A:
- where marginal cost reduces at higher levels of output, the aggregate of SRMCs will be lower than total traffic related costs; but
  - where marginal cost increases at higher levels of output, the aggregate of SRMCs will be higher than total traffic related costs.
150. In the case of operating, maintenance and renewals costs, it appears that the aggregate of SRMCs is lower than total traffic related costs. For example, analysis performed in respect of freight traffic suggests that for freight at least, if the increment to be considered were the removal of all traffic, unit SRIC would be roughly double SRMC<sup>44</sup>.
151. As seen below, this is not always the case, and for some cost categories, SRMC is above SRIC.
152. This is an inherent feature of an SRMC approach to cost allocation. The fact that SRMC may be either lower or higher than SRIC does not indicate that there is anything "wrong" with the costs that are allocated based on SRMC. It merely reflects the underlying marginal cost characteristics of the costs being allocated, and the fact that SRMC allocations will have different strengths and weaknesses from SRIC based allocations.
153. It should be borne in mind that SRMC and SRIC are simply alternative approaches to estimating variable costs. Under any foreseeable approach, variable costs remain a long way below Network Rail's overall revenue requirement, and the fixed costs required to make up that shortfall remain substantial. Therefore there does not appear to be any realistic prospect that alternative approaches to estimating variable costs could result in Network Rail's total income from charges exceeding its total costs.

### Maintenance and renewals costs - electrification assets

154. The SRMC of the maintenance and renewal of electrification assets is estimated and reflected in the Electrification Asset Usage Charge (EAUC), on the basis of vehicle miles or kgm for electrified traffic. The approach is in principle very similar to that for the maintenance and renewals costs captured by the VUC.

---

<sup>43</sup> Paragraph 16.104 to 16.120, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

<sup>44</sup> Paragraph 16.92, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

## Review of cost attribution and cost allocation approaches

155. Network Rail's approach to calculating the SRMC of this category of costs is to estimate how maintenance and renewal costs would vary as a result of a relatively small change in traffic levels.

### Coal spillage costs

156. Network Rail incurs costs attributable to the spillage of coal by freight wagons. These are assumed to comprise the costs of cleaning up points, accelerated renewals due to increased track wear and tear, and increased delay due to points failures. The incremental cost of all such costs (i.e. the costs that would be avoided absent all coal freight) is estimated and allocated and charged to coal traffic on the basis of kgm.
157. Strictly speaking this estimate is based on SRIC (with all coal freight as the increment) not SRMC (variation as a result of one coal freight train). However, it is not clear that there would be any material difference between the results of the two approaches.
158. The majority of coal spillage appears to occur at the start or end of each journey, very close to loading and unloading points<sup>45</sup>, and this is reflected in the way in which total coal spillage costs are estimated. This suggests that the amount of coal loaded or unloaded is the key cost driver for these costs. However, for the purposes of charging simplicity, costs are allocated and charged to coal freight on the basis of weight carried and distance travelled (kgm). This is likely to overstate costs for coal freight transported over long distances, and understate costs for coal transported over short distances.

### Schedule 8 costs

159. The "Schedule 8" performance regime aims to compensate train operators for the long term financial impact of unplanned service disruption (i.e. delays and cancellations), where performance is below predetermined targets. The regime also incentivises Network Rail and train operators to outperform targets, by providing a form of "bonus" payment equivalent to the rate of compensation.
160. Disruption can be attributable to:
- a) Network Rail, through incidents in its provision of the network infrastructure; or
  - b) other train operators, through incidents in their operations.
161. Network Rail makes contributions to the Schedule 8 regime for disruptions attributable to its own incidents. The scale of these contributions can vary with traffic, as higher levels of traffic can reduce Network Rail's ability to manage the disruption resulting from each individual incident.
162. Network Rail also acts as administrator for the Schedule 8 regime, using the "star model" to coordinate contributions from other operators causing disruption, and corresponding compensation payments to operators suffering disruption. There is however a potential shortfall between operator contributions and payments. Operator contribution rates are calculated on the basis of the historical impact of disruption, whereas operator payments are calculated on the basis of the actual impact of disruption. Higher levels of traffic may mean that each incident of operator disruption can have a greater impact on other operators, resulting in a gap between operator contributions and operator payments. Network Rail bears the cost of that shortfall.

---

<sup>45</sup> For example, at loading points where coal wagon rake cleaners have been installed, the incremental wear and tear on track appears to be considerably reduced (Paragraph 16.321, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013).

## Review of cost attribution and cost allocation approaches

163. Network Rail therefore incurs two costs as a result of the Schedule 8 regime: contributions in its role as infrastructure provider; and funding of the contribution/payment shortfall in its role as administrator. Both costs increase with higher levels of traffic, are therefore a traffic related cost for Network Rail.
164. The estimated SRMC of these costs, i.e. the marginal increase in Network Rail Schedule 8 costs resulting from the running of an additional train, is in principle allocated and charged to traffic through the "Capacity Charge"<sup>46</sup>.

### Geographical and temporal disaggregation

165. As with operating, maintenance and renewals costs, the SRMC of Schedule 8 costs is calculated on a highly disaggregated basis, both geographically (based on over 6,500 "Constant Traffic Sections") and temporally (based on 24 time bands). These SRMCs are then aggregated for charging purposes:
- until CP4, into operator "service groups"; and
  - from CP5, into more disaggregated operator "service codes" (where each service group consists of a number of service codes)<sup>47</sup>.
166. The knock on effects of incidents are typically greater on busy sections of track and at busy times of the day. SRMCs will therefore show a high degree of variation when calculated at a disaggregated level. Since service groups often cover both busy and quiet sections of track, and tend not to be differentiated by time of day, the CP4 capacity charge is likely to have significantly understated or overstated SRMC on specific sections of track and times of day. In a 2011 study for the ORR, NERA observed that:
- "in very congested parts of the network, the capacity charge may not reflect Network Rail's extra Schedule 8 payments in full, and this may affect Network Rail's incentives to provide additional capacity. Conversely, on other sections of route the capacity charge may significantly overestimate the extra Schedule 8 payments that Network Rail is likely to make"*<sup>48</sup>.
167. The move to service codes in CP5 would have improved the accuracy of the charge, albeit at the expense of greater complexity. However even service codes are at a relatively aggregated level. For example, they often do not vary by time of day, even though an incident during peak hours will clearly cause more disruption than one during off peak hours. It is therefore likely that material understatements and overstatements remain. However, any benefits from further disaggregation of charges would need to be weighed against the costs of introducing further complexity.

### Implementation for non-franchised passenger operators

168. The approach outlined above is applied to franchised passenger operators only, using SRMCs which were recalculated for CP5, and found to be significantly higher than those assumed for CP4.
169. In practice, different approaches are applied to other operators, meaning that not all capacity charges are based on the allocation of the best available estimates of SRMC to all traffic<sup>49</sup>:
- open access operators established before CP5 are charged at historical CP4 rates for services established before CP5, and at CP5 rates for new services;

---

<sup>46</sup> Paragraph 16.159, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

<sup>47</sup> Paragraph 16.159 and 16.163, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

<sup>48</sup> Footnote 6, *Using incentives to improve capacity utilisation*, NERA, December 2011

<sup>49</sup> Paragraphs 16.199 to 16.204, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

## Review of cost attribution and cost allocation approaches

- b) new open access operators and charter operators (who were not subject to the capacity charge in CP4) are charged at CP4 equivalent rates on services below a threshold, and at CP5 rates for additional services; and
  - c) freight operators are charged at slightly below historical CP4 rates, with a reconciliation mechanism designed to recover the additional cost associated with CP5 rates for additional services above a baseline level.
170. These differences were introduced at the start of CP5 in order to mitigate the commercial impact of higher CP5 rates on operators. They do not appear to be based on any differences on the costs these operators impose on Network Rail.

### Title of charge

171. The charge is titled the "Capacity Charge". However, from its inception in 2000, the charge was intended to recover only the costs associated with congestion related delay<sup>50</sup>. Now, as set out above, the charge reflects only Schedule 8 costs incurred by Network Rail, on an SRMC basis that is in principle consistent with the allocation of other costs incurred by Network Rail.
172. The charge does not, and is not designed to, reflect the full costs arising from capacity constraints, whether looked at from a short run or long run perspective:
- a) Under a short run perspective, as noted above, an important component of the short run costs associated with capacity constraints is the opportunity cost to operators who cannot run as a result of demand exceeding supply. These opportunity costs are not reflected in the Schedule 8 regime and do not therefore affect the Capacity Charge.
  - b) Under a long run perspective, the costs associated with capacity constraints are those costs that Network Rail incurs in enhancing the network to alleviate capacity constraints. These costs too are not reflected in the Schedule 8 regime and do not therefore affect the Capacity Charge.
173. Since the costs allocated under the Capacity Charge are only a partial reflection of the costs arising from capacity constraints, payment of the Capacity Charge cannot be considered payment for capacity. It is therefore not clear that the title "Capacity Charge" is helpful for the promotion of an understanding of the costs allocated through this charge.

### Relationship between SRMC and total traffic related costs

174. As with operations, maintenance and renewal costs, Schedule 8 costs are in principle estimated on the basis of SRMC, and allocated to all traffic<sup>51</sup>.
175. As noted in respect of operations, maintenance and renewals costs, the aggregate of SRMC based allocations can be lower or higher than total traffic related costs (SRIC based on increment of all traffic), depending on how marginal costs respond to levels of output.
176. In the case of operations, maintenance and renewals costs, it appears that SRMC based allocations are lower than total traffic related costs.

---

<sup>50</sup> *Assessment of capacity charges: Final report*, Symonds Group, July 2000

<sup>51</sup> Although the levels of aggregation used for cost estimation and charging differ between the two sets of costs.



