ARUP

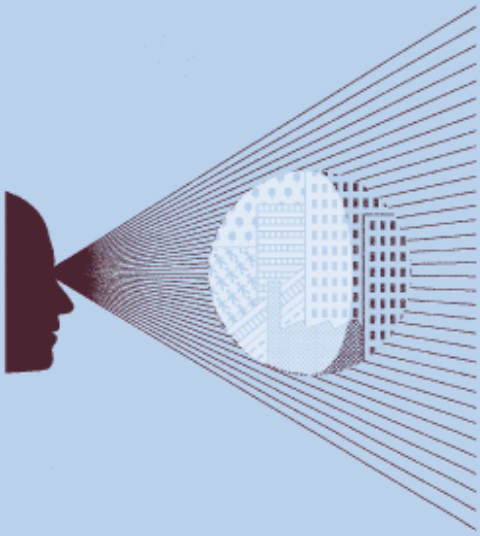
Oxera

How can a quality of service index be created?

Service quality report

Prepared for
the Department for Transport,
Transport Scotland, and
the Passenger Demand Forecasting Council

March 2010



Executive summary

Oxera and Arup have undertaken a study, 'Revisiting the Elasticity-Based Framework', by the Department for Transport (DfT), Transport Scotland and the Passenger Demand Forecasting Council (PDFC). As part of this study, which has focused on formulating a new elasticity-based rail passenger demand forecasting framework for Great Britain, the study team has created indices of service quality to be included as potential drivers of rail demand in its econometric analysis.

The purpose of creating service quality indices is to capture the 'soft' factors which affect the demand for passenger rail travel. In creating these indices, the objective was not to provide new estimates of sensitivities of rail demand to these factors per se, but to account for them in order to improve the estimates of the impact of the variables of primary interest (fares, socio-demographic and economic change, and modal competition) to the overarching 'Revisiting the Elasticity-Based Framework' study.

Separate service quality indices have been created for each market segment analysed for the overarching study, using data from the National Passenger Survey (NPS) and 'willingness-to-pay' data. These indices provide a new way of accounting for service quality directly in the econometric analysis.

As part of this study, Oxera has interviewed a number of NPS users from across the industry and would like to thank them for their input. These interviews highlighted a number of issues, a key one being the relationship between performance (ie, delays and cancellations) and responses to NPS questions. These issues have been taken into account during the creation of the indices.

The service quality indices remained in the final econometric models for two out of 28 market segments. This should **not** be interpreted as service quality being unimportant in explaining the demand for passenger rail travel, but may have arisen because of the relatively short time series of the data, and the relatively aggregate nature of the indices.

The indices created for this study have a range of further uses, from monitoring station quality to inclusion in revenue or cost models, or inclusion in future franchise specifications as a way to measure passengers' perceptions of service quality.

Contents

List of tables

List of figures

1 Introduction

Oxera and Arup have undertaken a study, 'Revisiting the Elasticity-Based Framework', by the Department for Transport (DfT), Transport Scotland and the Passenger Demand Forecasting Council (PDFC). Referred to in this report as the 'overarching study', this is concerned with re-estimating the fare and 'background growth' drivers (ie, socio-demographic, economic and modal competition) of the demand for passenger rail travel in Great Britain.

As part of this study, a number of reports have been produced, detailed below, which form key elements in the formulation of the overall final forecasting framework, and are referenced a number of times here.

Reports prepared by Oxera and Arup for the 'Revisiting the Elasticity-Based Framework' study:

- 'What are the findings from the econometric analysis?' (the *Findings* report)
  - 'Is the data capable of meeting the study objectives?' (the *Data capability* report)
  - 'How has the preferred econometric model been derived?' (the *Econometric approach* report)
  - 'What are the key issue for model specification?' (the *Model specification* report)
  - 'How has the market for rail passenger demand been segmented?' (the *Market segmentation* report)
  - 'Does quality of service affect demand?' (the *Service quality* report)
- 'How should the revised elasticity-based forecasting framework be implemented?' (the *Guidance* report)

During 2009, station quality was the focus of some discussion, and rose up the political agenda following the rail tour by Andrew Adonis, Transport Secretary (after which he noted that 'a few were excellent, including Sheffield, Manchester Piccadilly, and York, but others were poorly maintained'),[1] the subsequent appointment of station champions and their report,[2] and increased funding for stations.[3] The quality and amount of rolling stock in Great Britain have also changed considerably over the study period (1991–2008), with the replacement of old rolling stock, and the focus on increasing capacity being one of the objectives for the first High Level Output Specifications (HLOSs) by the Westminster and Scottish governments.

The invitation to tender for the overarching study required as many variables as possible to be estimated directly in the econometric analysis, in order to increase confidence in the final parameter estimates.[4] Therefore this study considers the demand drivers of direct interest—eg, fares, socio-economic and demographic variables, and modal competition—and both 'soft' and 'hard' aspects of service quality. Although the service quality factors are not key objectives of the overarching study, controlling for other factors should result in more accurate parameter estimates for the main variables of interest, and, hence, improved forecasts of the demand for rail travel.

[1] http://www.dft.gov.uk/press/railstations

[2] Green, C. and Hall, P. (2009), 'Better Rail Stations: An Independent Review Presented to Lord Adonis, Secretary of State for Transport', November.

[3] Department for Transport (2009), 'Government Response: Station Champions' Report–"Better Rail Stations"', November, p. 3.

[4] Department for Transport (2008), 'Rail Passenger Demand Forecasting: Revisiting the Elasticity-Based Framework Request for Proposal and Statement of Requirement', July, p. 2.

This section now provides a brief description of the process used to decide on the most feasible approach to accounting for service quality changes over time and between market segments in the overarching study's econometric analysis.

1.1 Background

Initial research suggested that the National Passenger Survey (NPS) may provide a source of time-series data on service quality for use in this study. The NPS is a national survey of rail passengers, conducted for Passenger Focus in the spring and autumn. It designed to provide a robust sample size for each train operating company (TOC), although not necessarily for each of the market segments used for the overarching study.[5] These market segments, and the process followed to derive them, are detailed in the *Market segmentation* report. In order to provide the most benefit to the overarching study, separate indices have been developed for each market segment. The matching of the service quality indices to the segments derived for this analysis is discussed in section 4 below.

A number of questions have been asked consistently throughout the history of the NPS, which cover the following areas of interest in the construction of the service quality index:

– ticket-buying facilities at the station;
– the provision of information about train times/platforms at the station;
– the upkeep/repair of the station buildings and platforms;
– the cleanliness of the station;
– the facilities and services available at the station (eg, toilets, shops, cafes, etc);
– the attitude and helpfulness of the staff at the station;
– the connections with other forms of public transport (eg, bus, Tube, tram, taxi, etc) at the station;
– the facilities for car parking at the station;
– the overall station environment;
– the frequency of the trains on the journey;
– the punctuality or reliability of the train (ie, the train arriving/departing on time);
– the length of time the journey was scheduled to take;
– the connections with other train services;
– the value for money of the ticket;
– the cleanliness of the train;
– the upkeep and repair (condition of seats, walls, tables, etc) of the train;
– the helpfulness/attitude of the staff on board the train;
– the space for luggage on the train;
– the toilet facilities on the train;
– whether there was sufficient room for all the passengers to sit/stand on the train;
– the comfort of the seats on the train;
– the ease of being able to get on and off the train;*
– the availability of catering (food/drinks) on the train;*
– the passengers' overall satisfaction with their journey.*

These questions cover a wide range of attributes of service quality over different waves of the survey, and provide a firm foundation on which to base service quality indices. Those questions which have not been used to create the service quality indices are marked with an asterisk. In some cases, this has been because of the lack of appropriate data to provide weights.

