THE MEASUREMENT OF PASSENGER PREFERENCES TOWARDS RAIL STATION AND ON-TRAIN FACILITIES

DAVID PEARMAIN

A thesis submitted in partial fulfilment of the requirements of the Council for National Academic Awards for the degree of Doctor of Philosophy

July 1993

Oxford Polytechnic

ABSTRACT

This study is concerned with the provision of passenger facilities, such as catering or information, at rail stations and on trains. It takes the premise that appraisal methods used by rail operators and planners for evaluating investment in such facilities are limited. The result can be under-investment in such facilities. It makes the case that such facilities are important devices for improving the quality of rail travel and that under-investment in facilities implies reduced demand for rail services.

It is suggested that stated preference (SP) techniques, already applied to this topic on a number of occasions, are effective methods for measuring the monetary equivalent values of passenger facilities. These values can be inserted into investment appraisal methods based on financial criteria or cost-benefit analysis. However, previous applications of SP techniques have exhibited a number of weaknesses, which have called into question the plausibility of some of the values that have been obtained.

This study reports on research on the London Underground, investigating the potential introduction of a range of improvements to passenger facilities on the Northern Line. In this original work, discrete choice SP methods were used to measure the values of these improvements in a way that will produce more plausible results, it is argued, than those derived from earlier SP applications. The research showed that appreciably lower valuations were obtained with this method.

ACKNOWLEDGEMENTS

I would like to acknowledge the financial assistance and cooperation of British Rail (InterCity) in the early stages of the study. I wish to thank London Underground Ltd and Steer Davies Gleave for their permission to use their data for my case study.

I also thank Professor Peter Jones (University of Westminster) for his critical appraisal of the work, and Mervyn Jones (Swansea University) for his support as my supervisor. Thanks are also due to the advice given by David Shilton (InterCity), Geoff Copley (MVA) and Tim Grosvenor (Accent Marketing).

DEDICATION

To Henny

THE MEASUREMENT OF PASSENGER PREFERENCES TOWARDS RAIL STATION AND ON-TRAIN FACILITIES

TABLE OF CONTENTS

Section 3: Case Study

LIST OF TABLES

LIST OF TABLES (Contd) Page

LIST OF TABLES (Contd) Page

LIST OF FIGURES

LIST OF FIGURES **Page**

1: INTRODUCTION

1.1 Introduction

This study takes the premise that an important priority for rail operators and planners is the need to understand travellers' preferences towards passenger facilities (for example, information facilities and catering services.) These are elements of a rail service that may be regarded as secondary to the main process of travelling by train but which collectively have an important influence on travellers' perceptions of the efficiency and convenience of a rail service. It is also maintained that there is under-investment in stations and rolling stock on many British inter-city and urban rail services. This is likely to have resulted in a loss of patronage to rail services, because (it is argued) station and on-train passenger facilities can offer significant benefits for rail travellers, with a consequent positive effect on demand.

This situation is seen as a direct result of the inability of established research methods to quantitatively measure the benefits which travellers derive from good quality passenger facilities. It is for this reason that the focus of the study is upon assessing approaches to the quantitative measurement of travellers' preferences towards passenger facilities and seeking to further develop current measurement techniques.

1.2 Policy Context

Transport operators require reliable procedures to assist them in assessing priorities for investment in transport infrastructure. While various decision-making approaches can and have been used (Ortuzar and Willumsen, 1991, pp9-14), the tradition among most transport operators in this country has been to seek a "substantive/rational" approach: one in which attempts are made to quantify the utilities or benefits related to alternative investment options and to model their impacts on travel demand. Under this scheme the option with the greatest perceived utility (or least disutility) and/or most desirable impact on travel demand is to be preferred, depending on the resources required to implement it.

To aid this process, and as far as is practical, these procedures should allow alternative investment options to be assessed on a common scale. In British Rail planning, financial performance, in terms of acceptable rates of return on investments, has been the chief criteria (Prideaux, 1984, p13.) Alternative approaches have been available in the form of social cost benefits, such as may be measured in terms of time savings (used in some rail projects jointly funded between rail operators and local or central government.) London Underground has often applied cost benefit analysis as well as financial criteria to the assessment of new rail schemes.

While these approaches have many critics, it is not considered within the scope of this study to argue the case for alternative decision-making procedures. Instead, it accepts that for the foreseeable future, a project's financial return or economic benefit will continue to be the predominant basis for evaluating rail investment. It is within this context that the study seeks to explore the ways in which techniques to aid investment appraisal can be improved.

A criticism of established approaches to the evaluation of rail investment is that the impact of some features of train services on the demand for rail travel (and therefore their financial performance) can be difficult to measure quantitatively, particularly on a monetary scale. The implication is that this can lead to certain aspects of rail services being starved of investment. It is in respect of station and on-train passenger facilities that these issues are explored in this study.

1.3 Principle Research Questions

The central research question which this study aims to address may be phrased as follows:

> "How best can the value to passengers of investment in station and on-train passenger facilities be measured, given the requirement to assess it on the basis of the financial returns which they will yield?"

Financial returns are dependent on the demand for rail services and travellers' willingness/ability to pay for them. The implicit aim of the research is therefore to understand the relationship between investment in passenger facilities and the demand for rail travel. Key elements to this aim are the assumptions that such a relationship actually exists and that variations in the standards of passenger facilities can be shown to have an impact on travel demand and fare pricing. These may be phrased as the following formal statements:

Statement 1

"The quality of passenger facilities at rail stations and on board trains affects travellers' perceptions of the overall quality of rail services and therefore:

(i) the benefit they derive from such services;
(ii) their willingness to use such services and
(iii) their willingness to pay for such services."

Statement 2

"The standards of rail services in the United Kingdom vary across different systems and regions, but there is room for improvement in the opinion of many travellers. Subsequently, this has a negative impact on the demand for rail services and their financial performance."

As later chapters will show, the opportunities to establish directly from market data a quantitative relationship between the quality of passenger facilities and the demand for rail services are very limited. It is more practical to infer the relationship through the impact of passenger facilities on travellers' perceptions of quality of service (as they might be measured on "attitude scales" or as categorical stated opinions.) This difficulty in measurement leads to a third statement:

Statement 3

"Though poor standards of station and on-train passenger facilities may partially reflect the limited availability of investment funds, they are also the product of limitations in the investment evaluation methods used by rail operators and related decision-makers."

Provided this criticism of existing evaluation methods can be substantiated and that any developments which promote the demand for rail services may be considered valuable, a final statement will suggest a strong case for more research in this area:

Statement 4

"Advances in market research and behavioural modelling techniques have led to recent improvements in procedures for evaluating investment in station and on-train passenger facilities, but further improvements are both desirable and attainable."

The call for further improvements makes the case for the original research reported in this study, in which certain evaluation techniques are developed in the context of a practical rail investment study.

1.4 Study Approach

The statements presented above define the structure of the study. Succeeding chapters examine the evidence to support these statements, explore the literature relating to the principal issues raised and present the results of research designed to extend the body of knowledge related to these issues.

The issues raised in the statements above are discussed in chapter two. The aim of that chapter is to establish a theoretical understanding of the relationship between passenger facilities and demand for rail services, examine evidence for the existence of such a relationship, assess the strength and quality of the relationship and identify the factors that influence it. When the first two statements are examined and substantiated, chapter two then goes on to explore the policy context and related issues raised in the third statement, concerning limitations in investment appraisal procedures. This leads on to the requirement and potential for the development of appropriate research techniques, noted in the fourth statement, to which the succeeding chapters of the study are principally addressed.

Chapter three identifies the main areas of development in the relevant research methods. It then begins an examination of the development of useful research techniques and the opportunity for further improvements. Its principal contribution is to identify, at a general level and on the basis both of theoretical and practical considerations, those techniques most applicable to the principle research question presented at the beginning of the previous section (1.3.)

Chapters four to seven focus upon those research techniques which are considered most applicable to the area of interest to this study. These belong to the family of research methods generally referred to in the transport literature as 'stated preference' (SP) techniques. They offer particular advantages over other methods for the evaluation of investment in station and on-train facilities and a number of useful applications are reported in the literature.

New and original applications of these techniques constitute the main original research contribution by this study and the aim of these four chapters is therefore to present a detailed examination of these techniques. Chapter four identifies the main types of SP techniques; chapter five discusses the issues relating to the design and application of the most widely applied SP techniques ('Conjoint Analysis'); chapter six discusses analytical methods and chapter seven reviews previous applications to the examination of station and on-train facilities.

Chapters eight, nine and ten report on a study which examined rail passenger valuations of improvements to passenger facilities on the London Underground (Northern Line.) The report of this case study has two complementary aims:

(i) to contribute to an increased understanding of rail users' priorities towards improvements in passenger facilities, particularly in terms of their willingness to pay for such improvements and, less directly, the potential impact of these improvements on the demand for rail services;

(ii) to contribute to the development of SP techniques and to assess their robustness and usefulness to the researcher, particularly in relation to the monetary evaluation of improvements to marginal service characteristics.

Chapter eleven provides a summary of the study and presents the conclusions that can be drawn from it. The main objective of this final chapter is to draw together the main points from the literature reviews and to identify the contribution of the original research presented in the three preceding chapters in relation to the existing body of knowledge. It argues the case for using the new developments in research techniques outlined in those chapters, but also attempts to identify the criticisms that may be levelled at certain aspects of the study. Finally, it suggests directions for future research, both at the level of understanding rail users' priorities towards improvements in passenger facilities and at the detailed technical level of SP survey design.

2 THE ROLE OF STATION AND ON-TRAIN PASSENGER FACILITIES: THEIR INFLUENCE ON THE DEMAND FOR RAIL TRAVEL.

2.1 Introduction

This chapter considers the first two statements in chapter one:

> Statement 1
>
> "The quality of passenger facilities at rail stations and on board trains affects travellers' perceptions of the overall quality of rail services and therefore:
>
> (i) the benefit they derive from such services;
> (ii) their willingness to use such services and
> (iii) their willingness/ability to pay for such services."
>
> Statement 2
>
> "The standards of rail services in the United Kingdom vary across different systems and regions, but there is room for improvement in the opinion of many travellers. Subsequently, this has a negative impact on the demand for rail services and their financial performance."

From a review of the literature, this chapter examines researchers' understanding of the relationship between passenger facilities and the demand for rail services. It then presents evidence of the impact of passenger facilities on rail users' perceptions of service quality and discusses the implications for their willingness to pay for and use services with improved passenger facilities. It will become apparent that the ability to obtain direct evidence of a relationship between passenger facilities and the demand for rail services, as opposed to the relationship implied by rail users' reported perceptions and attitudes, will be limited. This leads to an investigation in the next chapter of techniques that may be suitable for quantifying such a relationship.

The study is not primarily concerned with operational aspects of rail stations or rolling stock (for example, platform layouts or pedestrian flow management.) The focus instead is upon the way the provision of 'secondary' passenger facilities (eg buffets, information systems) can affect travellers' perceptions of the comfort and convenience of rail services and their propensity to pay for and use such services.

2.2 Rail Services and Their Relationship to Travellers' Perceptions of Rail Travel.

Rail services will only constitute one element of a traveller's journey between their origin and destination, but they may often represent the most important stages in terms of distance, journey time or route. In this respect, on-train facilities that provide for passengers' comfort (eg seating, toilets) and reassurance (eg announcements, signing) have an important role to play in influencing travellers' perceptions of the attractiveness of rail travel. Some on-train facilities (eg catering) can be revenue-earning and profitable in their own right.

The time spent on the train is only one part of a rail journey. There will always be a requirement for access and egress by some other mode or modes (including walking.) The ease with which travellers can transfer between other modes and rail services will also have a significant influence on travellers' perceptions of the attractiveness of rail travel. Transfer between different rail services is also important, the impact of which can vary considerably:

> "There are cross-platform or same platform interchanges with booked, reliable connections, which are only a modest deterrent to travel. At the other end of the scale, a journey across London from, say, Dover to Manchester, involves two interchanges between main line trains and Underground (metro) trains, each interchange including the use of two escalators, and for many people the purchase of a separate ticket for each stage of the journey. This is clearly a very major deterrent to travel, especially to people who do not know London well."
>
> Segal and Todd (1985, p120)

The frequency and timing of services will determine the amount of time that travellers will spend in waiting for train services to depart or, conversely, the ease with which travellers will be able to carry out different activities (such as buying a ticket) before they can board the train. As focal points of interchange to, from or between rail services, stations are therefore key elements in rail travel. It has been argued by a number of observers (Bell, 1976; Prideaux, 1984) that the standards of design and quality of service will have a significant impact on the size of the disbenefit associated with waiting time and the ease with which transfer can take place.

In examining these issues, it is useful to consider an important precept in transport research: most travel represents a disbenefit to travellers. Travel is not usually an activity to be undertaken for its own sake, as Ortuzar and Willumsen (1990, p3) suggest:

> "The demand for transport is derived, it is not an end in itself. With the possible exception of sight-seeing, people travel in order to satisfy a need (work, leisure, health) at their destination."

In this context, station and on-train facilities may be seen as devices for mitigating the disbenefit of rail travel, rather than as 'attractors' to rail services. Nevertheless, this distinction becomes blurred in such cases where the station environment becomes a destination in its own right (eg the Victoria Centre shopping development, London) or the rail journey itself becomes a leisure activity (eg the 'Orient Express'.) Despite these instances, it still remains the case that for many travellers, rail services impose a number of disbenefits which are fundamental to the nature of rail travel.

These disbenefits will include the basic cost and journey time associated with a rail service. On top of this will be added access time, egress time, waiting time, interchange and any additional costs such as car parking and other public transport fares. Therefore, the rail traveller would ideally hope for an accessible rail service which is affordable, reliable, fast, frequent, comfortable and convenient.

Within the rail carriage environment, passengers are likely to derive a benefit from being able to get a seat. The comfort and availability of seating, together with the quality of the suspension, the micro-climate (heating, air circulation, humidity) and other environmental characteristics (decor, sound proofing, cleanliness) of the carriage, are factors likely to increase in importance with the length of the rail journey being made. The same is likely to be true of toilet and buffet facilities. The latter, if provided on trains, allow the passenger greater flexibility in choosing when to have meals. Additional details such as tables and loudspeaker announcements add to an impression of a quality service. Luggage space will also be an important practical item.

Concerning the station environment, this has to be viewed in the context of rail travellers' unavoidable requirement to interchange to, from and sometimes between rail services. The elements of this inconvenience may be summarised (after Prideaux, 1984) as:

(i) an extension of journey time;

(ii) changes in the nature of the time involved, from in-vehicle travel to waiting, walking and queuing;

(iii) some physical discomfort (changes of level, changes of temperature);

(iv) additional anxiety from being in an often unfamiliar environment, sometimes without complete certainty of catching the intended rail service on time;

(v) possible 'out-of-pocket' costs, such as parking charges, and

(vi) disturbance of any activity with which a traveller may have been occupied (eg reading, sleeping) during his or her journey to the interchange.

If suitable facilities are provided to mitigate the effect of these inconvenience factors, it is reasonable to assume that the overall disbenefit which travellers associate with interchange will be reduced. If the penalty of interchange is sufficient to deter travellers from using rail, the implication is that station facilities which reduce sufficiently the disbenefits of interchange will stimulate demand for rail travel.

Despite this likely relationship between passenger facilities and the disutility of interchange, it is important not to overstate the importance of such facilities relative to other features of a rail service and other factors affecting journey decisions.

This point is made clearly in a research proposal submitted to British Rail:

> "Terminal facilities are unlikely to enter into the conscious decision process on mode choice and, in a simple ranking of the importance of different quality of service features, terminals will be given a low rating compared to more obvious attributes of service quality. This is not to say that they do not enter at all into the decision process, merely that terminals are not intrinsically attractive and people do not expect their experience of them to be one which is capable of giving pleasure. In other words, people do not have a high level of expectation about terminals."
>
> (MVA, Jan 1985c, p4)

This removed role of rail stations appears to be re-enforced as rail services, not surprisingly, tend to be the first concern of rail operators. This can lead to the perception that stations are neglected as a result:

> "...we are entitled to wonder whether the people who plan the splendid new trains think quite so hard about the not so splendid stations."
>
> (Roberts, in Height & Cresswell, 1978, p6)

Such perceived differences between train and station passenger facilities, where they may exist, are examined below.

2.3 The Importance of Station and On-Train Facilities to Rail Travellers

Given that the above section has asserted the importance of station and on-train passenger facilities as components of rail services, it is necessary to consider the evidence for this assertion. The issue being addressed here is rail travellers' perceptions of the importance of different passenger facilities and useful indicators of these qualitative assessments may be found in the results of attitudinal research.

Most attitudinal research is based on the premise that an individual's perception of each attribute of interest derives from two factors: the importance the individual attaches to a given attribute and the level of satisfaction he or she receives from the attribute. The value of the attribute to the individual is the product of his or her sensitivity to it and his or her satisfaction with it. Travellers' sensitivity to different rail service attributes (indicating their relative importance) is examined in this section. Satisfaction with rail service attributes is examined in the next section. These two dimensions of travellers' attitudes towards rail services will give an initial understanding of the values passengers place upon rail station and on-train attributes.

Table 2.1 shows the results of attitudinal research carried out by British Rail (InterCity.) The first InterCity results are an annual summary of a monthly monitoring survey which covers all the main InterCity routes throughout the United Kingdom: in response to a self-completion questionnaire handed out on their train, rail passengers rated attributes related to the rail service on a scale indicating importance. A similar method was used in the later InterCity Quality Analysis, though a number of station attributes were included. The figures from the InterCity Quality Analysis are for the month of February 1991 only.

Table 2.1 The Importance of InterCity Rail Service Attributes

InterCity Monitor (MAS, 1988)[1,2]

(All travellers)

Score (1-100)	
Ability to get a seat	95
Train running on time	93
Seating Comfort	91
Cleanliness of train	89
Fast journey time	87
Comfort of temperature in carriage	86
Cost of journey	85
Frequent train service	84
Courtesy of the staff	83
No need to change trains	79
Smoothness of ride	78
Quality of buffet/ restaurant car	76
Availability of buffet/ restaurant car	72
Loudspeaker announcements on train	66

InterCity Quality Analysis[1]
(SDG Research, 1991)

(All travellers)

Score (1-5)	
Punctuality of train	4.8
Ease of finding a seat	4.7
Journey time	4.6
Direct train	4.6
Courtesy/helpfulness of conductor	4.6
Frequency of trains	4.5
Cleanliness of toilet	4.5
Clarity of loud-speaker announcements	4.5
Seating comfort	4.4
Cleanliness of seating	4.4
Temperature in carriage	4.4
Usefulness of loud-speaker announcements	4.4
Service at stations[3]	4.4
Ease of ticket purchase	4.3
Queuing time	4.3
Cleanliness of stations	4.2
Smoothness of ride	4.2
Quality of on-train food and drink	3.9
Choice of on-train food and drink	3.8

1 Importance rating weighted on a 100 point scale:
very unimportant = 1, quite unimportant = 25,
neither important or unimportant = 50,
quite important = 75, very important = 100.

2 Importance rating weighted on a five point scale:
very unimportant = 1, quite unimportant = 2, neither important nor unimportant = 3,
quite important = 4, very important = 5.

3 A number of station attributes were included in this survey which do not have comparable attributes in the 1988 survey. These are highlighted in bold.

There are a number of differences in the ordering of the attributes which may reflect to a limited extent the differences in the way travellers' responses were interpreted in the analysis. As the Table footnotes indicate, different weights were used to obtain average ratings across the whole samples, which is likely to explain some of the differences between the surveys.

Despite this difference, the variation in the overall ranking of attributes could suggest that perceptions of the relative importance of service attributes can vary over time. Comparisons between the 1988 InterCity monitor results with the previous year (MAS, 1988, p45), for which the method of analysis was the same, suggested that the order could vary marginally. As already stated, perceptions of importance are only one element of travellers' attitudes (sensitivity.) The other element, perceptions of quality (satisfaction), is likely to interact to some degree with perceptions of importance. Where the quality of an attribute has declined, this may heighten perceptions of the importance of that attribute, because the need for improvement is more apparent. The relationship of perceptions of attributes' importance to perceptions of quality is explored at the end of the next section. Given that the InterCity Quality Analysis (SDG Research, 1991) uses data collected in the winter, it might be expected that some attributes could reflect some seasonal effect.

An issue which limits the sensitivity of the attitude scales used in these studies is the fact that all the average scores lie within a very narrow range. Fluctuations in the rankings of attributes across the two surveys are therefore not likely to be as dramatic as the re-ordering might imply, given that the scores only vary by a few points.

Despite the limitations, the figures in Table 2.1 provide a useful indication of the relative importance of some InterCity rail service and station attributes. Over all travellers, the ability to get a seat is valued highly in both surveys. In the InterCity Monitor, the comfort of the seating is also highly placed, together with cleanliness. In the InterCity Quality Analysis, these items are placed noticeably lower in the order, with operational characteristics such as journey time, train frequency and a direct service placed more highly. In contrast, staff courtesy has been placed more highly than in the InterCity monitor. The attributes related to stations, while placed higher than some train attributes, still come some way done the list given by the InterCity Quality Analysis, though the average score lies between "quite important" and "very important". This would indicate that stations are perceived by InterCity rail travellers to be an important element of the journey, but secondary to many train service characteristics.

The above examination of travellers' perceptions of the importance of different rail service attributes gives a useful indication of the relative importance of passenger facilities. Nevertheless, the methods that were used here to measure travellers' assessments of each attribute are limited. They are dependent on the use of arbitrarily defined response scales which, though they have meaning to the respondents, require a number of assumptions about their mathematical properties (ie they can be interpreted as interval scales) to enable aggregate analysis. As already seen, the assumptions about the weighting of the scales at the analysis stage is likely to alter the interpretation of the results.

In later chapters, more sophisticated market research techniques are examined and developed to explore the way in which the relative importance of passenger facilities can be quantified on meaningful scales (eg equivalent monetary values) and related directly to travel behaviour. These other techniques are themselves subject to certain limitations, but have the potential to provide a deeper understanding of travellers' perception and valuations of on-train and station attributes.

Findings from two surveys of travellers' attitudes to passenger facilities undertaken for London Underground are shown in Table 2.2. In the first, a five point scale was used, similar to those used in the InterCity surveys, but the information was summarised simply as the proportion of people identifying a facility as "quite important" or "very important". In the second, respondents allocated a limited budget across a range of possible improvements, from which cardinal rankings could be inferred. The two surveys are not strictly comparable, because the first investigates the intrinsic importance of attributes, whereas the second evaluates the importance of improvements to attributes and is therefore related to perceptions of current levels of service. Nevertheless, both serve to provide some indication of the priorities of passengers towards different features of the Underground.

These figures suggest some similarities with InterCity travellers: comfort and cleanliness attributes related to the carriage environment are high on the list. Nevertheless, there are also noticeable differences: seating comfort and crowding (which may be taken to relate to seat availability) are some way down the list of priorities compared to InterCity users. The comfort of seating attribute is likely to have a lower value, because of the much shorter average journeys on the Underground. The importance of crowding will of course vary with time of day, and since the surveys reported here, a large increase in the number of travellers on the Underground may well have heightened travellers' perceptions of the importance of reducing crowding. From the 1984 survey in Table 2.2, information and ticketing arrangements are the main concern at stations, though seating is also important.

Comparisons between the Underground and InterCity surveys can only be limited, because some attributes were defined differently and the Underground surveys did not include main journey attributes such as fare or train frequency (though the Transecon study of 1985 does show a small improvement in journey time to be less important than a number of qualitative service improvements.) Despite these limitations, it is intended that the above discussion has illustrated that many passenger facilities at stations and on trains are perceived to be important elements of rail services.

Table 2.2 The importance of London Underground Service Attributes

Steer, Davies & Gleave Ltd (1984d)
(All travellers)

	% Important/ very important[1]
Train information on platform[3]	95
Well ventilated carriage	94
Clean carriage	91
General information in booking hall	90
Well regulated heating	86
Ticket machines	83
Seats in platform areas	76
Smooth ride	73
Toilets	72
Uncrowded carriage	68
Comfortable seats	63
Train information in booking hall	60
Small shops in booking hall	54
Quiet carriage	53
Seats in booking hall	13
Vending machines on platforms	12

Transecon (1985)[1]
(All Travellers)

	Cardinal Rankings[2]
Improved train cleanliness	2.81
Improved heating and ventilation	2.63
More/better announcements/information	2.13
Smoother ride	1.91
Two minute faster journey time	1.85
Lower noise level	1.80
More comfortable seats	1.42
More handrails	1.40
More standing room	1.23
Opening windows	0.84
Better lighting	0.63
Painted exteriors	0.53
Carpets/coloured floors	0.44
Larger windows	0.39

1 Respondents rated importance on a five-point scale ranging from very unimportant to very important. The figures in this column indicate the proportion identifying each facility as important or very important.

2 Cardinal Rankings were established by passengers allocating a nominal sum of £20 across the attribute improvements

3 Station attributes are highlighted in bold.

2.4 Travellers' Perceptions of the Rail Passenger Facilities

An indication of the relative importance of passenger facilities has been obtained from the sources examined in the above section. It is now useful to compare passengers' satisfaction with these facilities as an indication of how the quality of passenger facilities as provided by rail operators matches the importance placed upon them. Such a comparison will show the extent to which rail operators, seeking to prioritise the expenditure of limited budgets on passenger facilities, have successfully responded to passenger preferences.

Using the same sources as Table 2.1, Table 2.3 shows travellers' assessments of the quality of service provided by the attributes examined previously. The scales used to measure respondents' satisfaction with each attribute, though weighted and summarised in the same way as before, cannot be interpreted on the same metric as the scales measuring importance. To assume that the average scores were comparable would be to assume that the semantic categories occupied the same position (eg "very important" = "very good"), which cannot be proved and would appear unlikely. Despite this, an association between the importance and quality of the attributes can be made by comparing the rankings of attributes. This comparison is included in Table 2.3.

The majority of the most important attributes appear to be ranked as highly in terms of perceived quality. The main exceptions in the InterCity Monitor are train cleanliness, train frequency and journey cost, which all come lower down the list. In the InterCity Quality Analysis, the clarity of loudspeaker announcements and the cleanliness of toilets are ranked very low compared to their importance ranking.

Table 2.3 Travellers' Perceptions of the Quality of Service Delivered by InterCity On-Train and Station Attributes

InterCity Monitor (MAS, 1988)[1,2]

(All travellers)

	Score 1-100	(Rank)[3]
Ability to get a seat	73	(1)
Comfort of temperature in carriage	72	(6)
Fast journey time	69	(5)
Seating Comfort	68	(3)
Train running on time	66	(2)
Courtesy of the staff	64	(9)
No need to change trains	63	(10)
Cleanliness of train	61	(4)
Availability of buffet/ restaurant car	61	(13)
Smoothness of ride	54	(11)
Frequent train service	51	(8)
Loudspeaker announcements on train	46	(14)
Quality of buffet/ restaurant car	42	(12)
Cost of journey	40	(7)

InterCity Quality Analysis[1]
(SDG Research, 1991)
(All travellers)

	Score 1-5	(Rank)
Ease of finding a seat	4.1	(2)
Punctuality of train	3.9	(1)
Direct train	3.9	(3)
Courtesy/helpfulness of conductor	3.9	(3)
Journey time	3.8	(3)
Frequency of trains	3.8	(6)
Ease of ticket purchase	3.8	(14)
Seating comfort	3.7	(9)
Cleanliness of seating	3.7	(9)
Queuing time	3.7	(14)
Temperature in carriage	3.6	(9)
Service at stations	3.6	(9)
Usefulness of loud-speaker announcements	3.5	(9)
Cleanliness of stations	3.4	(16)
Smoothness of ride	3.4	(16)
Quality of on-train food and drink	3.4	(19)
Clarity of loud-speaker announcements	3.2	(7)
Choice of on-train food and drink	3.2	(21)
Cleanliness of toilet	3.1	(7)

1 Quality rating weighted on a 100 point scale: very good = 1, good = 25, neither good or poor = 50, quite good = 75, very poor = 100.

2 Quality rating weighted on a five point scale: very poor = 1, quite poor = 2, neither good nor poor = 3, quite good = 4, very poor = 5.

3 Importance rankings from table 2.1 are given in brackets.

Another way of comparing satisfaction ratings with importance ratings is to compare the difference between the two scores for each attribute, relative to the difference between the two average scores over all the attributes. Table 2.4 shows that the average importance score from the InterCity Monitor ratings was 83 and that the average satisfaction rating was 59. This allows a way of relating to two scales to one another: the importance score has to be factored by 0.71 to be scaled onto the satisfaction score. This factor, when applied to each individual importance score, produces a predicted satisfaction score. Where there is no difference between the predicted score and the actual satisfaction rating, the difference between the importance and the satisfaction obtained from the attribute is no different than the average. Where the difference is positive, the attribute is better than average; where it is negative, the attribute is worse than average. The same method is applied to the InterCity Quality Analysis.

Table 2.4 confirms the poor performance of some of the attributes ranked low for satisfaction in comparison to the importance rankings and highlights more clearly some of the better performances, such as the availability of a buffet car, carriage temperature and the ease of ticket purchase.

A similar comparison of importance and satisfaction ratings for London Underground users, using the same sources as Table 2.3, is not possible. The satisfaction rating questions from the Steer, Davies and Gleave (1984) study were not compatible with those used in the importance ratings, while the Transecon (1985) study did not examine perceptions of quality. Table 2.5 summarises the satisfaction ratings that were obtained from the 1984 Underground study. In this question, a number of train service operational characteristics, such as reliability and frequency, were added. It would appear that the train operational characteristics are reasonably satisfactory, but passenger facilities and the general quality of environment are rated very poorly.

Table 2.4 Comparison of Travellers' Perceived Importance and Quality of InterCity Station and On-train facilities

InterCity Monitor (MAS, 1988)

	Importance	Quality	Expected	Difference
Cost of journey	85	40	61	-21
Quality of buffet/restaurant car	76	42	54	-12
Frequent train service	84	51	60	-9
Cleanliness of train	89	61	63	-2
Smoothness of ride	78	54	56	-2
Loudspeaker announcements on train	66	46	47	-1
Train running on time	93	66	66	0
Seating comfort	91	68	65	3
Courtesy of staff	83	64	59	5
Ability to get a seat	95	73	68	5
Fast journey time	87	69	62	7
No need to change trains	79	63	56	7
Availability of buffet/restaurant car	72	61	51	10
Comfort of temperature in carriage	86	72	61	11
Average	83	59	59	0

InterCity Quality Analysis (SDG Research, 1991)

	Importance	Quality	Expected	Difference
Cleanliness of toilet	4.5	3.1	3.7	-0.6
Clarity of loudspeaker announcements	4.5	3.2	3.7	-0.5
Usefulness of announcements	4.4	3.5	3.6	-0.1
Smoothness of ride	4.2	3.4	3.5	-0.1
Cleanliness of stations	4.2	3.4	3.5	-0.1
Punctuality of train	4.8	3.9	3.9	0.0
Choice of on-train food and drink	3.8	3.1	3.1	0.0
Temperature in carriage	4.4	3.6	3.6	0.0
Service at stations	4.4	3.6	3.6	0.0
Journey time	4.6	3.8	3.8	0.0
Seating comfort	4.4	3.7	3.6	0.1
Cleanliness of seating	4.4	3.7	3.6	0.1
Frequency of trains	4.5	3.8	3.7	0.1
Direct train	4.6	3.9	3.8	0.1
Courtesy/helpfulness of conductor	4.6	3.9	3.8	0.1
Queuing time	4.3	3.7	3.5	0.2
Quality of on-train food and drink	3.9	3.4	3.2	0.2
Ease of finding a seat	4.7	4.1	3.9	0.2
Ease of ticket purchase	4.3	3.8	3.5	0.3
Average	4.4	3.6	3.6	0.0

Table 2.5 Travellers' Perceptions of the Quality of Service Delivered at London Underground Stations and Trains

Steer, Davies & Gleave Ltd (1984d)
(All travellers)

	% Good/ very Good
Ease of finding platform	85
Reliability	64
Frequency	62
Value for money	57
General cleanliness	42
Train information on platform	41
Helpfulness of staff	39
Appearance of platforms	35
Smoothness of ride	30
Appearance of booking hall	27
Noise levels	22

2.5 A Conceptual Model of the Relationship Between the Features of a Rail Service and Travel Demand

To understand and quantify the relationship between station and on-train facilities and the demand for rail travel, the researcher needs to consider a theoretical framework that enables him or her to establish the way in which the disbenefits associated with the components of a rail journey influence the travel demand process. The remaining sections of this chapter consider this issue and the literature relating to it.