## Review of cost attribution and cost allocation approaches

177. In the case of Schedule 8 costs, SRMC based allocations are higher than total traffic related costs. This effect is particularly marked for Schedule 8 costs: since delays associated with existing traffic are reflected in the Schedule 8 benchmarks, Network Rail should in theory only incur net costs where traffic exceeds existing levels. In other words, marginal cost should be zero (or potentially negative due to "bonus" payments) up to existing traffic levels; and expected total costs at existing traffic levels should be zero. As a result, allocating SRMC to all traffic will result in total allocated costs that are considerably greater than total traffic related costs.
178. As noted above though, the fact that the aggregate of SRMC based allocations may differ from total traffic related costs is in principle an inherent feature of an overall cost allocation framework that is based on SRMC.

### *Operator short run marginal costs*

179. As noted above, operator short run marginal costs are costs to other operators resulting from accommodating an additional train on a network which is assumed to have a fixed capacity. These costs, which become significant when capacity is constrained, form an important part of SRMC. They include:
- a) the cost of increased disruption and delays to other operators; and
  - b) the opportunity costs to other operators no longer being able to access the network (sometimes referred to as "scarcity costs")<sup>52</sup>.
180. These costs are not fully reflected in the current cost allocation framework:
- a) The cost of increased disruption and delays is reflected, at least to a degree, through Schedule 8 costs. As noted by ORR, the existing Schedule 8 based capacity charge:  
*"has characteristics similar to that of a congestion charge because it is calculated to reflect the incremental revenue losses to other services resulting from reactive delay associated with a more congested network"*<sup>53</sup>.
  - b) However, the opportunity costs to other operators of no longer being able to access the network is not allocated anywhere in the current regime, either through the Capacity Charge (which as noted above is not designed to reflect these costs) or through any other mechanism.
181. Because opportunity costs are not included in the current regime, in circumstances where capacity is constrained, allocated costs do not fully reflect SRMC.
182. As noted above, where prices are below SRMC, a customer is incentivised to buy output, even where the cost of providing that output, to the firm and other customers collectively, exceeds that customer's willingness to pay. As a result, customers are incentivised to demand an inefficiently high level of output.
183. In the current context, the absence of an opportunity cost component to the current charging structure:
- a) encourages operator A to run a train;
  - b) even where that means that operator B is denied the ability to run a different train; and

---

<sup>52</sup> *Charging for Scarce Rail Capacity in Britain: A Case Study*, Johnson and Nash, March 2008

<sup>53</sup> Paragraph 7.45, *Periodic review 2013: Consultation on incentives*, ORR, December 2011

## Review of cost attribution and cost allocation approaches

- c) the revenues that operator A earns from running its train<sup>54</sup>; are less than
  - d) the revenues that operator B would earn from running its train.
184. In other words, operators are incentivised to run, even where the aggregate financial impact of doing so is negative<sup>55</sup>. This weakens the role that pricing mechanisms based on charges can play in the management of capacity constraints.
185. However, decisions relating to capacity constraints are not made purely on financial terms, and take account of non-financial considerations including societal benefits and franchise arrangements. A heavy reliance is therefore placed on administrative mechanisms, including the specification of franchises, the approval process for track access contracts, and the Long Term Planning Process, for decisions such as:
- a) the allocation of scarce capacity between existing operators with competing demands for track access, and by implication the denial of capacity to some operators;
  - b) the allocation of capacity to new services requesting access to track that is already congested, for example HS2 trains joining the existing "classic network"; and
  - c) the specification of the nature and extent of track, signalling and station enhancements to relieve capacity constraints.
186. It seems likely that the need for administrative mechanisms to manage these decisions has increased over the years. When the broad principles of the current charging structure were put in place in the mid-1990s, traffic on the railways was in slow decline. Existing capacity constraints were managed through an administrative mechanism, supplemented at the time by negotiated capacity charges which in principle would have been influenced by opportunity costs. Since then, traffic on the railways has doubled, and negotiated charges have been abandoned, both of which developments have left administrative mechanisms with more work to do.
187. The identification and allocation of operator opportunity costs could in theory support the administrative mechanism by providing better quality information to Network Rail, its funders, and operators:
- a) for capacity allocation decisions, this may improve understanding of the relative costs of different train paths at different times of the day, helping the consideration of alternative operator allocations; and
  - b) for network enhancement decisions, this may improve the quality of the cost benefit analysis.
188. As explored below, however, there may be significant difficulties with implementing such an approach.

---

<sup>54</sup> That is, the increased income, from fares and other sources, net of the increased running costs, for that operator as a result of running that train.

<sup>55</sup> As noted by NERA (2011), where capacity constraints exist, the factor which prevents potential services from running on the network is not the volume of traffic that actually runs on the network, but the volume of booked train paths for which capacity is reserved. In principle, any measurement of opportunity costs should take this into consideration.

## Review of cost attribution and cost allocation approaches

### Level of disaggregation

189. The level of disaggregation at which elements are defined for measuring opportunity costs is critical. Capacity constraints are very unevenly distributed over time and location: for example, some stations are a particular bottleneck in the network. Measurement based on broadly defined geographical and temporal elements could result in opportunity costs attributable to one constrained part of the network being allocated to traffic in unconstrained parts of the network. This could mean that cost allocations in unconstrained parts rise well above current cost allocations, even though this would not be warranted by local conditions.
190. For example, if opportunity costs were calculated for the entire Scotland operating route, and allocated to all traffic on that route, trains running during mid-afternoon in the West Highlands would be allocated a proportion of opportunity costs associated with capacity constraints during the morning peak around Edinburgh and Glasgow.
191. Such allocations could distort the information on which decisions are made, suggesting that access should be denied to some traffic on quiet parts of the network, even where the true cost of granting such access is low. This could result in an inefficient reduction in levels of network traffic.
192. The risk of inefficient reductions of this nature has been raised as a key concern in the past by the ORR<sup>56</sup> and NERA (1998).
193. The level of such a risk can in principle be reduced by the careful choice of multiple narrowly defined elements by time and location. In theory, with elements that are defined narrowly enough, opportunity costs would only cause cost allocations to rise above current cost allocations where traffic-related capacity constraints exist. However, the more narrowly that elements are defined, the greater the challenges associated with implementing and operating the cost allocation approach.
194. The key consideration is the balance, at the level of disaggregation that is realistically likely to be achievable, between:
  - a) the benefits of reflecting opportunity costs in constrained parts of the network, in terms of improving the efficiency of capacity allocation and enhancement decisions in those parts of the network; and
  - b) the costs of overstating opportunity costs in unconstrained parts of the network, in terms of inefficiently reducing levels of traffic.
195. The balance between these benefits and costs is far from obvious, as noted for example by NERA (2011). This could be an issue that would benefit from a pilot study analysis on a section of the network.

### Measurement

196. A key difficulty with allocating costs on the basis of opportunity costs is that opportunity costs are inherently hard to measure.
197. As noted above, opportunity costs reflect the value foregone, in terms of net revenue not earned, as a result of removing an existing train and/or denying a potential train from running on the network. But neither Network Rail nor the ORR has access to such information, especially at a disaggregated level; and operators are unlikely to volunteer such information, given its commercial confidentiality and the incentives created by the knowledge that providing such information will influence their own charges.

---

<sup>56</sup> *Access pricing in rail - Principles and structures*, Thomas and McMahon, 2005

## Review of cost attribution and cost allocation approaches

198. As a result, the information required to estimate opportunity costs is not observable (as for example noted by NERA (2011)), and must be estimated based on very limited information. This is likely to introduce a high margin of error into the estimates, with the potential for cost allocations to be materially misstated.
199. NERA (1998) stressed the difficulty of measuring opportunity costs and suggested the use of market mechanisms such as auctions, which could reveal operators' willingness to pay based on net revenue potential. However, in practice, the costs of implementing such mechanisms on anything but a limited basis could be prohibitive.

### Volatility

200. The level of operator opportunity cost depends on the relationship between capacity, existing output and the value of output at a single point in time. These can all vary from one year to another, which can make opportunity cost vary in response to changes such as:
  - a) the pattern of network use by the various services (i.e. the level of existing output); and
  - b) the impact of enhancement projects (i.e. the level of capacity).
201. This potentially introduces a degree of volatility into cost allocations.

### Scale

202. Previous analysis has suggested cost allocations ranging between £1 and £20 per train km, depending on the time of day, for a typical inter-city route to London<sup>57</sup>.

### Charges

203. In theory, there could be benefits from reflecting operator opportunity costs in the setting of variable charges, potentially reducing reliance on administrative mechanisms. This has been considered in several studies, generally under the title of "scarcity charging", as a potential amendment to the structure of charges, for example NERA (1998), CEPA (2010), and NERA (2011).
204. As highlighted by those studies, any consideration of charging would need to take a number of factors into account, including:
  - a) the potential for the disaggregation, measurement and volatility issues discussed above to be exacerbated;
  - b) the ability of operators to respond to the signals created, given the way in which services are specified for franchised passenger operators;
  - c) the need to reflect societal benefits in the allocation of capacity, which may not be fully recognised by pricing mechanisms;
  - d) the incentives that a short run approach to charging for capacity constraints could have on Network Rail's incentives to invest in enhancements<sup>58</sup>;
  - e) the compatibility of resulting charges with the EU regulatory framework; and
  - f) the administrative implications in terms of measurement and billing.

---

<sup>57</sup> Section 4.1, NERA (2011)

<sup>58</sup> NERA (1998) observed that the very high charges that could result potentially create a disincentive for Network Rail to invest, because charges will fall as soon as enhancement made (Section 3.4.2 and 6.3, *An examination of rail infrastructure charges*, NERA et al, May 1998)

### Long run marginal costs

#### *The long run perspective*

205. The focus in the UK railway industry is on the allocation of short run marginal costs. As noted above, in other regulated network industries, it is common to consider the long run rather than the short run perspective of costs. In the context of rail, this amounts to extending existing cost allocations to include the cost to Network Rail of traffic related enhancements.