The questions in the survey are answered on a five-point scale, with 1 corresponding to 'very satisfied' and 5 to 'very dissatisfied'.

[5] http://www.passengerfocus.org.uk/research/nps/content.asp

In addition to the core requirement—to reflect service quality accurately—in order for the service quality indices to be useful for the study, they need to be:

– easy to calculate and use in a demand forecasting framework;
– forecast in a straightforward and flexible manner;
– useful in the econometric analysis;
– sufficiently flexible to include or exclude performance and crowding;
– able to provide a final index that broadly reflects passenger satisfaction.

This section has described the information available from the NPS, and the requirements from an index of service quality. The remainder of the report is structured as follows:

– section 2 discusses the preliminary analysis undertaken, to provide reassurance that the service quality indices will add value to the study;
– section 3 discusses the meetings held with users of the NPS in relation to the use of the survey in the industry;
– section 4 considers the conceptual issues related to creating service quality indices from NPS data;
– section 5 details the NPS data and section 6 discusses how the data could be weighted to create each index;
– section 7 provides details on the construction of the indices;
– section 8 details the segment-specific indices;
– section 9 concludes.

In preparing this report, Oxera interviewed a number of people and would like to express its gratitude to Simon Blainey, Mike Boden, David Chilvers, Jeremy Clarke, Rose Costello, David Greeno, Matthew Lodge, Vince Lucas, Rob Sheldon, and Peter West.

2 Prerequisites for using a service quality index in the econometric analysis

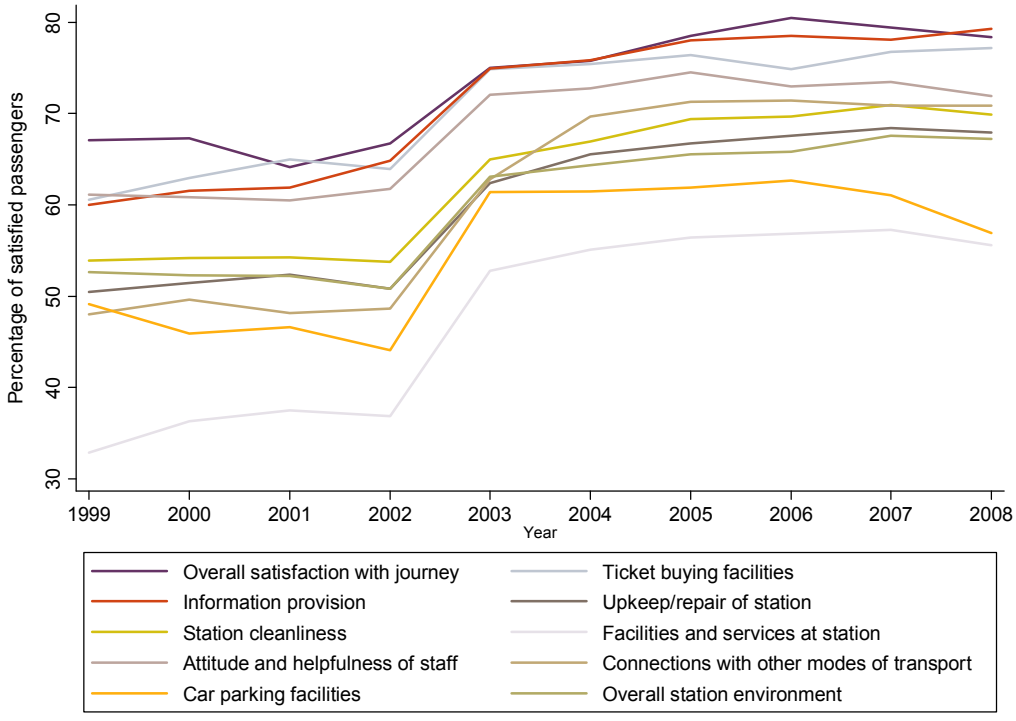
Before creating the service quality indices, the feasibility of using NPS data to create the indices was assessed, by analysing;

- the variability of the data;
- whether there was sufficient data available to provide weights for a service quality index;
- whether passengers' preferences were consistent across different studies (in order to enable this information to be used to weight each question response in an overall index);
- whether the sample size of the NPS was sufficient to provide confidence in any indices created;
- whether the survey has a sufficiently long time period to be useful for inclusion in the econometric analysis.

2.1.1 Data variability

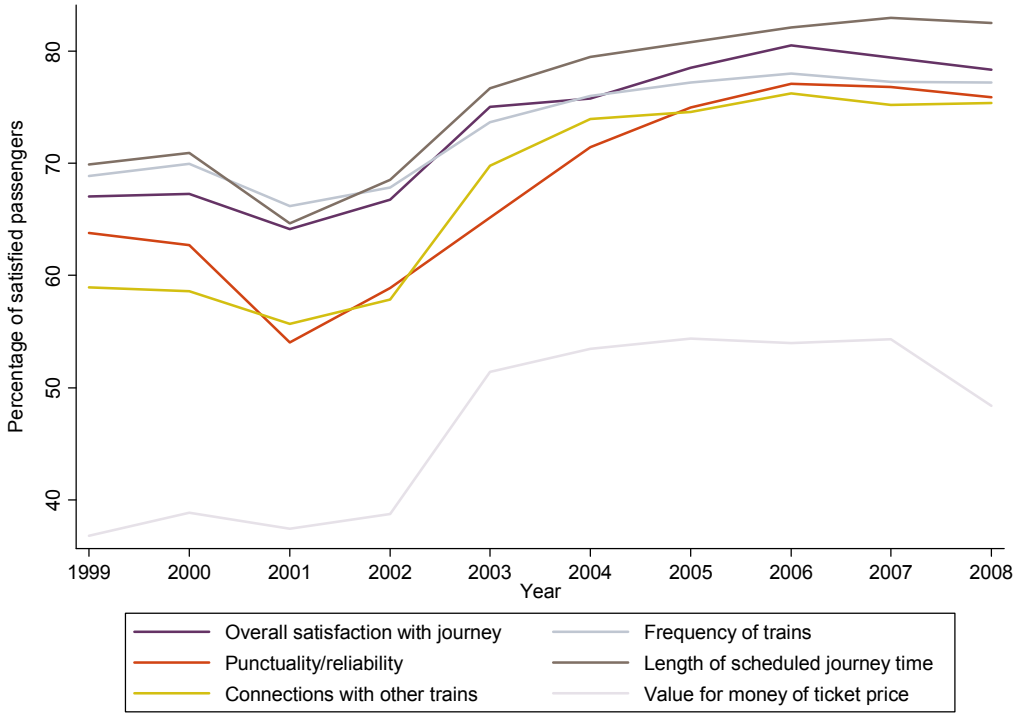
Figures 2.1–2.3 show how the key variables in the NPS have varied over time at a national level in order to inform an initial assessment of whether there is likely to be sufficient variation in the data for the NPS to provide a useful data source for the creation of service quality indices.

Figure 2.1 Respondent satisfaction: station attributes



Source: NPS.