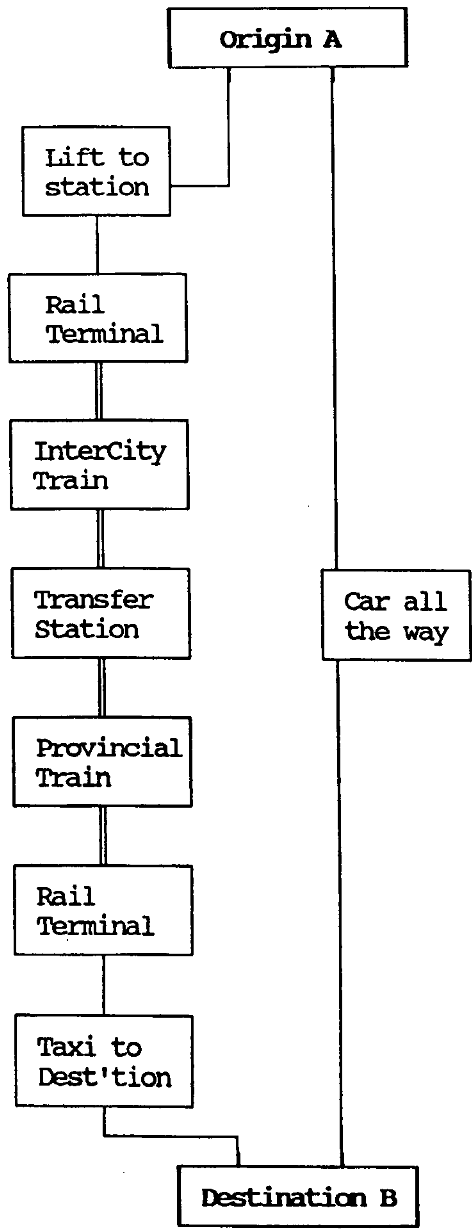
Where travellers have a choice between alternative ways of making a journey (mode, route, timing) or simply between making a journey or not making it, it is reasonable to assume that they will choose what they perceive to be the most attractive alternative. The most widely applied assumption in travel demand modelling is that individuals attach some 'utility' to an alternative, which represents the degree of satisfaction or benefit which they derive from it. As travel is invariably a necessary but inconvenient activity, the term 'generalised cost' is perhaps more appropriate, as it implies a measure of disutility. Such a view suggests that the individual chooses the option with the least disutility. As described here, this view of the travel decision process is simplistic. It is discussed more thoroughly in the next chapter.

Figure 2.1 illustrates a travel choice situation which might present itself to an individual. Here, the traveller has a choice between two modes: car and train. The principal costs associated with use of the car are the time it would take to travel from A to B and journey-specific monetary costs such as petrol and parking charges. In contrast, the train journey not only presents monetary costs (the fare) and in-train travel time, but also a series of other journey elements:

(i) access/egress time;

(ii) waiting/walking/queuing time;

Figure 2.1 A Travel Choice Between Car and Rail



(iii) interchange at the origin and destination stations (inter-modal transfer such as car to train) and once during the train journey itself (intra-modal transfer between connecting services;)

(iv) a downgrade in the quality of the rolling stock from an InterCity service in the first stage to a Provincial service in the second.

(v) a possible downgrade in the quality of the station at the rail/rail interchange point, compared to the origin terminus and the destination station.

The traveller attaches different degrees of importance to each journey element. Within the framework of a generalised cost model (see above) these elements combine, according to the weighting given to them by the traveller, to create the overall disutility associated with each of the travel options. Of course, the example in Figure 2.1 is a fairly extreme one, to illustrate how rail journeys can comprise many complex elements when compared to alternatives. Many travellers will only use one rail service in a journey and perhaps perceive a number of difficulties relating to the car journey that have not been identified. Examples may include access/egress problems (car availability; parking problems) and restrictions on activities possible while travelling.

Though many of the elements of the travel options described above can be measured on standard metric scales (money, time), it is possible that individuals' perceptions of the values of these elements will differ from the actual value. For example, Wilson (1983, p310) observed that train travellers tended to over-predict the length of experienced journey times by a proportion of some 8%, while walk times of less than two minutes duration were often not reported. This threshold on the perception of time is also noted by Wardman (1988, p80.) It is therefore important to recognise that in obtaining information on travellers' journey experience, there is likely to be some divergence between their reported journey characteristics and the real values as they might be measured by objective means.

Divergence in travellers' perceptions of time duration should not be confused with the different weighting they may place upon different elements of travel time. A wide range of travel studies show how the economic value of time spent in waiting at stations is considerably higher than for in-vehicle travel time, often by a factor of two or more (MVA et al, 1987.) There is also evidence that the mere fact of having to interchange, regardless of time spent in walking, queuing or waiting, imposes a sizeable cost, which is seen to vary widely according to the scale of the journeys involved. Bell (1976, p16) quotes equivalent penalties ranging from 2.5 to 9.5 minutes for urban interchange, while Segal and Todd (1985, p120) quote values for InterCity ranging from 10 to 40 minutes. These latter high values are to some extent inflated by British Rail's practice of not weighting wait time differently from in-vehicle time in their forecasting models.

There are likely to be a number of qualitative items implicit in figure 2.1: a modal bias incorporating perceptions of comfort-of-ride, reliability, safety and on-train facilities; interchange penalties reflecting a basic resistance to changing modes/train services and the perceived quality of the station designs and facilities. If such qualitative factors are included in generalised cost models, quantitative equivalent values (for example money) can be established. By measuring the value of each qualitative item on one generalised cost metric, the researcher establishes a common base for evaluating the relative importance of different attributes as influences on travel demand. The feasibility and accuracy of measuring the value of such qualitative items in generalised cost terms is of course a critical issue and one which this study has aimed to address.

2.6 The Costs of Rail Travel

Travel modelling techniques, as suggested, allow researchers to estimate the approximate magnitude of the generalised costs (or utility components) of rail travel: the relative merits of alternative modelling techniques for achieving this are discussed in the next chapter. In models of travel behaviour, factors influencing the magnitude of rail travel costs may be included as variables or, usually more appropriately, as criteria for segmenting the travelling population.

It is suggested from the literature (MVA et al, 1987) that factors affecting the costs of the different elements of rail journeys may be considered to include:

(i) The 'scale' of the journey being made - as the length of the journey increases, so the magnitudes of the costs and times will also increase. This will effect travellers' valuations of improvements to journey characteristics. For example, a traveller may value a 10 minute saving on a 30 minute rail journey more highly than a 10 minute saving on a two hour journey, because the impact on the generalised cost will be proportionally greater.

(ii) The nature of the access/egress modes - travellers' perceptions of the attractiveness of rail services will vary according to the access and egress modes that are available. Rail travellers arriving by car, for example, are able to optimise their arrival times at stations (traffic levels permitting), while bus users are dependent on the scheduling and reliability of their access mode. The greater the uncertainty with which a traveller believes he or she will be able to transfer to and from rail services with sufficient time, the higher the disbenefit of using rail. Uncertainty of this nature can lead to travellers allowing considerable time between arriving at a station and boarding their train when it arrives. Facilities that reduce the disutility of waiting time (eg buffets, shops, seating) are therefore particularly useful for these travellers.

There is also some influence on travel behaviour, as already inferred, from the way the 'image' of a rail service compares with that of the access/egress modes. For bus users, travelling by train may be regarded as a 'step up' in quality. For car users, rail may be generally regarded as basically an inferior mode, but this is outweighed by other factors, such as faster journey times. This distinction is somewhat generalised, but illustrates the way perceptions of rail services are influenced by its position relative to other components of the rail journey and alternative travel options.

(iii) The purpose of the journey - journey purpose is recognised across many studies of travel behaviour as a major factor influencing travellers' priorities. In the context of rail travel, journey purpose is likely to have some influence on the ease with which travellers may use rail services. For example, commuters are not likely to be burdened by luggage and will be using stations and services that are familiar to them. The opposite will be experienced by many non-work travellers. On the other hand, some commuters will be travelling within narrow time constraints, whereas some travellers on non-work journeys will have greater flexibility.

(iv) Travellers' personal characteristics - travellers' attitudes to rail travel will reflect a number of different characteristics. An obvious factor is the presence of a mobility handicap, which may require certain facilities such as lifts and room for a wheelchair. A traveller's confidence in using rail services will be influenced by whether he or she is travelling alone or in a group, or with children. Socio-economic characteristics are also likely to be associated with different weights applied to components of travel costs. Sex, age, employment status and particularly income are some of the factors which have served to distinguish different patterns of travel behaviour in numerous transport studies.

(v) The quality of the station and on-train facilities - this reintroduces the topic of principal interest to the study. Examples include: the provision of information facilities to inform travellers of the identity, location and timing of their trains; movement aids (eg escalators, lifts, signing, automatic doors); waiting facilities at stations (eg seating, waiting rooms); buffet and retail facilities at stations (eg seating, waiting rooms); buffet and retail facilities to provide refreshments and help to reduce the disutility of waiting and travel time.

Table 2.6 gives some examples of generalised cost and time values for components of a rail journey. These summarise the disbenefits travellers associate with the various elements of travel time and the high penalties placed on the need to interchange.

Table 2.6 Generalised Cost Values for Main Elements of Rail Journeys

	Walking time (per minute of in-vehicle time)	
Intra-urban services	0.6 - 2.3	(MVA et al, 1987)
	Waiting time	
Intra-urban services	0.7 - 2.5	(MVA et al, 1987)
	Value of in-vehicle time (pence per minute)	
InterCity services	2.0 - 5.1 pence/minute	(MVA et al, 1987)
Intra-urban services	2.4 - 4.0 pence/minute	(MVA et al, 1987)
	Interchange (per minute of in-vehicle time)	
Intra-urban services	2.5	(Bates & Roberts, 1986)
Intra-urban services	7.6	(Wilson and Bell, 1984)
Intra-urban services	2.5 - 9.5	(Bell, 1976)
InterCity	10 - 40[1]	(Segal and Todd, 1985)

1 Variation from a same/cross platform interchange to transfer between stations

2.7 Passenger Facilities and the Demand for Rail Services: Summary

The above sections suggested that passenger facilities have a significant influence on travellers' perceptions of the overall quality of rail services. It has found some support for this view in a discussion of how facilities might reduce the disutility of travel time and interchange, and in surveys indicating the importance of such facilities relative to other journey characteristics. Passenger facilities must clearly be regarded as 'secondary' factors in the travel decision process, but nevertheless have some contribution to make via their influence on perceptions of overall service quality.

Surveys of passenger satisfaction with service attributes also suggest that the quality of some facilities is regarded as poor, including some that are regarded as important (eg general cleanliness, cleanliness of toilet facilities and catering.) There appears therefore to be some opportunity for improving current passenger facilities generally, though in practical planning terms individual services and rail lines need to be examined in detail to determine the area of improvement required.

Later in this study (chapter seven), a review of research into the value of passenger facilities considers generalised cost and time values for improving them. Although technical problems are identified in the way some of these values are derived, it will be seen that even some of the more conservative estimates suggest benefits (or reduced disbenefits) from improved facilities equivalent to fare changes in the region of 4% to 6% for on-train attributes and 2% to 8% for station attributes.Considered within the framework of the conceptual model of section 2.5 and the magnitude of the journey costs identified in Table 2.6, this would suggest that any noticeable reduction in disbenefits which passenger facilities might bring about could have some importance in relation to the demand for rail services. For example, if British Rail demand elasticities ranging from 0.4 to 1.1 (Harris, in Harris and Godward, 1992, p13) are applied to these fare values, there is the suggestion that the demand for rail services could increase in the region of 1.6% to 7.2% in response to improved on-train facilities and from 0.8% to 8.8% for improved station facilities.

As regards the costs of providing good passenger facilities, these must be regarded as moderate, though as a proportion of the total cost of providing a station or train service, they can vary considerably. Simple halts may cost as little as £150,000 or less to construct, compared to £12 million for a new local train service with associated stations (Preston, in Harris and Godward, 1992, pp74-75.) The standard of passenger facilities inevitably improves with the size of stations and range of train services, reflecting the numbers of travellers that will benefit and the cost of the facilities in relation to overall expenses. Nevertheless, with potential traffic increases of the magnitude cited above, adequate expenditure on suitable facilities at any level of station or train service should be considered important.

Given the perception among many rail travellers that the quality of some facilities could be improved (section 2.3) and the relatively small capital costs associated with such improvements, the possible impacts on the demand for rail services suggest the importance of providing good quality on-train and station facilities. The requirement for operators is that the methods used to measure the benefits of such improvements should be adequate to the task, so enabling them to assess the value of investing in better passenger facilities. This is the primary aim of subsequent chapters: to consider the issues involved in the measurement of benefits from passenger facilities and to contribute to the development of robust research methods for carrying out such a task.

2.8 The Policy Context: Evaluation Procedures

The third and fourth statements provided in chapter one related poor quality passenger facilities to the weaknesses of investment evaluation methods and the measurement techniques available to rail planners and researchers. These statements are repeated here:

> Statement 3
>
> "Though poor standards of station and on-train passenger facilities may partially reflect the limited availability of investment funds, they are also the product of limitations in the investment evaluation methods used by rail operators and related decision-makers."
>
> Statement 4
>
> "Advances in market research and behavioural modelling techniques have led to recent improvements in procedures for evaluating investment in station and on-train passenger facilities, but further improvements are both desirable and attainable."

It has been suggested that the requirement by rail operators that the value of the returns on rail investment should be defined in wholly financial terms is likely to be biased against those items for which monetary values cannot be easily measured. As secondary elements of rail services, passenger facilities are included in this. It has been accepted that in attempting to measure the value of rail passenger facilities to travellers, the aim must be to obtain these in terms of values. This is because financial criteria and monetary valuations are likely to remain the basis for assessing nearly all rail investment for some considerable time. A brief discussion of the limitations imposed by financial investment criteria and other approaches to investment appraisal is made below.

Transport policy makers require a thorough knowledge of the factors which influence travel demand and a quantitative measure of their relative importance. They need this for two inter-related reasons:

(i) to evaluate priorities among different policy elements and

(ii) to establish accurate forecasts of behaviour in response to alternative policies.

The subject of the present study falls primarily into the first category, but all of the families of research approaches considered in the next section attempt to respond to both these requirements. The consequence of these motivations for understanding travel demand is the need to be able to model travel decision processes in such a way that the inputs (stimuli and constraints) and subsequent outputs (perceived benefits and behaviour) can be quantified. An important consideration is the criteria which operators should use to assess investment options and therefore in what units such measurements should be made.

A relevant debate currently dominating transport planning concerns approaches to the evaluation of transport systems not only in terms of their operational and economic performance, but also their wider impacts in social and environmental terms. The main area of contention relates to the dominance of quantifiable variables (eg economic value of travel time savings, financial returns) over qualitative items (eg community severance, visual intrusion) in established appraisal methods.

For rail operators, financial considerations often dominate; for road planners, cost benefits associated mainly with journey time reductions are the main influences. This has led to growing pressure to pursue a more balanced approach in two ways:

(i) to evaluate rail and road projects on the same basis;

(i) to give more representation to social and environmental impacts of transport investment.

Where competing road and rail schemes have been assessed on the same cost benefit criteria, rail schemes have often made a better case than the road, despite failing to meet current financial requirements (Prideaux, 1984.) Concerning the second point above, support for this has led to additional requirements for large transport appraisal projects to widen the scope of their analysis and consider a number of environmental impacts (eg the formal requirements of the Department of Transport's Manual of Environmental Appraisal.) Despite this, financial and cost criteria continue to dominate.

A range of solutions are being offered. At one end, this involves the complete replacement of current appraisal criteria with "objective-led" approaches, in which the aim of a transport project must be to meet certain minimum standards of quality, before financial considerations take effect (Buchan, in Goodwin, 1991, pp3-6.) A less radical approach is to modify cost benefit approaches to include some monetary values for environmental and other non-transport attributes, an approach recently recommend by the Department of Transport (1992.) Despite its limitations, cost benefit analysis is supported by many as a workable framework for evaluation, given that some advances in market and social research methods, of which stated preference (SP) methods are foremost, suggest that the range of attributes that may be valued in terms of monetary units can be increased (Rendell Planning & EAG, 1992.)

The issues in this debate are repeated within a smaller scope with regard to the evaluation of investment in rail passenger facilities. It has often proved difficult to derive monetary measures of the value of such facilities to passengers or to measure their impact on the demand for rail travel. This can lead to their being neglected while investment is channelled to service attributes whose benefits are more quantifiable in relation to fare levels and travel demand.

Policies akin to an "objectives-led" approach have sometimes attempted to ensure adequate standards of passenger facilities, usually encouraged and paid for by a third party. Examples include the construction of the Bury and Altrincham bus/rail interchanges in preparation for the Manchester Light Rail scheme (Evans, in Height and Cresswell, 1978, p32) and Birmingham International InterCity station, for which high quality decor and facilities were partially funded by the Airport authorities and the National Exhibition Centre (Hamilton, 1976, p117.)

Despite such examples as these, the provision of passenger facilities is dependent upon decisions made according to financial appraisals, to which their contribution is usually small. In the same way that proponents of modified cost benefit analysis extend valuations to include environmental attributes through the application of new research methods, it may be argued that passenger facilities can be given proper appraisal with similar modifications to financial appraisal methods. The approach will be to determine the relative values of facilities against primary journey characteristics such as fare and journey time, and by inference suggest their contribution to the financial performance of rail services. The next step is therefore to consider which research methods may be most appropriate for understanding the way travellers respond to different levels of rail service and in particular will provide a way of measuring the values of passenger facilities.

3: QUANTIFYING TRAVEL BEHAVIOUR

3.1 Introduction

This chapter examines current theory relating to how individuals make travel decisions, with a view to understanding the position of transport service characteristics in relation to all other factors likely to influence travel behaviour. A basic conceptual model of behaviour is presented to identify which of those processes determining travel choices are of principal interest to the study. Following this, the chapter conducts a general overview of the major developments in approaches to the understanding of travel behaviour. It is written with a view to their relevance to the present study which, in attempting to obtain quantitative measurements of passenger valuations towards station and on-train facilities, represents a very specific area of travel behaviour research.

3.2 A Conceptual Model of Travel Behaviour

A review of the literature (Willumsen and Ortuzar, 1991; Jones et al, 1983; Golob and Golob, 1982) demonstrates that a wide range of methods have been developed for the purpose of understanding travel behaviour. Most of these were developed over the last 20 years, in response to the widely perceived inadequacies of earlier approaches. The rather simplistic and sometimes mechanistic portrayal of the travel decision process in early studies was seen by a growing number of researchers to be unrealistic. In addition, the highly aggregate nature of the data that was usually used was perceived as too coarse for the complexity of the processes being modelled.

The numerous methods developed over the last two decades may be seen to reflect two main trends in transport research:

(i) A growing awareness that travel behaviour is the product of wider issues than simply the 'engineering' and other operational variables, such as travel time, service frequency and cost;

(ii) The travel decision process for an individual traveller is complex and unique, so that the way transport and other factors combine to influence behaviour cannot be properly captured by techniques which analyse such behaviour at an aggregate level.

Allied to these changes in the perception of transport analysts has been the greater involvement of researchers from academic fields which extend beyond the 'traditional' areas of transport, namely engineering and operations management. These new fields have included applied psychology, econometrics and market research.

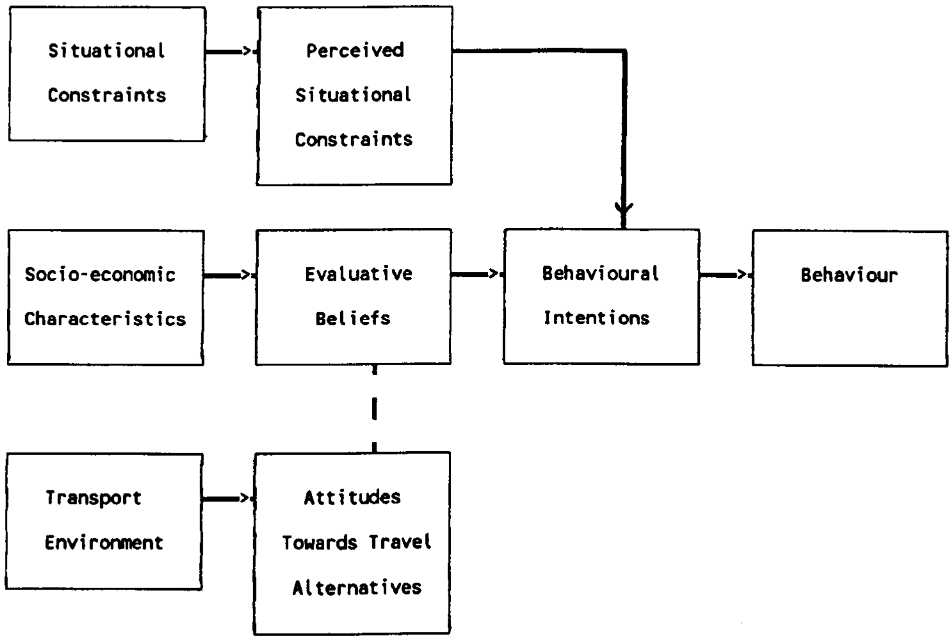
Before commencing a detailed examination of the development of travel modelling techniques, it is worth considering a general paradigm of travel behaviour against which the positioning and contribution of different techniques can be determined. A comprehensive conceptual model of the process by which an individual chooses to travel is provided by Kroes and Sheldon (1985, p310; based on Golob, 1980.) This is illustrated in Figure 3.1.

The variables included in the model of Figure 3.1 can be summarised as follows:

Observable Variables

(i) Situational constraints - factors related to an individual's personal needs (access to facilities; availability of substitutes), his or her interactions with others (joint activities with family/ household; sharing of transport) and geographical/temporal constraints on movement;

Figure 3.1 A Conceptual Model of the Travel Decision Process (Kroes and Sheldon, 1985, p310)



(ii) socio-economic characteristics - features of an individual that will give some indication of the constraints and personal preferences that will influence his or her travel choice;

(iii) transport environment - the travel opportunities open to an individual, determined by the quality of public transport services and the availability of private modes (a car particularly);

(iv) behaviour - an individual's response to the interaction of the three groups of variables outlined above, shaped also by his or her subjective assessment of those factors and his or her personal preferences and prejudices.

Unobservable Variables

(i) Perceived situational constraints - constraints on choice coloured and sometimes distorted by an individual's perception of them;

(ii) evaluative beliefs - personal assessments of the relative importance of different aspects of the transport environment, derived principally from experience and the information to which he or she has been exposed (eg advertising; peers);

(iii) attitudes towards travel alternatives - an individual's opinion regarding the relative quality of the travel options which he or she faces;

(iv) behavioural intentions - an individual's reasoned travel choice, which may be modified when he or she makes the actual journey (due to confrontation with factors previously ignored or misunderstood.)

The different approaches to understanding travel behaviour which are discussed below can to some extent be distinguished by relating them in terms of this paradigm. More complex versions may be considered, such as those which incorporate 'feed-back' from behaviour to travellers' experiences, attitudes and perceptions, but this model may be seen to identify the key elements. For the purposes of this study, techniques that can quantitatively relate elements of the transport environment (ie a rail service) to actual travel behaviour (or, at best, statements of intended behaviour) will be of primary use. Such approaches will suggest the trade-off value of passenger facilities to primary service characteristics (fare, travel time) and ideally the direct relationship between facilities and travel behaviour.

3.3 The Development of Travel Demand Modelling

Models of travel behaviour developed up to the mid-seventies were seen to be unreliable, chiefly for the following two reasons:

(i) they were based on paradigms of the travel decision process that had little in common with the actual decision process employed by travellers;

(ii) they used data on observed travel behaviour that was an analysed at a highly aggregate level.

A paradigm of travel behaviour used by many models is one that is usually referred to as the "four-stage" process. These stages are: trip generation (the amount of travel stimulated by different attractors such as work places and shopping centres), trip distribution (the subsequent allocation of trips across space), modal split (the way in which such trips are made) and trip assignment (the allocation of trips to transport networks.) Researchers do not usually suppose that individual travel decisions develop through this process and many have attempted modifications of this basic approach to reflect this. Nevertheless, the classic four-stage paradigm endures in many contemporary studies, as a practical basis for forecasting and policy evaluation.

The weaknesses of earlier models gave rise to a number of new approaches. As the four-stage model is still the regarded as the most practical basic paradigm for many transport studies, some these new approaches have tended to be used as enhancements to such models, rather than as complete replacements. They can provide inputs to an aggregate forecasting model which relate more directly to individuals' travel behaviour and to non-engineering variables.

Hartgen (1981) suggests that the techniques developed in response to the limitations of so-called value-of-time studies can be broadly classified into two groups. They can be distinguished by their 'behavioural' content: that is, the extent to which they attempt to take account of the personal and psychological variables influencing travel choices.

The first group of approaches employ advanced mathematical and statistical methods to create models based on economic concepts of 'utility' in travel. They relate actual behaviour (and more recently, intended behaviour or stated preferences) to the observable variables of Figure 3.1, through the use of appropriate mathematical functions. The coefficients estimated in the resulting models indicate the importance which travellers attach to each influencing factor entered into the calibration.

The second group of techniques take greater account of the psychological variables and the direct relationship between personal/situational constraints and behaviour. Travel behaviour is modelled as a direct function of such variables: the first group of techniques can only infer such a relationship. These truly 'behavioural' approaches therefore have the potential for a greater understanding of behaviour because they attempt to take account of more of the factors involved. On the other hand, they encounter greater problems in the measurement of such factors.

A comprehensive review of those travel analysis techniques which represent new approaches to the understanding of travel behaviour is given by Golob and Golob (1982), who identify them in terms of five broad categories:

(i) Disaggregate Choice models - travel behaviour replicated by stochastic mathematical models fitted to observed travel choices of individuals (revealed preferences) (Luce, 1959; McFadden, 1987, 1986);

(ii) Segmentation techniques - travel behaviour related to groups within the study population defined in terms of explanatory variables such as individuals' personal characteristics, situational constraints, 'transport environment', attitudes and preferences (Dobson, Heathington and Barnaby, Tye in Hensher and Stopher, 1979);

(iii) Attitude measures - travellers' subjective beliefs and preferences related to the travel options open to them and their observed behaviour (Dobson, 1975; Golob and Recker, 1979);

(iv) Activity studies - travel behaviour analysed as a function of individuals' and households' activity patterns, interaction between individuals (especially within households) and space-time constraints (Brog and Erl, 1981; Jones et al 1983);

(v) Controlled experiments - explanatory factors measured in terms of travellers' responses to experimental situations. In practice, such experiments are predominantly 'stated preference' approaches in which respondents state their preferences or behavioural intentions in response to hypothetical choice alternatives (Kocur et al, 1981; Journal of Transport Economics and Policy, Jan 1988; Bradley and Kroes, 1990a.)

The first two categories, choice models and segmentation, correspond to Hartgen's first grouping of new approaches, namely those techniques which relate observed behaviour (revealed preferences) to individuals' personal and transport characteristics. The third and fourth categories, comprising attitude and activity studies, fall into Hartgen's second grouping of more 'behavioural' approaches. The final category of controlled experiments has a place in both groupings. Experiments, whether carried out in real transport environments or 'laboratory' conditions (ie hypothetical options inducing statements of preference or behavioural intentions) provide data which appear most suitable for analysis with choice models. Nevertheless, unlike data representing observations of existing behaviour, data from experiments allows the researcher to examine more precisely the decision process behind travel behaviour. This greater potential for a 'behavioural' understanding of travel demand allows controlled experiments to be closely associated with Hartgen's second grouping of methods.

Most research applications straddle a number of these categories. Developments in disaggregate choice modelling have now made segmentation and stated preference experimental data common elements of travel studies. Studies which attempt to model behaviour directly on the basis of attitude measurement remain relatively rare in the field of transport research while activity-based travel studies represent a largely separate development. Despite this, some researchers have been able to define relationships between these two latter categories and the former, 'mainstream' approaches, seeking in some instances to develop practical hybrid research tools. Examples include attitudinal measurement with stated preference experiments (Kroes and Sheldon, 1985) and stated preference experiments with activity simulators (Ampt et al 1987.)

The aim of this chapter is not to exhaustively investigate every area of development in understanding travel behaviour. It has already been stated in the previous chapter that the present study is concerned with very specific elements in the traveller's decision-making process. The objective is to identify those developments in transport research methods which appear most relevant to the measurement of travellers' valuations towards rail station and on-train facilities. From the five categories taken from Golob and Golob (1982), the two categories of attitudinal measures and activity studies do not present themselves as appropriate methods for measuring the values of passenger facilities, despite the important contributions they have made in other areas of transport research.

Attitudinal methods consider beliefs and judgements concerning the transport environment: they are not suited to examining trade-offs between transport attributes, though they have been applied in some studies to forecasting travel behaviour (Timmermans et al, 1982.) Equally, activity studies, which have represented a particularly significant departure from established transport research methods, are focused on the constraints and inter-personal relationships that influence travel decisions, not on travel responses to detailed transport service characteristics (though they can provide a context for their assessment - see Bradley, 1988.)

The remaining categories of discrete choice models, segmentation and experiments are directly relevant to the measurement of the value of passenger facilities, because they are methods which aim to model travel behaviour (or behavioural intentions) as a direct function of the transport environment, of which passenger facilities are a small part.

3.4 Choice Models

As the earliest of the "new" approaches, discrete choice models offered some improvements over previous models. These were:

(i) a more plausible paradigm of travel choice behaviour (random utility theory) and;

(ii) application to less aggregate data, in which the separate choices of individuals were represented.

In such models, the choice process being examined is usually presented as an individual's choice of one item from among a group of specific alternatives. The choice is then considered to be influenced by a number of attributes attached to each alternative, representing their attractiveness to each individual. These attributes combine to determine the "utility" which an individual attaches to each alternative. This utility may be defined as the amount of satisfaction which an individual derives from each alternative. As mentioned in the previous chapter, the common parallel of utility in transport research is "generalised cost", in which each alternative presents some cost to the individual, perhaps in terms of travel time or financial expense.

The basis for constructing a model of choices requires assumptions to be made about two fundamental issues:

(i) the way in which the attributes combine to determine the utility attached to each alternative (eg will a simple linear additive relationship operate or are more complex forms applicable?)