#### *Approach*

206. In theory calculating LRMC would involve estimating the additional enhancement costs that would be required to accommodate an additional unit of traffic, over and above existing levels. In practice, the impact of single unit increases in traffic can be highly volatile, depending on the precise balance between traffic and capacity. It may therefore be more practical to consider the impact of a modest increase in traffic, over and above existing levels.
207. For example, one approach would be to estimate the additional enhancement costs that would result from a 20% increase in traffic over and above existing levels. This would effectively extend the approach currently used for the allocation of maintenance and renewal costs, where a 20% increment is used as a proxy for the measurement of marginal cost, to include enhancement costs.
208. Under such an approach, LRMC would only rise above current variable cost allocations on sections of track that are currently either capacity constrained, or are relatively close to being capacity constrained. Where capacity is not constrained, LRMC would not rise above current cost allocations, since additional traffic would not necessitate enhancement<sup>59</sup>.
209. This would bear some similarities with, and share some of the benefits of, the allocation of operator opportunity costs (as observed, for example, by NERA (1998)). This reflects the fact that long run and short run approaches are alternative but closely related perspectives on the costs associated with capacity constraints:
- a) short run approaches focus on opportunity costs to operators which result from capacity constraints which, under the short run perspective, are assumed to be fixed; whereas
  - b) long run approaches focus on costs to the providing firm of alleviating capacity constraints, i.e. enhancement costs incurred by Network Rail.
210. A cost allocation approach based on LRMC would also bear strong similarities with the approaches adopted in the electricity and gas transmission networks.

#### *Level of disaggregation*

211. The level of geographical and temporal disaggregation chosen for LRMC analysis is likely to be critical. Broadly defined elements would risk enhancement costs attributable to constrained parts of the network being allocated to unconstrained parts of the network, with the risk of overstated cost allocations and inefficient reductions in traffic levels.
212. Narrowly defined elements, i.e. to disaggregated sections of track and times of the day, could reduce this risk. However they would introduce significant complexity. The key consideration will again be whether, at the level of disaggregation that is realistically likely to be achievable, the benefits of reflecting enhancement costs in constrained parts of the network outweigh the costs of overstating enhancement costs in unconstrained parts of the network.

---

<sup>59</sup> This appears to be the type of long run approach implicitly considered in *Access pricing in rail - Principles and structures*, Thomas and McMahon, 2005, who observe that LRIC (as termed in that paper) "would be equal to SRIC where there is sufficient spare capacity or capability". However, it differs from the type of approach considered by NERA (2011), who claimed that LRIC would increase costs allocated to uncongested parts of the network and times of the day.

### Measurement

213. A significant advantage of an LRMC approach over an opportunity cost approach is that the principles for estimating the impact of traffic levels on long run investment costs are well established, and have been applied in a range of industries, including gas and electricity transmission.
214. Crucially, whereas Network Rail does not have access to information on operators' willingness to pay, which is necessary to estimate opportunity costs, it does in principle have access to information on the potential costs of enhancing its own network, which is necessary to estimate LRMCs. This is likely to mean that measuring LRMC could be substantially easier than measuring opportunity costs.
215. Since:
- a) allocating LRMC bears some similarities with, and shares some of the benefits of, allocating opportunity costs; and
  - b) LRMC is likely to be easier to measure than opportunity costs;
  - c) it may be worth considering LRMC as a potential alternative to identifying and allocating opportunity costs.
216. The implementation effort required for LRMC would still, however, be significant. The complexity and scale of the analysis required to estimate LRMC in a network as complex as rail should not be under-estimated, particularly at the disaggregated level required for meaningful results. In particular, the nature of and level of enhancement expenditure required to increase capacity in different parts of the network may differ markedly, and in some cases may require information that may not be readily available<sup>60</sup>.

### Volatility

217. LRMC is also, like an opportunity cost approach, likely to introduce some volatility into cost allocations. In particular, once enhancement expenditure to relieve a capacity constraint has taken place and capacity expands, the forward-looking cost of enhancement required to meet an increase in traffic over and above existing levels will fall. However, this does have the benefit of cost allocations which in principle can encourage the use of newly provided capacity.

### Conclusions

218. The current cost allocation framework for what the industry describes as "variable costs" is mainly based on SRMC. It appears to broadly reflect the SRMC of costs that Network Rail incurs, assuming a network without capacity constraints, i.e. where demand is consistently below available capacity.
219. However, with the growth in traffic that the network has experienced over the last two decades, demand is not consistently below available capacity, and capacity constraints are widespread. In the presence of capacity constraints, the theoretical efficiency benefits of SRMC require it to reflect:
- a) the short run costs that Network Rail incurs; *and in addition*
  - b) the short run costs that operators incur as a result of not being able to access a capacity constrained network as much as they would like, i.e. the value foregone by demand that cannot be accommodated by existing available capacity (sometimes referred to as "scarcity costs").

---

<sup>60</sup> See NERA (1998); *Access pricing in rail - Principles and structures*, Thomas and McMahon, 2005

## Review of cost attribution and cost allocation approaches

220. The current cost allocation framework does not measure or allocate this additional set of short run costs, and so does not reflect the cost of capacity constraints. SRMC and variable costs are therefore understated in parts of the network that are capacity constrained.
221. One specific consequence of this partial reflection of SRMC is that the current cost allocation approach can lead to perverse results. In particular, a disaggregated analysis of the costs that Network Rail incurs suggests that variable costs are lower than average in capacity constrained parts of the network. However, this ignores the fact that variable costs resulting from capacity constraints are higher than average in those parts of the network. Therefore a move to disaggregation, using an approach which reflects the first set of costs but not the second, could serve to worsen rather than improve the cost reflectivity of cost allocations.
222. More generally, since costs relating to capacity constraints are not reflected, the management of capacity constraints must rely heavily on administrative mechanisms. Decisions on allocating scarce capacity and on investing in enhancement expenditure to relieve constraints, have to rely on bespoke assessments in which Network Rail, funders, operators and other stakeholders consider a range of factors in the decision making process. Identifying and allocating the costs associated with capacity constraints may help improve the effectiveness of these administrative mechanisms, by giving decision makers better quality information for consideration.
223. In principle, it may therefore be worth considering expanding the analysis of variable costs in the cost allocation framework to include the marginal costs associated with capacity constraints.
224. This consideration should take account of the likelihood that in practice, there will be two main difficulties in implementing such an expansion:
  - a) Measurement: Measuring the short run costs associated with capacity constraints is inherently very difficult, as it requires estimation of commercially confidential operator information on the value of unmet demand. However, it may be possible to gain similar benefits from an easier to implement long run approach, based on the concept of LRMC. LRMC considers the marginal cost to Network Rail of enhancing the network to alleviate capacity constraints..
  - b) Complexity: The costs associated with capacity constraints vary significantly by location and time of day. If the benefits of allocating these costs are to be realised, this suggests analysis at a very disaggregated level. Careful consideration would need to be given to the likely scale of benefits given the level of disaggregation that is likely to be achievable in practice.

## **Fixed costs**

225. The preceding section considered what are generally referred to in the industry as “variable costs”, i.e. costs that vary in response to relatively small changes in traffic.
226. This section considers what are generally referred to as “fixed costs”, i.e. all other costs. From a cost causation perspective, the term “fixed costs” essentially comprises two components:
- a) Long Run Incremental Cost (LRIC); and
  - b) common costs.
227. This section considers each of these components.

## **Long Run Incremental Cost**

### ***Approach***

228. In theory, the increment used for the calculation of LRIC could be defined in many different ways. In practice, when LRIC is applied to inform cost allocation, the increment is typically defined in terms of the removal of an existing level of output.
229. This gives LRIC, under this interpretation, a different focus from LRMC:
- a) LRMC focusses on the impact of small increases in traffic on enhancements, i.e. on the future investment necessary to accommodate that increased traffic; whereas
  - b) LRIC focusses on the impact of existing traffic on investment already in place, i.e. on what elements and features of the current network could in the long run be avoided at lower levels of traffic (sometimes referred to as an “avoidable cost” approach).
230. As an example, consider a section of track with:
- a) 2 up/down pairs of parallel tracks (i.e. 4 tracks in total); and
  - b) an hourly capacity of 15 trains per track pair (i.e. a total capacity of 30 trains).
231. If the existing level of traffic is say 30 trains, both LRMC and LRIC will generate cost allocations that rise above existing variable cost allocations. LRMC might reflect the additional long run cost of the third track pair necessary to increase capacity, and LRIC might reflect the avoidable long run cost of the second track pair no longer required, in the long run, at lower levels of traffic.
232. However, if the existing level of traffic is say 25 trains, LRMC and LRIC will generate very different cost allocations. LRMC might not rise above existing variable cost allocations, since there is spare capacity and some increase in traffic can be accommodated without the need for enhancement expenditure. LRIC, on the other hand, might still reflect the avoidable long run cost of the second track pair no longer required at lower levels of traffic. As a result, LRIC could rise above existing variable cost allocations, even where existing traffic levels are well below available capacity.
233. This distinction seems particularly relevant to the rail industry, because in many cases investment in enhancement is particularly lumpy, exacerbating the potential differences between LRIC and LRMC.



## Review of cost attribution and cost allocation approaches

234. As a result, a LRIC or avoidable cost analysis is less likely to be of use in informing or refining allocations of variable costs in the sense that term is currently used in the industry. However, LRIC can be of considerable help in allocating fixed costs, because it can reveal an objective causal link between levels of traffic and long run levels of cost.

235. AEA Technology adopted a LRIC approach to identify long run avoidable costs in a 2005 study for the ORR<sup>61</sup>. The study was aimed at an improved approach to allocating fixed costs between franchised passenger operators. AEA defined "avoidable costs" as:

*"These are costs that would be avoided were a particular activity not to occur. They will comprise the relevant variable costs of the activity plus the activity-specific fixed costs. As an example, were Thameslink's services not to operate on the Midland Main Line then not only would the relevant usage (wear and tear etc) costs be avoided, but so too would the fixed costs necessarily incurred for the Thameslink operation - such as those associated with the third and fourth line south of Bedford and the electrification."*<sup>62</sup>

236. AEA estimated the number of track km that could be avoided with the loss of individual passenger operators, and on the basis of those estimates, calculated consequential avoided:

- a) track maintenance and renewal costs;
- b) structures renewals costs;
- c) signalling and telecoms costs; and
- d) operations costs.