Figure 2.2 Respondent satisfaction: journey attributes

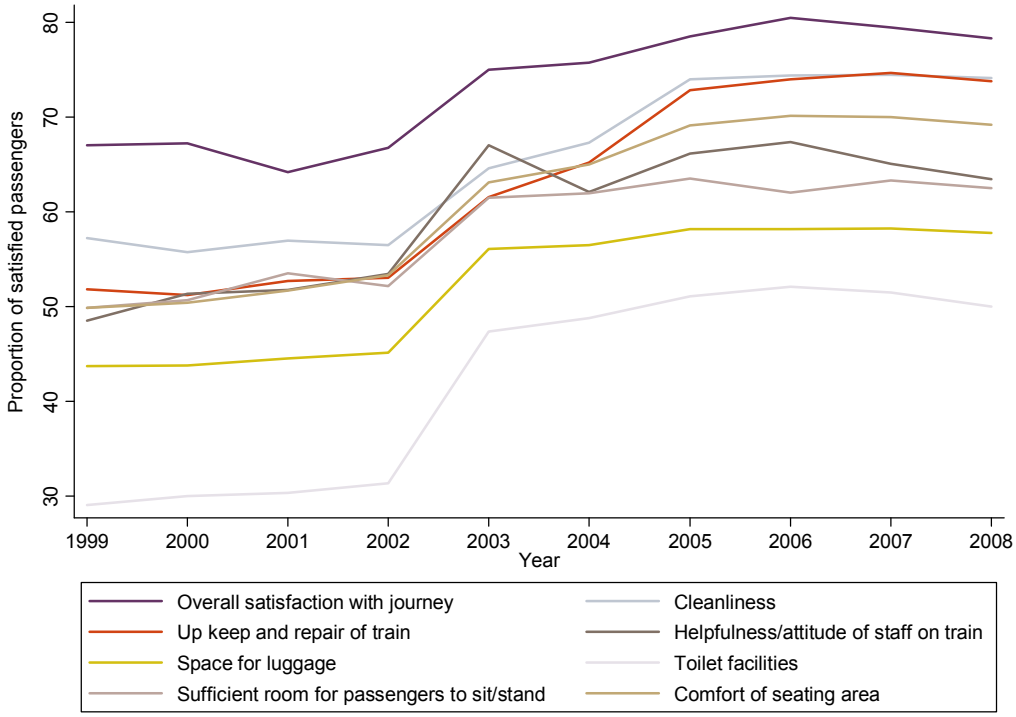


Source: NPS.

Although journey time and expected waiting time are captured in the econometric analysis in the overarching study in the generalised journey time (GJT) variable, passengers' perceptions of whether they are satisfied with the scheduled journey time, or the frequency of trains on the route, may be expected to differ systematically from the measured GJT. (For example, a wait of 30 minutes at a major station with a hot drink may be very different to a 30-minute wait on an unmanned platform with no facilities). Hence, a measure which reflects passengers' perceptions, as well as the GJT of the journey, is useful for inclusion in the econometric analysis. To guard against multicollinearity,[6] the study team has assessed the correlation between the derived service quality indices and GJT. Typically, this is less than 0.05, which means that collinearity between these variables is not a cause for concern. Similar arguments apply for the relationship between fare and value for money, and these factors are therefore included in the final indices.

[6] TO BE DEFINED.

Figure 2.3 Respondent satisfaction: train attributes



Source: NPS.

Figures 2.1–2.3 show that the general trend in the data is towards increased satisfaction. This corresponds to expected trends.[7] The variation in the service quality index will depend on the weights given to each attribute. While there are no hard-and-fast rules for what is 'enough' variation in a data series, the degree of variation present in the NPS data provided the study team and the study management group with sufficient comfort that the effects of changes in service quality are likely to be useful in explaining the demand for rail travel.

2.1.2 Data for weights

The study team's preference was to use data on passengers' willingness to pay (WTP) for alternative improvements in rail service quality to weight each question in the NPS (for which there is consistent data over time). The rationale behind this approach is that the more passengers are willing to pay for a given improvement relative to another improvement, the greater the weight this aspect is likely to have in determining passengers' perception of service quality. Oxera has examined a number of sources of data on passengers' WTP for improvements in service quality, including the PDFH v5,[8] Transport Research Laboratory (2004) (also known as TRL 593),[9] University of Southampton et al. (2008),[10] and a study for Passenger Focus (MVA Consultancy 2007).[11] These studies are considered in more detail in section 6. The University of Southampton et al. (2008) study used WTP data to generate station quality indices.

[7] Passenger Focus (2009), 'Fares and Ticketing Study Final Report', February, p. 6; and (2005), 'National Passenger Survey Autumn 2005', p. 36.
[8] Association of Train Operating Companies (2009), 'Passenger Demand Forecasting Handbook', August.
[9] Transport Research Laboratory (2004), 'The Demand for Public Transport: A Practical Guide', pp. 83–101.
[10] University of Southampton, Accent and ITS Leeds (2008), 'The Effects of Station Enhancements on Rail Demand: Phase 2', May.
[11] MVA Consultancy (2007), 'Passengers' Priorities for Improvements in Rail Services', June.

Table 2.1 contrasts the results of the relationships between overall satisfaction and the responses to the NPS questions with the factors with the highest passenger WTP in the report produced by MVA Consultancy. It is important to compare these different ways of reflecting underlying unknown passenger preferences; that they should both reflect the same underlying preferences—and hence their consistency (or otherwise)—is an important indicator of whether the indices accurately reflect passengers' perceptions of changes in service quality.[12]

Table 2.1 Relationship between MVA report and correlations

Ranking	MVA report	Correlations
1	Value for money of the ticket	Punctuality of the trains
2	Sufficient trains operating	Seat comfort
3	Punctuality of the trains	Scheduled journey time
4	Availability of a seat	Cleanliness of the train
5	Information about delays	Ease of getting on and off the train

Source: MVA Consultancy (2007), 'Passengers' Priorities for Improvements in Rail Services', and Oxera analysis of NPS data.

The top five responses between the two data sources are not particularly consistent; however, when the top *ten* factors are considered, six of them are common to both sources, which suggests that they both provide a reasonably consistent picture of passengers' preferences.

2.1.3 Sample size

The NPS is a national survey designed to provide a reasonable sample size for each TOC, and has well-known small sample sizes if the survey is split to a flow level.[13] However, the sample size of 25,600 is sufficient to disaggregate into the market segments set out in the *Market segmentation* report for the overarching study.

A key issue is the availability of NPS data over time. The survey began in 1999, suggesting that its inclusion could result in the econometric estimations undertaken for the overarching study being based on a restricted time series, since the demand data used for the overarching study begins in 1991/92. While this limits the value of any index created using the data, its availability enables the checking of the impact of including the index on elasticities estimated over a time-limited sample of the data. These conclusions are then transferred to the estimates obtained from the longer time series available from other variables. This is a similar approach to that adopted for the performance variable, where there is a trade-off between the level of detail and the length of the time series (see the *Data capability* report for more details).

This section has described the preliminary analysis undertaken to ensure that the construction of a service quality index was likely to be practical, thus adding value to the overarching study. The next stage in the study was to consult a number of NPS users to determine whether there were data or other important issues (eg, availability of NPS forecasts) to be aware of, as well as how the survey is used in different parts of the industry. The next section describes the results of these consultations.

[12] A common issue with stated-preference research (of which the MVA Consultancy report for Passenger Focus is an example) is that the context can affect the responses to questions. Hence, should the answers to these two methods appear inconsistent, the question of whether the inconsistency is 'real' or an artefact of the different contexts would need to be considered.

[13] MVA Consultancy (2007), 'Rail Passenger Demand Forecasting Research: Workstream 2—Review and Specification of Data Series', March, p. 2.7.