(ii) the way in which individuals choose between alternatives (ie do they consistently choose the alternative with the most utility or are other factors involved, which may be random or systematic.)

As the variables indicating choices in such models are categorical in nature (eg a person travels by car or does not), statistical procedures designed to model such choices have to interpret them in terms of probabilities. This allows a continuous scale for the dependent choice variables against which variations in the independent variables (eg time or cost) can be compared. For example, a group of travellers may each be making a discrete choice between car and bus. The researcher might then assume that, for this group, the average mode shares at different combinations of journey time and cost can be taken to represent the probabilities of a traveller choosing car or bus. It must also be assumed that all the dependent variables in the model can explain each individual's choice. The same approach may be used for each individual traveller, but multiple observations will be required (eg time series or stated preference data.) Data of this quality is rarely available, so that probabilistic discrete choice models are most applicable to the analysis of group travel behaviour. The data must be aggregated into segments for which travellers' choice behaviour is thought to be fairly homogeneous. In this way the data that is used remains aggregate, though not to the extent as in earlier types of model.

A discussion of the main types of discrete choice models and their distinguishing characteristics is given in appendix 1. If discrete choice models of this kind are to be of use in evaluation and forecasting, it has to be assumed that the preference weightings attached to the different independent variables in each alternative are stable. That is, an individual will attach similar weights in other choice situations of the type examined by the model. This requirement is of course appropriate to any modelling approach, but critics of discrete choice models draw attention to the wide range of conditions that must be satisfied, and frequently are not:

"(a) The choices concerned must be real ones;

(b) where choices exist, they must be fully perceived and there must be grounds for believing that individuals are aware of the alternatives available;

(c) the effects of all variables thought likely to affect choices must be explicitly considered;

(d) there must be perceptible differences between alternatives;

(e) the variables considered relevant must not be too closely correlated;

(f) the variables affecting choice must show a fair amount of variation in the sample;

(g) the sample under consideration must be assumed similar with respect to factors included explicitly in the analysis;

(h) the sample analysed must show a reasonable proportion choosing each of the relevant options;

(i) the number of choices explained by the analysis must be high."

(Atkins, 1984, p4)

Failure to satisfy even some of these requirements can raise doubts concerning the validity of choice models, but examination of the above items suggests that the issues revolve principally around the quality of data from which they are estimated, calibrated and applied, rather than the models themselves. For example, Brög and Erl (1981, p41) show that in a study of mode choice in West Berlin, only 12% of travellers perceived a competitive alternative mode to their present one. A model fitted to all travellers in such a situation would therefore fail to meet items (a) and (b), and would be weak in relation to (h) and (i.)

Despite the difficulties, the appropriate application of other new approaches suggest that the conditions set out above can be met. In the West Berlin study by Brög and Erl (1981) mentioned above, this demonstrated how appropriate segmentation and activity-based approaches can be used to greatly enhance the quality of the model. Travellers were classified by whether a real alternative could be identified, whether constraints made the alternative feasible, the information on alternatives they possessed and whether they perceived an alternative to be available. This process identified of the small proportion of travellers who actually considered an alternative mode to be available.

A large number of studies now exist in which experimental data (almost without exception stated preference data) are used instead of actual travel data. With this type of data, items (c) to (f) can easily be satisfied, though there is of course the problem that the data represents choices between hypothetical alternatives rather than real choices. In response to this, recent approaches which estimate discrete choice models from a combination of real and stated preference choice data appear to provide a way of including real-life constraints on choices, and therefore producing more reliable forecasts. These issues are covered in detail in the next chapter.

It has been stated that discrete choice travel models are dependent upon aggregating the data set into groups and developing separate models for each group. Such groups are defined by characteristics thought to be important influences on travel behaviour and likely to engender a degree of homogeneity between group members. This reduces the degree of variation in travel behaviour, so improving the fit of the models. The segments are usually defined on an "a priori" basis (categories defined before model estimation.) The most common segmentations of travellers take the form of mode used, trip purpose, car ownership and income. In the context of rail passenger facilities, a classifications were suggested in the previous chapter which would apply to any models examining preferences between different levels of train and station service. Examples include journey length and access/egress modes used.

While segmentation improves the accuracy of choice models, the process of partial disaggregation does not strictly improve their "behavioural" content. For a truly disaggregate model of behaviour:

> "The most correct approach is to predict the behaviour of each individual based on the individual values of the explanatory variables and aggregate predictions over individuals. Because of lack of data on the levels of explanatory variables for each individual in the population, the 'market segmentation' technique has evolved as a workable compromise between the aggregate and completely disaggregate forecasting procedures."
>
> (Tye, in Hensher and Stopher, 1979, p265)

Multiple behavioural observations for each individual are generally only available with experimental data (though some time series data may sometimes be suitable.) Even with this data, accurate models of individual behaviour are difficult to develop, because of the practical limits on the number of observations that can be obtained.

In the context of general travel behaviour outlined in Figure 3.1, it is possible that each stage of the process can provide a basis for segmentation. Some of the information is purely factual, such as the socio-economic characteristics of individuals and the transport environment (resulting in the common use of such variables for segmentation.) Other items are of a qualitative nature, such as perceptions of situational constraints, evaluative beliefs about the transport environment and travellers' attitudes towards travel alternatives. Thus, models can be developed in which travel behaviour is directly related to all of the characteristics identified as influences on travel choices. Nevertheless, it is important to recognise that the relationship between behaviour and the other variables is only inferred and not directly measured. A more practical limitation will be the need to have sufficient numbers of people in each segment, so that the more segmentations that are used, the larger the samples that will be required.

To summarise, the segmentation of data in the modelling process greatly improves the accuracy and relevance of choice models and provides proxy measures of relationships between behaviour and a whole range of factors that extend beyond simply transport characteristics. What it does not do is provide direct quantitative measures of the process by which these various factors exert causal influences over travel behaviour. This has led to a new approach to quantifying travel behaviour, namely stated preference experiments. These aim to link consumer judgements and behavioural intentions to the transport environment, allowing a direct relationship to be observed between transport environment factors and behaviour, though only by assuming that judgements and stated intentions will correspond to actual behaviour.

3.5 Stated Preference (SP) Experiments

This final group of new approaches defined by Golob and Golob (1982) covers a range of techniques designed to observe individuals' preference and travel decisions as direct responses to changes in the transport environment under experimental conditions. Individuals are offered choices between hypothetical options and asked to state their preferred options or their likely behavioural response. The most commonly applied stated preference (SP) techniques construct the hypothetical options on the basis of an experimental design. This allows the attributes of each option to be varied independently from one another. Other approaches adopt alternative means of constructing hypothetical alternatives, but with the same objective of isolating the individual effects of different transport (and other) attributes. In this way, variations in the stated preferences obtained from individuals can be directly related to variations in the values of the attributes, free of the problems of multi-collinearity and dominance between variables.

A variety of analytical techniques can be used with SP data, but most recent developments of SP techniques have aimed to produce data suitable for analysis with discrete choice models. As previously suggested, criticisms of discrete choice modelling approaches focus on the limitations of the data, but most of these can be eliminated through the use of experimental data. As real-life experiments are usually very difficult and expensive to conduct, SP experiments offer an affordable substitute.

The main short-comings of SP techniques are that the data produced are either travellers' judgmental evaluations of competing options or statements of intended behaviour. Coupled with these problems is the sensitivity of individuals' responses to the precise way in which the hypothetical responses are presented. On these counts, SP data cannot automatically be given the status of real (revealed preference) data.

Despite these limitations, a range of validation studies carried out in a number of fields (Louviere et al, 1981; Levin et al, 1983; Bates, 1986; MVA et al, 1987; Wardman, 1988) suggest that well-designed SP experiments are capable of producing accurate forecasts of behaviour. Moreover, they may represent the most realistic way of investigating likely behavioural responses to options that do not currently exist, and therefore cannot be examined in real circumstances.

The next section (3.6), argues that SP techniques appear the most appropriate approaches to use in the evaluation of station facilities, coupled with the use of discrete choice models and appropriate segmentation whenever the data collection resources allow it. For this reason, SP techniques are examined more thoroughly in the next chapter.

3.6 Quantifying Travel Behaviour: Analytical Techniques Relevant to the Evaluation of Rail Passenger Facilities

This chapter has aimed to give an overview of the principle strands of development in quantifying travel behaviour. It has sought to identify the main groups of analytical techniques that measure components of travel disutility and so are relevant to the valuation of passenger facilities, and identify those elements of the travel decision process to which they are addressed. This has been done by relating the subject area of each group of techniques to the conceptual model of travel behaviour established in section 3.2.

In assessing the suitability of these analytical techniques for the evaluation of rail passenger facilities, it is necessary to consider how such facilities are represented within such a model. They are obviously an element of the transport environment but have been identified as relatively small components of transport services. Nevertheless, it can be argued that they have a direct, if small relation to travel behaviour, subject to travellers' evaluative beliefs about them and their attitudes towards them.

Models based on aggregate data are clearly inappropriate for the fine measurements necessary for quantifying the impacts of secondary items such a passenger facilities. Disaggregate choice models can attempt to infer a direct measure of the relationship between transport facilities and travel behaviour, but there are problems relating to the sensitivity of such models, given limitations in the data that are available from observations of actual travel choices (revealed preferences.) Segmentation can enhance such models, but data which allows them to measure the influence of secondary transport characteristics on travel choices is very difficult to obtain. This is because sufficiently subtle variations in secondary transport facilities, coupled with sufficiently sensitive variations in travel behaviour (which at best can only be measured in terms of a few discrete mode choice or other categories), are not available.

With respect to the limitations of models derived from observations of real travel behaviour, techniques which explore travellers' preferences and intended travel behaviour directly in relation to specific elements of the transport environment offer a richer source of data. In stated preference (SP) experiments, the relationship between selected explanatory variables and travellers' choices is more easily defined.

Given the problems of setting up real transport experiments, in which it would be costly and impractical to carry out variations in say, fares, travel times and the quality of passenger facilities, stated preference (SP) techniques appear the more practical approach. Transport attributes can be varied systematically across a number of hypothetical situations represented by verbal and/or visual descriptions, and sensitive variations in travellers' expressed judgements or stated intentions obtained. It may be hoped that the confidence derived from validation studies with primary transport characteristics (fare, time) may also be warranted for SP studies of secondary characteristics, for which empirical validation against revealed preference information is not likely to be practical (given the likelihood that the impacts of such items will be swamped by primary characteristics and other factors.)

Most studies carried out for British Rail and London Underground, relating to the valuation of passenger facilities, have used SP techniques, because they represent such a sensitive approach to measuring secondary items of this nature. It is worth noting that such techniques are now recommended by the Department of Transport (1991, p4), specifically those referred to as 'conjoint analysis' techniques. In recognition of the potential of SP techniques in this area, the next two chapters are devoted to their appraisal, which will include an examination of the advantages offered by discrete choice models and segmentation approaches, where appropriate.

4: STATED PREFERENCE TECHNIQUES

4.1 Introduction

The previous chapter identified the position of stated preference (SP) techniques relative to the travel decision process and suggested that they are particularly suited to isolating the values of secondary items whose impacts are normally difficult to detect with other methods. This chapter now examines SP techniques in more detail.

The most widely used SP techniques are those often referred to as 'conjoint analysis' (Journal of Transport Economics and Policy, 1988, vol 22) and most of chapter five is devoted to this family of techniques. Other approaches which may also be classed as SP techniques are 'transfer pricing' (Bonsall, 1983; 1985), 'Johnson's trade-off technique' (Johnson, 1974) and the 'priority evaluator method' (Copley and Bates, 1988; Hoinville, 1977.) These less commonly applied techniques are examined, but for reasons which will be identified below, the focus of this and the next chapter is upon conjoint analysis techniques.

4.2 Stated Preference (SP) Techniques: A definition

The term "stated preference technique" refers to any survey method which uses peoples' statements on how they would respond to hypothetical situations as a basis for quantitatively measuring their priorities towards different factors and modelling their resulting behaviour. A very simple transport research example of one type of SP exercise is shown in Figure 4.1. This is a hypothetical situation that might be offered to an individual traveller, with the sort of response he or she might give.

Figure 4.1 Simple Example of An SP Choice Situation

Travel Situation A		Travel Situation B		Which Do you Prefer?	
Cost	Travel Time	Cost	Travel Time	Prefer A	Prefer B
£0.40	30 mins	£1.00	15 mins		

In this example, it may be seen that a time saving of 15 minutes (the difference between option 1 and option 2) is valued more highly than a fare increase of 60 pence. If a range of such situations were offered to the individual (respondent), and repeated for a number of individuals, the researcher could build up a picture of how possible time savings could off-set cost increases (or vice versa.) This information shows the dependency of mode choice on time and fare values. The researcher therefore not only has information on the relative importance of time and cost but may also construct statistical models of mode choice behaviour with time and cost as the independent variables.

Whichever SP technique the researcher chooses, he or she will have to decide which attributes should be presented to respondents, and which levels of those attributes should be included. Some attributes may be of interest to the researcher but not to respondents, and vice versa. The omission of a key attribute from the exercise might undermine the realism of the choices that respondents make. The use of inappropriate attribute levels may also undermine the realism of the exercise as well as reducing the usefulness of the information obtained (eg the attribute 'parking cost', if presented at low levels throughout an SP exercise, may not give any information on how car users might switch to public transport in response to higher pricing policies.)

To summarise, the principal features of SP techniques may be identified as follows:

(i) they involve the presentation to individuals ('respondents') of hypothetical but plausible choice situations;

(ii) these situations can represent changes to the levels of a single attribute, the opportunity to define an optimum combination of attribute levels or, most commonly, choices between pre-defined 'packages' of attributes which represent a definable 'product' or service;

(iii) the definitions of the attributes in each option and the attribute levels used are pre-specified by the researcher according to some systematic procedure (eg the use of an experimental design) and usually presented in the context of respondents' present situations;

(iv) the respondents **state their preferences** towards each option by either ranking them in order of importance, rating them on a scale indicating strength of preference, allocating a constrained budget or simply choosing the most preferred option from a pair or group of options.

A review of the literature shows that four main groups of SP techniques have been developed in recent years. These are, in order of progressive complexity and sophistication:

(i) Transfer pricing - the systematic alteration of a single attribute (in transport research, most often cost or travel time) against which a respondent is asked to give some rating of utility or to identify a change in choice behaviour (eg a switch from car to rail travel.)

(ii) Johnson's trade-off technique - a step-by-step comparison of all pair-wise combinations of attributes and attribute levels that have been selected by the researcher; the respondent in each pair is asked to identify the preferred attribute change.

(iii) The priority evaluator method - an exercise in which a respondent is shown all the attributes and attribute levels in a single display and told that to improve each attribute from a particular position, he or she must allocate an imaginary constrained budget. The budget ensures that the quality of some attributes will have to be foregone to allow improvements in others, thus allowing trade-offs between the attributes to be observed.

(iv) Conjoint analysis - a family of survey methods in which respondents are presented with pre-defined packages of attributes, often constructed on the basis of an experimental design. Respondents are asked to prioritise between the different packages, by ranking, rating or simply choosing between pairs or sub-groups.

It will be useful at this stage to give a fuller description of each of the different types of SP techniques mentioned above.

4.2.1 Transfer Pricing and Trade-off Techniques

As previously stated, these techniques present attributes in isolation from one another. Their main attraction is in the simplicity of the response task (though many repetitions may be required), but their main disadvantage is that they will be offering respondents less realistic choice situations, because real life choice behaviour is rarely related to individual attributes in isolation from one another. The counter to this is that the evaluation of attributes individually reduces the risk of 'information overload' for respondents.

Transfer Pricing

This technique attempts to measure directly the utility associated with changing from one level of an attribute to another. In transport research this is usually the monetary cost of an item, but it can be any continuous variable. Response scales may take the form of utility ratings or the identification of a threshold (eg "At a journey time increase of 15 minutes I would switch from car to rail.") An example of a transfer price question is shown in Figure 4.2.

Figure 4.2 Example of a Transfer Price Question

By how much would your present fare have to increase before you chose not to travel by train?

(tick appropriate box)

50p	£1.00	£1.50	£2.00	More

Transfer pricing has the advantage of simplicity compared to techniques which require respondents to assess a number of attributes simultaneously analysis and in certain applications may be adequate. Respondents might easily imagine a situation where all attributes remain constant except for one, such as public transport fares. The problem with systematically altering one attribute in isolation is that it is likely to cause the respondent to give undue attention to it. There is a danger that some 'policy response bias' will occur as respondents realise that the upper limit of, say, a price or travel time increase is wholly dependent on their response.

Johnson's Trade Off Technique

This technique involves a series of paired comparisons of attributes. An example of such trade off situations is given in Figure 4.3. This approach requires the respondents to analyse a large number of paired comparisons in order that enough information may be produced for the analysis. The main advantage is that each paired comparison requires a simple decision on the part of the respondent. The principle disadvantage is that like transfer pricing respondents may be called upon to give too much attention to certain items and to consider them out of context with other attributes.

Trade-off has not been widely applied in transport research in the United Kingdom, though it has enjoyed widespread use in the United States, in more general applications to market research. It has obtained its most sophisticated development in the form of computer-based interviewing methods (Sawtooth Software, 1989.)

Figure 4.3 Example of Johnson's Trade-off Technique

For each pair, please state which item you prefer...

1.	A 50 pence reduction in fare	v	A 10 minute reduction in travel time
2.	A 70 pence reduction in fare	v	A 10 minute reduction in travel time
3.	A 50 pence reduction in fare	v	A 15 minute reduction in travel time
4.	A 70 pence reduction in fare	v	A 15 minute reduction in travel time

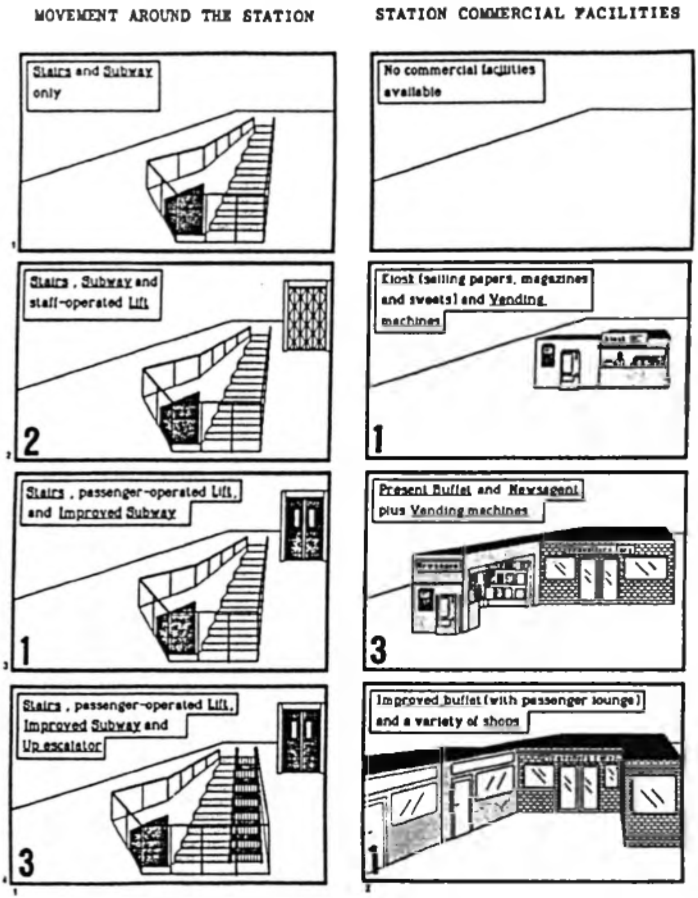
etc...

4.2.2 The Priority Evaluator Method

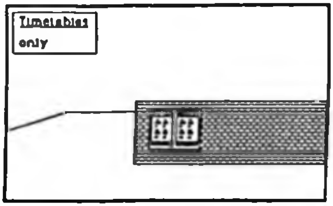
In contrast to the above techniques, the priority evaluator method is a survey instrument that requires individuals to construct optimum combinations of attributes within the limits of a 'budget'. By this process, the respondents infer their relative preference values towards each variable without having to explicitly consider such values (which in many cases may be impossible for respondents to consciously do). The researcher can control precisely the budget constraints to which the choices are subject. This technique dates back to the early seventies (Hoinville and Berthoud, 1970a-c) but has only received limited application in transport research, almost exclusively in recent years to the valuation of rail passenger facilities (MVA, 1985; Copley and Bates, 1988.) For this reason, it remains a technique of particular relevance to this study, though it will be argued that conjoint analysis techniques should be the preferred approach.

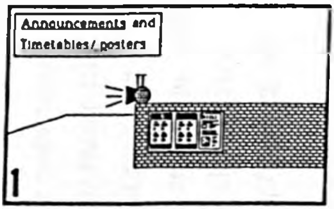
An example of a priority evaluator board is shown in Figure 4.4. To summarise, the basic task of the respondent is to weigh up the value of each attainable variable level, consider the attractiveness of the price attached to each variable level in relation to this value and define the most attractive combination of the options within the limited budget available. In addition, most applications have considered a prior exercise of respondents identifying their present situation on the game board an essential element of the priority evaluator method.

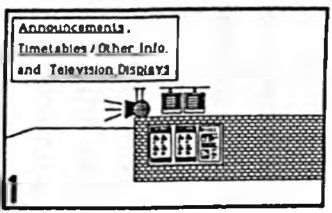
Figure 4.4 Example of a Priority Evaluator Game Board

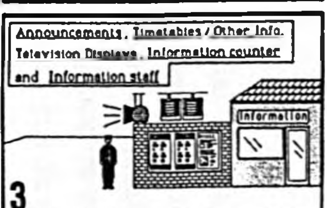


INFORMATION AT THE STATION









JOURNEY TIME

Journey time:

Journey time:
3

Journey time:
3

Journey time: Present time
3

Journey time:
3

Journey time:
3

Journey time:
3

This SP technique approaches the valuation of attributes from the direction of respondents constructing their optimum package of attributes. This is in strong contrast to conjoint analysis and most other trade-off/pricing techniques which present pre-constructed packages to respondents. While such packages represent the way individuals are normally presented with choices in real life, the requirement of the priority evaluator for respondents to construct their own packages may avoid any possible appearance of manipulation by the interviewer which could otherwise influence responses. The only constraints operating on an individual are a 'budget' and a 'pricing structure', which determine by how much they can improve an attribute.

Details vary across applications of the priority evaluator, but the following essential features apply:

(i) attributes of interest to the researcher are presented on scales of two or more levels (qualitative variables are represented by suitable verbal or graphical descriptions);

((i) each attribute level is assigned a value, which represents the "cost" of improving the attribute from the preceding level to that current level (this value may be in recognisable units, such as money, or defined in terms of an arbitrary point system);

(iii) the respondent is given a 'budget' sufficient for him or her to construct a range of combinations, but constraining the choice so that improvements to some attributes can only be made by foregoing improvements to others.

The definition of the budget constraints can be made in more than one way. A universal budget (or range of budgets) may be set for all respondents; alternatively, each respondent may be asked to identify what they perceive to be the combination of attribute levels most representative of the current situation. From this perceived current situation, small budgets may be offered to the respondent to make marginal improvements.

An apparent attraction of the priority evaluator method over alternative SP techniques is that it involves less repetitive choice tasks for respondents. It appears to have the potential to examine many more attributes and attribute levels than would be possible using an experimental design (as used for constructing conjoint analysis options.) Nevertheless, the task of the researcher in designing the survey instrument seems to be less straightforward than for other SP methods.

A complication for users of the priority evaluator is the need to define "prices" for each attribute level, something that is not applicable to other SP techniques. A particular uncertainty is how the combination of these values and the budgets used can affect the measurement of respondents' preferences. Little work has been done to examine this issue, though there is evidence to suggest that the results can be very sensitive to different value/budget combinations (Pearmain, 1989.) The sensitivity of the values obtained for respondents' preferences in relation to design characteristics of the survey instrument is of course relevant to other SP techniques and this issue, in relation to conjoint analysis, is discussed later in chapter five.

Compared to conjoint analysis, the priority evaluator has not enjoyed widespread application in the field of transport research. Nevertheless, it was recently introduced into the evaluation of rolling stock and rail station improvements by the MVA Consultancy and therefore presents itself as a technique relevant to the current study. Before these applications of the priority evaluator to the assessment of rail passenger facilities, analytical techniques developed for this technique were seen to be very naive. They lacked any error theory relating to the way respondents were understood to identify their priorities and allocate their budget in an optimal way. This therefore produced measures of relative importance for each attribute model which were not capable of being subjects to tests of statistical significance or model "goodness-of-fit". Only comparatively recently (Copley and Bates, 1988) have advanced statistical models been successfully fitted to priority evaluator data, and this it appears exclusively in the valuation of passenger facilities. This is given more attention in chapter seven.

Despite the development of the priority evaluator, it can be seen that there are still serious criticisms to be made which, it is argued, suggest that it is an inferior method to conjoint analysis and other "trade-off" techniques. These criticisms are:

(i) The principles of good design relating to the imposition of suitable price structures and budget constraints are unclear.

(ii) Analytical Methods have generally lacked sophistication and despite the introduction of advanced statistical models, it is still uncertain how particular prices and budgets will influence the importance weightings for each attribute level that can be derived.

(iii) The basic assumption that budget allocations are a suitable way of getting individuals to express consumer preferences has to be questioned. The fact that in the real market place, most consumers are used to assessing competing packages of attributes rather than constructing their optimum combination of attributes, suggests that the priority evaluator is more removed from reality than conjoint analysis.

The last point concerning realism is extended when one considers that the priority evaluator can only obtain individuals' judgements concerning the relative value of attributes. In contrast, some other techniques, most notably conjoint analysis, can obtain both judgements and, more realistically, discrete choices analogous to real life choices (eg mode or route choice.)

4.2.3 Conjoint Analysis

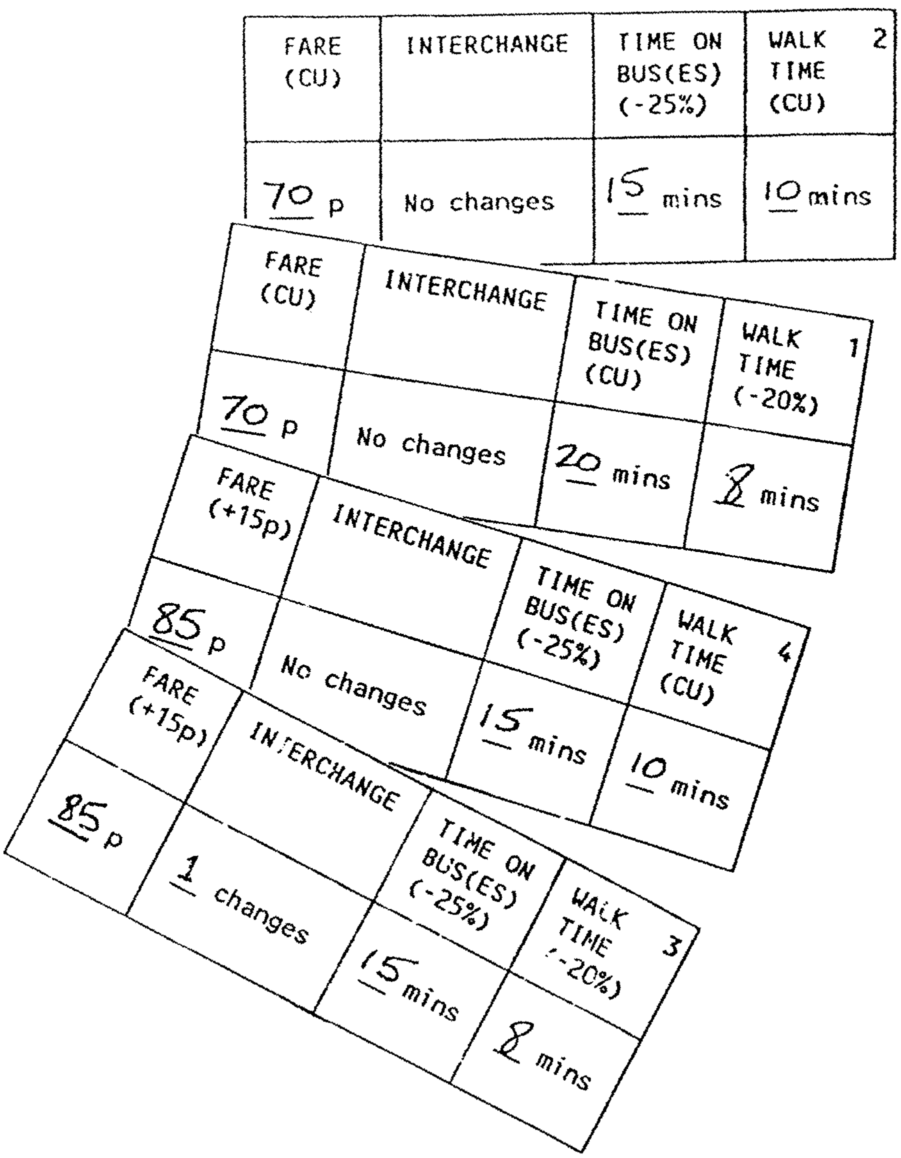
As already suggested, this is a family of techniques which are by far the most widely applied types of SP techniques. Though the techniques in this category are similar in their approach to the design of hypothetical options they are different in they way they measure individuals' responses.

Conjoint analysis methods form a distinct sub-group of SP techniques because they involve the presentation to individuals of 'packages' of attributes which may later be 'decomposed' in the analysis to derive separate part-worth utilities for each attribute. This is the origin of the term 'conjoint'. As suggested, in the construction of these attribute packages they are largely but not always characterised by the use of experimental designs. Below is a discussion of the main types of conjoint analysis technique, which can be categorised by the type of response scales that they use.

Conjoint Measurement: Rating Responses

This type of SP technique, like functional measurement (rating responses, discussed below), has its roots in applied psychology (Krantz and Tversky, 1971) and has enjoyed widespread application in transport research. The approach presents all the hypothetical options at once to the respondents, who are then required to rank the hypothetical options in order of preference, so implying a hierarchy of utility values. An illustration of this approach is given in Figure 4.5.

Figure 4.5 **Example of Conjoint Measurement (SP Ranking) Options**



The attraction of this type of response is that all the options are considered together. The exercise is also fairly straightforward for the respondent to carry out. Nevertheless, there is a limit to the number of options that can be offered without the respondent becoming fatigued. More importantly, the researcher needs to be aware that the data provided by this method represent **judgements** by respondents, which do not necessarily correspond to the kind of choices they face in real life. This is also a problem with rating responses (discussed below) and of course with the techniques already discussed. It is partly for this reason that conjoint analysis methods which use discrete choice responses are enjoying increasingly widespread application.

In the analysis of conjoint measurement data (rankings), the researcher has recourse to a number of modelling approaches: Monotonic Analysis of Variance ('MONANOVA'), multiple regression (using transformations of the rankings to a metric or log-linear scale) and 'exploded' logit (in which each choice is modelled as the preferred option from a sub-set of options - ie all those ranked below it.) Each of these approaches has its own merits and defects. MONANOVA is able to produce part-worth utilities for each individual respondent, but lacks an error theory with which to substantiate the models. Multiple regression is a widely available and robust technique, but the transformation of rankings to a metric scale inevitably requires some 'a priori' assumptions which are not easily tested. Logit or probit models are the most attractive approaches, but have the disadvantage that preference weights can only be derived for groups of respondents rather than individuals. Little can therefore be surmised about the distribution of individual part-worth utilities around a mean group value.