237. LEK Consulting also adopted a LRIC approach to identify long run avoidable costs for freight operations in a 2013 study for the ORR<sup>63</sup>. The study was aimed at informing charges levied on freight operators. However, only a proportion of costs identified were translated into charges, and only for certain commodities, due to long run avoidable costs being considered a mark-up on charges for costs "directly incurred"<sup>64</sup>. LEK defined avoidable costs as:

*"the theoretical long-run annual cost saving, over 35 years, which would result from removing commercial freight traffic from the network in its entirety on a permanent basis"*.<sup>65</sup>

238. The details of LEK's approach to estimating LRIC differed from AEA's approach, but both were aimed at a similar overall analysis. LEK identified as avoidable:

- a) long run costs associated with Freight Only Lines, being the costs of lines that would close in the absence of all freight services, including segments of branch lines used only by freight traffic and terminal lines<sup>66</sup>, and the costs of freight specific property that would no longer be required;
- b) long run costs associated with track on mixed use track sections that would no longer be required (including predominantly freight lines, loops, sidings, and parallel tracks); and

---

<sup>61</sup> *Recovery of Fixed Costs*, AEA Technology, October 2005

<sup>62</sup> Section 1.4, AEA (2005)

<sup>63</sup> *Estimating Freight Avoidable Costs*, LEK Consulting, May 2013

<sup>64</sup> Paragraph 16.377, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

<sup>65</sup> Page 11, LEK (2013)

<sup>66</sup> Paragraph 16.343, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013. Although CP5 incorporates separate Freight Only Line and Freight Specific charges, they were kept separate mainly for transparency reasons, and in principle Freight Only Line costs are simply a subset of Freight Specific costs (Paragraph 16.344, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013).

## Review of cost attribution and cost allocation approaches

- c) enhancement expenditure that could be avoided.
239. As highlighted by LEK, in principle LRIC should reflect all of the costs that could in the long run be avoided by the loss of traffic. This includes not only the cost of elements of the network that would no longer be required, but also any reduction in costs for elements of the network that would still be required, but could be configured at a lower cost. The key examples given by LEK were the potential for lower track categorisation and lower criticality banding for remaining track, and reduced inspection regimes for structures<sup>67</sup>.
240. It is possible to extend this principle further. The theoretical concept of the “long run” places no constraints on the degree to which networks can be redesigned to realise all of the avoidable costs that might theoretically be achievable by the loss of traffic. Avoidable costs could therefore also reflect more fundamental changes to the network. For example, as highlighted by the Passenger Transport Executive Group (PTEG) in a 2014 report:
- “Each type of train will have different infrastructure requirements and different impact on the infrastructure itself. For example, in order to reach consistently high speeds, inter-city services need straight alignments and gentle gradients, which require a large number of tunnels, cuttings, embankments and viaducts.”*<sup>68</sup>
241. In theory, this suggests that the LRIC of high speed services might reflect the avoidability not only of track and its maintenance, but also of structures created to manage alignments and gradients, i.e. cuttings, embankments, etc, which would no longer be required to accommodate only low speed services. Reflecting effects of this kind in LRIC would raise the allocation of avoidable costs for services which place greater demands on the design of the network.
242. In practice, it is common in LRIC modelling to place some constraints on assumed network redesign under alternative traffic scenarios, in some recognition of the difficulties that network operators would in reality have in re-optimising network configurations. In particular, a “scorched node” constraint is sometimes applied which requires the location and function of network nodes to be held constant when traffic losses are considered. The interpretation of the scorched node concept to rail would require more detailed consideration, but might for example preclude the removal of some structures.

### **Increment definitions**

243. As noted above, the LRIC increment is typically defined in terms of the removal of an existing level of output. For the railway network, two alternative definitions appear to be of particular interest:
- a) a removal of the traffic associated with individual operators; and
  - b) a removal of total traffic across all operators.

244. Each is considered in turn below.

### **Operator traffic increments**

245. Under an operator traffic LRIC approach, the increments considered would be a removal of all traffic from each operator individually. For any given section of track, LRIC would be equal to the long run avoidable costs associated with removing traffic associated with one operator, but retaining the traffic associated with other operators on that same section of track. This would equate to the difference in cost between:
- a) the long run costs of the current section of track; and

---

<sup>67</sup> Pages 92 to 95, LEK (2013)

<sup>68</sup> Page 10, *A heavy load to bear? Towards a fairer allocation of rail industry costs for regional rail*, PTEG, July 2014

## Review of cost attribution and cost allocation approaches

- b) the long run costs of the track required over that same section to service the remaining operators only.
246. Where a current section of track serves only one operator, then removing that operator will allow that section to be removed in its entirety in the long run, and all of the costs of that track will be included in LRIC.
247. Where a current section of track serves two or more operators, removing one operator may allow a reduction in the cost of track in the long run, depending on levels of operator traffic and traffic capacities. Returning to the example above of a section of track with 2 track pairs, each with a capacity of 15 trains, now assume that there are two operators on that section of track:
- a) operator A with 20 trains; and
- b) operator B with 10 trains.
248. The LRIC of operator A would equal the costs that could be avoided absent operator A, i.e. with track serving operator B alone. Operator B has 10 trains, which could be accommodated on a single track pair with a capacity of 15 trains. This implies that one track pair, and the costs associated with it, could in the long run be avoided. So the LRIC of operator A would equal the total long run fixed costs of the second track pair.
249. The LRIC of operator B would equal the costs that could be avoided absent operator B, i.e. with track serving operator A alone. Operator A has 20 trains, which could not be accommodated on a single track pair with a capacity of 15 trains, and a second track pair would still be required. This implies that the second track pair could not in the long run be avoided. So the LRIC of operator B would be zero.
250. It is notable that this approach can lead to relatively volatile results in response to small changes in patterns of operation. Returning to the example above:
- a) if operator A has 16 trains, and operator B has 14 trains, operator A will be allocated all of the costs of the second track pair, and operator B none of them; whereas
- b) if operator A has 14 trains, and operator B has 16 trains, operator A will be allocated none of the costs of the second track pair, and operator B all of them.
251. AEA (2005) adopted an operator traffic approach to its analysis of avoidable costs. It considered individual sections of track and estimated:
- "whether the number of running lines could be reduced if each TOC in turn were removed from the network."*<sup>69</sup>
252. AEA gave a specific example of the potential implications of an operator traffic approach:
- "the stretch from Paddington to Old Oak Common currently has 6 tracks; it was judged that this could be reduced to 2 tracks if the Greater Western TOC services (76% of the trains on this section) were removed, leaving Heathrow Express and very little else; but that it could not be reduced to less than 6 tracks if any other single operator were removed."*<sup>70</sup>

---

<sup>69</sup> Section 2.2.2, AEA (2005)

<sup>70</sup> Section 2.2.2, AEA (2005)

## Review of cost attribution and cost allocation approaches

253. As observed by AEA in its study, the results of this approach are highly dependent on the level of service aggregation. That is, on whether increments are defined, as above, in terms of all of the services provided by an operator, or whether they are defined in terms of smaller sets of services<sup>71</sup>.
254. This does however point to a potential weakness of such an approach. This is perhaps best illustrated by reference to AEA's example. At the time the study was conducted, the "Greater Western" operator/franchise in that example comprised the services originally provided by three separate operators/franchises: Great Western; Thames Trains; and Wessex Trains.
255. The (correct) implication of AEA's observation is that the level and pattern of LRIC based cost allocations, and ultimately the implied levels of government funding, would have been different:
- a) had those three operations/franchises still been separate, i.e. had the analysis proceeded on the assumption of not two but four separate operators/franchises (the three Greater Western operators/franchises plus Heathrow Express); or
  - b) had Great Western merged with Heathrow Express instead of Thames Trains (in which case, the fixed costs allocated to services to Heathrow would have been higher, but the fixed costs allocated to services to Greenford would have been lower).
256. However the merger of the operations/franchises was essentially an administrative matter of corporate restructuring and franchise consolidation. It is not clear that corporate ownership structures and the grouping of franchise services, which have no inherent impact on the costs caused by the train services that are running on tracks, should have a significant effect on the allocation of physical infrastructure costs.
257. LEK (2013) also adopted an operator traffic approach to its analysis of avoidable costs. It considered individual sections of track and estimated not the removal of individual freight operators, but the removal of freight operations in their entirety.
258. It is not clear why LEK's analysis aggregated freight operators, rather than treating each operator as an individual increment, in line with the approach taken by AEA. As AEA observed, it is likely that an individual operator approach would have led to lower identified avoidable costs.
259. Nor is it clear that LEK's aggregation of operators is necessarily wrong, if the issue it is designed to inform is the allocation of costs to freight as a whole. The slight tension between the LEK and AEA approaches is perhaps simply an additional illustration of the reliance that the operator traffic approach places on the choice of service aggregation.

### Total traffic increment

260. An alternative approach is to define the LRIC increment not in terms of the traffic of individual operators, but in terms of total traffic from all operators. Under a total traffic approach, total LRIC would be equal to the long run avoidable costs associated with removing all current traffic from a section of track, but retaining some minimal capability track on that section. This would equate to the difference in cost between:
- a) the long run costs of the current section of track; and
  - b) the long run costs of a section of track providing the same connectivity, but designed to service only minimal traffic levels.