3 Meetings

To gain a better understanding of the NPS and how it is used throughout the industry, Oxera held meetings with NPS users in TOCs and the DfT, and with Passenger Focus and its market research company, Continental Research.

The meetings covered a range of topics, including whether and how the NPS was used in the business of the person being interviewed, and whether there were any issues with the survey that may be of relevance to this study.

3.1 Use of NPS across the industry

The use of NPS data within TOCs varied considerably, with some not using the NPS measure in great detail, and others using it as a basis for their own (larger) surveys. There was agreement among those the TOC representatives interviewed that the sample size of the NPS meant that the TOCs needed to undertake their own surveys with larger sample sizes—in particular, to identify issues at a station or flow level. (Apart from a few large flows and stations, the NPS results were judged to be insufficiently robust when disaggregated below TOC level.)

Little work has been carried out on linking NPS scores to the demand for passenger rail travel, although one respondent commented that there was ongoing work within their company in this area. A more common application is for the TOCs to use the NPS and their own surveys as a performance management/improvement tool. For example, one TOC has conducted work to investigate the impact of certain initiatives on NPS scores, but not how this feeds through to demand. Another TOC uses the NPS as a benchmark for their internal service quality survey.

Therefore, in a number of cases the focus was on the optimal deployment of current resources, rather than making a business case for investment. One respondent commented that return on investment was calculated using traditional demand forecasting models.

The DfT is looking to increase the use of quality measures (such as the NPS) in franchise specifications; for example, explicit targets for NPS scores have been included in the new Southern franchise.

3.2 Data issues

Other issues that arose included a seasonal pattern in responses, with autumn scores generally lower than spring scores, possibly reflecting leaf fall timetables and typically worse performance scores.[14]

One pattern consistently identified in the interviews, which has been the subject of work by Passenger Focus, is that performance is an important determinant of passengers' perceptions of service quality—ie, there is a high correlation between performance and NPS scores. This is important for the creation of service quality indices because performance data is available for the econometric analysis in the overarching study as a separate variable. Therefore, the impact of performance needed to be removed from the service quality index to avoid double-counting. This issue is considered in more detail in section 4.2.

[14] For example, see ORR (2009), 'National Rail Trends 2008–2009 Quarter 1', Table 2.1a.

Another issue discussed was the existence of non-linearities in responses (ie, whether the move from ‘satisfied’ to ‘very satisfied’ is the same as from ‘very dissatisfied’ to ‘dissatisfied’), although respondents were unsure about this. This issue, which has been carefully considered during this study, is discussed further in section 4.3.

One respondent commented that they only forecast the NPS scores for the next wave (the impact of performance on the scores is so great that, without such a forecast of performance, it is not possible to forecast the NPS scores). This issue is considered carefully in the *Guidance* report, which describes how to use the forecasting framework developed in the overarching study.

This section has discussed the outcomes from a range of meetings which were held with users of the NPS from across the industry. These meetings highlighted a number of factors which were subsequently investigated further. The next section discusses the conceptual approach to generating indices of service quality.

4 Conceptual approach

The NPS data could not be used in its 'raw' form (changes in passenger responses over time) as it is a survey of passengers' perceptions. There are multiple responses for each market segment. To enable inclusion of the indices directly within the econometric estimation of the overarching study, these different responses need to be aggregated to provide a single number for each market segment.

Oxera has carefully considered the questions that have been asked consistently in the NPS (see section 1). Some of these have been rejected because they are not central to the quality of service being offered, while others have been excluded because they do not offer a sufficiently long time series to be useful in the index.

This study uses a linear aggregation rule to create the index:[15]

$$QI_{it}=\omega_1 x_{1,it} + \ldots + \omega_n x_{n,it} \qquad \text{Equation 4.1}$$

where QI_{it} is the index for segment i at time t, x_j (j=1,…,n), and represents the different variables for use in constructing the index for segment i at time t, and ω_j (j=1,…,n) is the weight attached to that variable.

An aggregation rule of this type provides for the creation of a wide range of indices, depending on the weights used. The weights within such an index are implicitly trade-offs between the different variables that make up the index. This implies that an increase in one variable can offset a decrease in another. While this may be appropriate for small changes (eg, a small increase in the comfort of the seats may offset a small decrease in the cleanliness of the train), this may not be the case for large changes, and care should be applied in the interpretation of the index.

This approach has been adopted for its combination of simplicity and flexibility. The following sections discuss how to assign the weights (ω_j, j=1,…,n) and the variables (question responses) used to create the indices.

4.1 Weights

There are several approaches to developing index weights and considerations to be taken into account. Two possible approaches are discussed below.

Exogenous weights

One method of providing the weights is to impose them from an external source, and one such source for this study is WTP data. This data has the advantage that the weights used in the index are consistent with their statistical meaning. Hence the interpretation of the indices is robust because the weights reflect their true statistical meaning of allowing trade-offs between the different elements of the index to be understood.[16] Data is available from existing research such as the work on station enhancements undertaken by the University of

[15] Other approaches are possible, such as non-linear aggregation rules. However, the application of such rules is often complex and interpretation is not straightforward. Given that one of the requirements of the service quality indices is that they should be easy to calculate, a linear aggregation rule was considered more suitable for this study.

[16] This is rarely the case in composite indices. See Munda, G. and Nardo, M. (2005), 'Constructing Consistent Composite Indicators: the Issue of Weights', European Commission Joint Research Centre and Institute for the Protection and Security of the Citizen.

Southampton et al.,[17] or MVA Consultancy's work on the value which passengers place on improvements to service quality.[18] These studies are discussed in more detail in section 6.

One disadvantage of this method is that the weights are not necessarily consistent with the underlying patterns in the data, as they are imposed from an external source at one point in time (whereas the NPS data to be used in developing indices changes over time). However, both the WTP data and the NPS reflect underlying passenger preferences and should therefore be largely consistent. This consistency arises as the underlying passenger preferences, which generate both the NPS and WTP data, are assumed fixed at a given point in time; consequently, weights derived from the two sources should be broadly comparable.

Principal components analysis

Principal components analysis (PCA) is a statistical method used to analyse patterns within data. The underlying principal is that aspects of the data which drive the largest proportion of the variation in the data are allocated the largest weight (see Box 4.1).

Box 4.1 Principal components analysis

The principle is that the data can be decomposed into vectors (known as 'principal components'), which are uncorrelated with each other but explain the variation in the data. In this way, a number of variables are transformed into uncorrelated vectors. The vector explaining the largest proportion of variation in the data (known as the 'first principal component') is used to provide the weights for the index.

The advantage of this method is that the weights are assigned such that they are consistent with the underlying patterns in the data. However, the weights are not necessarily consistent with an economic rationale and are likely to change whenever the data alters—ie, when new waves of the survey are included in the dataset. Section 7.1 compares the two approaches and provides conclusions on which is preferred for this study.

4.2 Correlations

It is important that the index avoids any double-counting of factors that are included elsewhere in the forecasting framework (eg, performance). Given that the impact of performance on the NPS responses was highlighted in a number of the interviews with industry members undertaken for this study, this relationship has been considered carefully.

The study team has considered two options in order to view how performance may affect the NPS scores of other factors. In the first option, the response for a given factor depends on the quality of that factor, but the overall perception of service quality is dominated by performance:

$$\text{response (factor)} = f(\text{factor}) \qquad \text{Equation 4.2}$$

Performance solely affects overall satisfaction by having the largest weight in the index: passengers' responses to the other (non-performance) NPS questions are not affected by performance. In this case, removing the impact of performance from the index is accomplished by adjusting the weights applied when creating the index. The adjustment of the weights is discussed in section 7.2.