Whichever analytical technique is applied, there still remains the problem that rankings may provide an inconsistent measure of preference. It is reasonable to suggest that those options placed by an individual towards the top of the list are clearly the better options in his or her opinion. In contrast, those options placed in the middle ranks are likely to be harder to differentiate between, because differences in their utilities will be marginal; those at the bottom may be seen as so unrealistic or indistinguishable from one another that their relative positioning is not seriously considered. One solution with 'exploded' logit (where the rankings are converted to choices from sub-sets) is to only analyse the top half of the rankings, but this is a somewhat arbitrary approach.

Functional Measurement: Rating Responses

This form of conjoint analysis requires respondents to express the strength of their preferences on numerical or 'semantic' scales (in the latter case, the points of the scale are defined by phrases such as "definitely choose option A"; "probably choose option B"; "cannot choose between A or B".) The respondent may be asked to express relative preferences for each option by indicating a particular 'score'. Alternatively, he or she may be asked to express the strength of preference between a pair of options.

Where the respondent gives a separate score to each item, the researcher not only has information on the order of preferences between the options, but also the strength of those preferences. This assumes that the respondents can consistently rate the options. As this is not a procedure they carry out when choosing between real options, this method, like ranking approaches, can be criticised for the artificial nature of the exercise. Potentially, this approach provides the richest type of response data, if one can assume that the scores have the property of an interval scale. The power of the technique improves with the fineness of the scales used.

Some studies have used scales as large as 1 to 100 but such a fine scale is of little use if respondents cannot make judgements to the same degree of precision. Scales of 1 to 20 and 1 to 10 are more common. An example of this type of rating exercise is shown in Figure 4.6. Multiple regression is a suitable analytical technique for this kind of approach, though the researcher must be aware that the assumption of an interval scale is not entirely defensible. One may argue, for example, that some individuals tend towards the mid-point of the scale while others tend towards the end-points. Some appropriate weighting strategy may therefore have to be considered to compensate for this effect.

Where respondents make choices between pairs of options, scales of only five levels are most common. Paired choices offer simple choice exercises to respondents but can become repetitive if too many are presented. An example of such a choice is shown in Figure 4.7. The response scale indicates likelihood of choice. The implied scores can then be used as they are are or preferably transformed into plausible probabilities of choice (eg score of 1 = 0.1; score of 3 = 0.5; score of 5 = 0.9) to construct a multiple linear regression model (see Bates, 1986.)

If the researcher is concerned as to the reliability of the scores obtained from paired choices, he or she has the option of 'collapsing' the scores into simple binary choices, so that all scores below the mid-point of the scale represent a statement in favour of one option, all those above the mid-point are in favour of the other option. This introduces one form of the remaining type of conjoint analysis technique: discrete choices.

Figure 4.6 Example of Functional Measurement (Ratings) Exercise

Your Bus Service

Journey Time: 25 mins

Comfort of Ride: Poor

Frequency: Every 30 mins

Reliability:
One in twenty is cancelled

How would you rate this service?

Very Poor			Average	✓	Very Good	
1	2	3	4	5	6	7

Figure 4.7 Example of a Paired Choice Rating Exercise

WHICH OF THESE ALTERNATIVE NORTHERN LINE TRAINS WOULD YOU CHOOSE ?

OPTION 1	or	OPTION 2
This train is....		This train is....
BRAND NEW		BRAND NEW
CLEAN		DIRTY & VANDALISED
It has....		It has....
AIR CONDITIONING		"FORCED AIR" VENTILATION
NO GANGWAYS BETWEEN CARRIAGES		NO GANGWAYS BETWEEN CARRIAGES
* The FARE is: £ 2.86		* The FARE is: £ 2.60

Which option would do you choose ?

Definitely Option 1	Possibly Option 1	Cannot decide	Possibly Option 2	Definitely Option 2	Neither of these

Discrete Choice Methods

In designing a SP survey, the researcher will wish to ensure as much realism and simplicity as possible in the exercises that are presented to the respondents. Discrete choice designs, in which the individual simply selects the most preferred option from a pair or group of options, come closest to achieving this goal. The development of suitable analytical procedures, such as the logit model, has enabled these particular types of SP approaches to come to the fore. The simple example of Figure 4.1 illustrated such an approach.

Exercises that present options in pairs can be constructed in two ways. The first is to define options as one would a ranking exercise and present every combination of each option with every other in the choice set. By simply choosing the preferred option in each case, the respondent implies the rank order of all the options. Unfortunately, the number of choices required may be very large: nine options, for example, when presented in every possible combination of two at a time, will require the respondent to consider 36 pairs. One solution is to present a random sample of these pairs in each interview, but this introduces a loss of information and control by the researcher. A recent development of computer-assisted SP interviews (discussed in chapter five and reported in chapter eight) allows some pairs to be removed on the basis of assumed 'dominance' of some options over others. A further necessary requirement for this approach is that respondents are assumed to be transitive in their choices.

An alternative to the pseudo-ranking approach discussed above is to define the exercise so that characteristics of competing alternatives (eg car v train) are defined together in the same option. In this way, the exercise is very similar to the last rating exercise described above, with the scale being collapsed into simple binary responses.

An alternative form of discrete choice exercise is reported by Hensher and et al (1988), in which respondents are required to choose one option from a group of options. This approach requires a 'double experimental' design. One experimental design is used to construct the options, the other is used to construct the groups from which the respondent chooses his or her preferred option.

Determining the Preferred Conjoint Analysis Technique

Each type of conjoint analysis technique has features to recommend it and there is no consensus in the literature for one method over another. Ranking and rating methods offer the richest form of data but offer less realistic choice exercises. Discrete choice approaches offer simpler and more realistic exercises, but the information acquired is more limited. A range of established analytical techniques are available for each method, though it may be argued that discrete choices offer a firmer theoretical underpinning for the analysis, because less a priori assumptions are required for interpreting the responses.

In deciding the method appropriate to the research topic, the researcher should be guided by the type of choice situations to be offered. Where the options are different versions of the same item (eg different quality bus services), all forms of conjoint analysis may be applicable (ie ranking, rating or choices.) Where the choices are to involve competing alternative items, such as car travel versus a rail service, paired choice comparisons more closely represent the kind of choices faced in respondents' real life experiences.

4.3 Conclusions

The wide application of conjoint analysis to transport appraisal and its advantages over other SP methods commends itself as a suitable approach for pursuing the aims of this study. Chapter five is therefore devoted to the design, application and analysis of conjoint analysis techniques. This provides an entry into the examination of SP applications for the valuation of rail passenger facilities in chapter seven and a basis for the case study in chapters eight and nine. While transfer pricing, trade off analysis and priority evaluator have been identified as less effective techniques than conjoint analysis, the priority evaluator remains a technique used a number of times in the analysis of passenger facilities and therefore receives further coverage in chapter seven.

5: CONJOINT ANALYSIS METHODS

5.1 Introduction

The previous chapter identified conjoint analysis techniques as the most suitable set of methods for analysing travellers' preferences toward secondary items such as passenger facilities. This chapter discusses the detailed issues that relate to the design of conjoint analysis instruments, the main principles of which have been applied in the case study of chapter eight.

5.2 Design Issues

Most conjoint analysis techniques are characterised by the use of experimental designs to construct hypothetical alternatives presented to respondents. A factorial experimental design (in which all combinations of attribute levels are included) ensures that the attributes presented to respondents are varied independently from one another. That is, the design is 'orthogonal'. The result is that the effect of each attribute upon responses is more easily isolated in the analysis. This avoids 'multi-collinearity' between attributes, which is a common problem with revealed preference data (ie observations of actual behaviour.)

5.2.1 Experimental Designs

Consider the example of an experimental design in Figure 5.1. Here, the researcher wishes to examine respondents' preferences towards three attributes of a transport service (fare, travel time and service frequency), each with two levels. Normally the researcher would wish to include more levels that this but for simplicity they have here been limited to two. The combinations are shown in descriptive and numerical terms (the latter representing standard notation for experimental designs.) It can be seen that the eight options represent different quality public transport services, which respondents would be asked to evaluate (eg rank in order of importance.) Figure 5.2 gives an example of how the set of options might be presented in a survey.

The experimental design presented in this example is known as 'full factorial'. This is because every possible combination of attribute levels is used. For the example used here, the number of combinations is the result of the number of levels raised to the power of the number of attributes. Thus, eight options = 2^3 (2 levels each, 3 attributes.) If attributes with differing numbers of levels are used, the values are simply multiplied together. For example, a design with two three-level attributes and two two-level attributes would have $3^2 \times 2^2 = 36$ options.

When the number of options to be evaluated becomes high, it is likely that respondents become fatigued in carrying out the exercises, so increasing the response error. Likewise, too many attributes or levels may lead to some items being ignored by the respondents. Kroes and Sheldon (1988, p14) suggest a range of 9 to 16 options as acceptable.

The practical limit on the number of options to present in one exercise will to some extent depend on the context of the survey in which the conjoint analysis exercise is introduced. On-vehicle interviews, for example, benefit from as small a number of options as possible, whereas home-based interviews, with less time constraints and a more comfortable environment, allow a larger number of options. Self completion surveys, where the degree of guidance will be minimal, also benefit from simple designs and limited numbers of options.

Figure 5.1 Example of a Stated Preference Design For Three Attributes, Each at Two Levels

		Attributes		
		Fare	Travel Time	Frequency
Options	1	Low	Fast	Infrequent
	2	Low	Fast	Frequent
	3	Low	Slow	Infrequent
	4	Low	Slow	Frequent
	5	High	Fast	Infrequent
	6	High	Fast	Frequent
	7	High	Slow	Infrequent
	8	High	Slow	Frequent

(Numeric Representation:)

		Attributes			
		1	2	3	
Options	1	1	1	-1	(-1 = poor 1 = good)
	2	1	1	1	
	3	1	-1	-1	
	4	1	-1	1	
	5	-1	1	-1	
	6	-1	1	1	
	7	-1	-1	-1	
	8	-1	-1	1	

Figure 5.2 Survey Show Cards Developed From the Experimental Design of Figure 5.1

Public Transport Service			(Option 1)
Fare = £0.30	Travel time = 15 mins	Frequency = every 30 minutes	

Public Transport Service			(Option 2)
Fare = £0.30	Travel time = 15 mins	Frequency = every 15 minutes	

Public Transport Service			(Option 3)
Fare = £0.30	Travel time = 25 mins	Frequency = every 30 minutes	

Public Transport Service			(Option 4)
Fare = £0.30	Travel time = 25 mins	Frequency = every 15 minutes	

Public Transport Service			(Option 5)
Fare = £0.50	Travel time = 15 mins	Frequency = every 30 minutes	

Public Transport Service			(Option 6)
Fare = £0.50	Travel time = 15 mins	Frequency = every 15 minutes	

Public Transport Service			(Option 7)
Fare = £0.50	Travel time = 25 mins	Frequency = every 30 minutes	

Public Transport Service			(Option 8)
Fare = £0.50	Travel time = 25 mins	Frequency = every 15 minutes	

(Note: Experiment attributes are:

Fare: Low = £0.30; High = £0.50

Time: Fast = 15 mins; Slow = 25 mins

Frequency: Frequent = every 15 mins; Infrequent = every 30 mins)

Clearly, if the number of options is to be constrained in this way, the use of full factorial experimental designs severely limits the number of attributes and levels that can be examined. If a limit of nine options is imposed, for example, only the following full factorial designs can be used:

(i) with two attributes: $2^2 = 4$ options: 2 attributes with 2 levels
$2^3 = 8$ options: 3 attributes with 2 levels

(ii) with three attributes: $3^2 = 9$ options: 2 attributes with 3 levels

(iii) with mixed level attributes: $2^1 \times 3^1 = 6$ options: 1 attribute with 2 levels; 1 attribute with 3 levels;
$2^1 \times 4^1 = 8$ options: 1 attribute with 2 levels; 1 attribute with 4 levels.

This is a severe limitation because in most studies a minimum of three attributes will be desirable, to place the choice exercises in a realistic context. In the example given earlier, the researcher may only be interested in preferences towards travel time and fare, but the inclusion of service frequency makes the exercise more realistic. If only travel time and fare were presented, the likelihood is that respondents will give undue prominence to these items. They are also perhaps more likely to imagine that the options represent a particular strategy being considered by the policy makers (eg that fare will rise) and so attempt to influence that strategy through their responses. This is the problem of 'policy response bias' discussed earlier.

For attributes of particular interest to the researcher, more than two levels are advisable in cases where the preferences are likely to be non-linear (an issue discussed later in this chapter.) Thus the requirements for a minimum of three attributes and perhaps at least one with three or more levels moves the number of options into double figures (the smallest full factorial design which meets this criteria is $2^2 \times 3^1 = 12$ options.)

5.2.2 Strategies for reducing the number of options

To resolve the problem of many options, the researcher needs to consider ways of reducing their number. A review of the literature suggests five main approaches which can be taken to achieve this:

(i) remove those options that will 'dominate' or be 'dominated' by all other options in the choice set;

(ii) separate the options into 'blocks', so that the full choice set is completed by groups of respondents, each responding to a different sub-set of options;

(iii) carry out a series of experiments with each individual, offering different attributes, but with at least one attribute common to all, to enable comparisons;

(iv) use 'fractional factorial' designs;

(v) define attributes in terms of differences between alternatives (eg journey time of train = journey time of car + 10 minutes.)

If required, these approaches can be used in conjunction with one other.

Removing 'Dominated' Options

This first approach allows limited reduction of options. It uses the proposition of **dominance** among the options. In the example provided earlier in Figure 5.1, it will be seen that in a ranking exercise, option 2 is better than all the other options (ie it 'dominates' the other options in the choice set) because fare is low, travel time is fast and frequency is high. Option 7 is the worst option and is 'dominated' by all the other options. These two options could be omitted from the choices set, with the assumption that respondents would always choose option 2 first and option 7 last. Six options would then remain.

One problem with this approach is that any respondents choosing randomly or illogically will not be easily identified from their responses. Thus if at least one of these best/worse options remains in the choice set, their logical or illogical positioning by each respondent provides some indication of the reliability of the responses. Another problem is that the omission of such options reduces the orthogonality of the experimental design, so reducing the accuracy of the analysis. The analyst could always insert these options back into the data set with artificial but logical responses. A criticism of such an approach is that the insertion of such 'correct' data artificially reduces the error in the real data.

This approach can be taken a step further by identifying dominated sub-sets within the overall choice set being used. An example might have four options: A, B, C and D which are presented in pairs. Option A dominates B; option C dominates D. If a respondent therefore prefers A to C, the researcher may assume that A will also be preferred to D. The respondent may not therefore need to be presented with a choice between A and D. Alternatively, if C is preferred to A, it may be assumed that C would be preferred to B, in which case the C versus B choice may be omitted instead.

In such instances, the respondent is assumed to be **transitive** in his or her choice behaviour. That is, his or her preferences for one option over another will remain the same regardless of any other options that may be available. This may be considered a reasonable assumption and the researcher may therefore be justified in omitting choices between options for which the result can be predicted on this basis. Removing choices as a result of a respondent's earlier responses can be difficult to implement in a conventional questionnaire or show card format. Computer based applications offer a solution.

As before, the result of omitting dominated or dominating choices is the possibility that any respondent not exhibiting transitivity in his or her choice behaviour will not be detected. In such a case, his or her assumed responses between the omitted options will therefore be incorrect.

Using a 'Blocked' Design

This second approach, which requires the division of the choice set into sub-sets (known as 'blocks') retains the original experimental design chosen by the researcher but divides the task over a number of respondents. The success of this approach rests on the assumption that the preferences across the sample of respondents will be sufficiently homogeneous such that the responses can be combined over the sub-sets of the options. Inevitably, differences between individuals will increase the error associated with the results. This strategy was adopted in the case study of chapter eight as a way of enabling a limited number of interaction effects (see below) to be investigated.

Using Common Attributes over a Series of Conjoint Exercises

By carrying out a series of experiments with each individual, the researcher can keep the number of attributes and options to a manageable number in each experiment. As each experiment is individually less demanding than one large and complex design, there is the potential for obtaining more observations from each respondent before he or she becomes fatigued. The inclusion of at least one attribute common to all the exercises carried out (eg fare or travel time) allows comparison of relative preferences over all the attributes being investigated. A useful example of this approach is to be found in a study undertaken by Steer Davies and Gleave Ltd (1986; also Andersen et al, 1986.) The approach and findings of this study are discussed in chapter seven. This use of separate designs linked by a common variable was also used in the case study of chapter eight.

Fractional Factorial Designs

This fourth approach to reducing the number of conjoint options is the most commonly applied procedure. This is because it allows the examination of appreciably large numbers of attributes and levels, without necessarily the need to divide a choice set among individuals or use more than one experimental design, though these strategies may also be applied. The approach rests on the assumption that some or all of any interaction effects between attributes are negligible.

Significant interactions may be described as the effects of two or more attributes which, when acting together, have an influence different from the sum of the individual effects of each attribute. Consider the example choice set used at the beginning of this chapter in figures 5.1 and 5.2. It is possible that for some individuals, the effect of fare and journey time, when they are expensive and slow respectively, is more negative than the separate influences of each attribute added together. Such a relationship is illustrated in Figure 5.3.

This illustration shows how an individual's response alters with changes in the combinations of two attributes. As attribute 1 moves from 'bad' to 'good', the response becomes more positive. When there is no interaction, the response also becomes more positive when attribute 2 goes from a bad level to a good level, but the **rate** of improvement in relation to attribute 1 remains the same. When there **is** interaction between the attributes, the rate of improvement due to attribute 2 is **not** the same in relation to attribute 1: the combined effect of both attributes being at a good level is greater than the individual effects.

If such interaction effects are considered in advance of a survey to be insignificant, on the basis of earlier research and the reasoning of the researcher, the opportunity then exists to simplify the experimental design and reduce the number of options. An example is given in Figure 5.4. Here, the full factorial experimental design used earlier is shown with interaction levels. A reduced fractional factorial version of the design is shown below it.

Figure 5.3 Illustration of an Interaction Effect Between Two Attributes

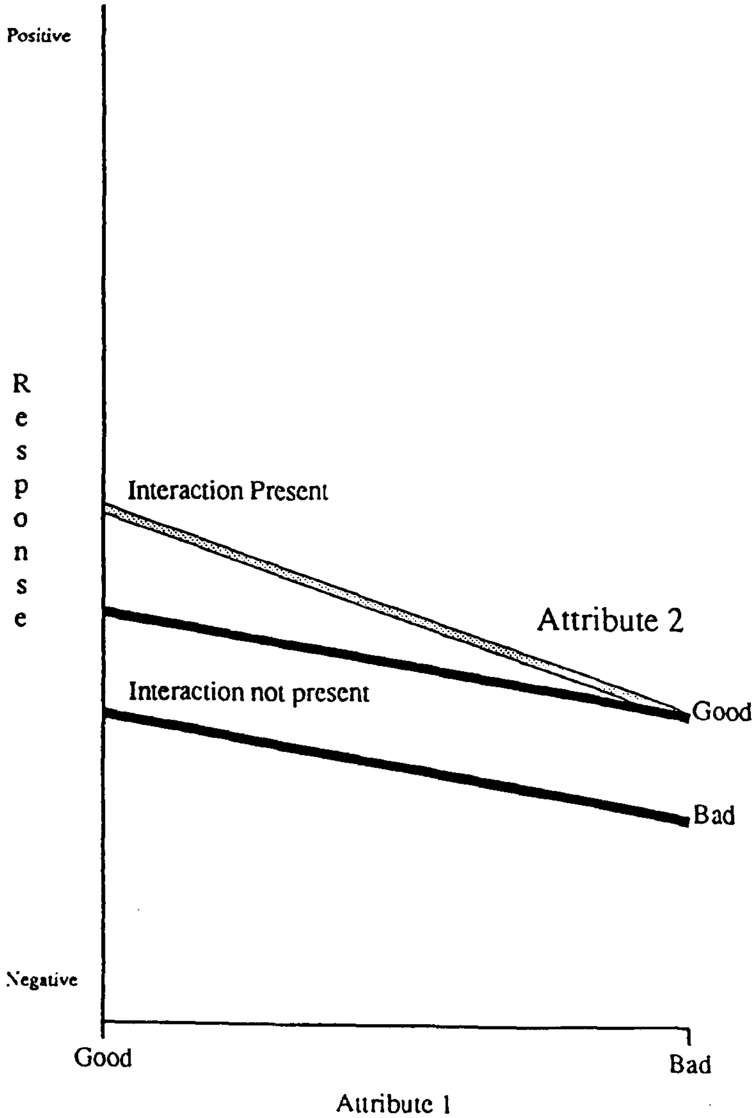


Figure 5.4 Example of a Fractional Factorial Design Derived From a Full Factorial Design

Full Factorial Design

	Attributes			Interactions			
				(Two-way)			(Three-way)
	1	2	3	1 x 2	1 x 3	2 x 3	1 x 2 x 3
Options:							
1	+1	+1	- 1	+1	- 1	- 1	- 1
2	+1	+1	+1	+1	+1	+1	+1
3	+1	- 1	- 1	- 1	- 1	+1	+1
4	+1	- 1	+1	- 1	+1	- 1	- 1
5	- 1	+1	- 1	- 1	+1	- 1	+1
6	- 1	+1	+1	- 1	- 1	+1	- 1
7	- 1	- 1	- 1	+1	+1	+1	- 1
8	- 1	- 1	+1	+1	- 1	- 1	+1

Fractional Factorial Design

2	+1	+1	+1	+1	+1	+1	+1
3	+1	- 1	- 1	- 1	- 1	+1	+1
5	- 1	+1	- 1	- 1	+1	- 1	+1
8	- 1	- 1	+1	+1	- 1	- 1	+1

The second design in Figure 5.4 allows each attribute level to be varied independently from each other attribute level, as before, but the interaction terms are not all independent. The second two-way interaction term, for example, is perfectly correlated with the second attribute. The three-way interaction shows no variation and is therefore perfectly correlated with the intercept (or 'grand mean'.)

If the interaction terms are insignificant, accurate measures of preference weights on each attribute can be obtained. However, if they are significant, their effects in a fractional factorial design will be loaded onto the individual main effects, so giving erroneous results (which becomes worse as the magnitude of the interactions increases.) In such a case, the main effects are confounded with interaction effects. The benefit of fractional factorial designs is that the number of options can be greatly reduced, in this case from eight to four (note that in the simple example used here, the fractional factorial design would probably be unnecessary, as the original number of options is manageable.) The advantage of this approach becomes clearer when one considers that that a 3^4 design (ie three attributes with four levels each), requiring 81 options, can be reduced to just 9 options. This assumes that no significant interaction effects are evident.

There are stages by which a full factorial design can be reduced, which allow the investigation of some (but not all) interaction effects or at least ensure that the main effects of attributes are independent from the influence of significant interactions. Kocur et al (1981, p183-184) list the different types of fractional factorial designs that may be employed. These are sometimes classified as 'resolution level' plans, which reflect their power to discriminate interaction effects.

The range of designs is as follows:

(i) Full factorial designs: allow estimation of all main effects and all interaction effects independently from one another;

(ii) Fractional factorial designs (1); 'Resolution V plans': assume three-way attribute interactions or higher are negligible, but allow estimation of main effects and all two-way interactions. Sub-sets of resolution V designs are:

(iia) those designs that allow estimation of main effects and two-way interactions between selected attributes and **all** other attributes. All other attributes are assumed negligible;

(iib) those designs that allow estimation of main effects and **some** selected two-way interactions. All other interactions are assumed negligible;

(iii) Fractional factorial designs (2); 'Resolution IV plans': allow estimation of main effects only, but **independently** of two-way interactions. All other interactions are assumed negligible. If significant two-way interactions exist, their effect does not distort measurement of the main attribute effects;

(iv) Fractional factorial designs (3); 'Resolution III plans': allow estimation of main effects only, assuming all interactions are negligible. Otherwise, all significant interactions will be confounded with main effects.

Clearly, the more interaction effects which the researcher can show or assume to be insignificant, the greater the opportunity for reducing the size of the experimental design. As a guide to what may be lost when a fractional design is used, Louviere (1988, p40) suggests the following:

> "In almost all cases involving real data, the following generalisations hold about significant effects:
>
> (a) Main effects explain the largest amount of variance in response data, often 80% or more;
>
> (b) Two-way interactions account for the next largest proportion of variance, although this rarely exceeds 3%-6%;
>
> (c) Three-way interactions account for even smaller proportions of variance, rarely more than 2% - 3% (usually 0.5% - 1%) and;
>
> (d) higher order terms account for minuscule proportions of variance."

The researcher may therefore conclude that main effects only and other fractional designs are valid tools, but wherever possible, care should be taken to use designs that avoid 'confounding' interaction effects with main effects and include all interactions likely to be significant. In some cases, fractional designs alone may not be sufficient, and instead can be used in conjunction with the option-reduction strategies already identified.

Defining Attributes in Terms of Differences Between Alternatives

Where the alternatives are to be presented as paired choices (eg a journey by car versus a journey by train), the attributes can be defined as the differences between the alternatives. For example, instead of defining the cost of car and the cost of train as separate attributes in an experimental design, a single attribute representing the difference between cost of train and cost of car could be used. One alternative (eg car) is defined as the base option. The levels of such an attribute might then be defined as "five minutes more than by car"; "ten minutes more than by car", etc. In this way, two attributes are represented by one attribute in the experimental design. To the respondent, of course, they are still presented as separate items.

For qualitative attributes, such as comfort of ride, a similar process can be applied, with descriptions presented as contrasts: eg good car comfort versus poor train comfort; good car comfort versus good train comfort. Again, two attributes (comfort of car; comfort of train) are represented by a single attribute: ie difference in quality of comfort. Designs that define attributes in terms of differences have been referred to as 'correlated' designs, because if the values of the base alternative are altered, the values of the other alternative(s) are altered in the same manner, while the difference between them is still varied independently.

An example of how a simple correlated design can reduce the number of attributes, and therefore options, is shown in Figure 5.5. It is possible to include further alternatives (eg bus, in addition to car and train), for which the attributes are also defined as differences from the base mode. This has the advantage of providing more information from a single response, but also complicates the respondent's task.

The main drawback of using correlated or 'difference' designs is that the researcher must assume that the values for the attributes are 'generic' across the alternatives. For example, a respondent may value the cost of travel by car differently to the cost of travel by train. This would probably reflect some perception of 'value for money' associated with each mode or differences in the method of payment. A correlated design would not be able to isolate these different values.

Figure 5.5 Example of Attributes Defined As Differences Between Alternatives

Option	Cost difference	Time difference	Comfort difference
1	Car Cost + 20p	Car time - 10 mins	Good car - poor train
2	Car Cost + 20p	Car time - 20 mins	Good car - good train
3	Car Cost + 50p	Car time - 10 mins	Good car - good train
4	Car Cost + 50p	Car time - 20 mins	Good car - poor train

If car cost = £2.00; car time = 50 mins, choices would be represented as:

	Car			versus	Train		
	Cost	Time	Comfort		Cost	Time	Comfort
1	£2.00	50 mins	Good		£2.20	40 mins	Poor
2	£2.00	50 mins	Good		£2.20	30 mins	Good
3	£2.00	50 mins	Good		£2.50	40 mins	Good
4	£2.00	50 mins	Good		£2.50	30 mins	Poor

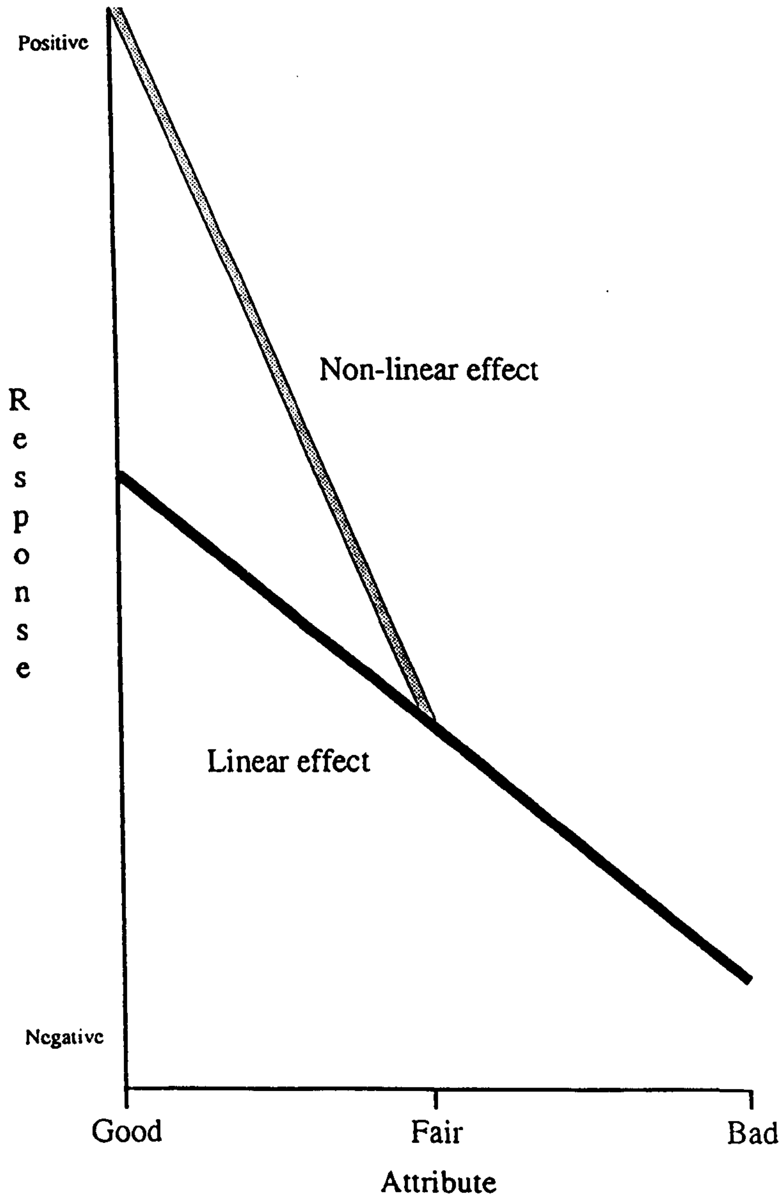
5.2.3 Examining Non-linear Effects

Concerning attributes of most interest to the researcher, it has already been suggested that more than two levels should be used. This is relevant in cases where the effects of an attribute on response are likely to be non-linear. Figure 5.6 illustrates a non-linear effect in relation to a linear effect. In this example, the move from 'bad' to 'fair' has about half the effect on response as a move from 'fair' to 'good'.

It was stated in the previous chapter that the most commonly applied utility models are linear additive. In a situation where an attribute is non-linear, this form of model is no longer adequate and the researcher should consider an alternative such as a polynomial (where attributes are raised to the power of 2 or more.) Measures of 'goodness-of-fit' (such as the rho bar^2 statistic in logit analysis) will provide an indication of how suitable alternative model forms are. Likewise, the use of 'dummy' variables in the modelling of continuous variables allows the researcher to observe the influence of each change in attribute levels, thus identifying non-linear shifts.

The issue of alternative model forms is important, because many studies fail to explore this issue, preferring instead to use simple linear models. Only when attributes of more than two levels are used can a full range of alternative models be examined. Consequently, as many levels as is practical should be used for each attribute.