---

<sup>71</sup> Section 5.3, AEA (2005)

## Review of cost attribution and cost allocation approaches

261. Total costs avoided would then be allocated equally to all traffic, i.e. without making any distinctions between the unit costs allocated to different operators.
262. As noted above, BT adopts a similar approach in its modelling of the LRIC of products. Traffic associated with products (rather than operators) is removed to very low levels, but the existing geographical reach of the fixed telecoms network is maintained.
263. There are also parallels with Ofcom's modelling of the LRIC of mobile telecoms voice call termination. Mobile network costs are allocated and charged to operators using termination services by estimating the avoidable costs associated with the removal of all termination traffic from all operators as a single increment, assuming no change in the level of geographical coverage. These costs are then allocated equally to all traffic.
264. As an illustration of the total traffic approach, returning to the example above of a section of track with:
- a) two track pairs;
  - b) an hourly capacity of 15 trains per track pair (i.e. a total capacity of 30 trains); and
  - c) two operators, A with 20 trains, and B with 10 trains.
265. If all traffic were removed, the second track pair would no longer be required and could in the long run be avoided. So, under a total traffic approach, total LRIC would be the difference between:
- a) the long run costs of two track pairs carrying 30 trains; and
  - b) the long run costs of one track pair carrying no trains.
266. This closely mirrors BT's network "thinning" approach to the LRIC modelling of its duct network. BT reduces the number of bores in an existing section of duct in response to falling traffic; here, the total traffic approach would reduce the number of track pairs in an existing section of track in response to falling traffic.
267. In contrast to the operator traffic approach, the total traffic approach allocates estimated avoidable costs equally to all traffic, so that each train would be allocated an equal estimate of traffic related cost, regardless of the operator to whom that train belonged, or to the corporate structure of operators.
268. Another difference between the operator traffic approach and the total traffic approach is the proportion of costs identified as avoidable with respect to traffic, i.e. as traffic-related. As demonstrated by the examples above, tracks in excess of the minimum required for low levels of traffic will, under the operator traffic approach, only be identified as traffic-related under certain patterns of train ownership by operators. Under the total traffic approach, such tracks will always be identified as traffic-related. Therefore, the total traffic approach is on average likely to lead to a higher proportion of costs being identified as traffic-related.
269. One of the notable conclusions of the AEA study was the observation that operators forming a small proportion of the total traffic on routes they used, received relatively low cost allocations<sup>72</sup>. It appears that this may have been heavily influenced by the choice of an operator traffic approach, and that an analysis based on a total traffic approach might have come to a materially different conclusion in this regard.

---

<sup>72</sup> Section 5.3, AEA (2005)

### [Comparison of operator traffic and total traffic approaches](#)

270. Both the operator traffic approach and the total traffic approach have merit:
- a) The operator traffic approach is based on a question that is clearly relevant to the allocation of costs to an operator, namely “what costs would be avoided absent this operator?”. However, it does appear to be prone to volatility, and to be influenced by corporate ownership structures which have limited relevance to the train services that use the infrastructure.
  - b) The total traffic approach identifies all traffic-related costs more completely, and is not affected by corporate ownership issues. However, it can result cost allocations to specific operators that differ from the costs that would be avoided absent those operators.

### *Non infrastructure costs*

271. The analysis required for LRIC has been described above principally by reference to the costs directly associated with network infrastructure. However it can in principle also extend to elements of centrally managed costs, particularly where cost drivers influencing the level of such costs can be identified, and where linkages can be made between the level of cost drivers and traffic levels.

### *Level of disaggregation*

272. As with opportunity costs and LRMC, the level of geographical and temporal disaggregation chosen for the LRIC analysis is again likely to be critical. For example:
- a) A route might include a two track pair section and a single track pair section, with different balances of total and operator traffic, and different avoidable costs, in each section. A disaggregated analysis which treated each section separately would be able to allocate the avoidable costs of each section in line with the traffic balances on that section. Under an aggregated analysis which treated the entire route as a single unit, the allocation of avoidable costs associated with one section would be affected by traffic on a different section.
  - b) A track section might have very different balances of traffic during peak and off-peak periods. A disaggregated peak v off-peak analysis would recognise the traffic balances during the periods relevant to cost avoidability.
273. Disaggregation is again however likely to introduce significant complexity. The AEA study, for example, disaggregated the network into 4,300 track sections in order to conduct its avoidability analysis<sup>73</sup>.
274. The key consideration will be whether, at the level of disaggregation that is realistically likely to be achievable, the benefits of more accurate cost allocation will be worthwhile.

### *Measurement*

275. The principles for measuring LRIC are well established and have been applied in a range of industries. In practice, however, the complexity and scale of effort required to estimate LRIC in a network as complex as rail should not be under-estimated, particularly if the analysis is conducted at a highly disaggregated level. Both factors are well illustrated by the AEA and LEK studies.

---

<sup>73</sup> Section 2.2.2, AEA (2005)

## Review of cost attribution and cost allocation approaches

276. The level of complexity will also be influenced by the choice of LRIC approach. A total traffic approach is likely to be simpler to implement: the total amount of traffic to be removed from a track section will always equal the total amount of traffic on that section; and the costs avoided will always be allocated across the same total traffic. In particular, there is no requirement to consider the specific pattern of traffic between operators. A total traffic approach might also allow the analysis to rely on existing traffic metrics, or variants of existing traffic, albeit at a disaggregated level. This might allow a graduated evolutionary path from the current FTAC approach to allocating long run total costs.

### Common costs

277. A LRIC approach will identify a proportion of so-called "fixed costs" as avoidable in the long run, and provide a means for attributing avoidable costs to operators. If fixed costs in total are to be allocated, the proportion of fixed costs that remains, i.e. non-avoidable or common costs, will also need to be allocated between operators on some basis.

### Approaches

278. Where a section of track serves only one operator, then all costs will be identified as avoidable under the operator traffic LRIC approach. Under the total traffic approach, the "minimal traffic" track will technically be left as non-avoidable or a common cost, but such costs should clearly be allocated to the sole operator on that section of track.
279. More generally, under both the operator traffic and total traffic LRIC approaches, common costs will arise on sections of track which serve two or more operators:
- a) under the operator traffic approach, common costs will be equal to the difference between total existing costs and the sum of the LRICs suggested by the individual removal of each operator; and
  - b) under the total traffic approach, common costs will be equal to the difference between total existing costs and the costs avoidable by scaling back to a "minimal traffic" track.
280. These common costs will need to be allocated between the operators that share that section of track.
281. There is no generally acknowledged single correct method to make such an allocation. In circumstances where cost allocations translate directly into charges, there are theoretical arguments for reflecting differences in operators' ability to pay, and allocating a greater share of common costs to operators with greater ability to pay - an approach known as "Ramsey pricing". PTEG (2014), for example, suggested such an approach for some common costs<sup>74</sup>. However, those circumstances do not apply in the case of the fixed costs of the railway network. Since a large proportion of fixed costs are financed through government funding, there is no direct link between fixed costs and charges. It is therefore not obvious that the theoretical benefits of Ramsey pricing are relevant in this context.
282. In circumstances where a LRIC approach has been applied and avoidable costs identified, a frequently adopted approach is to allocate common costs pro-rata with identified avoidable costs. So for example:
- a) if total costs are £100, the LRIC for operator A is £30, and the LRIC for operator B is £10; so that
  - b) the sum of LRICs is £40 (£30 + £10), split 75% (£30/£40) to operator A and 25% (£10/£40) to operator B; and

---

<sup>74</sup> Page 17, PTEG (2014)

## Review of cost attribution and cost allocation approaches

- c) common costs are £60 (£100 - £40); then
  - d) common costs are allocated 75% to operator A, and 25% to operator B, i.e. £45 to A and £15 to B; leaving
  - e) operator A with a total LRIC + common cost allocation of £75 (£30 + £45), and operator B with a total cost allocation of £25 (£10 + £15).
283. This is often referred to as an Equi-Proportional Mark-Up (EPMU) approach. The AEA study used an EPMU approach to allocate common costs between operators for cost categories where its LRIC analysis had identified significant avoidable costs: primarily maintenance and renewals costs.
284. One important feature of the EPMU approach is that all that is required to calculate an operator's share of total fixed costs, is that operator's share of avoidable costs. Under certain circumstances, this can significantly simplify the calculations required for allocating fixed costs. Returning to the example above of a section of track with 2 track pairs, each with a capacity of 15 trains, and two operators on that section of track:
- a) operator A with 20 trains; and
  - b) operator B with 10 trains.
285. Under the operator traffic approach, the LRIC of operator A was estimated as the long run cost of the second track pair, and the LRIC of operator B was estimated as zero. Since operator A has a 100% share of the avoidable costs, under the EPMU approach, it will end up with a 100% share of total fixed costs. It is not necessary to estimate what costs could in principle be avoided by the loss of the second track pair.
286. Under the total traffic approach, the LRIC of the loss of traffic was equal to the costs that could be avoided by the loss of the second track pair; and this would be allocated equally across the 30 trains, i.e. 2/3 to operator A and 1/3 to operator B. Under the EPMU approach, operator A would be allocated 2/3 of total fixed costs, and operator B 1/3. Again, it is not necessary to estimate what costs could in principle be avoided by the loss of the second track pair.
287. Although frequently adopted, EPMU is not the only possible approach to allocating common costs. Other approaches can be considered, including the kind of traffic metrics used in the existing FTAC allocation approach. However, these could be more complex to implement than an EPMU approach, as total avoidable costs (the cost of the second track pair in the example above) would need to be estimated.
288. There will be some categories of cost, for example some operating and centrally managed costs, where avoidable costs cannot be identified<sup>75</sup>. It is impossible to apply the EPMU approach directly in such cases, and some other approach is necessary. In the case of support costs, this can be related to the LRIC analysis: for example, head office management costs which are related to maintenance can be allocated between operators in line with their share of network maintenance costs. Alternatively, it may be that the traffic metrics continue to be the most appropriate in such cases (the AEA study for example allocated many operating costs on the basis of traffic metrics). Other options may be worth considering; for example, PTEG (2014) suggested allocation in line with operator revenues<sup>76</sup>.

---

<sup>75</sup> Often, this is not so much because no avoidable costs do not exist in such cases, but more because identifying them would require disproportionate effort. For example, a commonly cited example of a cost category with no avoidable cost element is "the Chairman's salary". In principle, the salary paid to Chairmen is not entirely independent of the scale of a company, so there may in principle be avoidable cost elements to that cost category. In practice, it may be too difficult to estimate avoidable costs of this nature with any reliability.