In the second option, the response for a given factor depends on both performance and the factor:

[17] University of Southampton, Accent and ITS Leeds (2008), 'The Effects of Station Enhancements on Rail Demand: Phase 2', May.
[18] MVA Consultancy (2007), 'Passengers' Priorities for Improvements in Rail Services', June.

response (factor) = f (performance, factor) Equation 4.3

Removing the impact of performance from the responses in this setting is more complex. The study team, in conjunction with Professor Khumbakar, academic adviser to the team, has considered a number of approaches to undertaking such a task, but concluded that such approaches were not feasible given the data available. Discussion with the academic advisers suggested that simultaneous regression techniques, using data on responses to NPS questions and on performance, might provide a solution. However, the study team did not have sufficiently detailed performance data. Such analysis would require data on performance at the level of the NPS responses—ie, to know the performance of the journey in relation to which the passenger is providing responses in the NPS. It may be possible to use performance at higher levels of either spatial or temporal aggregation, but this would need to be considered carefully before use. Therefore, the first approach has been adopted.

4.3 Linearity of responses

One issue to be considered in the analysis of categorical data (such as the NPS data) is how responses such as ‘satisfied’ or ‘dissatisfied’ relate to ‘very satisfied’ or ‘very dissatisfied’. This is important, not least because their relationship dictates whether responses can be aggregated without introducing aggregation bias.

This study tested different model specifications to determine whether aggregation of responses was feasible or desirable. The analysis included investigating the variety in the indices using different levels of aggregation (see section 5 below).

A number of conceptual issues related to the use of NPS data in indices of service quality have been discussed in this section. The data is examined in more depth next.

5 Data

The NPS is a national survey conducted in spring and autumn, and designed to provide a robust sample size for each TOC.[19] However, the sample size of 25,600 is also sufficient to disaggregate into the market segments used in this study. During the meetings with NPS users for this study, it was highlighted that the levels of the autumn responses tend to be lower than the spring responses due to a number of factors, as explained in section 3.2.

Oxera has had access to the first 18 waves of the survey, from autumn 1999 to spring 2008. When matching the data to The Oxera Arup Dataset (TOAD), the average of the responses across the two waves per year was taken to provide a single reading for each year since the dataset for the overarching study contains annual data only.

Following the input from industry representatives, described in section 3, the study team investigated the correlation between different questions and punctuality/reliability and overall journey satisfaction, to determine which questions are most affected by punctuality and, hence, how the responses to the questions may be affected by changes in punctuality.

Table 5.1 Correlations

	Punctuality/reliability (train arrives/departs on time)	Overall satisfaction with journey
Station		
Ticket-buying facilities	0.0390	0.0834
Provision of information about train times/platforms	0.2534	0.2847
Upkeep/repair of the station buildings/platforms	0.2185	0.2780
Cleanliness of the station	0.2241	0.2892
Facilities and services at the station (eg, toilets, shops, cafes, etc)	0.1167	0.1445
Attitudes/helpfulness of the station staff	0.0753	0.1268
Connections with other forms of public transport (eg, bus, tube, tram, taxi, etc)	0.0809	0.0854
Facilities for car parking	*0.0018*	0.0323
Overall station environment	0.2509	0.3292
Journey		
The frequency of the trains on that route	0.5065	0.3603
Scheduled journey time	0.5958	0.4144
Connections with other train services	0.1087	0.0788
Punctuality/reliability	1.0000	0.5129
The value for money of the ticket	0.3597	0.3234
Train		
Cleanliness of the train	0.2863	0.3993
Upkeep and repair of train (condition of seats, walls, tables, etc)	0.2800	0.3755

[19] MVA Consultancy (2007), 'Rail Passenger Demand Forecasting Research: Workstream 2—Review and Specification of Data Series', March, p. 2.7.

	Punctuality/reliability (train arrives/departs on time)	Overall satisfaction with journey
Helpfulness/attitude of onboard staff	0.0923	0.1372
Space for luggage on the train	0.0966	0.1377
Toilet facilities on the train	0.0452	0.0868
Sufficient/insufficient room for passengers to sit/stand	0.2811	0.3690
Comfort of the seats	0.2876	0.4321
Ease of being able to get on and off the train	0.3041	0.3991
Availability of catering (food/drinks) on the train	0.0414	0.0598

Note: Correlations which are not statistically significantly different from zero at the 5% level are shown in italics.
Source: Oxera analysis.

This table indicates that the responses to the questions regarding the journey, in particular, are highly correlated with performance. This may imply that adjusting the responses for the impact of performance may be preferable to adjusting the weights. However, for the reasons set out in section 4.2, this has not been possible in this study.

Oxera has investigated the impact of aggregating 'satisfied' responses with 'very satisfied' responses, and 'dissatisfied' responses with 'very dissatisfied' responses. The use of different values for the responses has also been investigated to determine whether the assumption of linearity in the data responses affects the estimated elasticities.[20] These investigations have revealed that they do not change the underlying patterns in the data. Therefore, the specification of the indices has assumed both that the responses are linear and that they can be aggregated.

This section has described the data available to create the 'x' in the aggregation rule (see Equation 4.1). The next section considers how the different parameters can be generated.

[20] The assignment of numerical responses to the survey answers is an assumption; for example, the scales 1–5 can be replaced by any other numbers, so long as the ordering of responses is maintained.

6 Parameters

As explained above in section 4.1, this study has considered two ways in which the parameters can be assigned: exogenously, or determined from the data. Section 7.1 provides details of the service quality indices, where the weights are assigned using both methods.

Oxera has reviewed a number of studies (detailed in Table 6.1) to establish the most appropriate source of data for exogenous weights.

Table 6.1 Valuation of service quality data

NPS question	PDFH v5	TRL593	University of Southampton et al.	MVA for Passenger Focus
Overall satisfaction				
Station facilities				
Ticket-buying facilities			x	
Provision of information: train times/platforms		x	x	x
Upkeep/repair of station buildings/platforms				x
Cleanliness	x			x
Facilities and services	x		x	x
Attitudes and helpfulness of the staff				x
Connections with other forms of public transport				x
Facilities for car parking			x	x
Overall environment				x
Personal security while using	x		x	x
Availability of staff				x
Station staff handling of requests				x
Train facilities				
Frequency of trains on that route		x		x
Punctuality/reliability (ie, train arriving/departing on time)				x
Scheduled length of journey (speed)		x		x
Connections with other train services				x
Value for money (for the price of the ticket)				
Upkeep and repair of the train	x			x
Provision of information during the journey	x			
Helpfulness and attitude of staff on train				
Space for luggage				x
Toilet facilities				x
Sufficient room for all passengers to sit/stand		x		x
Comfort of the seating area	x			x
Ease of access (being able to get on and off)				
Personal security while on board	x			x
Cleanliness of the inside of the train	x			x
Cleanliness of the outside of the train				x
Availability of staff				x
Management of delays (how well these were handled by the train company)	x			x

Sources: See section 2.1.2.

The following section provides a brief discussion of the studies reviewed by Oxera.

Passenger Demand Forecasting Handbook v5

The PDFH v5 considers non-service-related quality in Chapters B7 and B8. Although it does not contain values for some of the questions asked in the NPS, it provides a useful discussion in relation to measuring the impact of service quality on passenger rail demand.