Figure 5.6 Illustration of a Non-Linear Effect on Response



5.2.4 Departure from Wholly Orthogonal Designs

The designs discussed above ensure that the attributes are completely 'orthogonal'. That is, the options are varied independently of one another, such that their individual effect on respondents' preferences can be completely isolated. The discussion on fractional factorial designs showed how orthogonality between the attributes could be maintained even at the expense of reducing independence from interaction effects. By presenting every combination of the levels of one attribute with those of another, complete independence between them is maintained. One of the drawbacks of presenting every combination of the attribute levels is that some of these combinations may be unrealistic when presented to the respondent. The issue of realism is discussed later on, but this particular aspect is relevant to the topic of experimental design. It is important that the respondent considers an option to be believable. Thus, a particular combination of attributes in which, for example, the total journey time by a current public transport service decreases while the distance increases would appear implausible.

One solution may therefore be to eliminate such an option from the choices offered to the respondents. This avoids offering unrealistic situations, but reduces the orthogonality of the design. Taking the example of journey time and distance, only options where both increase together will be acceptable and so these two attributes will be correlated.

Combinations which might presently be thought unrealistic can be made believable by introducing some new factor. For example, a shorter journey time by a longer route might be possible if an express bus replaced a stopping service, or congestion on the shorter route was said to rise. Care should be taken if such items are introduced to the respondent, because they may have an additional effect on choice.

If it is not appropriate or even possible to provide an explanation for normally unrealistic options, these options could be eliminated from the design. The likelihood in such cases will be that less efficient models will result. This may be acceptable if the result is greater realism in the responses to the conjoint analysis exercise. The researcher can investigate the effect of eliminating options, and therefore reducing orthogonality in the design, by the use of simulated data. This is discussed in a later section of this chapter.

5.2.5 Determining Attributes and Attribute Levels

The need to limit the number of options which respondents are required to evaluate, together with the nature of experimental design discussed above, in turn limits the number of attributes and attribute levels that may be presented in any set of options. Even if a particular design allows a lot of attributes to be presented, it is advisable to limit the number to avoid confusing respondents. An upper limit of some six or seven attributes may be all that is practicable, though designs with 10 or more have been used (Kocur et al, 1981; Andersen et al, 1986.) If some of the attributes are currently unfamiliar to respondents or complex to define, a lower number may be preferable. For example, a new travel service such as an automated ticketing system or a completely new mode such as light rail will require lengthy descriptions for the respondents to absorb.

How the researcher should prioritise between the attributes or choose to present descriptions of the attribute levels is considered in section 5.3 below. Nevertheless, in the context of basic experimental design procedures it is appropriate to discuss guidelines for defining the quantitative values of continuous attribute levels (eg fare or distance.)

Concerning the definition of the attribute levels, the researcher must consider the following points:

(i) They must appear plausible - this was discussed in the previous section;

(ii) they need to relate to the respondents' experience of each attribute;

(iii) the values attached to the attributes should ensure that competitive trade-off decisions are presented;

(iv) the values attached to the attributes should present trade-offs that cover the range of valuations held by each respondent.

To present the conjoint analysis exercise in terms that are easily and realistically understood by respondents, it is often the practice to have them make choices in the context of a journey that is familiar to them. Therefore, to satisfy the second requirement above, the attribute levels used in the experimental design can be defined as variations relative to the attribute levels of an existing journey. Figure 5.7 illustrates how this might be done, first using absolute additions and subtractions to a respondent's present values and then using proportional changes.

It is debatable as to which method of defining attribute levels is preferable. If it can be assumed that 20 pence saving is worth the same to a traveller paying a £10.00 fare as one pay £1.00, then absolute changes may be acceptable. Against this there is some evidence to suggest that 'thresholds' exist for the way individuals value such items as cost or time savings. This might mean that 20 pence from a £10.00 fare has less perceivable benefit than from a £1.00 fare. In this case, proportional changes may be more appropriate.

Figure 5.7 A Definition of Attribute Levels Dependent on the Characteristics of a Respondent's Actual Journey

	Cost	Journey Time	Frequency of Service
Respondent's Actual journey	£1.00	20 mins	Bus every 20 mins

Definitions of Attribute Levels

Stated preference Alternatives	Cost	Journey Time	Frequency of Service
(As absolute changes)			
1	+30p	+10 mins	- 10 mins
2	- 20p	- 5 mins	+20 mins
(As proportional changes)			
1	+30%	+50%	- 50%
2	- 20%	- 25%	+100%

Presentation of Choices

	Recent Journey Cost	Time	Frequency	Alternative Journey Cost	Time	Frequency
1	£1.00	20 mins	1 every 20 mins	£1.30	30 mins	1 every 10 mins
2	£1.00	20 mins	1 every 20 mins	£0.80	15 mins	1 every 40 mins

Whichever may be considered the more suitable alternative between absolute and proportional changes, it is perhaps easier to argue that the rate at which an individual will trade one attribute against another will remain fairly constant (eg as defined by monetary values of travel time.) Thus if the likely ranges of the journey characteristics are known in advance, the attribute levels for the conjoint analysis choices may be varied in such a way as to include the sort of trade-off rates anticipated by the researcher (eg from values of time established from previous research.)

This discussion leads on to the remaining issues listed above. The need to ensure competitive choices can be met by the presentation of attribute levels that offer likely 'boundary values' in the trade-offs that are presented to respondents. Such boundary values represent the points at which individuals will switch between one option and another, given the attribute levels offered to them.

Figure 5.8 offers an illustration of boundary levels defined by the way a respondent switches between options, given the changes in the trade-off rate between fare, journey time and comfort of ride. In the first pair of options, the respondent has the opportunity of a 10 minutes time saving in return for a 45 pence fare increase (trade-off = 45 pence/10 minutes = 4.5 pence per minute.) This is too high a rate to induce him to switch from option 1 to option 2. In the second pair of options, a 15 minute time saving is offered at a higher cost increase of 50 pence (trade-off = 50 pence/15 minutes = 3.3 pence per minute.) This lower rate of trade-off, despite the higher price, induces him to switch to option 2. As the standard of comfort remains constant across these pairs of options, the responses to the first two pairs allows the researcher to surmise that:

(i) The respondent's value of time >= 3.3 pence/minute;

(ii) The respondent's value of time < 4.5 pence/minute.

Figure 5.8 Example of Boundary Values Causing a Respondent to Switch Between Options

Pairs	Option 1			Option 2			Response
	Fare	Time	Comfort	Fare	Time	Comfort	
1.	£1.00	40 mins	Average	£1.45	30 mins	Average	"option 1 preferred"
2.	£1.00	40 mins	Average	£1.50	25 mins	Average	"option 2 preferred"
3.	£1.00	40 mins	Average	£1.50	30 mins	Good	"option 1 preferred"
4.	£1.00	40 mins	Average	£1.45	25 mins	Good	"option 2 preferred"

From this it is clear that a smaller change in the trade-off rate, as one passes from the first to the second pair of options, would have given us a better estimate of the respondent's value of time. thus, given that the researcher will have some knowledge from previous research of where the boundary values are likely to lie, he or she must ensure that:

(i) the trade-offs presented to the respondents cover a sufficient range to include likely boundary values between attributes and

(ii) the trade-offs are close enough to each other to allow a sufficiently accurate estimate of the boundary values.

The researcher will need to try a number of different values for the attribute levels before he or she can be sure that the two conditions are properly satisfied. The less guidance which the researcher is able to obtain on likely boundary values, prior to designing the conjoint analysis experiment, the greater the importance of exploratory research in advance of the main fieldwork. This may necessitate a series of pilot surveys to ascertain the likely locations of the boundary values.

Returning to Figure 5.8, an example of a quantitative trade-off against a qualitative attribute is given. Estimation of the third pair of options reveals a situation similar to the first pair. The difference now is that option 2 offers a better standard of comfort than option 1. The effect has been to switch the response from option 1 to option 2. This demonstrates that there is some value attached to an improvement in comfort. More significantly, as the value of time could be as low as 3.3 pence per minute (so that the respondent may not be prepared to pay more than 33 pence for the 10 minute time saving) the minimum value for the improvement in comfort could be as much as 12 pence (45 pence minus 33 pence.) Nevertheless, from only examining this additional pair of options, the likely value for comfort is not well enough defined, because an upper limit for the value of improved comfort is not known.

An upper limit for an improvement in comfort is provided by the fourth pair, where the situation is similar to the third pair, but this time the fare increase in option 2 is 60 pence, in return for 10 minutes time saving and a higher standard of comfort. This has caused the respondent to switch back to option 1. The researcher may now conclude that the upper limit to the value of an improvement in comfort, given that the value of a 10 minutes time saving lies between 33 pence and 44.9 pence, is between 15 pence (60 pence - 44.99 pence) and 27 pence (60 pence - 33 pence.) From these four pairs a broad range of monetary values for journey time and comfort can be established, but clearly the range in this instance is wide. The inclusion of more levels, which reduce the range between trade-offs, will produce closer estimates of the boundary values.

Little has been written on the subject of defining the attribute levels in conjoint analysis experiments. Fowkes and Wardman (1988) provide a useful discussion, where they emphasise the importance of examining a number of alternative values and designs which should then be tested using simulated data (see also Fowkes, 1991 and Holden et al, 1992.) They also make the point that attribute levels should represent changes that are perceived by the respondent to be of at least some importance. Otherwise, the attribute is likely to be ignored. In the value of time study for the Department of Transport (MVA et al, 1987), changes to travel times of less than 2 minutes were ignored by respondents. A similar result was observed for savings to waiting time in a study by Steer Davies and Gleave Ltd for Leeds City Council (Steer Davies and Gleave, 1989.)

Another useful approach put forward by Moore (1985) is the use of unequal increments between attribute levels, where more than two levels are used. Consider a fare attribute with three levels: £1.00, £1.20, £1.40. this may be traded with a journey time attribute of three levels: 20 minutes, 25 minutes and 30 minutes. If all combinations of the attribute levels are used, the following trade-offs may be offered to respondents:

(i)	(£1.20 - £1.00)/(25 - 20) mins = 4p/min;
(ii)	(£1.40 - £1.20)/(25 - 20) mins = 4p/min;
(iii)	(£1.40 - £1.00)/(25 - 20) mins = 8p/min;
(iv)	(£1.20 - £1.00)/(30 - 25) mins = 4p/min;
(v)	(£1.40 - £1.20)/(30 - 25) mins = 4p/min;
(vi)	(£1.40 - £1.00)/(30 - 25) mins = 8p/min;
(vii)	(£1.20 - £1.00)/(30 - 20) mins = 2p/min;
(viii)	(£1.40 - £1.20)/(30 - 20) mins = 2p/min;
(ix)	(£1.40 - £1.00)/(30 - 20) mins = 4p/min;

Although trade-offs are offered over a range of levels, the rates of trade-off (which indicate the location of the respondent's boundary values) only vary over three values: 2p/min; 4p/min; 8p/min, because many of the fare and time differences are the same. Consider the same attributes with slightly different increments: fare has levels of £1.00, £1.25, £1.40; journey time has levels of 20 mins, 27 mins, 30 mins. The resulting wider range and larger number of trade-off rates will be:

(i)	(£1.25 - £1.00)/(27 - 20) mins = 3.57p/min;
(ii)	(£1.40 - £1.25)/(27 - 20) mins = 2.14p/min;
(iii)	(£1.40 - £1.00)/(27 - 20) mins = 5.71p/min;
(iv)	(£1.25 - £1.00)/(30 - 27) mins = 8.33p/min;
(v)	(£1.40 - £1.25)/(30 - 27) mins = 5p/min;
(vi)	(£1.40 - £1.00)/(30 - 27) mins = 13.33p/min;
(vii)	(£1.25 - £1.00)/(30 - 20) mins = 2.5p/min;
(viii)	(£1.40 - £1.25)/(30 - 20) mins = 1.5p/min;
(ix)	(£1.40 - £1.00)/(30 - 20) mins = 4p/min.

5.2.6 Using Simulated Data to Test Alternative Experimental Designs

It has been suggested that the way in which attribute levels are defined in conjoint analysis experiments will influence the accuracy with which the researcher can model individuals' preferences. As an aid to the development of an efficient conjoint analysis design, the researcher may consider the use of simulated data (Fowkes and Wardman, 1988; Holden et al, 1992; Swanson et al, 1992.) This involves the construction of an artificial set of utility values specified by the researcher. This artificial set of preferences is used to simulate responses to a particular conjoint analysis design. The researcher may then model the results and compare the implied utilities with those of the original artificial data set. When the researcher alters the conjoint analysis design and re-runs the simulation, he or she can observe the improvement or worsening in the implied utilities, relative to the known utilities. On this basis the researcher can choose the preferred combination and definition of attribute levels.

The simulation process may be summarised as follows:

(i) The researcher creates a set of artificial utility values representing the likely range of values possessed by the sample to be surveyed;

eg
```
Case 1 = 1.5 + 1.5*Cost + 6.0*Time + error (= 5)
Case 2 = 1.5 + 1.5*Cost + 6.0*Time + error (= -2)
Case 3 = 1.5 + 1.5*Cost + 6.0*Time + error (= -10)
etc...
```

(ii) The researcher then creates responses to the conjoint analysis design for each case of utility values created in (i);

eg in option A, Cost = 2, Time = 5,
in option B, Cost = 5, Time = 3,

=> Utility of A - utility of B =
(3-7.5)+(30-18)+error = 7.5+error

=> Response for error > -7.5 = 'prefer A'
(utility A > utility B)

=> Response for error = -7.5 = 'cannot choose'
(utility A = utility B)

=> Response for error < -7.5 = 'prefer B'
(utility A < utility B)

Therefore: Case 1 response = 'prefer A'
Case 2 response = 'prefer A'
Case 3 response = 'prefer B'
etc...

(iii) A model of the simulated responses is then estimated (using for example multiple linear regression or a discrete choice logit model) to produce estimates of the artificial utility values.

(iv) The researcher compares model estimates (implied utilities) against the original artificial utilities. The conjoint analysis design may then be altered and steps (ii) to (iv) repeated until the researcher is satisfied that an efficient design has been developed.

The importance of testing experimental designs for SP exercises has not always been appreciated and is rarely reported in the literature (all the applications to measuring the value of passenger facilities reported in chapter seven did test designs in this way and neither did the case study of chapter eight.)

5.3 The Presentation of Conjoint Analysis Techniques as Survey Instruments

Applications of conjoint analysis in the form of survey instruments will be seen to differ in a number of ways, which chiefly reflect:

(i) the location in which the interviews are carried out;

(ii) the time constraints likely to act on each respondent;

(iii) the required level of guidance for respondents;

(iv) the familiarity and complexity of the conjoint analysis exercise to respondents.

The survey designer will need to consider each of these factors when deciding the form which the survey instrument should take. The following options are available:

(i) The exercise may be administered by interviewers or 'self-completed' by respondents. Alternatively, a combination may be preferred. Examples include an interviewer explaining a conjoint analysis exercise and leaving the respondent to complete it; another is where the interview is conducted by telephone, the conjoint analysis options being sent to the respondent in advance.

(ii) The exercise may require one of a variety of choice tasks: ranking; rating; discrete choices.

This list is not intended to suggest a limit to the approaches to conjoint analysis that may be undertaken. Instead, it is provided to indicate the principal strategies that researchers have considered to date and provides a structure for the remainder of this section.

5.3.1 Survey administration

In considering whether to use interviewers or a self-completion approach, the factors likely to influence the decision may be:

(i) the complexity of the conjoint analysis exercises and the length of the questionnaire;

(ii) the detail with which alternatives need to be described;

(iii) the circumstances in which interviews would take place or questionnaires would be completed;

(iv) the ways in which individuals would be able to respond.

Note that the first consideration will interact with the remaining items, such that a proposed experimental design may need to be simplified and the questionnaire shortened in the light of the remaining factors, regardless of how the survey will be administered. Nevertheless, each item has a particular influence on the type of administrative approach. One further practical consideration is the geographical spread of the sample to be contacted. The more diverse the locations of the respondents, the higher the cost of sending interviewers out. Postal charges will remain the same, at least within the United Kingdom, so that this form of survey becomes more attractive as the geographical area to be covered increases.

The more complex and lengthy the conjoint analysis experiment and questionnaire, the greater the need for respondents to be guided through the stages of the exercise. Equally, the more complex or unfamiliar the alternatives to be considered in the conjoint analysis choices, the more description is required. Both these issues point to the value of interviewer-administered surveys. Interviewers can explain detailed aspects of the exercise and 'customise' the values used in the conjoint analysis choices to make them relevant to respondents' current experiences. Skilled interviewers will be able to probe respondents to ascertain their level of understanding concerning the alternatives.

As most conjoint analysis experiments are fairly complex (with generally at least three attributes presented over eight or nine options), the use of interviewers is preferable. Nevertheless, the remaining considerations listed above can support a case for self-completion exercises. Interviewers placed on public transport vehicles, for example, can target particular respondents (eg bus users) but are not likely to have sufficient time or comfortable circumstances in which to undertake the conjoint analysis exercise. In such a case, it may be better to ask respondents to take a questionnaire and complete it in the more appropriate environment of their own homes or work places.

Another advantage of self-completion questionnaires is that they offer a lower unit cost per usable questionnaire. Their main drawback is a lower response rate than for interviewer-administered surveys, in which sizeable groups (eg pensioners, people with low standards of literacy) may be poorly represented. Perhaps the most appropriate use of self-completion surveys is in circumstances where respondents are likely to be strongly motivated regarding the subject being studied and are likely to be confident in attempting to complete the conjoint analysis design. A number of examples of successful self-completion conjoint analysis surveys have been carried out. Notable examples are the surveys carried out as part of the Department of transport's Value-of-Time study (Wardman, 1988), a study of rail season ticket holders (SDG Research, 1989) and a series of large scale travel surveys reported in Kocur et al (1981.)

While interviewer-administered conjoint analysis surveys are to be preferred over self-completion surveys, a number of conditions will need to be satisfied if interviewers are used. An essential feature must be the use of personnel who are well-trained and well-informed as to the purpose and mechanics of the survey. Particular care will be necessary in 'customising' the options to respondents' present situations, although well designed computer-based applications will remove the need for interviewers to carry out this often complicated operation. Where computers are not used, it is the usual practice to have interviewers write in values on each card, according to the features of respondents' present choices and pre-specified levels of change.

One alternative approach is the use of pre-constructed choice sets, which vary over the most likely range of attribute levels. Respondents can simply be given the choice set that comes closest to their present situations. There is obviously some imprecision in this approach, but it can be useful in situations where existing attribute levels are confined to a limited number of values. In postal self-completion surveys the values must of course be pre-set, unless the researcher has been able to contact the respondent in advance, to obtain details upon which the attribute levels can be based.

5.3.2 Conjoint Analysis Interview Design

The form in which the conjoint analysis survey is presented to respondents will have some relevance to the ease and accuracy with which they interpret and assess the options. The use of printed cards is particularly useful when respondents are to rank the options because they will be able to physically arrange them in order of preference. If paired choices are being offered, separate cards are also useful to define each option individually, or at least each pair of options separately. The order of the pairs and/or the combinations of options within the pairs can then be varied randomly before each interview, so avoiding any systematic bias. This format also helps to emphasise that paired options are to be assessed independently from one another, which is an important assumption in the analysis.

Paired options may alternatively be presented in a 'fixed format' as an integral part of a questionnaire. Figure 5.9 presents an example from SDG Research (1989.) This approach involves fewer material components than one using separate cards, which is perhaps attractive for self-completion surveys. Unfortunately, the order of choices is not varied, unless the researcher takes pains to produce different versions of the questionnaire with options ordered differently. This still does not avoid the possibility that respondents may link the different paired choices, attempting to discern some pattern as they progress through the exercise.

Figure 5.9 Paired SP Options Presented in a 'Fixed' Format (SDG Research, 1989)

Choice 1

	Option A			or	Option B			Which option would you choose? (please tick one box only)					
	Price of Your Ticket	Type of Train	Type of Timetable		Price of Your Ticket	Type of Train	Type of Timetable	Definitely A	Probably A	No preference	Probably B	Definitely B	I would stop travelling by rail
Q10	£1740	Improved Network SouthEast	Limited Timetable		£2088	High Speed Train	Full Timetable	5	4	3	2	1	6

Choice 2

	Option A			or	Option B			Which option would you choose? (please tick one box only)					
	Price of Your Ticket	Type of Train	Type of Timetable		Price of Your Ticket	Type of Train	Type of Timetable	Definitely A	Probably A	No preference	Probably B	Definitely B	I would stop travelling by rail
Q11	£3132	High Speed Train	Full Timetable		£1740	Current Network SouthEast	Limited Timetable	1	2	3	4	5	6

Choice 3

	Option A			or	Option B			Which option would you choose? (please tick one box only)					
	Price of Your Ticket	Type of Train	Type of Timetable		Price of Your Ticket	Type of Train	Type of Timetable	Definitely A	Probably A	No preference	Probably B	Definitely B	I would stop travelling by rail
Q12	£2610	High Speed Train	Full Timetable		£1740	Improved Network SouthEast	Limited Timetable	1	2	3	4	5	6

Choice 4

	Option A			or	Option B			Which option would you choose? (please tick one box only)					
	Price of Your Ticket	Type of Train	Type of Timetable		Price of Your Ticket	Type of Train	Type of Timetable	Definitely A	Probably A	No preference	Probably B	Definitely B	I would stop travelling by rail
Q13	£1740	Current Network SouthEast	Limited Timetable		£2088	High Speed Train	Full Timetable	5	4	3	2	1	6

Choice 5

	Option A			or	Option B			Which option would you choose? (please tick one box only)					
	Price of Your Ticket	Type of Train	Type of Timetable		Price of Your Ticket	Type of Train	Type of Timetable	Definitely A	Probably A	No preference	Probably B	Definitely B	I would stop travelling by rail
Q14	£1740	Current Network SouthEast	Limited Timetable		£1914	High Speed Train	Full Timetable	5	4	3	2	1	6

For ranking exercises, it is suggested that a 'fixed format' approach is less suitable than using separate cards, as without the opportunity to physically establish a rank-ordering, the task of relating the options to one another becomes much more difficult for the respondent.

Illustrations may be a considerable advantage when presenting qualitative attributes, especially when they represent a combination of different factors. Visual stimuli will be more rapidly understood than verbal descriptions. Nevertheless, the researcher needs to pay special attention to the choice of illustrations. These should highlight the factors of interest, with a minimum of distraction from other items. For this reason, photographs may be less effective than graphic art work, because they can include irrelevant but distracting details.

There is no conclusive evidence that the use of visual aids gives different responses than when verbal descriptions are used in a conjoint analysis exercise. Louviere et al (1987) compared results from one conjoint analysis survey, in which a proportion of respondents were given descriptions in verbal form while the rest received combinations of verbal and visual descriptions. No significant difference was detected between the two groups' responses. Bradley and Bovy (1986, p296) nevertheless suggest that photographs help to clarify qualitative attributes and indeed Louviere et al (1987) do not claim their findings to be conclusive. It is clear that some factors will benefit immeasurably from illustration (eg alternative colour schemes for private transport vehicles; space provision on board vehicles.)

5.3.3 Defining the Choice exercise

An important goal of all conjoint analysis designs should be to achieve as much realism as possible in the choice exercises. This relates to the chief criticism that may be levelled against the use of SP data, namely that it represents statements about behaviour in response to hypothetical choice situations. The researcher is obliged to present choice situations that represent closely the type of choices that respondents may face in reality.

As a guide to the main points which the researcher should consider, the following quote from Jones (1989, p10-11) is of particular use:

> "Improved realism can be built into the **context** of the exercise, the **options** that are presented and the **responses** that are permitted, in a number of ways:
>
> (i) Focusing on very specific rather than general behaviour; for example, asking how respondents would have reacted to a new product on a given occasion, rather than in general.
>
> (ii) Using a realistic choice context - one in which respondents have had personal experience, or feel they might be placed.
>
> (iii) Using existing (perceived) levels of attributes within the exercise, so that options are built around existing experience, where this is feasible.
>
> (iv) Using respondents' perceptions of what is possible to set limits on the attribute ranges to be varied in the experiments (eg if looking at improved bus services, don't include options where the bus is closer to home than is physically possible.)
>
> (v) Wherever possible, incorporate checks on the answers given (eg that the reported walk, wait and in-vehicle times approximately match the door-to-door time.)
>
> (vi) Allow for the effect of day-to-day variability on choices (eg when considering the importance of waiting time.)
>
> (vii) Ensure all relevant variables are included in the analysis - especially important when developing travel choice models and not simply measuring the relative importance of attributes.
>
> (viii) Simplify the presentation of the choice options and the nature of the choice process that is being used (eg by highlighting the attribute levels which have changed between options.)
>
> (ix) Ensure constraints on choice are taken into account (eg fixed arrival time at work) - again especially important in travel choice applications.
>
> (x) Allow respondents to opt for a response outside the set of the experimental alternatives (eg in a mode choice exercise, if all the options become expensive then in reality the respondent may decide not to travel at all - so in a pair-wise choice allow a 'neither' response.)"

If all these requirements are to be met, the emphasis is on the use of interviewers for all but the simplest of conjoint analysis exercises. Nevertheless, even the most capable interviewers are likely to find some of the above tasks difficult. This is one reason why portable computers are being used increasingly for conjoint analysis interviewing.

5.3.4 Piloting the Survey Instrument

While the use of simulated data, discussed earlier, will usefully test the efficiency of alternative conjoint analysis designs, it will not provide reliable guidance on the way respondents will **actually** respond. At least one 'pilot' survey is therefore essential. This will test not only the suitability of the experimental design, but also the adequacy of the way in which it is presented (the clarity of the guidance provided; the explanatory power of the visual aids used.) A pilot survey will also highlight practical management issues, such as the likely response rates that will be achieved and the proficiency of any interviewers used to carry it out.

The pilot survey should ideally involve lengthier interviews that those planned for the main survey, in which respondents are probed in depth about their understanding of the exercise presented to them. A particularly thorough approach is reported by Olshavsky and Acito (1980), who employed 'protocol analysis' to investigate the choice processes behind individuals' responses. This involves principally the progressive questioning of respondents as they work through a series of conjoint analysis exercises. It is useful to pilot a self-completion questionnaire by having respondents complete it in the presence of the researcher, with the opportunity to comment at any stage.

Another established development procedure is the use of 'group discussions' with individuals who have the same characteristics as those likely to take part in the main survey. Group discussions require an experienced group moderator who can encourage the selected group of individuals to discuss a number of topics relevant to the study. Though the information obtained is highly qualitative in nature it can provide detailed information on the decision processes behind choice behaviour. It will also help the researcher to identify the kind of phrases suitable for describing the SP alternatives to respondents. The researcher may also test the questionnaire and conjoint analysis design at such a meeting, with the opportunity to discuss possible improvements.

5.3.5 Sampling

Having decided on the design of the survey instrument and tested it, the researcher must also determine the size and structure of the sample of individuals to be interviewed. The issues concerning sampling for conjoint analysis surveys are largely the same as for other market research surveys. The researcher is concerned with obtaining a representative group of people in the area with which he or she is concerned. The researcher needs to identify suitable sub-groups (or 'segments') of the population of interest and seek to obtain sufficient numbers in each. Main segments for a travel survey may include the mode used, trip purpose and car availability. Once representative numbers of interviews have been obtained for each segment, the models developed from them can be applied to these segments in the population as a whole, using other data such as the Census to weight them. Standard sampling theory (Bhattacharyya and Johnson, 1977, p250) suggests that a minimum of 30 respondents in each segment should be interviewed, though representativeness is greatly enhanced by higher numbers.

This issue of a minimum segment size is very important. The first issue relates to the danger of ascribing too much importance to small samples in conjoint analysis surveys. This can result from the fact that conjoint analysis exercises produce multiple responses. For example, 20 individuals could respond to 10 hypothetical situations, giving 200 observations. This might be sufficient to produce a model with robust goodness-of-fit statistics. Nevertheless, these 200 observations cannot be regarded as equivalent to the single real choices of 200 individuals. Instead, they only represent the **opinions** of 20 individuals.

A more serious development, which challenges the assumption of a minimum sample of 30 being adequate for each segment, is the view put forward by Bradley and Kroes (1990a, p13):

> "It seems to have become an unwritten law in transport SP research that the minimum acceptable sample size is 30 to 50 respondents in each market segment for which separate models will be estimated. While this size may often be adequate for SP preference tasks[1], the use of SP choice[2] tasks introduces greater variability into the data from exogenous factors and, as a result, larger samples are needed for reliable estimates. As a rough estimate, at least 75 to 100 respondents in a segment are necessary to give adequate estimates from SP choice data - depending somewhat on the heterogeneity of the market and the number of responses obtained per individual."

Such observations relating to sample sizes appears to be substantiated by more recent work (Swanson et al, 1992) examining discrete choice designs. This has important implications for the commercial attractiveness of discrete choice conjoint analysis studies. This is because it reduces one of its advantages over revealed preference: the ability to use smaller samples and therefore to produce more cost-effective surveys. Further empirical work is required to establish the relationship of survey accuracy to sample size, but the indication is that the comparatively lower information content of discrete choice data relative to ranking and rating (preference) data needs to be compensated by larger samples.

1 Ranking and rating tasks, in which respondents express judgements only.

2 Discrete choice tasks.

One further factor to consider is the importance of "choice-based" sampling. The structure of the sample should take into account the current travel behaviour of respondents, an issue implicitly raised in the above discussion on segmentation and sample sizes. For example, a conjoint analysis study of traveller priorities towards small-scale design improvements to buses need only be targeted at current bus users, because such issues are outside the present range of experience for non-bus users and are marginal to the mode choice process. However, a study investigating reductions in bus fares would achieve a more representative picture if non-bus users were also interviewed, as this would give some indication of the way other travellers might switch to the bus.

5.4 Applications With Micro-computers

The use of computers to administer conjoint analysis interviews goes back as far as 1980 (Johnson, 1980), but it is only recently that their use has become widespread (Ampt et al, 1986; Jones et al, 1987; Bradley, 1988; Polak et al, 1989.) The recent development largely reflects the increasing cheapness and quality of portable computers.

Most of the commercial software developed for interviewing with computers translate 'conventional' (ie non-computer based conjoint analysis) questionnaires directly onto the screen. They allow multiple-choice responses, numerical and alphanumeric input and different routing through the questions. For the conjoint analysis options specifically, they tend to be limited to pair-wise choices or rating exercises. Ranking exercises, in which all the options are presented at once are difficult to display on the screen, but they may be inferred through paired comparisons or ratings (Fowkes and Tweddle, 1988.) At the time of writing, the principle software available in this country are "ACA" (Sawtooth Software); "Alastair" (Steer Davies Gleave); "IASP" (Institute for Transport Studies); "MINT" (Hague Consulting Group) and "SP_ASK" (Peter Davidson Consultancy.) Not all these packages are 'adaptive', a major development in computer applications which is discussed below.

5.4.1 Advantages

Perhaps the most obvious advantage of using computerised interview techniques is that responses are recorded immediately, so that data entry costs are eliminated completely. Some programs incorporate a data analysis feature so that results may be presented as data is recorded. In such cases the capability is usually limited but provides useful summaries of initial results.

Steps may also be taken within the computer questionnaire design to ensure that the quality of data recorded is much higher than with manual techniques. Logical routing through the questionnaire, error traps and range checks can all be incorporated in the interview program. A consequence of this is that interviews may be more complex than would be practicable with manual techniques.