<sup>76</sup> Page 23, PTEG (2014)



### Level of disaggregation

289. As with other aspects of cost allocation in a network as varied as Network Rail's, the level of disaggregation at which the analysis is conducted is critical. A disaggregated analysis is likely to generate much more accurate results; albeit at the cost of additional complexity of calculation.
290. For example, as noted above, the common costs associated with a section of track should be allocated only between the operators that use that section of track. However, a route might consist of say three sections, with:
- a) section 1 used only by operator A;
  - b) section 2 used by operators A and B; and
  - c) section 3 used by operators B and C.
291. If common costs are allocated at the overall route level, aggregating the three sections, then some of the common costs associated with section 3, but not attributable to either of the sections used by operator A, will end up being wrongly allocated to operator A.
292. Similarly, where operating and centrally managed costs can be attributed to disaggregated parts of the network, disaggregated analysis can also help allocation.
293. Some disaggregation for the allocation of common costs is already explicitly recognised in the current regime. For example:
- a) the current approach to allocating the Fixed Track Access Charge (FTAC) allocates the costs, including common costs, attributable to each operating route only to the services using that operating route<sup>77</sup>; and
  - b) the current approach to allocating costs to the services provided on HS1 allocates the costs of that route only to services using that route.
294. It would appear that AEA's analysis could be improved upon in this respect. As noted above, AEA conducted its avoidable cost analysis at a very granular level, disaggregating the network into 4,300 track sections, and identifying LRICs for each track section. AEA also adopted an EPMU approach to the allocation of the common costs left behind by that analysis. However, the EPMU approach does not seem to have been conducted at the individual track section level, but rather at "*national, territory or area level*"<sup>78</sup>.
295. This mismatch between the level of disaggregation used for the identification of common costs, and the level of disaggregation used for the allocation of common costs, may have introduced material distortions into the study's results.

### Current approach to allocating fixed costs - FTAC

296. As part of the current approach to setting FTACs, the fixed costs of each operating route are allocated to operators on that route using traffic metrics such as train miles. Costs are allocated to franchised passenger operators, but not to open access operators or freight operators.

---

<sup>77</sup> The consistency with which this general principle is applied is considered further in the following section.

<sup>78</sup> Section 2.2.14, AEA (2005)

## Review of cost attribution and cost allocation approaches

297. There does not appear to be any explicitly stated linkage between the use of traffic metrics to allocate fixed costs between operators, and the costs that operators cause on the network. This is in contrast to a LRIC based approach, which as noted by AEA (2005), aims at an objective and transparent allocation of fixed costs between operators, in a way which reflects long run patterns of cost causation.
298. A LRIC based approach may bring benefits in terms of objectivity and transparency. It is therefore instructive to compare that approach with the current approach to allocating fixed costs.
299. It is important to stress that this remains a consideration of cost allocation, not of charging. Given the presence of government funding to finance the gap between fixed costs and charges, there need be no direct connection between the allocation of fixed costs, and the level and/or existence of fixed or variable charges for operators. However, a cost reflective LRIC based allocation of fixed costs could provide a clearer understanding of:
  - a) the allocation of government funding between operators; and
  - b) the pattern of support provided by the different providers of government funding.

### *Use and choice of traffic metrics*

300. The scale of difference between the current traffic metric approach and a LRIC based approach will vary, depending on which version of a LRIC based approach is considered.
301. Under a total traffic LRIC approach with an EPMU allocation of common costs, then as noted above, because identified LRIC is allocated equally to all traffic, i.e. in line with each operator's share of traffic, and because the EPMU allocation simply allocates common costs in line with this same share, total fixed costs can end up being allocated in line with traffic.
302. Therefore, there may not necessarily be a great deal of difference between the results that would emerge from a LRIC based approach of that nature, and allocation based on traffic metrics, as long as those metrics are appropriately selected, so that they reflect the long term drivers of incremental capacity.
303. It is not clear how well such drivers are reflected by the traffic metrics currently applied as part of the FTAC approach. It may be worth considering a review of the traffic metrics applied with this in mind. For example, a greater emphasis on traffic during peak periods, which is likely to drive many incremental costs, rather than overall traffic, may move the current traffic metric approach closer to a total traffic/EPMU approach.
304. The difference between the current traffic metric approach and other LRIC based approaches may be more significant. In particular, under an operator traffic LRIC approach, total fixed costs will often end up being allocated in quite a different pattern from the share of traffic, due to differing operator sizes and the lumpiness of avoidable costs. However, as discussed above, it is not clear that other LRIC based approaches are clearly superior to a total traffic/EPMU approach. Therefore, there may be merit in considering as separate issues:
  - a) the degree to which the gap between the existing traffic based approach and a total traffic LRIC/EPMU approach can be closed via a refined choice of traffic metrics; and
  - b) the degree to which a gap remains between that a total traffic LRIC/EPMU based approach and other LRIC based approaches, the complexity of analysis required to close that gap, and the likely benefits from doing so.

## Review of cost attribution and cost allocation approaches

305. The use and choice of traffic metrics is not the only likely source of differences between the current approach and a LRIC based approach. Differences of potentially comparable significance may be the result of way in which the current traffic metric approach is implemented, particularly in respect of:
- a) level of disaggregation;
  - b) consistency of operating route treatment; and
  - c) consistency of operator treatment.

### *Level of disaggregation*

306. As stressed above, analysis at a disaggregated level is critical for the accurate application of a LRIC based approach. Allocations under the current FTAC approach, however, are not generally performed at a level of disaggregation below the operating route level.
307. This suggests that there may be considerable opportunities to improve the current approach by allocating fixed costs at a more disaggregated level where possible, bringing the current approach closer to a LRIC based approach. Clearly this would involve significant effort, but increased disaggregation is something that could be implemented in a graduated evolutionary fashion.

### *Consistency of operating route treatment*

308. A LRIC based approach would apply the same cost allocation principles to all sections of track. However, the current approach to allocating fixed costs is not consistently applied across all operating routes. Specifically:
- a) the costs of the Scotland operating route are allocated entirely to the Scottish franchised passenger operator<sup>79</sup>, even though a share of traffic on that operating route is attributable to English and Welsh operators; and
  - b) the costs of each English and Welsh operating route are allocated between English and Welsh franchised passenger operators only, even if a share of traffic on that operating route is attributable to the Scottish operator.
309. This allocation approach appears to reflect an agreement between Transport Scotland and the Department for Transport at the point at which the Network Rail's Regulatory Asset Base was disaggregated between England & Wales, and Scotland. However there does not appear to be any objective cost allocation rationale for distinguishing the Scotland operating route from other operating routes in this way. It would seem more appropriate to treat every operating route consistently, and allocate fixed costs based on operator activity on that route.
310. A clearer and more consistent allocation of costs could help an improved understanding of the costs attributable to operators running services on both sides of the border, and of how those services are supported by funders on both sides of the border. It would also move the current approach closer to a LRIC based approach.

---

<sup>79</sup> A simplified explanation to explain the issue; strictly speaking this is only true for residual costs after deducting variable charges. It is worth noting, with respect to that deduction, that variable charges do not always reflect calculated SRMCs. For example: the ORR noted in its 2013 final determination that VUC charges were below SRMC (paragraph 16.136); and as noted above the capacity charges levied on non-franchised passenger operators are not consistently based on the application of SRMC to all traffic. While any decision to depart from cost reflective charges is properly a matter of policy, it should not affect the allocation of costs. It would therefore be better for residual costs to be calculated using SRMC estimates, not variable charges.

### *Consistency of operator treatment*

311. The FTAC cost allocation approach reflects Network Rail's methodology for allocating fixed costs between operators. That methodology is based on the physical use that each operator makes of each operating route. It allocates all fixed costs without making assumptions as to how those costs are ultimately recovered. In particular, Network Rail does not assume any specific splits of cost recovery between operator fixed charges and government funding, noting that decisions on such splits are matters for the ORR<sup>80</sup>.
312. A clear distinction between the approach to cost allocation and decisions on cost recovery is appropriate. LRIC based approaches, for example, apply the same cost allocation principles to all users of network infrastructure, based on estimated patterns of long run cost causation. Those patterns are not affected by how costs are recovered.
313. The distinction between cost allocation and cost recovery seems particularly appropriate in an environment where a significant proportion of fixed costs are recovered through government funding, and where there is therefore no essential requirement for fixed cost allocations to affect operator charges.
314. The FTAC cost allocation approach does not however appear to apply this distinction consistently. In particular, although the fixed costs of an operating route are allocated to an operator based on the physical use that operator makes of that operating route, that allocation is only made to operators that are franchised passenger operators. No allocation is made to open access or freight operators, regardless of the physical use they make of operating routes.
315. Network Rail's rationale for restricting the allocation only to franchised passenger operators appears to be that no fixed charges will be levied on open access and freight operators<sup>81</sup>. However, this is an assumption that relates to a cost recovery decision. Allowing such an assumption to influence the cost allocation approach is inconsistent with Network Rail's effort to distinguish the approach to cost allocation from decisions on cost recovery.
316. From a LRIC perspective, restricting the allocation of fixed costs only to operators that pay fixed charges effectively assumes that it is only operators that pay fixed charges that cause costs to be incurred in the long run. As demonstrated by the 2013 LEK study, which identified significant fixed costs that would be avoided by the loss of freight operations, such an assumption is invalid. As a result, the fixed costs allocated to franchised passenger operators are overstated.
317. More generally, the co-existence of two incompatible and contradictory approaches to fixed cost allocation, one for the setting of FTACs and another to inform the setting of the Freight Specific Charge, is likely to create a confusing and unclear picture of the distribution of fixed costs in the industry. The fact that the two cost allocation approaches are used to inform different charging decisions is not a good reason for them to adopt incompatible methodologies, because neither cost allocation methodology should be influenced by decisions on charging.
318. The inconsistent treatment of operators was also highlighted by PTEG (2014):

*"The effect of leaving freight rail freight out of cost allocation is to overestimate the actual level of public support going towards regional services while underestimating the value for money achieved from that subsidy (which should include the external benefits from rail freight)."*<sup>82</sup>

<sup>80</sup> Section 2.5, *Periodic Review 2013: Fixed Charges in CP5 - Conclusions*, Network Rail, March 2013

<sup>81</sup> Network Rail notes, for example, that costs are not allocated to open access operators because it has assumed such operators will not be levied fixed charges (Section 4.5, *Periodic Review 2013: Fixed Charges in CP5 - Conclusions*, Network Rail, March 2013).