The PDFH notes that different passengers will have different needs and perceptions of service quality (and hence different degrees of WTP for changes in service quality). The PDFH approach is to recommend that changes in rolling stock quality are treated as modifiers to 'in-vehicle time', rather than a direct effect (which is recommended to changes in station quality). Unlike previous versions of the PDFH, the PDFH v5 does not recommend capping the maximum impact of packages of improvements.

The Demand for Public Transport: A Practical Guide (TRL 593)

This compilation of research studies details various methods for quantifying the effects of quality of service.[21] It also provides information on WTP for service quality improvements for a number of bus- and rail-related attributes. The rail-related attributes were investigated to determine their usefulness for the service quality index.

The Effects of Station Enhancements on Rail Demand: Phases 1 and 2

Undertaken by Atkins for the PDFC, Phase 1 reviews existing research to identify 'deficiencies in scope and methodology'.[22] This study also highlights the importance of differentiating between the provision of a new facility and the upgrade of existing facilities.

Further research is carried out in Phase 2 by the University of Southampton et al.[23] This combines a literature review, analysis of the NPS and ticket sales (LENNON) data, and the use of survey data to create a station quality index for use in econometric work.

The literature review demonstrated that many of the studies over-valued station facilities owing to 'non-commitment and part-whole biases'.[24] Such bias can arise in stated-preference studies where respondents give a different response to how they would actually behave in reality. For example, they might say that they would be prepared to pay for an extra service, or more for the sum of the individual components of a package than for the package as a whole. This implies that, when using the results in the literature, care should be taken not to over-value the impact of station quality enhancements on demand.

Analysis of the NPS data suggested that enhancements at major stations could increase passenger satisfaction by up to two percentage points, while analysis of the LENNON data indicated that station enhancements at smaller stations could increase 'traffic' by 7%. The study also discovered a decay effect—whereby the demand uplift from enhancements decreased over time—and that an impact on demand of the disruption caused during the enhancements could reduce 'traffic' by 4%.

Station quality indices were calculated for 18 station enhancement schemes and subsequently used in econometric analysis. These indices were calculated using details of the enhancements made at each station, together with surveys, which were used to provide WTP data for the enhancements. It was found that the average station enhancement more than doubled the station quality index and resulted in demand uplifts of over 8%.

The study suggested that the impact of improvements to service quality on the demand for passenger rail travel could be substantial. However, it would not be practical for this current

[21] Transport Research Laboratory (2004), op. cit.

[22] Atkins (2006), 'The Effects of Station Enhancements on Rail Demand. Phase 1: Current Knowledge and Recommendations for Phase 2 Work', November 8th, p. 1.

[23] University of Southampton et al. (2008), op. cit.

[24] Ibid.

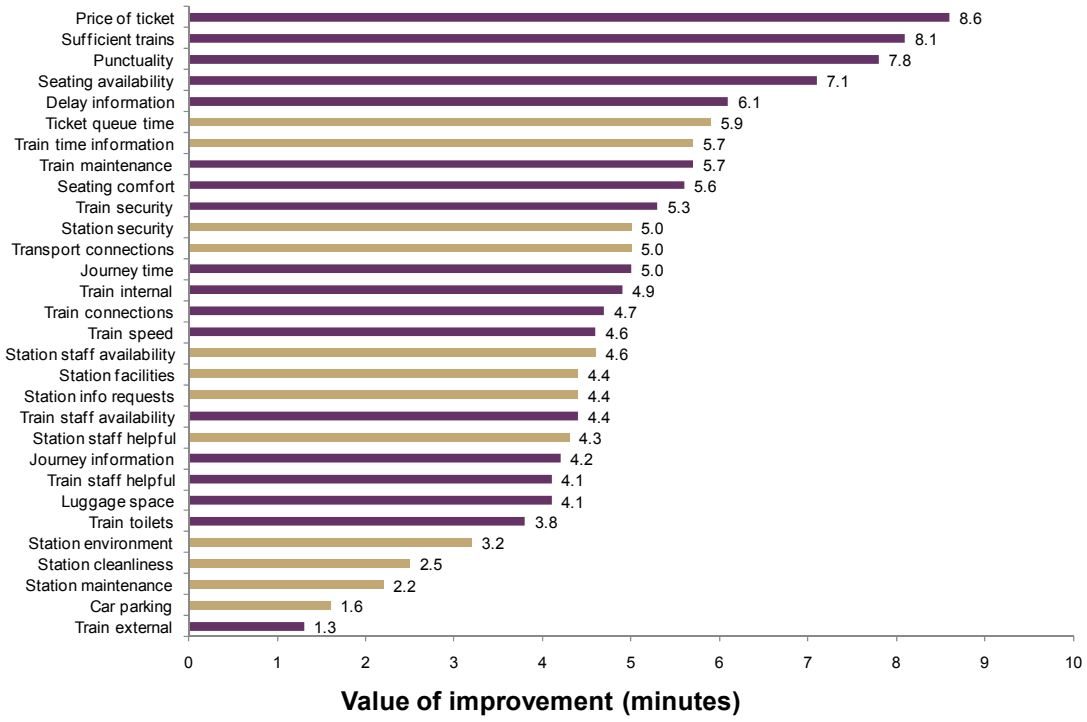
study to produce a separate station quality index for each station in the dataset, based on detailed information about the enhancements made at the station. Therefore, an alternative approach has been considered, as detailed in section 7.

Passengers' Priorities for Improvements in Rail Services

This report, produced by MVA Consultancy for Passenger Focus, aimed to identify passengers' expectations about their rail services, and the *relative* importance they attach to different attributes of their journey.[25] The report used stated-preference techniques to identify these trade-offs.

For this study, the most important aspect is the consistency with the NPS, with values available for almost every NPS question. MVA Consultancy assigned a value to each surveyed attribute *relative* to a five-minute improvement in journey time (see Figure 6.1).

Figure 6.1 Value of improvements



Source: MVA Consultancy (2007), op. cit., p. 24.

One of the criticisms often made about stated-preference analysis is that it produces accurate relativities between questions, but is often less accurate in producing absolute values.[26] In this case, this criticism is not crucial, as it is the relativities between the values that are important for producing weights (see section 7.2 for details).

These studies all provide information on passengers' WTP for service quality improvements. The most comprehensive (and consistent with the NPS) is the research undertaken by MVA Consultancy for Passenger Focus, and the values from that report have therefore been used to provide the exogenous weights for the NPS questions. This study was reviewed by Rob Sheldon, Managing Director of Accent Market Research and an adviser to the study team, to ascertain both compatibility with this particular service quality workstream, and the robustness of the analysis. This process has provided expert judgement that the use of the values from this study is robust in providing weights for the service quality indices.

[25] MVA Consultancy (2007), 'Passengers' Priorities for Improvements in Rail Services', June.

[26] See University of Southampton et al. (2008), op. cit.

In order to establish whether each method—principal components or exogenous weights—varies substantially from the other (given the data available), test indices were created at the national level. This process is detailed in section 7 below.

7 Construction of indices

There are two main steps in creating a service quality index:

– generating the variables to be weighted—ie, the 'x' from Equation 4.1;
– applying the weights to the variables.

As demonstrated in Figures 7.1–7.2 below, the correlation between train and station quality indices (irrespective of how the weights are generated) is very high—often greater than 0.9. Including both indices in an econometric model would result in problems of multi-collinearity, which could lead to increased standard errors and problems identifying the 'true' parameters. Therefore, a combined train and station service quality index was created and included in the econometric model.