These advantages might apply to any computerised interview system. As far as conjoint analysis is concerned, there are a further number of specific and significant advantages. It is usual practice to tailor a conjoint analysis interview so that it is based on the direct experience of the respondent (see 5.3.3 above.) Although a fixed experimental design may be used, the levels for each attribute will usually be based on those values reported by the respondent. In 'show card' exercises this means that the interviewer has to calculate the values to use and write them on the cards. This will take time and introduces a greater likelihood of error. In contrast, a computer will carry out the calculations almost instantly and without error.

In some surveys it may be necessary to use several different experimental designs to suit the circumstances of the respondent. For example, in a mode choice exercise different designs may be needed for each base mode available. This is burdensome for an interviewer using cards, while a computer-based questionnaire may have as many designs available as are required. A similar advantage is to be found with the use of 'blocked' experimental designs, where only a portion of the overall design is to be shown to each individual.

The computer may be instructed to choose randomly between the different blocks, so ensuring an even presentation of the blocks over the whole sample. A computerised approach can also automatically randomise the order in which SP options are presented and, in the case of the "MINT" system (Hague Consulting Group, 1990), the order in which the attributes are presented on the screen.

Experience suggests that the use of a well designed computerised interview stimulates greater interest on the part of respondents (Ampt et al, 1986, p15.) The usual approach is for interviewer and respondent to observe the computer screen together. It appears that the computer is sometimes regarded as a third, impersonal participant in the interview, so that the respondent does not associate the question being put with the interviewer. The effect seems to be that respondents are more willing to answer personal questions about income, or age, since they are being 'asked' by the computer, not another person.

To summarise, computer-based approaches offer a number of practical advantages:

> "(i) an interesting and flexible presentation format;
>
> (ii) a format which is consistent across interviews and respondents;
>
> (iii) automatic question branching and prompting;
>
> (iv) automatic data coding and storage;
>
> (v) the ability to incorporate checks to avoid inconsistent or wrongly entered answers."
>
> (Bradley, 1988, p134)

5.4.2 Disadvantages

Two principal disadvantages may be identified with the use of computers for conjoint analysis interviewing. The first is the cost of buying the computers and the second is the investment in software development. Normally, portable computers must be used, although a 'hall test' environment for a survey could allow larger machines to be used. Light and powerful machines may now be purchased for under £1000, with some new models claiming 30 to 60 hours battery life.

Software development or purchase may also be a significant investment. Although many early computerised conjoint analysis studies were carried out using software written specifically for them, this is not economical and most interview software packages are written at a 'general' level so that they can handle many different interview types and conjoint analysis formations.

These initial investments, once made, may be discounted over the number of surveys that are subsequently carried out since no further costs are involved, other than the normal outlays for fieldwork. In the long run, they may prove to more cost effective than conventional surveys, due to reductions in the cost of questionnaire production, data entry and the average length of interviews. Market research interviewers generally make the conversion from conventional to computer-based surveys successfully. Nevertheless, one potential problem is for the interviewers to become used to the computer structuring the interview and therefore to neglect to record unusual responses or comments which they would normally note down in a conventional interview. For this reason it is advisable to allow the interviewers the opportunity to record 'open-ended' answers within the questionnaire framework.

A small but obvious constraint in the use of computers to administer SP surveys is the size of the screen (note the inappropriateness, for example, of trying to replicate a ranking exercise as it would be constructed using cards.) The use of colour is a possible enhancement, but this requires investment in more expensive computers.

5.4.3 Adaptive SP Techniques

The above discussion outlined the presentation on portable computers of conjoint analysis exercises that would normally be administered by an interviewer using a conventional questionnaire, albeit developed to a more complex level. Computer-based conjoint analysis techniques begin to offer much more significant advances when they allow the choice presented to respondents to be constructed on the basis of earlier responses made by the same individual. To some extent, this may involve departure from orthogonal experimental designs. Adaptive SP techniques require sophisticated software because the researcher needs to specify complex procedures for 'adapting' the generation of new choice situations on the basis of previous responses.

One approach to adaption is to infer rankings of SP options from paired choices. As already mentioned, it is unfeasible to display more than two or three options from an experimental design on a computer screen, so that a direct ranking of options as in a conventional exercise would be impractical. The alternative is to have respondents choice between every possible pair of options in the SP experimental design. Unfortunately, this soon becomes impractical: a design with eight options would require 28 pairs; a design with nine options would require 36 pairs. With an appropriate adaptive computer algorithm, this problem can be reduced to a manageable size. The "Alastair" software developed by Steer Davies Gleave and the "MINT" software developed by Hague Consulting Group offer such an approach, though to the knowledge of the author other commercially available packages do not.

Firstly, the software used to present the SP exercises can monitor for pairs with one option that clearly dominates and a logical preference between options is therefore obvious. For some attributes, the order of preference between levels may not be obvious, so that it is useful to have respondents state their order of preference in advance of the SP exercise. This approach is used by the "Alastair" package. Secondly, the software can monitor implied dominance. Pairs may be identified where preferences can be inferred from information already given by the respondent. These pairs do not therefore have to be presented.

With the "Alastair" software, a respondent who chooses entirely logically will not be shown more than eight or nine pairs for a design of eight options. However, most individuals are likely to show some inconsistency in their response. Where this is detected by the software, additional pairs may have to be shown in cases where implied dominance can no longer be established. The software can continue until a full set of rankings have been produced or some pre-defined cut-off point is reached. An example of adaptive ranking is presented in chapter eight.

Another adaptive approach can be considered for use in 'choice' designs, where in a pair of options the packages presented relate to different things (eg car versus bus, as opposed to different versions of the same bus journey, which would be the case in a ranking exercise.) The advantages of an adaptive SP technique in this case are potentially great.

To illustrate this, figures 5.10 and 5.11 respectively compare a conventional approach using a fixed experimental design with an adaptive approach. In the first diagram, it can be seen that a respondent may only 'trade-off' between a few of the options. For much of the design, the response is predictable: one alternative (rail) easily dominates the other (coach.)

The problem here is that the range of fare and time used allows little scope for the respondents to switch from rail to coach. Careful testing with simulated data and pilot surveys as described earlier, will help the researcher to produce designs which partly overcome this problem. Nevertheless, in each case some proportion of the choices will be redundant, because the area covered by the different attribute combinations will generally be rather broad. A much more efficient approach would be one that allowed the researcher to 'home in' on respondents' trade-offs. Such an approach is suggested in Figure 5.11, where a number of steps lead on from the illustration in Figure 5.10.

Figure 5.10 Example of a Set of SP Options Based on a Conventional Design, Placed Against a Particular Boundary Value.

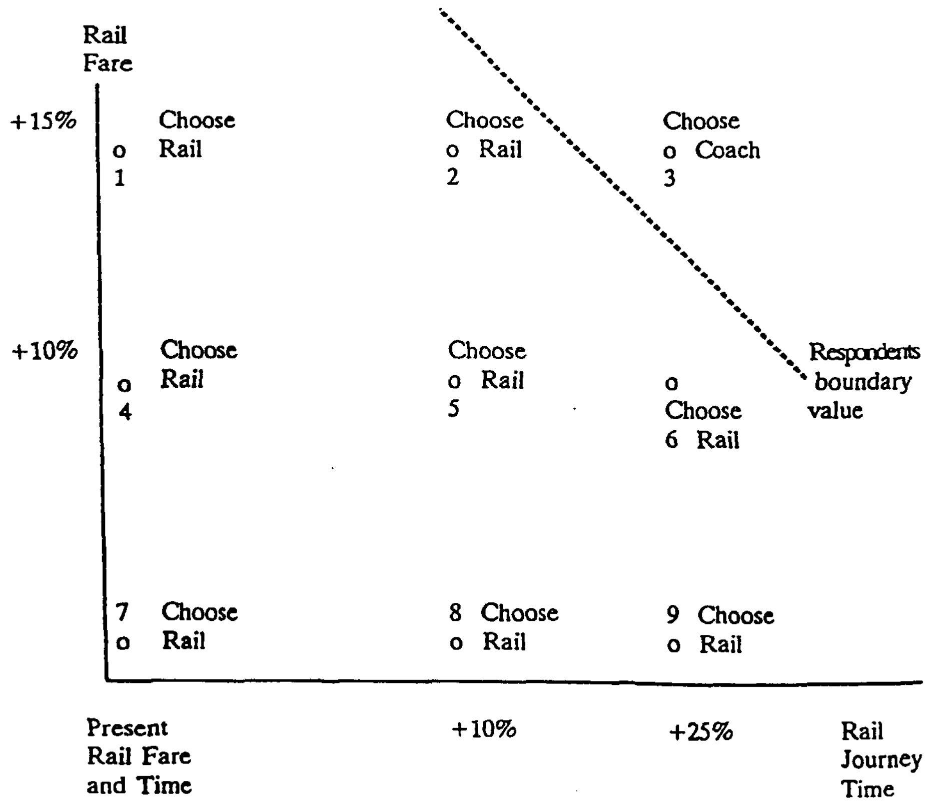
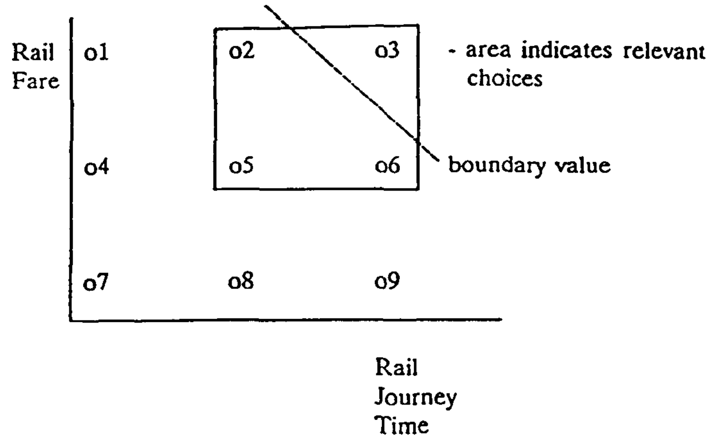
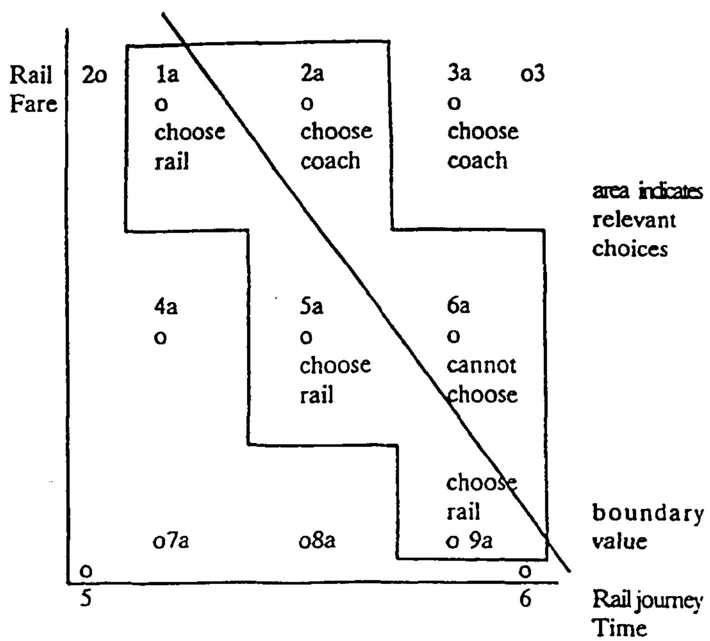


Figure 5.11 Example of a Set of SP Options Based on an Adaptive Design, Placed Against a Particular Boundary Value.

Step 1



Step 2



Further steps would be taken to present an even closer-spaced sub-set of attribute combinations, or sub-sets that move further along the line of the boundary value between coach and rail choices.

A computer program that could select sub-sets of relevant attribute combinations in this way, while ignoring irrelevant combinations, would produce responses that define the boundary values much more precisely than a conventional approach. This adaptive approach can be generalised to a larger number of attributes, all being traded against one another. With an adaptive approach such as this, respondents may only have to consider 10 or so choice situations, similar to the amount required by a conventional fixed experimental design approach. The difference will be that each observation will yield more information on boundary values.

Software that is capable of carrying out this kind of adaptive approach is still very much at the development stage, though applications are growing in number. Polak et al (1989) reports on an approach which establishes a large 'grid' containing a wide range of attribute combinations. The computer program then offers combinations near the limits of this grid. It uses an algorithm which allows the choices to progress by selecting sub-sets of combinations within the grid, using preceding responses to select each sub-set, very much in the way discussed above. This ensures the presentation of choice situations that are always likely to be competitive.

A further development in adaptive techniques is to use some statistical modelling procedure on which to base new choice situations. Johnson (1985) developed a procedure which created choice situations on the basis of results from a multiple regression routine included in the survey program. As respondents progress through a series of situations, the multiple regression program develops a model of the respondent's preference structure. Further choice situations are then created with an aim to focus on those attributes for which coefficients are statistically weak within the regression model. Bradley (1988) reports on an application of the technique for examining the travel behaviour of urban bus passengers. A rating scale to describe the degree of preference between pairs of choices is necessary for the regression procedure, though this can be a simple five-category response.

An experimental approach investigated by Steer Davies Gleave (1990) also uses regression to adapt choices, but the dependent variable is defined by one of the attributes, against which all others are regressed. The individual's responses provide the data points around the boundary 'surface' between two competing alternatives (eg coach v train) which the regression attempts to estimate. The main problem experienced with this approach has been the increased complexity of the analysis required. Correlations between attributes are likely to be stronger because of the departure from an experimental design format and the analysis of groups of individuals becomes difficult. This is because the responses given are no longer attached to a common choice set: the adaptive algorithm allows a very wide range of options to be presented.

Adaptive SP techniques represent the most advanced area of conjoint analysis and as such have still to be established as the preferred approach. Conventional conjoint analysis, administered by post, interviewer and increasingly by portable computer still accounts for the great majority of conjoint analysis studies in transport.

5.5 Conclusions

This chapter has reviewed the principal issues relating to the design and application of conjoint analysis surveys. It can be seen that the use of such techniques requires a number of critical issues to be considered, for which in some cases the judgement of the researcher must cover for the absence of rigorous guidelines. This is particularly the case concerning issues of presentation.

To summarise, the researcher must consider the following issues when embarking on a conjoint analysis study:

(i) the attributes to be examined;

(ii) the number and definition of attribute levels;

(iii) the appropriate response task (ranking, rating or choices);

(iv) compromises necessary in the construction of the attribute packages;

(v) testing of the experimental designs and the presentation materials (simulated data; qualitative/literature research; pilot surveys)

Once assured that an effective survey instrument has been developed, the researcher must also decide on a rigorous method of analysing the SP data. This issue is discussed in the next chapter.

6: THE ANALYSIS OF CONJOINT ANALYSIS TECHNIQUES

6.1 Introduction

The previous chapter discussed in depth the issues relating to the design and application of SP methods, specifically conjoint analysis techniques. The use of experimental designs and the collection of multiple observations from each respondent allows for a degree of flexibility in the analysis of conjoint analysis data. This chapter examines the underlying principles which guide the main analytical approaches currently in use. A review of the most common analytical methods is made, beginning with the simplest "graphical" methods and progressing towards the most advanced stochastic modelling techniques. Given the importance of discrete choice models, in view of the technical advance they represent over other methods, a section is devoted specifically to the theoretical issues relating to their application to SP data. Finally, a short discussion covers issues relating to the reliability of SP data and the models that can be derived from them.

6.2 Objectives in the Analysis of SP Data

Conjoint analysis techniques can borrow much from analytical approaches developed from revealed preference (RP) data and most applications use modelling procedures well established in disaggregate data analysis. The aim of the analysis will be to 'decompose' the preferences expressed in the SP data into 'part-worth' utilities attached to each attribute presented in the choice exercises. As a basis for this analysis, the researcher needs to infer a utility construct. Using the example of a simple linear additive utility model, the part-worth utilities can be identified with the model coefficients that may be estimated:

$$U_i = a_1X_1 + a_2X_2 + \ldots.. a_nX_n \qquad (6.1)$$

where U_i = utility of option i;
$X_i \ldots X_n$ = attributes;
$a_i \ldots a_n$ = model coefficients = part-worth utilities.

The part-worth utilities yield the following information:

(i) They indicate the relative importance of the attributes.

(ii) They infer relative monetary and other quantitative values if attributes such as cost and travel time are included. The established value-of-time measure can be deduced. This is enabled through the identification of an indifference curve between two or more attributes. For example, equilibrium is implied if:

c.COST = t.TIME

where COST, TIME = values of attributes presented in the conjoint analysis exercise;

c, t = coefficients of COST and TIME respectively; (=> COST/TIME = t/c)

(iii) They specify utility functions to be used for forecasting models, based either on the present survey data or a comparable database.

6.3 Practical Approaches for Model Estimation

As conjoint analysis has developed, a wide range of approaches have become available for analysing the SP data which it yields. Pearmain et al (1991, p72) suggest that they may be categorised as follows:

(i) 'Naive' or graphical methods;
(ii) Monotonic Analysis of Variance;
(iii) Multiple regression techniques;
(iv) Discrete choice models, principally 'Logit' and 'Probit' models.

Before considering each of these types of analytical procedures, it is important to consider the quality of information that may be provided by a conjoint analysis exercise, as this has an important bearing on the type of analytical technique that may be considered appropriate. The scale on which the responses from a conjoint analysis exercise are measured need to be assessed in relation to a hierarchy of measurement scale types defined by their level of information content.

This hierarchy of scale definitions may be summarised as:

(i) nominal data - the response is simply classified in one of a number of alternative categories, or discrete choices;

(ii) ordinal data - responses indicate the order (ranking) of preferences between options through their rank ordering, but no inference can be made about the degree of preference of one option over another;

(iii) interval data - responses indicate not only order of preference but also degree of preference between options. For example, a score of 3 is two intervals higher than a score of 1. This is not to say that a score of 3 has three times as much value as a score of 1;

(iv) cardinal data - responses indicate order and absolute degree of importance relative to some base value, so that a score of 3 can be regarded as three times as important as a score of 1. Hence the term ratio scale.

The higher the position which the researcher can assume the response scale to inhabit, the greater the flexibility in his or her choice of analysis technique. The danger lies in attributing to a response scale too high a place in the above hierarchy and therefore applying an analysis technique which cannot be strictly supported by the quality of data being investigated. As responses to most conjoint analysis are merely nominal or ordinal at best, analytical techniques that implicitly recognise the paucity of the information provided by the responses are to be favoured.

'Naive' (or Graphical) Analysis

Much information on respondents' priorities and choice behaviour can be inferred from SP data without recourse to advanced statistical procedures. Many experimental designs ensure that every attribute level is presented with similar frequency across the hypothetical situations presented to respondents. The researcher is therefore able to directly observe a relationship between the response variable and the variations in those attribute levels. One basic approach for analysing responses defined on rating scales is outlined in Kocur et al (1981, p55-61):

(i) "We can calculate how a factor in the experiment influences that person's choice by taking the difference between the mean scores a person gives to each level of that factor."

For example, attribute (or factor) 'A' has two levels which appear equally frequently across an experimental design offering eight hypothetical choice situations. Respondents indicate preferences on a five-point scale where 1=low opinion; 5=high opinion. The mean difference between scores is calculated as follows:

Level of Attribute	No. of Situations	Scores	Total Scores	Mean Score	Difference in Mean Score
Poor	4	1+3+1+3	8	2.00	
Good	4	4+3+4+4	15	3.75	1.75

(ii) "The difference in the mean scores a group of respondents gives to two levels of a factor in an experiment is the coefficient of the factor when treated as a variable in that group's utility function." Such a function will be linear additive:

$$\text{Response} = a_0 + {}_{i=1}\Sigma^n a_i . X_i \qquad (6.2)$$

where a_0 = constant
a_i = (difference in mean score)/(change between levels)
X_i = attribute i

(iii) "The coefficient of a variable describes the change in utility with respect to that variable, holding everything else constant. In other words,

$$\text{Coefficient} = \delta \text{ Response} / \delta \text{ Variable} \qquad (6.3)$$

The constant is derived from the following equation:

$$\text{Constant} = \text{Average Response} - {}_{i=1}\Sigma^n a_i . Z_i \qquad (6.4)$$

where a_i = (difference in mean score)/(change between levels)
Z_i = average value of attribute i."

The resulting utility function prefigures the type of utility construct that may be estimated through multiple linear regression. Although the method is described for a rating scale response, average scores representing rankings or frequency of choice could be substituted. The advantage with such a 'naive' method as this is the simplicity and transparency of the calculations. The weakness, of course, is that no error theory is incorporated in the analysis. The researcher cannot introduce the concept of random utility or test the accuracy of the estimation technique, other than to observe its reliability in providing utility functions with which the researcher may attempt to reproduce the original response scores. From such an approach as this, multiple regression is the next logical development.

Monotonic Analysis of Variance

In early applications of conjoint analysis, in which rankings were the most commonly collected SP data, a number of estimation techniques were developed which use a heuristic approach to find a utility function consistent with the rankings. This 'trial and error' approach could be based on a number of alternative paradigms, reflected in a number different computer programs which became commercially available to researchers. PREFMAP and LINMAP are common examples of this type of technique, but the most widely applied for conjoint analysis has been MONANOVA (MONotonic ANalysis Of VAriance), reported in Kruskal (1965.)

In this approach an iterative algorithm is used, in which the first iteration is the solution by a 'naive' form of analysis similar to that outlined above (in which the dependent variable is scaled rankings.) If the part-utilities from this first iteration can reproduce the original rankings, the results are retained and the algorithm finishes. If the rank order produced by the first iteration are not the same as the original rankings, the part-utilities are systematically varied in order to improve the correspondence between predicted and observed rank order until some optimum is reached. The way in which the algorithm is able to improve the predicted rankings can be quantified by a measure of 'stress' (ie the difference between predicted and observed rankings.)

The problem with this approach is that the only measure of goodness-of-fit ('stress') is a relative one. Worse still, it is possible for more than one type of utility function to reproduce the same set of rankings and no measures will be available to determine which function is the more robust. Finally, Monotonic analysis is restricted to linear additive utility functions, which constrains the options open to the researcher.

For these reasons, together with the decrease in conjoint analysis applications using ranking responses, MONANOVA and similar packages are now rarely used. One advantage that can be cited for this type of analysis is the fact that estimation has to be conducted for each respondent separately, so providing individual utility functions and removing the effect of inter-personal taste variations. Such disaggregate information is difficult to obtain from more sophisticated methods, though as Carmone et al (1978, p300) observe: "Non-metric algorithms, such as MONANOVA, are expensive to use in scaling individual respondent data that may involve sample sizes of 500 or more".

Multiple Regression

Multiple Regression techniques are widely established in transport research. In the analysis of SP data, the principle is to use 'least squares' to breakdown preferences into part-utilities for each attribute level presented in the conjoint analysis survey. Commonly, the response variable is a metric scale as described for 'naive' methods above, but transformed rankings have also been used. Discrete choices may conceivably be analysed on the basis of the distribution of preferences over categories, but this requires moving towards a probabilistic model of the log-linear variety and therefore belongs to the logit/probit types of model discussed below. 'Ordinary least squares' (OLS) is the most commonly applied multiple regression technique, though some conjoint analysis studies (Bates, 1984) report the more advanced application of 'weighted least squares' and 'generalised least squares'.

Multiple regression has the attraction that it is widely available as a component of popular statistical software packages, such as SPSS and SAS. It often represents the limit of knowledge for many transport practitioners in their familiarity with statistical models based on RP travel data. Multiple linear regression is based on well established statistical theory and provides a range of goodness-of-fit statistics which guide the researcher in his or her pursuit of a reliable estimate of utility functions. However, it assumes that the response scale occupies a high place in the hierarchy of measurement scales discussed above, having the at least the properties of an interval scale.

A crude regression approach for analysing rankings is to interpret the ranks as scores and analyse them as an interval scale. This requires the researcher to make an assumption about the intervals between the rank positions, which cannot easily be tested. A more appropriate approach for ranked data, suggested by Moore (1985, p30-31), is to count the frequency with which each alternative is ranked first. Provided no dominant options are included in this analysis, a numerical scale of relative importance can be derived from the frequency, which is then regressed against the attributes in the exercise. As this only uses the first choices and ignores the lower rankings, this approach does not utilise the full value of SP data, with its multiple observations for each individual respondent. More sophisticated regression approaches are available using the logit model as the basis of the analysis and this is discussed below in connection with discrete choice models.

Discrete Choice Models

In recent years, considerable advances have been made in the development of analytical packages capable of estimating utility functions from discrete choices (that is, nominal data, the simplest form of measurement scale.) This has led to the now wide application of the 'logit' model, which can be estimated with advanced computer packages as ALOGIT, BLOGIT and MVLOGIT, all of which use the statistical principal of maximum likelihood estimation. The logit formula aims to predict the probability with which a group of individuals will choose one option from all the options available (see appendix one.)

The key assumptions which underpin the application of the logit model are:

(i) the error component of random utility is independently and identically distributed across each of the alternatives;

(ii) the error terms are scattered according to a double negative exponential 'Gumbell' (sometimes incorrectly referred to as 'Weibull') distribution, which produces the distinctive logit curve.

(iii) the error term is small relative to the size of the model coefficients, so that a large portion of the responses can be attributed to the independent variables and not random error.

The more general method offered by 'probit' models removes the assumption of a Gumbell distribution of errors and instead assumes the less obscure normal distribution. Unfortunately, it requires much more complex calculations than even logit models and consequently commercial software is scarce. Developments in the direction of a widely available and efficient probit modelling package may result in this model form becoming the preferred approach to the analysis of discrete choice data, but for the present the logit model remains the main option.

The logit model represents the most flexible approach to analysing conjoint analysis data for the following principal reasons:

(i) it can analyse a full range of responses, from simple categories to rankings and rating scales (subject to modification);

(ii) it can analyse a wide range of utility function forms, not just linear additive functions as is the case for MONANOVA. Such forms include interaction effects and quadratic variables;

(iii) it can incorporate a range of choice constraints.

The latter characteristic refers to the possibility that respondents may ignore some of the alternatives in the choice set, not perceive some choices to be available or 'nest' the alternatives within some hierarchical or 'tree' preference structure. The usual assumption is that a 'multinomial' logit model is appropriate. This assumes that all alternatives are considered when respondents make their choices and that preferences towards one alternative are not affected by the presence or absence of competing alternatives. If this is not the case, the researcher will need to restructure the model into a 'nested' form, in which alternatives with close associations in the respondents' minds (eg bus and train services) are clustered separately from distinctly individual alternatives (eg car.)

Ranking data can be analysed with a logit model through the 'exploded' logit procedure outlined by Chapman and Staelin (1982.) In this instance, each option is taken to be the preferred choice from a subset composed of itself and all options ranked beneath it. The estimation is achieved through the progressive working through of these subsets until they are exhausted. Because lower ranked items are likely to be less easily differentiated by respondents and will therefore be open to greater error in their rank positioning, it has become the usual practice only to 'explode' through half the rankings.

Rating scale data may be analysed with the logit model by interpreting the scores as indicators of rankings and analysing in the above manner. With pair-wise choices, the scale may be collapsed into binary choices and a binary logit model calibrated against the discrete categories.

This discussion of the logit model introduces the possible further applications of multiple regression. A refinement of simply regressing against the metric response scale for a choice between two alternatives is suggested by Wardman (1988, p75; also Bates and Roberts, 1983.) As an alternative to collapsing the scale into binary choices, this approach seeks to use the greater information provided by the scale. It does this by having the researcher assign arbitrary but sensible probabilities of choice to each response score (eg on a five point scale, 1 = 0.1; 3 = 0.5; 5 = 0.9.)

The binary form of the logit model may be written as:

$$P_i = 1/[1+\exp^{(U_i-U_j)}] \qquad (6.5)$$

where P_i = Probability of choosing alternative i
U_i = Utility of alternative i
U_j = Utility of alternative j

Assuming a linear additive utility function and 'generic' coefficients (that is, coefficients for common variables such as cost and time are the same for each alternative), the difference in utility between alternatives i and j can be expressed in the following way:

$$U_i - U_j = a_0 + a_1(X_{1i} - X_{1j}) + a_n(X_n - X_n) \qquad (6.6)$$

where a_0 = constant
$a_1 \ldots a_n$ = coefficients
$X_{1i} \ldots X_{ni}$ = alternative i attributes
$X_{1j} \ldots X_{nj}$ = alternative j attributes

Linear transformation of the binary logit model allows the combined formulas 6.5 and 6.6 to be rewritten as:

$$\mathrm{Log}[P_i/(1-P_i)] = a_0 + a_1(X_{1i} - X_{1j}) + a_n(X_n - X_n) \qquad (6.7)$$

A probabilistic metric scale is now achieved to which multiple regression can be applied.

The logit model and multiple regression can also be combined for the purposes of analysing rankings. A regression approach which uses all of the information provided by ranking exercises is reported by Louviere (1980, p13.) The assumption behind this approach is that if an alternative is ranked first it will always be chosen from any combinations where it is present. If it is chosen second then it will be chosen from half the choice sets in which it is present, as the one preferred alternative will have already been chosen in the other half. The next will be chosen in a quarter of the choices sets and so on.

As an example, a set with three alternatives (a, b, c, indicating order of preference) produces seven of choice sets:

abc, ab, ac, a = a chosen 4 out of 4 times = 1;
bc, b = b chosen 2 out of 4 times = 0.5;
c = c chosen 1 out of 4 times = 0.25.

This is a perfect logarithmic series (1, 0.5 0.25 etc) and is represented by the formula:

$$Rf_i = 1/2^{(ri-1)} \tag{6.8}$$

where Rf_i = Measure of relative preference
r_i = rank of alternative i

Given the standard logit formula and assuming Rf_i represents probability of choice P_i, then:

$$Rf_i = \exp^{(Ui)}/_{j=1}\Sigma^N \exp^{(Uj)} \tag{6.9}$$

where U_i = Utility of alternative i
$_{j=1}\Sigma^N \exp^{(Uj)}$ = some of all Utilities in the choice set

As $_{j=1}\Sigma^N \exp^{(Uj)}$ remains a constant K for each set of alternatives, then:

$$Rf_i = \exp^{(Ui)}/K \tag{6.10}$$

The researcher can now take logarithms on both sides of the equation and regress the logarithm of the relative frequency measure against the attribute values together with a term representing the choice subsets. For designs using the same set of choice alternatives, this term can be ignored, because it is constant. If a range of choice sets are used, as in the case of blocked designs, dummy variables can be inserted.
The opportunity to use multiple regression in the context of a logit model formulation is attractive chiefly for its familiarity and wide availability to transport researchers. Nevertheless, statistical packages which use maximum likelihood estimation remain the superior techniques, because they are truly probabilistic in their approach to calibration and require minimal scale information (ie discrete choices only) from SP data.

6.4 Theoretical Issues Specific to Discrete Choice SP Data

The above discussion of analytical techniques shows how the analysis of SP data borrows much from the analysis of revealed preferences, particularly in the application of discrete choice models. It has covered the more basic forms of analysis. It also identifies the need for practical modifications to RP analyses that arise out of the particular characteristics of non-discrete choice conjoint analysis data (ie rankings or ratings.) Nevertheless, once the discussion moves on to discrete choice models, there is a requirement to investigate more fully the fundamental ways in which SP data differ from revealed preferences. These differences have important implications for the application of established discrete choice methods in SP analysis.

The differences between stated preferences and revealed preferences relate to the errors contained within each type of data. Bates (1988, p62) identifies three sources of error in RP data:

"(a) The 'unobservable' factors which affect choice. These may be specific to the individual, representing prejudices in favour of certain alternatives.