<sup>82</sup> Page 15, PTEG (2014)

## Review of cost attribution and cost allocation approaches

319. Regardless of whether the cost allocation approach is based on the existing traffic metric approach, a LRIC based approach, or some hybrid between the two, there would be considerable merit in applying a single consistent approach to all operators. This would allow a clearer and more objective allocation of fixed costs, and a more transparent understanding of the distribution of government funding, without requiring any change to the level of charges paid by operators.
320. The fact that fixed cost allocations need not directly affect charges is clearly reinforced by the rationale behind the Freight Specific charge. Only a proportion of freight avoidable fixed costs identified by the LEK study are translated into charges, and only for certain commodities, due to these avoidable costs being considered a mark-up on charges for costs "directly incurred"<sup>83</sup>.

### *Previous consideration of a LRIC based approach*

321. Network Rail considered a LRIC based approach during PR08 in the wake of the AEA study, but raised three objections:
- a) it would be complex and unwieldy, and incapable of being "automated";
  - b) it would not necessarily be cost reflective, as it would consider only the costs that would be avoided if that operator were to cease services, which may not be the basis on which decisions are made; and
  - c) it could not fully incorporate all of the building block components for the calculation of the net revenue requirement<sup>84</sup>.
322. These concerns have some merit; on the other hand:
- a) while undoubtedly more complex than the current approach, the steps set out above may offer the opportunity to move closer to a LRIC based approach in a manageable fashion;
  - b) while the basis on which decisions is important for efficient decision making, it does not necessarily determine the questions that are relevant for the issue of cost reflectivity; and
  - c) the possibility that LRIC based approach cannot be applied to every building block component does not necessarily reduce the benefits that would flow from applying it to some building block components.
323. Given the potential benefits from a LRIC based approach, and the potential for the current approach to move closer to a LRIC based approach in an evolutionary fashion, it may therefore be worth exploring further.

## Current approach to allocating fixed costs - GB rail industry financial information

324. Fixed costs are also allocated between operators in the ORR's annual report, "*GB Rail industry financial information*".
325. The cost allocation process is independent from that applied in setting FTACs: a further demonstration of the principle of independence between cost allocation and charging in the railway network.

---

<sup>83</sup> Paragraph 16.377, *Periodic Review 2013: Final determination of Network Rail's outputs and funding for 2014-19*, ORR, October 2013

<sup>84</sup> *Periodic Review 2013: Fixed Charges in CP5 - Conclusions*, Network Rail, March 2013

## Review of cost attribution and cost allocation approaches

326. In broad terms, the approach adopted by the ORR has historically been similar to that adopted in setting FTACs: that is, fixed costs have been allocated, on the basis of traffic metrics, to franchised passenger operators, but not to open access or freight operators. As such, the observations made above with regard to the FTAC approach apply equally to the ORR's analysis.
327. While the overall traffic metric approach has been similar, the details of the ORR's approach have differed from that used in setting FTACs in a number of respects, principally because the ORR's approach is done on a much simpler and more aggregated basis. It is not clear why such differences exist, or that the scale of their impact could not be significantly reduced without undue effort. A more consistent approach between the two processes would contribute to a more transparent understanding of cost allocations in the industry.
328. The ORR does appear to acknowledge the differential treatment of operators in the cost allocations. It notes:
- "our analysis for the industry... does not include freight and open access train operators' income and expenditure as, financially, these are smaller components of the industry than franchised train operators, information is less readily available for freight and open access train operators and there is a lower need for public accountability as they are not direct recipients of government support."*<sup>85</sup>
329. These do not seem particularly convincing reasons for excluding open access and freight operators from the cost allocation process. Specifically:
- a) the 2013 LEK study identified freight avoidable costs of up to £428m a year<sup>86</sup>, around 7% of total Network Rail expenditure of £5,984m in 2012/13<sup>87</sup> - while this is a small proportion, it is large enough to have a material impact on the allocation of costs to other operators;
  - b) the traffic metric approach to allocating costs is a relatively high level exercise - it is not clear why the necessary information to extend this to open access and freight operators would not be available; and
  - c) for any particular operator, the existence or absence of government support is a function of the relationship between the charges paid by that operator, and the costs objectively attributable to that operator - it does not seem appropriate to pre-judge that relationship by making specific assumptions on government support in the cost allocation process itself.
330. PTEG (2014) suggested a number of improvements to the ORR's historical cost allocation approach. Of particular relevance to the issues considered above is PTEG's suggestion that maintenance and renewals costs would be better allocated on the basis of track wear and tear, rather than traffic metrics<sup>88</sup>.
331. PTEG's wear and tear suggestion has been adopted in the most recent version of the ORR's annual report, although the ORR does recognise that there are different views on how such costs should be allocated and advises care in the interpretation of its analysis<sup>89</sup>.

---

<sup>85</sup> Paragraph 5.35, *GB rail industry financial information 2012-13*, ORR

<sup>86</sup> Paragraph 16.388, PR13 Final Determination, ORR

<sup>87</sup> Table 4.1, *GB rail industry financial information 2012-13*, ORR

<sup>88</sup> Page 22, PTEG (2014)

<sup>89</sup> Paragraph 6.24, *GB rail industry financial information 2013-14*, ORR

## Review of cost attribution and cost allocation approaches

332. It is of course important that the cost allocation regime reflects the costs caused by track wear and tear. Such costs are in principle reflected in variable cost allocations. If all remaining fixed costs were truly common between operators, there could be an argument for allocating such costs in line with wear and tear costs, or variable costs more generally, using the logic of the EPMU approach.
333. However, as set out above, on a long run perspective, it is not the case that all fixed costs are common between operators. Many fixed costs can be attributed to operators using a LRIC avoidable cost analysis. Moreover, the driver of these long run avoidable costs is not wear and tear, but traffic. However, the effect of PTEG's suggestion is to allocate these costs on the basis of wear and tear. This does not seem appropriate<sup>90</sup>.
334. As to the common costs that remain after a LRIC analysis, the EPMU approach set out above suggests that these should be allocated in line with LRIC avoidable costs. It is not clear that allocating in line with wear and tear would be a better approach, but it could be argued that the EPMU approach should be based on the aggregate of LRIC avoidable costs and wear and tear costs. This might be worth considering, and could be achieved if, instead of conducting the LRIC analysis on fixed costs, the LRIC analysis is applied to total costs as a first step, with the variable cost analysis applied as a second step<sup>91</sup>. Indeed, this appears to have been the approach applied by AEA (2005).

## Conclusions

335. The current cost allocation framework for what the industry describes as "fixed costs", i.e. all costs not identified as variable, is based on allocating costs to franchised passenger operators in line with traffic metrics.
336. There does not appear to be any explicit linkage between the use of traffic metrics to allocate fixed costs between operators, and the costs that operators cause on the network. This is in contrast to a Long Run Incremental Cost (LRIC) or "avoidable cost" based approach. Such an approach aims at an objective and transparent allocation of fixed costs between operators, in a way which reflects long run patterns of cost causation as far as possible. Moving closer to a LRIC based approach, including a revised approach to the allocation of residual non-avoidable or "common" costs, may therefore bring benefits in terms of objectivity and transparency.
337. It is important to stress that this remains a consideration of cost allocation, not of charging. Given the presence of government funding to finance the gap between fixed costs and charges, there need be no direct connection between the allocation of fixed costs, and the level and/or existence of fixed or variable charges for operators. However, a cost reflective LRIC based allocation of fixed costs could provide a clearer understanding of:
- a) the allocation of government funding between operators; and
  - b) the pattern of support provided by the different providers of government funding.
338. The current use of traffic metrics as an allocation approach is not necessarily incompatible with some versions of a LRIC based approach, as long as those metrics are appropriately selected, to reflect the long term drivers of incremental capacity on the network. It may be worth considering a review of the traffic metrics applied with this in mind. For example, there may be a case for a greater emphasis on traffic during peak periods, which is likely to drive many incremental costs, rather than overall traffic.

---

<sup>90</sup> The fact that not all fixed costs are driven by wear and tear has been highlighted in the past by Network Rail (Section 3.2.2, *Periodic Review 2013: Fixed Charges in CP5 - Conclusions*, Network Rail, March 2013)

<sup>91</sup> An additional benefit of including wear and tear in the EPMU cost base would be that the EPMU approach could still be applied to any sections of track with two or more operators but no avoidable fixed costs.

## Review of cost attribution and cost allocation approaches

339. The difference between the current traffic metric approach and other LRIC based approaches may be more significant. However, it is not clear that these other approaches would be clearly superior, and it may be appropriate to consider the benefits of moving to such approaches against the effort that would be involved.
340. The use and choice of traffic metrics is not the only likely source of differences between the current approach and a LRIC based approach. Differences of potentially comparable significance may be the result of way in which the current traffic based approach is implemented, particularly in respect of:
- a) level of disaggregation;
  - b) consistency of operating route treatment; and
  - c) consistency of operator treatment.
341. Allocations under the current FTAC approach are generally performed at operating route or more aggregated levels. There may be considerable opportunities to improve the current approach by allocating fixed costs at a more disaggregated level where possible. This would bring the current approach closer to a LRIC based approach. Clearly this would involve significant effort, but increased disaggregation is something that could be implemented in a graduated evolutionary fashion.
342. There does not appear to be any objective cost allocation rationale for the current distinction between the treatment of the Scotland operating route from other operating routes in England and Wales. It would seem more appropriate to treat every operating route consistently, and allocate fixed costs based on operator activity on that route.
343. The current approach to allocating fixed costs appears to conflate the issue of cost allocation with the issue of charging. Fixed costs are allocated to franchised passenger operators only, and not to open access and freight operators.
344. The distinction between operator types does not appear to be based on any objective cost allocation considerations. Rather, the distinction appears to be based on the charging arrangements in place for different operators, and in particular on the fact that only franchised passenger operators are subject to fixed charges. However, given the existence of government funding, there is no necessary link between cost allocation issues, and charging issues.
345. Regardless of the cost allocation methodology adopted, the consistent treatment of all operators by that methodology would therefore allow a clearer and more objective allocation of fixed costs, and a more transparent understanding of the distribution of government funding, without requiring any change to the level of charges paid by operators.