The combined indices used the same source data as the separate indices—ie, weights derived from the same WTP data were used.[27]

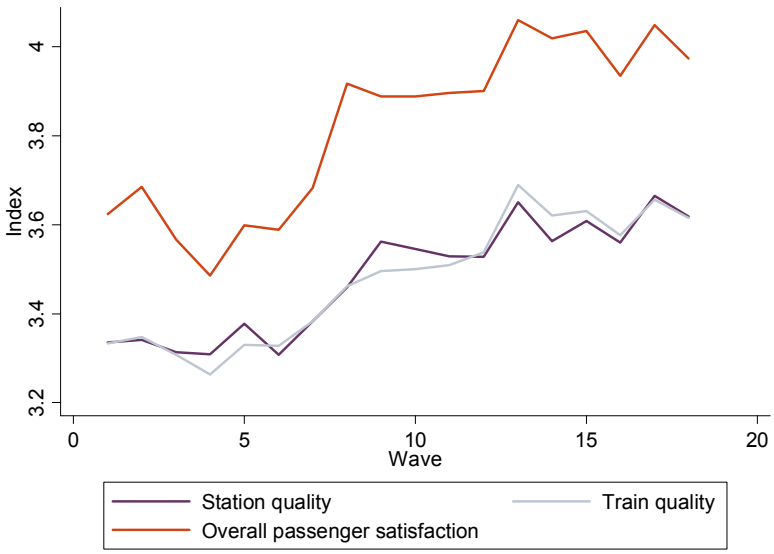
7.1 Weighting

As discussed in section 4.1, two main methods can be used to assign the weights to the data; the analysis undertaken to choose between these two methods is described below.

7.1.1 Different approaches to creating the index

As can be seen from Figure 7.1, which depicts the principal components index and the aggregated responses across all surveys in each wave to the NPS overall passenger satisfaction question, the individual indices both track overall passenger satisfaction, with the train index tracking overall passenger satisfaction more closely than the station quality index. This is not surprising given both the greater correlation between overall satisfaction and train attributes, and the greater WTP to improve train attributes, relative to station attributes.

Figure 7.1 Principal components

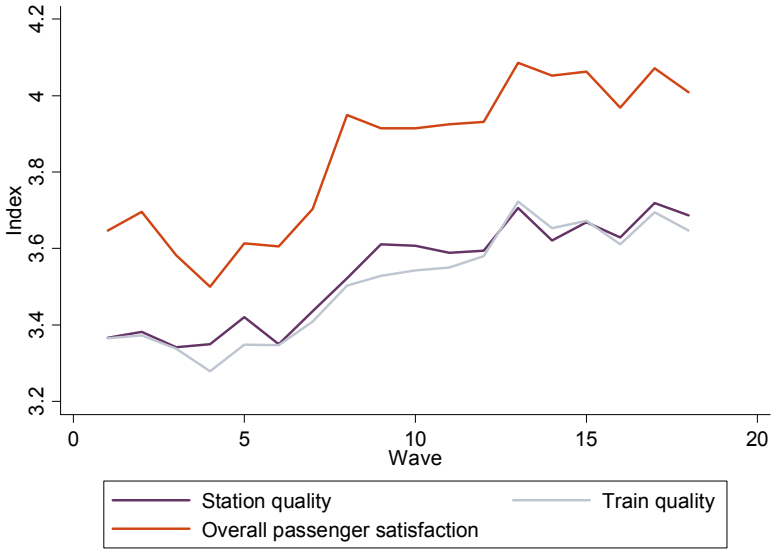


Source: Oxera analysis.

[27] MVA Consultancy (2007), 'Passengers' Priorities for Improvements in Rail Services', June.

Figure 7.2 shows the train and station quality indices generated through the use of exogenous weights; an increase in the index corresponds to an improvement in service quality.

Figure 7.2 Exogenous weights



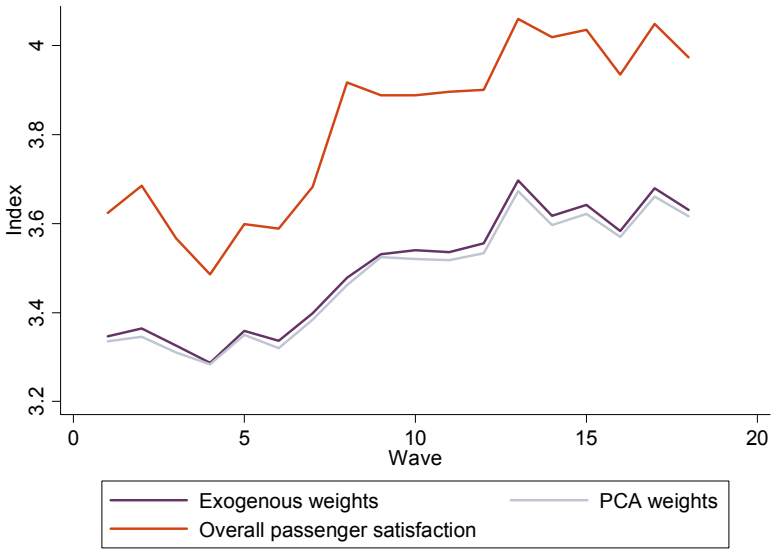
Source: Oxera analysis.

The train quality index appears to be improving at a similar rate to the station quality index, and the overall trend in all three indices is similar. The overall picture is also very similar to the indices when the weights are determined using principal components.

Comparison and summary

Figure 7.3 shows a comparison of the indices created using different methods.

Figure 7.3 Comparison between methods



Source: Oxera analysis.

As can be seen, the two different methods both closely follow the overall passenger satisfaction measure—indeed, the correlation coefficients between the three measures are in excess of 0.9.

Table 7.1 summarises the objectives of the service quality indices and the extent to which each of the two methods can satisfy these objectives.

Table 7.1 Objectives

Objective	Principal components	Exogenous weights
End result easily usable	Y	Y
End result can be forecast	Y (with caveats)	Y (with caveats)
Index needs to display sufficient variation over time and segments need to be useful to the econometric analysis	Y	Y
Methodology needs to be flexible enough to include or exclude performance and crowding	Y	Y
End result broadly reflects overall passenger satisfaction	Y	Y

Source: Oxera.

Given that the underlying preferences of passengers are, by assumption, fixed for a given point in time, and that the two types of index are offering an approximation to these preferences, the two approaches should provide similar (although in practice not identical) answers. The more straightforward interpretation of the WTP-based index has meant that it has been used in the subsequent analysis. However, given the degree of correlation between the indices, this choice does not have a material impact on the estimated elasticities from the econometric analysis.

7.2 Deriving the indices

The previous section considered how indices could be weighted; this section now details how the indices have been constructed in practice.

The raw data from the NPS cannot be used directly in creating the service quality indices because there are too many responses: therefore, the responses must be aggregated. This is done by summing the responses (on a scale of 1 to 5) for each segment and then dividing by the number of survey responses. This provides a single number for each question for each year, which can then be weighted to generate the overall index.

The weights, as previously noted , are derived from two distinct sources: the dataset (using principal components) and WTP data. The weights are derived from the WTP data by:

$$\frac{WTPi}{\sum_{t=1}^{N} WTPi} \qquad \text{Equation 7.1}$$

where WTP_i is the willingness to pay to improve aspect i and the denominator is the sum of the relevant WTP data (eg, WTP for an improvement in seat comfort added to WTP for an improvement in the cleanliness of the train, etc).

7.2.1 Adjusting weights for performance

To adjust the weights for the impact of performance, the denominator of Equation 7.1 is adjusted to exclude the WTP for improvements to performance, and the weights are then recalculated.