(b) Measurement error in the explanatory variables entering the function for V[1]. In transport, examples are calculations of journey times based on coarse zoning systems, and the rounding of reported times to the nearest five minutes.

(c) Model specification errors. These arise not only in the decision on which variables should enter the formula for V[1], but also in the way in which they are entered."

SP data may be seen to be free of the first two types of error. The attributes influencing the respondents' choices are explicitly defined in the conjoint analysis exercises and any remaining unidentified systematic factors influencing choices but not readily identifiable can be represented by an alternative-specific constant. The clear definition of the SP attributes and the researcher's knowledge of their characteristics removes the problem of 'unobservables' and measurement error, though it is still possible that respondents' perceptions of the attributes vary from one to another.

1 V represents the estimate of utility, where Random utility $U_i = V_i + e_i$ (error term)

The third type of error is still present in the analysis of SP data. While the variables to enter into the model formulation are more easily identified than in RP data, the utility function which applies to them is not. This stresses the importance of including in conjoint analysis designs sufficient scope for testing alternative utility functions (eg allowing interactions and non-linearities.) Perhaps more fundamental is the assumptions that the researcher can make about the error term component of random utility, as this determines the type of probabilistic model that may be used. As already noted, the logit model assumes independent and identical Gumbell distributions for errors attached to each alternative. These assumptions may break down in the face of certain choice behaviour (eg preferences for one alternative being dependent on the presence/absence of another), so that alternative forms may have to be considered (eg nested logit.)

Though SP data may be seen to have few errors relating to the inclusion of all relevant variables and the measurement of attribute values, additional errors may be seen to relate to the measurement of the response variable which are not likely to occur in RP data. In RP surveys, which usually seek details of recent travel or other behaviour, a simple response is required. For example, respondents may simply have to identify the mode they last used and the best alternative. Errors, as already intimated, are more likely to occur in their perceptions of journey characteristics (eg the cost and length of the journey) than in recalling their choice behaviour.

Simple discrete choice responses in conjoint analysis exercises will bear close resemblance to the kind of alternative behaviour strategies which respondents might face in reality. This alone is a strong argument against using more complicated and unrealistic ranking or rating responses, even though these often contain more information then discrete choices. Unfortunately, even discrete choice SP data are likely to possess some additional error.

There are a number of ways in which individuals' responses to SP exercise may be subject to error. The following list is compiled with reference to Wardman (1987, p11-14), Bonsall (1983, p50) and McFadden (1986, p289.)

(i) Justification bias

It is possible that some consumers attempt to justify their purchasing behaviour when viewing it retrospectively. This is the concept of 'cognitive dissonance'. In RP data, it can lead to favourable estimates of attribute values relating to the chosen option and unfavourable estimates for the rejected alternative. Nevertheless, it is not likely to affect the process of identifying which they chose and which they rejected.

In SP data, respondents are not describing past real behaviour but instead are asked to predict hypothetical behaviour. There is a likelihood therefore that respondents may consistently choose an alternative which they have favoured in real situations, not necessarily because it is the best alternative but because they wish to justify their present behaviour to themselves and/or the interviewer. Extreme cases will stay with one alternative all through the conjoint analysis options and in so identifying themselves could conceivably be eliminated from the analysis. This would be at the risk of discarding unbiased respondents who instead, for example, have very high or very low values of time.

Some justification bias may exist in the way people make actual decisions and is therefore acceptable if present in conjoint analysis responses. Genuine habit and inertia may exist in the way a person makes decisions and this should be captured by alternative specific constants. Justification bias is only a problem when its effect is due to respondents believing they have to rationalise their behaviour, when in real situations such decision processes are not normally exposed.

If the respondent is trying to 'impress' the interviewer, the result may more properly be identified as 'affirmation bias', which Bonsall (1983, p50) identifies as:

> "The tendency of respondents to detect the underlying philosophy of the interviewer (and of the analyst who defined the survey instrument) and to respond in such a way as to parrot those views."

(ii) Unconstrained Response Bias

There is a great danger with conjoint analysis and other SP techniques that respondents are asked to make choices without fully considering the constraints that may exist if such choices were available in reality. This emphasises the importance of setting conjoint analysis scenarios within the context of a recent experience. This concern that respondents should recognise realistic constraints led to the hybrid conjoint analysis and Household Activity Travel Simulator (HATS) reported by Ampt et al (1987).

(iii) Policy Response Bias

Some respondents may be tempted to give false responses in the hope that this will influence policy makers' decisions. Like affirmation bias, it may depend on respondents perceiving, rightly or wrongly, a particular strategy behind the conjoint analysis scenarios. Certain precautions may be taken to minimise this possibility, such as not disclosing who has commissioned the SP survey until after the interview is complete and avoiding if possible controversial policies implied in the conjoint analysis choices. This latter problem may be difficult to avoid in such sensitive areas as public transport fares or parking cost increases, but if respondents are at least offered some compensatory improvements, this form of error might be reduced.

(iv) Non-Compensatory Decision Making

As Wardman (1987, p12) observes:

"The neo-classical economic approach to consumer behaviour assumes that individuals trade-off the utilities associated with various attributes in identifying that option with greatest overall utility. Whilst this theory may be generally applicable, there are alternative theories of choice such as the elimination by aspects or lexicographic choice processes which may apply to some individuals. .... For individuals who possess such choice processes, the model derived from utility maximisation theory, such as the logit model, are no longer appropriate...".

Only some non-compensatory decision making will be easily identifiable in SP data. For example, a respondent to a ranking game may place all the cheapest options first, and within that group arrange fastest options highest, and so on (ie lexicographic choice behaviour.) Such responses, if small in number, could conceivably be removed from the data set, but as in the instances of justification bias, may lead to the removal of compensatory decision makers who have not been presented with trade offs around their indifference curves. This form of bias again supports the attractiveness of discrete choice conjoint analysis, because non-compensatory decision making may result from a respondent's attempt to make manageable a complex response task (eg ranking, rating) rather than reflect his or her actual choice behaviour.

(v) Repeated Observations

A principal feature of SP data has already been identified in the section on sampling, namely that the observations represented in the data set do not correspond to the same number of individuals making independent choices, as they would in RP data. They represent repeated observations, albeit under different hypothetical or 'laboratory' conditions.

This has important implications for the errors in the data: observations belonging to the same individual are likely to be subject to similar error terms, unless there is some systematic error. Choices made later on in the progression of the conjoint analysis exercise may be affected by factors such as boredom, fatigue, increasing familiarity with the response task ('learning' effects or, in the case of ranking data, indifference between lower order options - see above.)

These errors may be identified as additional elements of intra-personal taste variation not represented by the error term in random utility (where $U_i = V_i + e_i$.) They are likely to be quite different in nature to those errors representing differences in tastes between observations from different respondents (inter-personal taste variation.)

The problems of boredom and fatigue affecting choices may be minimised by keeping conjoint analysis exercises fairly short (generally 8 or 9 options in each choice set.) Their effect on the responses to each option, together with 'learning' effects, may at least be randomly distributed across groups of respondents by randomly altering the order in which the options are presented in each interview. This replaces one form of systematic error across groups of respondents with random error, but intra-personal systematic error remains.

For the analysis of SP data, the presence of repeated observations needs to be taken into account in as much as it affects the goodness-of-fit statistics. The analyst may reasonably expect intra-personal error to be less diverse or sizeable than inter-personal errors (individuals are likely to be internally more consistent in their choice behaviour than groups of individuals), despite the problems discussed above. This cohesiveness between the repeat observations is likely to inflate goodness-of-fit indicators, such as the t-Statistics (the ratio of the coefficient to its standard deviation.) As Bradley and Bovy (1986, p298) suggest, some adjustment should therefore be considered.

The existence of errors associated with the response variable in SP data has important implications for the way in which the model estimation can be utilised by the researcher. The concept of random utility, already discussed, recognises the existence of errors associated with the measurement of the components of utility (the researcher's selection and definition of the independent variables; the respondents' perceptions and mis-perceptions of the attributes), intra-personal taste variations due to inconsistency in individual choice behaviour and inter-personal taste variations due to different individuals having different preferences. In this respect, RP data and SP data are similar, with SP data having the advantage of clearer definitions for the independent variables and a more obvious relation between the dependent variable (behavioural response) and the attributes which compose the utility equation.

The new problem for the analyst which results from the use of SP data is how to incorporate the additional error associated with the response variable. As Bates (1988, p64) suggests:

> "...we are not getting a true estimate of U [true utility], but rather some pseudo-utility which we may call $\hat{u}$, where the general linking framework is
>
> $$U_i = V_i + e_i = \hat{u}_i + n_i \text{."} \qquad (6.11)$$

If the researcher can assume properties for the error distribution n_i similar to those assumed for e_i, they may be conflated by a repositioning of the above formula, such that:

$$V_i = \hat{u}_i + (e_i - n_i) \qquad (6.12)$$

At this point, a clear division is established between analytical methods that can establish reliable estimates of coefficient ratios and those that can establish accurate forecasts of behaviour. In equation 6.12, it will be seen that the size of the coefficients in $\hat{u}_i$ will not be the same as in V_i due to the presence of the error components, unless $e_i = n_i$, which is unlikely. Given no systematic relationship between either of the error terms and the components of $\hat{u}_i$, size is the only difference.

Thus, ratios between the coefficients in $\hat{u}_i$ will be the same as in V_i (the random utility function.) If the researcher wishes only to establish quantitative measures of respondents' relative valuations of attributes (eg value of time, monetary equivalents for passenger facility improvements) then the type of models established for analysing RP data remain suitable for analysing stated preferences. To this end, the objectives of the present study may be seen to be met by established RP analytical techniques, though critical awareness of their shortcomings and the limits of their underlying assumptions is still important.

The difference between the sizes of coefficients in $\hat{u}_i$ and V_i is only of consequence if the researcher wishes to forecast behaviour on the basis of the preference weightings derived from the SP data. The problem is that the dependent variable indicating probability of choice in discrete choice models is directly related to the scale of the coefficients. As Bates (1988, p65) argues:

> "..we will be making estimates of the pseudo-utility $\hat{u}$ rather than the 'true' utility U: in other words, we are making estimates of relative preferences as expressed in a SP experiment rather than of what would occur in the market. ... An understanding of the magnitude of n_i is thus of crucial importance to the use of SP in forecasting. Only if n_i is insignificant relative to e_i can the estimated model be used directly to give forecasts: in all other cases some kind of scaling of the coefficients of V_i relative to the random terms is required, and in general the knowledge of how to do this is lacking."

This marks the point at which this discussion of analytical approaches for SP data may be drawn to a close. The complex issues relating to the use of SP data for establishing forecasts of future consumer behaviour are not of direct relevance to the present study. Since the above quotation from Bates (1988), some advances have been made in the utilisation of SP data for forecasting, most notably the use of nested logit models to constrain estimates from SP data against RP observations (Morikawa, 1990, Bradley and Kroes, 1990b.) While these new approaches have been developed for forecasting purposes, they also have implications for the derivation of relative quantitative valuations, and these are discussed in the concluding stages of chapter ten.

6.5 The Reliability of SP Data

Having discussed the analytical methods appropriate to conjoint analysis data and the theory behind the use of discrete choice models with SP data, there remains the issue of reliability. In section 3.5 of Chapter three, reference was made to studies which have observed a favourable comparison between forecasts derived from models from SP data and models from RP data.

This section will now consider the evidence from such studies concerning the reliability of SP data. Reliability can be seen to have two components in this context:

(i) Consistency - this comprises the ability of respondents to choose consistently within an SP exercise and the consistency of their choices when asked to repeat a conjoint analysis exercise after some period of time. This last point is an important test for any experimental approach: the stability of results across repeated trials.

(ii) Validity - in some circumstances, SP data may be directly compared with disaggregate RP data, in which the model results from the former may be compared with those from the latter, either in terms of the parameter values estimated or the forecasts they produce. The RP data may represent contemporary choice information or, ideally, choices being made after the introduction of the changes investigated in SP surveys. More often, SP techniques are applied because adequate RP data is not available or even impossible to collect, so that validation from comparisons between different data types is not appropriate. In such instances the assessment of the plausibility of the model results, in terms of their consistency with established theory, and the use of internal statistical measures will be the only guidance available to the researcher.

6.5.1 Consistency

Within a single conjoint analysis exercise, a number of respondents can usually be identified making inconsistent responses to the SP task required of them. This is most easily identified in ranking exercises, but may also be detected with less sensitivity in discrete choice designs (reflecting the lower information content of the response scale.) Inconsistency in responses may result from an individual not fully understanding the exercise or some of the attributes presented, from disinterest in the attributes or from fatigue. Other sources of error have been identified in section 6.1.2. The proportion of a sample who are obviously inconsistent in their responses is likely to vary with the quality of the SP design, the marginal nature or otherwise of the attributes investigated and the way the SP instrument is applied. Self-completion exercises are perhaps less likely to motivate respondents to think deeply about the choices presented to them compared to exercises administered by a conscientious interviewer.

Concerning the proportions of respondents to SP surveys that have been identified as being inconsistent in their responses, some conjoint analysis studies have reported figures of 7% (SDG Research, 1989; p35), 8% (Hensher et al, 1988, p55) and 11% (Steer, Davies and Gleave, 1989.) Such proportions, it is suggested, are acceptably small.

Studies which have required the same individuals to complete the same conjoint analysis exercise after a sizeable time interval (Acito, 1977; Best, 1979; Parker and Srinivasan, 1976 and Leigh et al, 1984) have generally shown a high degree of correlation between the responses. Even if there can be inconsistency within some individual sets of responses, it appears that most respondents will be consistent in the way they have arrived at their choices.

6.5.2 Validation

The validity of a measurement technique such as conjoint analysis can be determined in two ways: internally and externally. Internally, tests of statistical significance, goodness-of-fit measures and most important, the signs and magnitudes of the parameters all serve as indicators of the plausibility of the model results that can be obtained from SP data. These issues has been discussed above in the context of modelling procedures.

Externally, there may sometimes be the opportunity to compare findings from SP data with those derived from real market behaviour. Such circumstances are often rare in the case of commercially sponsored projects, because if adequate market data is available, the need to undertake additional SP surveys is not obvious. Where such data is not available, SP techniques may be used as a more cost effective method than undertaking new RP surveys. In circumstances where a completely new service or product is being used, the opportunity for collecting RP data does not exist. Validation exercises for SP techniques are therefore to be found in academic studies or government-sponsored projects, for which direct commercial/planning use of the results is not the main aim and a better understanding of the techniques is.

Louviere et al (1981), Bates (1986) and Wardman (1987; 1988) all report comparative studies of models fitted to SP and RP travel data which produced comparable weightings for journey variables such as cost and travel time. In a review of 16 conjoint analysis studies, Levin et al (1983) list a set of findings which together make a fairly strong case for the validity of SP models.

While this evidence for the external validity of SP techniques is encouraging, it has to recognised that studies relevant to the issue have a number of shortcomings:

(i) none of the reported validity research is done in a systematic way;

(ii) most of the reported validity research is carried out as a by-product of a practically-oriented study;

(iii) some of the validity studies in Levin et al (1983) are based on incorrectly applied prediction methods;

(iv) typically the reported validity research only concerns the reproduction of existing behaviour of the investigated sample; hardly any studies are known which deal with the generalisation of predictions to whole populations, and very few of them look at the ability to predict behavioural changes in response to changed circumstances.

From this discussion, it has to be concluded that further systematic validity research is needed before really conclusive answers and general guidelines can be given. Despite this, it can be taken as reassuring that almost all the evidence that is available seems to suggest a high validity for SP data.

7: APPLICATIONS OF STATED PREFERENCE APPROACHES TO THE VALUATION OF PASSENGER FACILITIES

7.1 Introduction

The last three chapters have considered in detail the development of SP experiments as instruments for the measurement of consumer preferences. It will be seen in this chapter that a number of studies have used these approaches in the valuation of rail passenger facilities. This is in response to the need for the quantitative measurement of consumer preferences towards such items, as identified in chapter two.

In the studies reported below, conjoint analysis SP experiments have been used to evaluate individual facilities (London Transport, 1984; Steer, Davies and Gleave Ltd, 1984c-d, 1985, 1986a; Sheldon et al, 1985; Anderson et al, 1986) and 'packages' of facilities (the MVA Consultancy, 1985a; Wicks and Beswick, 1986; Steer, Davies and Gleave Ltd, 1986b; SDG Research, 1989.) An alternative SP approach is to be found in the priority evaluator method, a form of which was used in London Transport studies (Hatch and Flack, 1974; Bradley and Maw, 1975, 1977a-b; Maw , Muir and Hendry, 1976; Bradley, Maw and Muir, 1977; Maw, 1977; Maw and Bradley, 1977) and a more advanced form later for British Rail and Dutch Railways (MVA Ltd, 1985b-d; Copley et al, 1987; Copley and Bates, 1988.)

The SP approaches to the valuation of passenger facilities may be placed in two groups, the first relating to studies for London Transport, the second to overland rail services. The reason for this categorization is only partly due to the different nature of the transport systems involved. More importantly, each group represents distinct lines of development in the valuation of passenger facilities. The first group, composed of studies by the London Transport Executive and later by Steer Davies Gleave, observes early applications of a basic version of the priority evaluator method which are then superseded in later studies by use of detailed conjoint measurement approaches. That is, an SP experiment application is presented specifically as a more sophisticated and preferred alternative to the priority evaluator method.

The second group, comprising studies by the MVA consultancy, presents a development of the priority evaluator method that is intended to compliment SP experiments and indeed replace the latter approach in applications where it is considered weak. The following review is an appraisal of the survey instruments used in the different studies and the final results obtained. From this it is intended that some guidance may be given as to the way subsequent studies could proceed in the use of SP techniques for valuing passenger facilities.

7.2 Studies of London Transport Services and Passenger Facilities

An Early Application of the priority evaluator method

Over the period from 1973 to 1975, three surveys were carried out by Opinion Research Centre on behalf of London Transport into customer attitudes towards London Underground trains and stations. Each survey involved approximately 1,000 home interviews. Despite some modifications to the approach from one survey to the next, the interviews conformed to a pattern of basic steps, which may be summarised as follows:

(i) The respondents were asked to rate different aspects of their underground train service and station on a five-point scale and suggest improvements that they would like to see.

(ii) Improvement options were presented in a priority evaluator format. Respondents indicated those of interest to them and then selected their optimum package of these facilities within the constraints of notional costs for each improvement and a limited budget.

(iii) The respondents were asked how much of a fare increase they would be prepared to accept in return for the improvement packages. They were then asked how small a fare reduction they would accept in return for not implementing the improvements. Finally, they indicated any desired improvements missing from the exercise.

The priority evaluator exercise was employed for the following reasons:

> "The aim of this technique is to force respondents into the same position as that of London Transport decision makers, namely, choosing between competing projects within a constraint, in this case a fixed sum of money. The results are more likely to predict future behaviour, in that respondents have expressed their true priorities, than are those surveys using traditional, attitudinal research methods."
>
> Maw and Bradley (1977, p28)

In relation to the measurement of priorities on attitudinal scales, the ability of the priority evaluator to obtain trade-off decisions from travellers is certainly attractive, but the earlier discussion in chapter four of the shortcomings of the technique have to be borne in mind. Certainly the assumption here that forcing respondents into the role of planners is an advantage has to be brought into question.

The options presented to respondents were priced in terms of notional supply costs and respondents were asked to allocate a budget of £500,000 (sometimes £750,000 in the first survey.) The choices given to respondents are listed in Table 7.1. Each was represented by written descriptions on 'show-cards'. All the options are presented as two-level variables (the improvement is chosen or it is not.) The respondents were first asked to indicate which improvements they were interested in. The interviewer added the values of the items chosen and if they greatly exceeded the budget limit, the respondents had to readjust their selection. If the total value was below, they were not required to spend to the maximum.

This process was described as a simpler alternative to the SCPR priority evaluator method (see chapter four):

> "This priority evaluation method presents respondents with a yes/no decision for the inclusion of each project in the selection of projects to meet the budget specified, compared with the more complex requirement for respondents to trade-off alternative levels of improvement in individual environmental factors described by Hoinville (1973.)"
>
> Flack and Hatch, (1974, p7)

Table 7.1 Choices offered to Underground Travellers in the Application of the Priority Evaluator Method Over Three Surveys, 1973 to 1975 (Maw, Muir and Hendy, 1976, app 2)

	1973	1974	1975
BUDGET:	£500,000 or £750,000	£500,000	£500,000
Station Improvements	£000's	£000's	£000's
Lifts instead of stairs	700	——	——
Escalators replacing stairs	300	——	——
Modernising lifts	300	——	——
Escalators replacing lifts	700	——	——
Eliminating island platforms	700	——	——
Extending platform canopies	——	40	——
Providing platform canopies	——	100	——
Extra booking clerk	——	40	——
Reducing draughts	200	——	——
Reducing dust	10	——	——
Heated waiting room	——	40	——
Quieter stations	100	——	——
Cleaning station	——	40	100
Cleaning and repainting station	——	100	——
Station facelift	200	250	250
Complete station modernisation	——	400	400
Improved station lighting	40	40	40
Security guard	40	40	40
	——	——	——
Train Improvements	——	——	——
	——	——	——
All new trains	300	——	——
More new trains	200	——	——
Painting train doors red	——	10	——
Faster trains	——	——	10
Quieter trains	——	——	40
Improved train suspension	10	——	10
Cleaner trains	10	40	40
More rush hour trains	200	100	150
More off-peak trains	40	100	100
Later trains at night	100	100	100
Improved information about delays	——	——	40
Fewer minor delays	10		——
	——	} 40	——
Fewer major delays	100		——
Improving train indicators	10	——	——
Providing train indicators	——	10	10
	——	——	——

Sums in £000's indicate that the project was offered in the appropriate year. Not all projects were offered at all stations in any one year.

In the first stage of the interview, in which respondents suggested improvements and rated the quality of their present station and train service, a markedly higher satisfaction was noted towards surface stations than for those underground. Suggested priorities for improvement were redecorating, cleaning and better lighting at most stations, improved access for deep level stations and better waiting rooms on surface stations. This section identified a greater demand for improvements to stations as opposed to train services, with only service frequency improvements in the rush hour a major demand.

The priority evaluator in stage two of the interviews elicited the responses displayed in Table 7.2. Only those improvement options being chosen by at least 40% of the respondents in one year are presented. It is interesting that some train improvements not mentioned by respondents in the first stage are chosen frequently in the second stage, perhaps indicating that respondents were reminded of possible improvements by the exercise. They may also have been attracted to spending surplus wealth on the less expensive items, such as cleaner trains, even though an allocation of the full budget was not required.

In the final stage of the interview, respondents specified the value of their chosen improvement package in terms of fare changes. This exercise was included to obtain predictions of the traffic generation that might result from station improvements. Established fare elasticities would be multiplied against these trade-off rates to produce the demand forecasts. The respondents were offered fare increases and then decreases in 5% increments, until they would not progress further. Though not referred to as such, this exercise was a straightforward transfer price question. Fare increases were introduced as the means by which improvements would be financed. Respondents were generally unwilling to accept large increases, because they did not wish to support a policy of raising fares. Fare reductions were presented as the result of increased Government subsidy, but were seen to lack credibility in a time of rapid inflation (reductions were in fact omitted from the 1975 survey for this reason.) As it was public money they were spending, the majority preferred to spend on improvements rather than fare reductions.

Table 7.2 Underground Service Improvements Chosen by at Least 40 % of Sample in One or More Years (Maw & Bradley, p30)

	1973 %	1974 %	1975 %
Station Improvements			
Lifts instead of stairs (one station only)	75	—	—
Access Escalators replacing stairs	49	—	—
Lift modernisation	47	—	—
Extra booking clerk	—	40	—
Dust reduction	49	—	—
Heated waiting room (open air stations)	—	64	—
Cleaning station	—	77*	50*
Cleaning and repainting station	—	67*	—
Station facelift (renovation)	42*	53*	50*
Improved lighting	45	68*	64*
Train Improvements			
Security guard	42	31	50
Cleaner trains	54	42	47
More rush-hour trains	49	53	68
More off-peak trains	54	47	54
Later trains on Friday and Saturday nights	46	49	47
Improved information on delays	—	—	42
Providing train indicator (where not already available)	26	35	45

*These figures show the total demand for the improvement indicated, by summing the proportions opting for that project with those opting for larger projects within which it was incorporated. (Thus, 77% for cleaning is made up from 10% who chose it in isolation, plus 23% for complete modernisation, 30% for facelift and 14% for cleaning and repainting together.)

There is a problem at this stage which was to be identified in later applications of the priority evaluator to studies for British Rail. That is, the packages of improvements for which the fare value was derived were the optimum packages for each individual: no information was provided on how each valued a common set of improvements. Another deficiency is that no analysis seemed to be carried out on the way the overall fare value obtained at the final stage related to each improvement selected in the second stage of the interview. This would require analysis at the most disaggregate level.

Because the fare increase exercise was considered to be biased against choosing improvements and the fare decrease was biased in favour, the results were averaged. Table 7.3 shows the range of percentage fare changes that respondents were prepared to accept in return for improvements at each station surveyed. Using a fare elasticity of 0.15, traffic generation was predicted to range from about 2% (0.15 x 14%) to 4.5% (0.15 x 29%.)

Further analysis found that the proportion of acceptable fare change fell as the base fare level rose, implying that longer distance travellers did not attach a much higher absolute value to improvements, so giving lower percentage values. Linear regression was therefore employed to establish a relationship between journey lengths and package values. From this, factors were produced to modify package values according to the system-wide average journey lengths to produce a standardised figure for each station.

If the answers to the increase and reduction questions are used separately, variations in the values established for tube stations range from 25% below to 50% above the figure originally derived. In addition, a confidence interval of 15% at the 95% confidence level, due to the limited samples used, extends the margin of error further.

Table 7.3 Mean Values for Improvement Packages in terms of Percentage Fare Change Equivalents (Maw, Muir and Hendry, 1976, p7)

1973 Survey (by station type)	
Lift stations	21%
Tube stations with escalators	18%
Surface stations	16%
1974 Survey (by station)	
Shepherd's Bush (Central)	20%
Bayswater	19%
Manor House	18%
East Acton	15%
Preston Road	14%
North Ealing	14%
1975 Survey (by station)	
Shadwell	29%
Stepney Green	24%
Shepherd's Bush (Metropolitan)	24%
Leicester Square (Northern)	21%
Totteridge	21%
Goodge Street	21%
Boston Manor	16%

These issues were also to be encountered in a survey of traveller attitudes towards facilities at London Transport bus stations (Bradley and Maw, 1977a; 1977b; Maw, 1977), carried out in 1976. It followed the exact same procedure as the Underground survey of 1975, except that prior to allocating a fiscal budget, the respondents were asked to rate with scores out of ten the priority evaluator options in terms of their importance.

The facilities presented in the survey where of a very detailed nature, including items such as vending machines and clocks, offered as separate options. The difference in prices was less extreme than in the Underground studies, although some items such as an overall roof for a station could still consume most of the budget allowed. Percentage fare values and traffic generation estimates could be made for individual items and are shown in Table 7.4. The realism of such figures must be questioned, as it would seem optimistic to suggest that the provision of station clocks on their own would compensate for a 5% fare rise or an increase passenger traffic by 0.8%, as suggested by the results.

In addition to the problems identified above, Maw (1977) notes other uncertainties related to the use of the results from all these applications of the priority evaluator. The first is a suggestion that the projects with smaller point scores biased responses towards the smaller and cheaper options, on the basis of 'value for money' rather than the benefit alone. Secondly, priorities established in the survey are based on the respondents' perceptions of the items as offered in the priority evaluator exercise and may differ when the projects are actually implemented.

As Maw (1977, p7) suggests:

> "Market Research surveys can never adequately measure attitudes to a hypothetical situation, as it is impossible to tell whether all the respondents have the same idea of what any particular improvement will actually mean to them. Quite often, previous expectations are not satisfied when the improvement takes place, so that there remains a residual demand for things to be even better."

Table 7.4 Traffic Generation Estimates Associated With the Provision of Individual Passenger Facilities at London Transport Bus Stations (Maw, 1977, p6)

Improvement	Predicted % increase in demand	% Fare Value
Shelters with draught screens	2.0%	13.3
Travel Enquiry Office	1.6%	10.7
Public Address	1.6%	10.7
Public Toilets	1.4%	9.3
Overall roof	1.2%	8.0
Seats	1.2%	8.0
Public Telephones	1.0%	6.7
Heated Waiting Room	0.8%	5.3
Clocks	0.8%	5.3
Improved overhead lighting	0.6%	4.0
Snack Bar	0.4%	2.7
Vending Machines	0.2%	1.3

This argument identifies two difficulties associated with all SP techniques: the reliability of using descriptions of attributes in the SP study as accurate portrayals of real changes to the transport environment and the opportunity to validate findings against real market data. The practicality of the suggestion that the estimates can be calibrated against observed demand after the projects are carried out must be questioned, considering the difficulties associated with the use of revealed preferences to measure valuations towards secondary service changes. The point is also made that the valuations are specific to the sites to which the exercises were related.

Application of conjoint measurement

Later studies of Underground trains and station facilities by London Transport (1984) and Steer, Davies and Gleave Ltd (1984d; Sheldon et al, 1985) were to use conjoint measurement (rankings) as the preferred survey method. This was due primarily to the difficulty in establishing actual consumer valuations from the priority evaluator survey, particularly when the respondents' answers were often influenced by the supply costs attached to each improvement option. A ranking experiment, in which respondents arranged pre-constructed packages of improvements in order of preference, would remove such cost considerations.
Another attraction of conjoint measurement was seen to be the number of observations that could be obtained from each individual, offering the potential for preference weightings measured at the disaggregate level (using MONANOVA.) These individual valuations could then be combined as a whole sample or in market segments. Variance from the mean valuations for each group could then be estimated, a measure not available from aggregate approaches. Later work would show that MONANOVA had drawbacks which largely over-rode this advantage; the chief weakness was the unreliability of its estimates. More sophisticated approaches, such as the logit model, would be used instead, though these would have to be applied to groups rather than individuals.

A study undertaken by London Underground Ltd (London Transport, 1984), with the assistance of Steer, Davies and Gleave Ltd, examined passenger valuations of a number of train characteristics and facilities. A conjoint measurement exercise was carried out, with attributes being traded against a 10% fare change. Each attribute had two levels, one corresponding to an improvement, the other to the current situation. The implied valuations are shown in Table 7.5.

In an SP experiment used in a later survey (Steer Davies Gleave, 1984d; Sheldon et al, 1985), fourteen two-level variables were presented in combinations conforming to full profile fractional factorial designs. To avoid respondent fatigue, the procedure was divided into three ranking exercises, with eight combinations in each. Interaction effects could be estimated for some variables and as fare was included in each experiment, monetary values could be calculated from the ratio of each item's coefficient to that of the fare variable.

It was considered important to set the exercise within the context of an Underground journey familiar to each respondent, care being taken not to educate them about the service. This was to ensure that their responses were made at the level of knowledge with which they currently made their travel choices. An additional twelve items were also presented to respondents, to be rated separately on attitude scales. Mixed with the items were some of the variables included in the SP experiment. This was done so that fare values established in the experiment could be used to infer values for the items presented only in the rating exercise, according to the ratios of their scores. All the variables used in the conjoint measurement and rating exercises are listed in Table 7.6 and Table 7.7 respectively. The monetary values derived directly from the conjoint measurement exercise are displayed in Table 7.8.