### Next steps

346. The current cost allocation framework does not appear to be always well understood. A better common understanding would help consideration of potential opportunities to refine the framework. It may be therefore be beneficial to develop a document which summarises the framework, including a description of the legal and regulatory context, details of current cost allocations and their inter-relationships, and explanation of the rationale for cost allocation approaches. Some of this is already summarised in the RDG's *Charges and Incentives User Guide*, which could form a good starting point.
347. The clearest opportunity for refining the current framework, in terms of likely benefits compared with implementation effort, would appear to be an improvement in the consistency and clarity of fixed cost allocation. This would be based on the current definition of fixed costs, i.e. including within the scope of variable costs, only Network Rail's SRMC.
348. As a first step, this could involve extending Network Rail's existing methodology for allocating fixed costs so that it is consistently applied to all operators and all operating routes. The impact of such a change could then be assessed and consulted on before considering potential implementation
349. Other steps might include:
- a) clarifying the relationship between the freight avoidable cost analysis and the fixed cost allocation approach; and
  - b) improving the alignment between Network Rail's methodology and that applied in the ORR's *GB rail industry financial information* publication.
350. There also appear to be significant opportunities, again based on the current definition of fixed costs, to improve the accuracy of Network Rail's existing methodology for allocating fixed costs, and closing the gap with LRIC based approaches, by:
- a) refining the choice of traffic metrics to better reflect drivers of incremental network capacity; and
  - b) applying traffic metrics at a more disaggregated level.
351. Realisation of these opportunities could require significant implementation effort in terms of data gathering and analysis; and questions remain as to alternative approaches to measuring LRIC. It may be best to explore these issues by conducting and consulting on a pilot study on a small section of the network before deciding whether to implement any changes more widely.
352. The remaining main area of opportunity relates to the possibility of expanding the analysis of variable costs to include the LRMCs associated with capacity constraints. The key issue here relates to the practicality of achieving the level of disaggregation necessary to deliver the potential theoretical benefits of such an expansion. Again, this may be best explored through a pilot study on a small section of the network, before deciding on potential implementation.
353. Should all of these potential opportunities be implemented, the resulting cost allocation framework would comprise:
- a) "variable costs", which would be allocated based on Network Rail's LRMC (i.e. the existing approach of Network Rail SRMC, expanded to include long run as well as short run marginal costs), resulting in a higher level of variable costs in aggregate; and

## **Review of cost attribution and cost allocation approaches**

- b) "fixed costs" (which based on (a) would be at a lower level than is currently assumed), which would be allocated based on a revised traffic metric approach more closely reflecting a LRIC or "avoidable cost" based approach, including a revised approach to the allocation of residual non-avoidable or "common" costs.

## Annex A - Relationship between SRMC and total traffic related costs

354. As noted above:

- a) where marginal cost reduces at higher levels of output, the aggregate of SRMCs will be lower than total traffic related costs; but
- b) where marginal cost increases at higher levels of output, the aggregate of SRMCs will be higher than total traffic related costs.

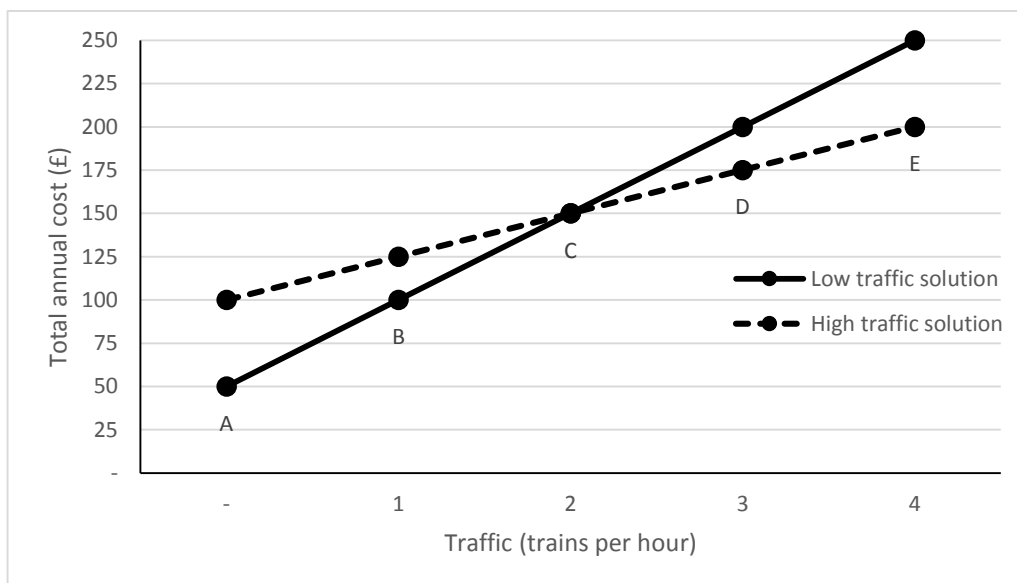
355. This annex explains these relationships using illustrative worked examples.

### Marginal costs which reduce at higher levels of output

356. Suppose, for a section of track, there are two alternative engineering solutions which have different maintenance cost characteristics:

- a) a "low traffic solution", with total maintenance costs that are relatively low at low levels of traffic, but rise rapidly as traffic rises; and
- b) a "high traffic solution", with total maintenance costs that are higher at low levels of traffic, but rise more slowly as traffic rises.

357. The two alternative solutions are illustrated in the diagram below. The "low traffic solution" is optimal (i.e. least costly) at traffic levels up to 2 trains per hour, above which the "high traffic solution" is optimal:



358. The total cost curve across all levels of traffic is defined by the kinked line ABCDE. At each point along that line, SRMC is the additional cost caused by increasing traffic by one train per hour. So SRMC at point A is the increase in cost between point A at £50 and point B at £100, i.e. £50. It can be seen that the kink causes SRMC to reduce at higher levels of output:

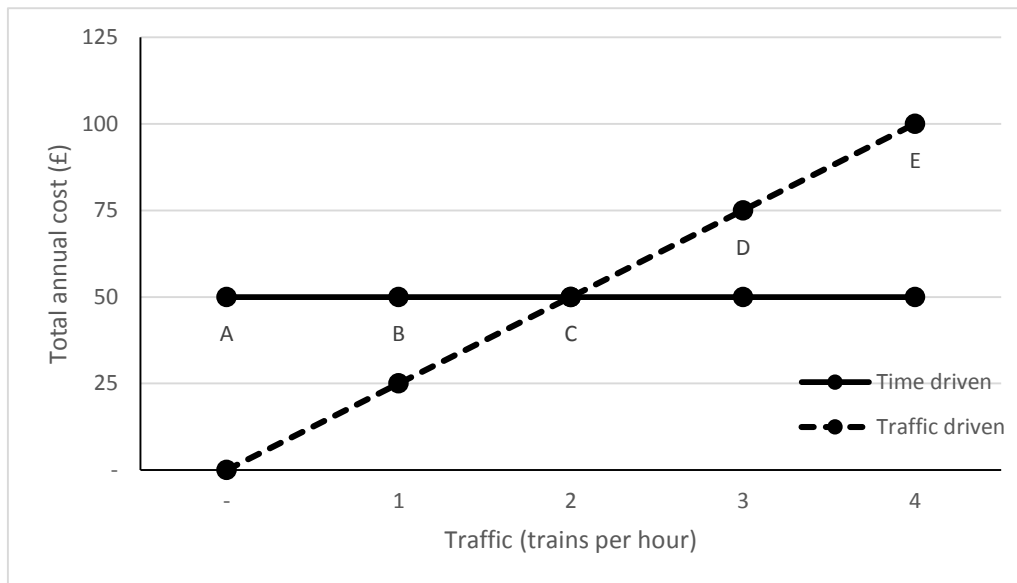
- a) SRMC is £50 at 0 and 1 trains per hour; and
- b) SRMC is £25 at 2 and 3 trains per hour.

## Review of cost attribution and cost allocation approaches

359. If the existing level of traffic is 3 trains per hour:
- allocating the SRMC of £25 to every one of the three trains results in a total allocation of £75; and
  - total traffic related costs, i.e. the costs that would be avoided absent all traffic, are £125 (£175 at point D, less £50 at point A).
360. Thus, with SRMC reducing at higher levels of output, the aggregate of SRMCs will be lower than total traffic related costs. This is because SRMC reflects only the gradient of the cost line at point D; it takes no account of the fact that at lower levels of traffic, the gradient is steeper.

### Marginal costs which increase at higher levels of output

361. Suppose, for a section of track, maintenance costs are driven by two factors:
- the passage of time, which causes a certain level of maintenance expenditure regardless of traffic levels; and
  - traffic levels, which cause additional wear and tear.
362. The two factors are illustrated in the diagram below. Maintenance costs are driven by the time factor at traffic levels up to 2 trains per hour, above which the traffic factor dominates:



363. The total cost curve across all levels of traffic is defined by the kinked line ABCDE. At each point along that line, SRMC is the additional cost caused by increasing traffic by one train per hour. So SRMC at point A is the increase in cost between point A at £50 and point B at £75, i.e. £25. It can be seen that the kink causes SRMC to increase at higher levels of output:
- SRMC is zero at 0 and 1 trains per hour; and
  - SRMC is £25 at 2 and 3 trains per hour.
364. If the existing level of traffic is 3 trains per hour:
- allocating the SRMC of £25 to every one of the three trains results in a total allocation of £75; and

## Review of cost attribution and cost allocation approaches

- b) total traffic related costs, i.e. the costs that would be avoided absent all traffic, are £25 (£75 at point D, less £50 at point A).
365. Thus, with SRMC increasing at higher levels of output, the aggregate of SRMCs will be higher than total traffic related costs. This is because SRMC reflects only the gradient of the cost line at point D; it takes no account of the fact that at lower levels of traffic, the gradient is flat.