As WTP for improvements to performance can be assumed to be positive, the exclusion of this factor has the effect of decreasing the denominator, and in turn increases the weights attributed to the other factors (eg, train cleanliness). A similar process could be followed to remove other factors, should this be considered desirable.

This section has set out the method used to create the indices at a national level. The next section discusses the results of the indices for each of the market segments.

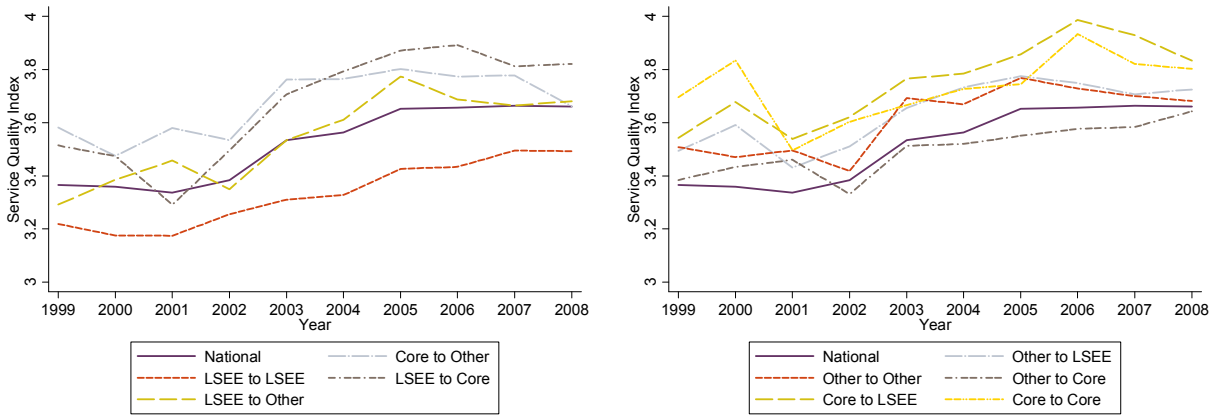
8 Segment-specific indices

The process of creating separate indices for each market segment is very similar to that used to create a national index. Due to data limitations, the indices for the 'to airports' segment were not robust and were therefore not used in the modelling.

The weights used are the same for the different market segments. While it would be preferable to use different weights to reflect local priorities, it is not possible to do this with the WTP data. Although it should be possible to produce different weights for different market segments using the principal components approach, in order to provide consistency between approaches, this has not been attempted. The differences in the indices are therefore due to differing responses to the questions in different market segments.

The matching between index and market segment has been done by using only those responses from a particular area for a market segment—ie, only those responses recorded on non-London Core cities to non-London Core cities services were included in the creation of the index for that segment—and this provides a unique index for each segment. Due to data restrictions, the indices do not differ by ticket type. Figure 8.1 shows how the service quality indices vary by segment.

Figure 8.1 Segment-level indices



Note: To facilitate comparison, the solid (purple) line ('National') represents the national service quality index.
LSEE, London, South East and East of England.
Source: Oxera analysis.

Figure 8.1 shows a number of patterns in the service quality indices for different segments: however, the general upward trend after 2001 implies that passengers' perceptions of service quality have improved since then. The differing impact of the Hatfield derailment across the market segments is interesting, with some segments (eg, Core to Core) falling substantially, while other sectors (eg, Other to Other) are hardly affected. Other features of potential interest include the much lower score of the LSEE to LSEE segment, which may reflect the degree of crowding prevalent in this market segment.

The expected sign of the variable in the econometric analysis is positive, as an increase in the quality of the service would—all else being equal—be expected to increase the demand for the service.

The next section discusses the results obtained from using the service quality indices in the econometric analysis.

9 Results and conclusions

The matched service quality index, included in the general econometric model, is set out below:

$$\text{Journeys} = \text{Journeys}_{t-1} + \text{fare} + \text{population} + \text{income} + \text{employment} + \text{prop. no car} + \text{car cost} + \text{car journey time} + \text{GJT} + \text{performance} + \text{service quality index}$$

A general-to-specific modelling approach has been adopted, whereby the model is estimated and the least statistically significant variable is removed. The model is then re-estimated, and this process continues until the remaining variables are statistically significant and are of economically meaningful signs and magnitudes.

The service quality index remains in the 'specific' model in two of the 28 ticket-type market segments: LSEE to non-London Core cities, full fare tickets; and Other to LSEE, full fare tickets.

For LSEE to non-London Core cities, the estimated elasticity of demand with respect to service quality is 1.34 after three years. This implies that, if the service quality index is increased by 1%, the demand for passenger rail travel (in this segment) would increase by 1.34%. Any change in the service quality index would be driven by changes in passengers' responses to NPS questions.

A similar interpretation applies to the Other to LSEE full fare ticket segment, where the estimated elasticity is 1.28 after three years. This implies that a 1% increase in the service quality index is likely to result in a 1.28% increase in the demand for passenger rail travel in this segment after three years. There is, however, uncertainty surrounding these estimates because of noise in the data; hence, the actual effect may be larger or smaller than these 'point estimates'.

The inclusion of the service quality index means that the demand impact of 'soft' factors (such as the cleanliness of the train) has been quantified for the first time, at the same time as the demand impacts of changes in incomes and rail fares. Although the impact has not been identified as being statistically different from zero in many cases, the testing of this reflects a substantial change from the treatment of these factors in previous demand forecasting frameworks. In particular, it seeks to ensure that the parameter estimates for other drivers of rail demand are not biased as a result of service quality factors being excluded from the analysis, as may have been the case in previous estimations.

9.1 Next steps

This report details how NPS data has been used in the econometric analysis for this study, providing a direct estimate of the demand impact of 'soft' factors, such as cleanliness, etc. Although the service quality index has only been included in the 'specific' model in two cases, the fact that it has been included in the econometric analysis improves the parameter estimates for other drivers of rail demand. The indices could be further improved by:

- updating the indices to include the latest waves of the NPS;
- incorporating weights which change across segments, although this would require substantial new stated-preference work to be undertaken;
- producing forecasts (on the basis of implicit or explicit targets from franchise agreements, or based on specific initiatives, see below);
- understanding how operational adjustments affect the recorded NPS scores. One respondent commented that they are already undertaking this research within their TOC;

– making further changes to adjust the responses for the impact of performance (eg, by adjusting the responses, rather than the weights).

Regardless of these possible improvements, the indices that have been developed may have a wider range of uses—specifically, within econometric analysis, these may include models of demand, revenue or costs, etc, while, outside of econometric analysis, the indices may be useful to monitor service quality (eg, passengers' perceptions of stations).

Given the structure of the NPS, it should be possible to derive specific indices for TOCs, franchises, or areas within a franchise. Such indices could be created using either the sources of WTP data presented in this report, or more local research to reflect more precisely local priorities. Before such indices could be included, it would be important to gain a better understanding of how NPS scores could be forecast and/or how operational changes feed through into such indices.

Park Central
40/41 Park End Street
Oxford OX1 1JD
United Kingdom

Tel: +44 (0) 1865 253 000
Fax: +44 (0) 1865 251 172

Stephanie Square Centre
Avenue Louise 65, Box 11
1050 Brussels
Belgium

Tel: +32 (0) 2 535 7878
Fax: +32 (0) 2 535 7770

Thavies Inn House
7th Floor
3/4 Holborn Circus
London EC1N 2HA
United Kingdom

Tel: +44 (0) 20 7822 2650
Fax: +44 (0) 20 7822 2651

www.oxera.com