While the use of attitude scales to extend the SP valuations to other attributes is a pragmatic solution to the difficulty of handling large numbers of variables, the reliability of the extended valuations has to be questioned. The attitude scales will not be very sensitive and it is even doubtful whether they can be used reliably as interval scales.

Table 7.5 Monetary Values for London Underground Train Services as Derived Directly from a Conjoint Measurement Exercise (London Transport, 1984)

Improvements	Pence/ Journey*	Percentage Fare Value
Well ventilated carriage	5.8	11.6
Clean carriage	5.4	10.8
Well regulated heating	5.1	10.2
Smooth ride	4.8	9.6
Uncrowded carriage	4.6	9.2
Comfortable seats	4.6	9.2
Quiet carriage	4.4	8.8

*Average fare 50 pence

Table 7.6 Underground Train Service and Station Variables Presented in a Conjoint Measurement Exercise (Sheldon, Bottom and Golob, 1985, pp145-146)

First Exercise: Access/Station Environment

Cleanliness
Modernisation
Ticket Purchase Time
Ticket Office Staff Plus Machines
Real Time Train Arrival Information (on Platform)
Fare (+20%)

Second Exercise: Train Environment

Presence of Guards
Noise Level
Presence of Drivers
On-train Location Information
Fare (+20%)

Third Exercise: Train Operation

Reliability
Frequency
On-Train Information on Delays
Fare (+20%)

Interactions capable of Measurement in the Above Exercises

Clean and Modern Station
No guard and No Driver
No Guard and No Directional Information in Carriage
Low Reliability and No On-train Information About Delays
Low Reliability and Crowded Train Conditions

Table 7.7 Underground Train Service and Station Variables Presented in a Conventional Attitude Scaling Exercise (Sheldon, Bottom and Golob, 1985, pp148)

At the Station

Ticket Office Staffing
Information About Next Train Arrival in Booking Hall
Seats in the Booking Hall
Seats in the Platform Area
Toilets in the Concourse Area
Real Time Train Arrival Information on Platform
Vending Machines on Platforms
Small Shops in Booking Hall
General Information About Current Services in the Booking Hall

On the Train

Cleanliness
Crowdedness
Seating
Heating
Smoothness of Ride
Noise Level
Ventilation

Variables in **bold face** were also included in the trade-off research allowing the results from the two exercises to be linked. Two reference points for each rating exercise were included.

Table 7.8 Monetary Values for London Underground Train Services and Stations as Derived Directly from a Conjoint Measurement Exercise (Sheldon, Bottom and Golob, 1985, p151)

Main Effects	Pence/ Journey*	Percentage Fare Value
Clean station	7	14
Electronic arrival information	5	10
Less crowded train	5	10
Information on train about delays	5	10
Less noisy train	4	8
Ticket office staff and machines	3	6
Modernised station	3	6
Reliable service	2	4
Faster ticket purchase (by 50%)	2	4
Directional information on train	1	2
Interaction Effects		
Clean and Modernisation	4	8
Not unreliable and crowded at the same time	4	8
Not without guards and without directional information	4	8
Not unreliable and without delay information	3	6

*Average fare=50 pence

The application of the priority evaluator method as outlined here may be seen to possess a number of major weaknesses which relate both to the purpose of the exercises and the actual design of the instrument. A fundamental weakness suggested by the proponents of the later conjoint measurement studies is the assumption that respondents placed in the planning role are able to clearly infer their priorities as consumers. The observation in the earlier studies that responses could be distorted by supply costs in part recognises this problem. In justifying the use of conjoint measurement, the earlier approach is criticised explicitly for this weakness:

> "The main advantage that it [conjoint measurement] has over more traditional 'bag of money' approaches is that it overcomes, during the data collection stage, the main deficiency attributed to such approaches, which is that consumers are asked to act as quasi-planners. Within this process, the respondent, as a passenger and as a user of the system, is lost."
>
> (Steer, Davies and Gleave Ltd 1984d, executive summary)

In presenting priority evaluator improvement options as separate facilities (ie two-level variables), the London Transport approach is simpler than some other priority evaluator approaches. Allowing respondents to spend less than the maximum budget, or a little more, introduces flexibility into the respondents' choice options. Unfortunately, these developments are compromised by a lack of consideration for the way the prices and budget levels constrain the responses. The analysis similarly lacks justification as to why the use of the cost of each item expressed as a proportion of the overall budget should reflect individuals' actual valuations. There is also no recognition of the difficulties associated with the aggregation of responses from individuals spending budgets of different sizes.

In short, the priority evaluator exercise used in the London Transport studies fails as a method for measuring consumer valuations not simply because it places them in the role of transport planner. It fails because of shortcomings in its basic design and subsequent analytical approach. The constraints imposed by the price structure and budget will inevitably limit the respondents' ability to express their personal priorities. This is not the same as saying that by identifying priorities for investment, respondents cannot express their own preferences as consumers. The requirement is for the exercise to be placed firmly in the context of their own experience of the transport system and to provide them with sufficient choice through which they may infer their priorities. With this in mind, a conjoint analysis approach appears the more flexible.

As an alternative to the priority evaluator method applied for London Transport, the later conjoint measurement approaches are indeed an improvement. They reflect a stronger methodological base with a clear experimental design procedure and analytical approach. Monotonic analysis has limitations, which have been discussed elsewhere. Nevertheless, the opportunity for disaggregate analysis and estimation of interaction effects between variables are significant advantages. The concern of these later studies to place the improvement options in the context of respondents' current journeys must also greatly improve the realism of the responses. Nevertheless, an important problem with the values derived from the use of conjoint measurement is that when totalled together, the overall value appears implausibly high. If all the values from Table 7.8 were summed, the implied overall benefit of 52 pence over an average fare of 50 pence (104% of fare) seems unrealistic. The use of only one fare level must also limit the sensitivity of the results.

7.3 Application of SP Techniques to the Study of British Rail and NS (Dutch Railways) Train Services and Stations

Use of Conjoint Measurement Alone

The conjoint measurement approach developed in the studies for London Underground was also to be used in a study of various InterCity train service characteristics undertaken for British Rail by Steer, Davies and Gleave Ltd (1984c.) 468 travellers were interviewed, each being asked to complete two conjoint measurement exercises which were linked by a common fare variable. A variable representing the general quality of their terminal was included. Photographs and verbal descriptions indicated to respondents the different levels of terminal quality. The preference weightings established by the application of 'MONANOVA' to the responses, together with the subsequent monetary values (at 1984 prices), are displayed in Table 7.9. Note the fairly small values for items such as terminal quality and on-train catering, compared to seat density and punctuality. The small values are important, in the light of later studies reported below.

A study by Steer Davies and Gleave Ltd (1985) used a similar approach to that discussed above to assess improvements to Bristol Parkway station. This station, then functioning as a basic "parkway" road/rail interchange, had minimum passenger facilities. It was planned that rail/rail interchange would also become possible with the re-scheduling of some services. This increase in passengers waiting for trains suggested that some improvements to the station would be necessary.

In advance of undertaking SP surveys, almost 400 semi-structured interviews were undertaken: 300 at Bristol Parkway and a further 94 rail/rail interchanging passengers at other InterCity stations. These interviews sought to establish travellers' current perceptions of the station and their priorities for improvement.

Table 7.9 Preference Weightings and Monetary Values for InterCity Service Characteristics (Steer, Davies & Gleave Ltd, 1984c, p36)

Variable	Change being assessed	Importance (derived from MONANOVA)	Monetary Values (% of Fare)
Fares	10% increase	1.8	-
Journey time	20% faster	1.8	10.0%
Interchange	1 extra change	2.2	12.2%
On-train service/ 'club class'	provided v not provided	0.6	3.3%
Meal on-train	provided v not provided	0.2	1.1%
London peripheral stops	provided v not provided	-0.4	-2.2%
Terminal station quality	new/clean v old/dirty	0.3	1.7%
Seat density	high/low	1.5	8.3%
Punctuality	30 mins late or as expected	2.5	13.9%

For the SP surveys, conjoint measurement was used in a similar way to that described for the London Underground studies above. Three experiments were used to present marginal improvements, linked to one another by the inclusion of a fare variable. The respondent was to complete all three, but as a way of assisting this task, the improvements were grouped according to their relationship to different aspects of the station. These groupings are summarised in Table 7.10.

One noticeable feature from this table is the very small fare change (5%) used in the first three exercises. It is possible that some respondents might give a small weight to such a small change. As in the earlier study for London Underground, the use of one fare change is a weakness, as it does not allow the examination of variations in the trade-off rates across a range of fare levels.

So far, the approach used was the same as the London Underground studies. Where this exercise was to differ from previous applications was in the inclusion of a fourth conjoint measurement exercise, in which a single variable representing all the station improvements at once was introduced together with journey time as well as fare.

The package of improvements was included in the fourth exercise in recognition of the problems associated with trying to obtain a total valuation from the sum of individual values:

> "This exercise was included to provide a scaling parameter. Importance values can be obtained for a number of individual station improvement features to undertake project evaluations, but clearly there is a limit to what people would be prepared to pay for station investment (if a major upgrading exercise is conducted) vis a vis changes in other aspects of the rail product. By relating changes in design of the station to other primary characteristics (such as fare and time) a realistic estimate of passengers' "preparedness to pay" can be obtained."

(Steer, Davies and Gleave Ltd 1985, p46)

Table 7.10 Improvements to Bristol Parkway Presented in a Conjoint Measurement Exercise (Steer, Davies & Gleave Ltd, 1985)

First Group (Concourse)

Travel centre
Double number of Ticket Windows
TV screen information
Direct access to trains
Telephones
Fare (+5%)

Second Group (Waiting)

All-weather shelter
Fulling enclosed waiting room
Twice number of seats on platform
Separate cafeteria
Toilets on platform
Fare (+5%)

Third Group (Access to platforms)

Escalators to platform
Ramp to platform
Lift to platform
Trolleys available
Fare (+5%)
(Interaction: trolleys and ramp)

Fourth Group

New Station Design (all above facilities)
Fare (+10%)
Journey time (+20%)

This is the first recognition that the presentation of improvements to passenger facilities as a single package, together with major journey attributes in addition to fare, could give more plausible estimates of their total value than summing the values of each improvement introduced separately. Table 7.11 shows the implied fare values for each improvement, produced from the first three exercises, and the value for a single package representing all improvements produced from the fourth exercise.

The interpretation of these results provided in the study report is that the individual valuations are a correct reflection of the amount an individual will pay for any one improvement, but when the improvements are introduced all together, thresholds on willingness to pay are encountered. The low value for the package of improvements compared to the sum of the individual improvements recalls the low value for terminal quality obtained in the earlier InterCity study (Steer Davies Gleave, 1984c), though here the implied value is larger.

Regarding the total value derived from the fourth exercise:

> "This value is greater than the monetary value that passengers attach to any of the specific investment packages examined during games one to three. ..... As expected, diminishing returns do appear to apply, and by comparing major station investment with changes in other aspects of the rail product we have been able to determine the appropriate scaling parameter in the context of major upgrading."
>
> (Steer, Davies and Gleave Ltd 1985, p54)

This is the only recognition of a threshold effect in the report and the possibility of scaling individual values in the light of such an effect is not explored. Instead, the comparatively low valuation of the total station investment package is seen mainly as a result of negative connotations attached by travellers to the idea of a "modern" station, which acts against the benefits derived from better facilities, and the likelihood that the respondents considered the facilities within this total investment less carefully than in the first three exercises.

Table 7.11 Monetary Values Obtained for Improvements to Bristol Parkway Station (Steer, Davies & Gleave Ltd, 1985)

Improvement	Monetary* Value	Percentage Fare Value
Escalators to platform	57p	7.4%
All-weather shelter	40p	5.2%
Lift to platform	42p	5.5%
Double number of Ticket Windows	24p	3.1%
Fulling enclosed waiting room	24p	3.1%
Ramp to platform	24p	3.1%
Trolleys available	23p	3.0%
Travel centre	23p	3.0%
TV screen information	23p	3.0%
Twice number of seats on platform	18p	2.3%
Direct access to trains	16p	2.1%
Separate cafeteria	16p	2.0%
Telephones	12p	1.6%
Toilets on platform	10p	1.3%
Interaction: Trolleys and ramp	9p	1.2%
New Station Package (all above)	60p	7.8%

*Average fare £7.70

These are valid points, but their emphasis appears out of proportion compared to the more pertinent question of how this lower value should affect the interpretation of the values derived individually for each facility. The suggestion that the individual values represent the benefit of a facility when introduced on its own appears reasonable, but there is a question as to how these benefits combine within the overall package of improvement. An across-the-board application of a scaling factor (the ratio of the value of a package to the sum of individual values) may be one approach, but could a more complex set of relationships exist? None of these issues are explored in the report and only begin to be partly addressed in subsequent studies.

A later project undertaken for Danish National Railways by Steer, Davies and Gleave Ltd (Anderson, Moeller and Sheldon, 1986; Steer, Davies and Gleave Ltd, 1986) presented specific groups of station and on-train facilities using conjoint measurement as before. The variables presented to respondents are listed in Table 7.12. As in earlier applications, each group represents a different ranking exercise, linked by a fare variable. Unlike the Bristol Parkway study, an additional exercise with a total investment package variable is not included. The sample totalled 623 travellers and the monetary values they attached to the station facilities are listed in Table 7.13.

Despite this study being undertaken after the Bristol Parkway study by the same researchers, no exploration is made of possible scaling effects from limits on travellers' willingness to pay for packages of improvements. The importance of this issue does not therefore seem to have been recognised, with the problem still persisting as to how these individual values can be aggregated to give plausible total values.

Table 7.12 Variable Sets Offered to DSB Passengers in a Series of Conjoint Measurement Exercises (Andersen, Moeller and Sheldon, 1986a, p265)

General		Train/Ferry		Station	
(1)	(2)	(3)	(4)	(5)	(6)
Journey time	Access/ egress	Catering	Environment	Catering	Weather Protection
Frequency	Baggage conveyance	Special carriages	Toilet facilities design	Ticket booth	Toilets
Punctuality	Seat reservations	Seat density	Cleanliness	Waiting facilities	Cleanliness
Journey Planning Information	Bikes (on train)	Coach design	Adjustable seat	Information at station	Car parking
Interchange	Timetable format	Crowdedness	Ferry catering	Station shops	Platform accessibility
	Train/ ferry link				
Fare (+15%)	Fare (+10%)	Fare (+10%)	Fare (+10%)	Fare (+10%)	Fare (+20%)

Table 7.13 Monetary Values Attached to Passenger Facilities by DSB Passengers (Steer, Davies and Gleave Ltd, 1986a)

	Kr*	% of fare
Station Facilities		
Improved toilet (location and cleanliness)	25	18.1
Improved waiting facilities	23	16.7
Stairs and lift to platform	18	13.0
Improved baggage conveyancing facilities	18	13.0
Improved station cleanliness	15	10.9
Open ticket sales counter	15	10.9
Enclosed heated waiting room	14	10.2
Roof and bus shelter on station platforms	8	5.8
Standardised information at station	8	5.8
Station cafe (rather than mini-bar)	7	5.1
Open station shops	6	4.4
Drinks and both hot and cold meals	5	3.6
Station restaurant (rather than cafe)	2	1.5
Train Facilities		
Improved toilet	23	16.7
Bikes allowed on train	20	14.5
Drinks & cold meals	19	13.8
Family carriages	18	13.0
Adjustable seating	15	10.9
Improved environment	14	10.1
Only half seats occupied (crowdedness)	9	6.5
More cleaning	7	5.1
2+1 seating (First class only)	6	3.4**
All seats reserved	-10	-7.2
Open plan coaches (Second class only)	-7	-5.6***

* Average fare (all classes) 138kr
** Average First class fare 178Kr
*** Average Second class fare 126Kr

A later study of rail/rail interchange (The MVA Consultancy, 1985a; Wicks and Beswick, 1986) included some packages of station facilities in an SP experiment. The study was concerned specifically with travellers' preferences at stations where they change trains. The results indicated the relationship of the interchange penalty in a journey, measured in terms of travel time, with station layout, information and waiting facilities. Interchange time was valued in a separate SP experiment against travel time and fare. Both exercises were administered in a reply-paid self-completion questionnaire.

The results, derived from the application of a logit model in the data analysis, are shown in Table 7.14. The values attached to each group of facilities represents the disutility respondents associate with a drop to those levels of service from a position of high quality.

It will be seen from this study that the presentation of station variables with different levels of time spent at the station could be the most appropriate way of obtaining quantitative values for such facilities, compared with other factors, such as journey time or fare. This is because respondents are able to evaluate the facilities directly as factors that reduce the disutility of waiting time. The variations in waiting times can be imagined in the case of connecting services, but would not be realistic with passengers who are able to optimise their arrival and departure times at stations (eg travellers accessing services by car and by foot.) Journey time and fare, it may be suggested, have a more obvious relationship with on-train facilities, as the disutility of these two primary elements of the journey will be reduced by enhanced train facilities.

Two further studies undertaken for British Rail (Steer, Davies and Gleave Ltd, 1986; SDG Research, 1989) used conjoint analysis to measure respectively the overall values of improvements to Oxford station and new rolling stock for Network SouthEast services between Oxford and London.

Table 7.14 Valuations of Interchange Station Facilities in Terms of Interchange Time (The MVA Consultancy, 1985a; Wicks and Beswick, 1986, p22)

Interchange type		Facilities		Information	
Medium	Difficult	Medium	Poor	Medium	Poor
8.9	26.7	9.4	17.9	0.3	6.6

Key:

Easy change	-	same/cross platform interchange
Medium change	-	change of platform
Difficult change	-	change of station
Good facilities	-	modern station with good buffet and adequate waiting room
Medium facilities	-	small, cramped buffet; limited waiting facilities
Poor facilities	-	no buffet; old, uninviting waiting room
Good information	-	frequent announcements; full timetables; automatic displays; information office
Medium information	-	occasional announcements; summary timetable; no automatic displays; staff available
Poor information	-	few or no announcements; summary timetable; only booking clerk to consult.

In the study of Oxford station, a conjoint measurement offered trade-offs between three alternative new station designs, fare (+5%) and journey time (+10%.) An interesting exercise included a comparison of visual and written descriptions of the proposed station improvements, but with commuters only. The results are shown in Table 7.15. MONANOVA was used to obtain these results. Regarding the use of visual and written descriptions, the study report suggests the following:

> "The study has revealed that the careful use of visual material in informing passengers of the outcome of the improvement scheme is broadly neutral compared to the use of written material alone. This is an encouraging result since, when considering other investment schemes whose outcome may be more difficult to express only in written terms, the results are unlikely to be affected by respondent bias"
>
> (Steer, Davies and Gleave Ltd, 1986, p30)

The visual descriptions all used photographs, so this conclusion does not relate to the use of artist's impressions. Also, the strength of this conclusion is limited by the small sample sizes used in the study (125 respondents overall, divided into journey purpose/description type segments of about 30 each.) Another weakness is that although measures of dispersion could be obtained for the valuations (by calculating the standard deviations across individual valuations produced by MONANOVA) these were not calculated, so that no formal tests can be carried out on the figures to suggest whether the difference in valuations is non-significant. Note also the potential problem of having used a very small fare change, as noted for the Bristol Parkway study.

A study of travellers' choices between InterCity and Network SouthEast services running on the same route was carried out on the Oxford-London line in 1989 (SDG Research, 1989.) A self-completion questionnaire was sent to season ticket holders in which a discrete choice conjoint analysis exercise was used to present choices between competing "125" and local services. The trade-off was between faster InterCity and cheaper Network SouthEast services, the objective being to establish a premium price for InterCity trains that would reduce current crowding and make better use of comparatively under-utilised local services.

Table 7.15 Monetary Values Obtained for Improvements to Oxford Station (Steer, Davies & Gleave Ltd, 1986b)

Improvement	Segment	Description	% Fare Value
Rearrangement of facilities	Non-commuter	Written	4.1
Similar building on new foundations	Commuter	Written	3.2
	Commuter	Visual	3.4
	Non-commuter	Written	4.8
Completely new building	Commuter	Written	4.7
	Commuter	Visual	4.8
	Non-commuter	Written	6.4
	Non-commuter	Written	5.0
Final Weighted Results			
Rearrangement of facilities			4.1
Similar building on new foundations			4.2
Completely new building			5.2

As an additional factor likely to influence choices, respondents were also presented in some cases with choices that included new Network SouthEast rolling stock (the so-called "Turbo stock".) Over the sample of 419 respondents, a value of 4% of fare was obtained for this improvement, though it must be remembered that this figure only applies to season ticket holders.

Applications of the Priority Evaluator, with Conjoint Measurement

In contrast to the use of conjoint analysis to measure the values of individual improvements, a series of studies by the MVA Consultancy used the priority evaluator method as the preferred alternative, this being considered more effective for situations where choices are numerous and marginal. Qualitative research was to suggest that 'station quality,' as a distinct variable, could be presented as a factor in respondents' travel choices. It therefore seemed appropriate to present it in combination with primary journey factors in a conjoint analysis SP experiment, in the same way that it had been used in the earlier study of Bristol Parkway. However, the use of conjoint analysis SP experiments for the valuation of components of the 'station quality' variable (eg level of information; waiting facilities), which individually are secondary factors in the travel choice decision, was seen to possess limitations:

> "Inevitably, the number of secondary attributes to be assessed is such that they cannot be dealt with in a single SP exercise and resort has been made to linked exercises. In a number of cases, primary attributes such as fare or journey time have been included in the exercise as the variables to be traded. There is a real danger in this situation, because the secondary attributes may be of very limited value in comparison with the primary attributes. In such circumstances, they are either completely dominated by the primary attributes so that no useful valuations are obtained, or, particularly when only one primary attribute is included, their importance may be over-valued, merely for the sake of carrying out the experimental task. When the linked exercises are integrated, the implied overall value of improvements to rolling stock or stations may be unrealistic and out of keeping with SP valuations of the entire package."
>
> (Copley and Bates, 1988, p1)

This is a clear criticism of the Bristol Parkway study method. As an alternative to conjoint analysis SP experiments in the valuation of station facilities, the priority evaluator method was to be presented very much in the format established by the earlier SCPR studies. It was in the development of the analysis that significant developments would be introduced.

In 1985, the MVA Consultancy was to carry out an extensive survey for British Rail on the value of station facilities to users in the London and SouthEast region (now 'Network SouthEast'.) The research was undertaken in three stages, beginning with a series of in-depth interviews with rail travellers. This first stage sought to elicit detailed opinions on the role of stations as a component of rail journeys as well as attitudes and expectations about the service they currently delivered. The second stage (the MVA Consultancy, 1985c) involved a large attitudinal study in which almost 3,000 travellers returned self-completed questionnaires. Nineteen station facilities were rated on satisfaction and importance scales. General measures of satisfaction and priorities towards facilities were established from the scores that resulted and factor analysis was used to identify market segments according to the respondents' preferences. Finally, the third stage involved the application of the priority evaluator method together with a SP experiment in a series of face-to-face interviews. The object of this third survey was to obtain two types of consumer valuation, stated as:

> "(i) the valuation which customers place upon overall improvement of their station environments - that is both local and termini stations, and
>
> (ii) the valuation which customers place upon the improvement in or provision of individual facilities at stations."
>
> The MVA Consultancy (1985c, p15)

An SP experiment was used to obtain the first types of valuations with station facilities grouped to form different levels of a single 'station quality' variable presented with other journey variables. In approaching the second type of valuation, involving a number of facilities, an SP experiment was considered unsuitable:

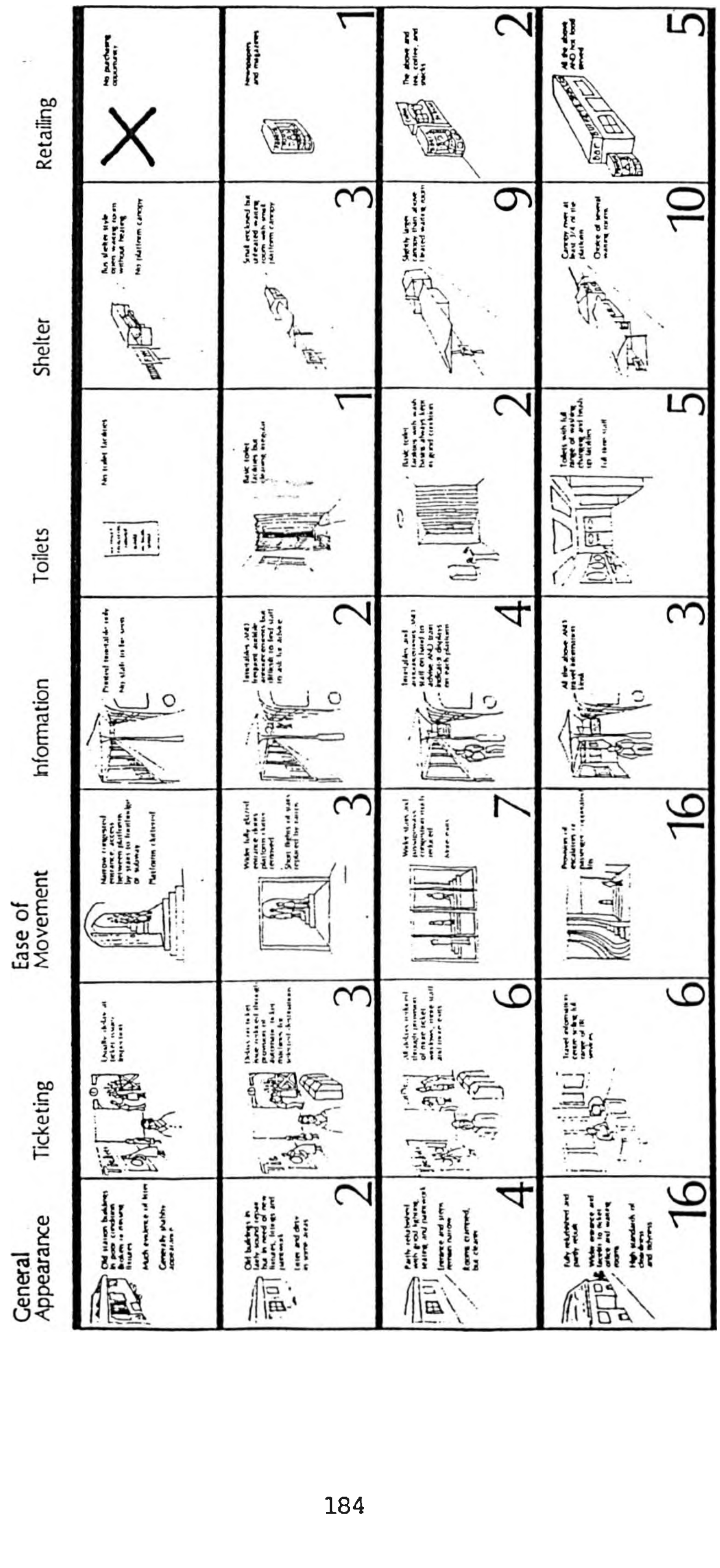
> "...the number of attributes cannot be reduced sufficiently to include all of the reduced set within an experimental design which still permits other major influences in the travel choice to be represented. This is the case even if we consider a separate design for each of the primary market segments. Therefore we must make use of the method originally proposed for the stage three surveys; the use of a priority evaluator board in combination with an SP experiment."
>
> The MVA Consultancy (1985c, p42)

The priority evaluator method applied in the third stage of this study was very similar in its design and presentation to the more complex instruments developed by SCPR (as opposed to that used by London Transport.) The nineteen variables examined in the stage two attitudinal surveys were combined through factor analysis into six clusters, labelled:

(i) General Appearance
(ii) Movement
(iii) Information
(iv) Toilets
(v) Waiting
(vi) Purchasing

A seventh category, ticketing, was introduced as the issues of ticket office staffing and automatic machines were of particular interest to British Rail. Within these groupings, the facilities were presented on a game-board, illustrated in Figure 7.1. The prices attached to each level are neutral units, representing approximately the same proportion of cost estimated to provide each level.

Figure 7.1 Priority Evaluator Game-board Used in the Measurement of Passenger Valuations Towards British Rail Station Facilities (The MVA Consultancy, 1985d)



The respondents were asked to identify the variable levels that most closely represented the present situation as they perceived it. They were then asked to allocate first a 'medium investment' budget of 15 points and then a 'high investment' budget of 25 points from their present situation rather than base levels. They were asked to allocate as much of the budget as possible.

Once respondents were satisfied with their allocations, they then responded to the conjoint measurement SP experiment, in which the three-level station quality variable varied from the present level of facilities through the individuals' medium investment allocation to the high investment allocation. Other variables used in the experiment were journey time, fare and service frequency. Using an "exploded" logit model, the overall station quality valuations in terms of fare were 4.5% for a shift from medium investment to high investment and 5% for a shift from the present situation to high investment. Thus the medium investment level was considered to be of insignificant value while the high investment package made fare sacrifices worthwhile.

To achieve the second objective of this final survey, that of establishing monetary values for individual facilities (or at least groups of facilities), the value of each group was taken to be represented by the proportion of budget allocated to it. The normalised values that resulted are shown in Table 7.16. These values were taken to be proportions of the trade-off value established in the SP experiment. It was recognised in the study that the levels of investment offered in the SP experiment were those chosen by the individual. Therefore, a modernisation scheme if implemented at a particular station would be deficient in some respects for some individuals and excessive for others. The optimum budget allocations indicated by each individual were therefore compared with the average allocation of all respondents at the one station.

Table 7.16 Normalised (Percentage) Expenditure by Priority Evaluator Attribute and Implied Values as Percentages of Fare (The MVA Consultancy, 1985d)

	% of Budget	% Fare Value
Medium Investment		
Appearance	11	0.50
Ticketing	22	0.99
Movement	13	0.59
Information	11	0.50
Toilets	12	0.54
Shelter	24	1.08
Purchasing	5	0.23
Total	**100**	**4.50**
High Investment		
Appearance	11	0.55
Ticketing	21	1.05
Movement	15	0.75
Information	10	0.50
Toilets	11	0.55
Shelter	24	1.20
Purchasing	7	0.35
Total	**100**	**5.00**

The lowest value for each attribute was used, so that on average the overall value of station improvements would be 60% of the value if all respondents received their optimum improvement package. This figure was considered low because it would not include any additional utility individuals may derive from 'over-provision' of facilities. Thus a midway figure of 80% of optimum package value was used. The final modified valuations for the high investment package listed in Table 7.17. A final step involved the application of known fare elasticities to predict traffic generation from station improvements. It was suggested that due to a degree of in-elasticity in the demand curve, station improvements alone would not generally recoup investment from increased demand. A combination of station improvement and fare increases would maximise gross revenue, sufficient to cover the costs of modernisation.

Despite the attractiveness of the basic concept, the application of the priority evaluator method described above may be seen to possess a number of weaknesses. These may be summarised as:

(i) The priority evaluator game board groups the station facilities as attributes of station quality. Though the earlier attitudinal studies provide substantial guidance for this grouping and ordering of facilities according to the mean average, there is a likelihood that many respondents will have different perceptions from these averaged combinations of items. Some respondents will be relating variables differently from these groupings and achieving their priorities with varying degrees of ease or difficulty.

(ii) Because respondents will vary in their perceptions of the present situation, they face different choices. A respondent who begins, for example, at level three on each attribute scale may face less permutations than someone beginning on level two of each attribute.

Table 7.17 Calculation of Percentage Fare Value for Rolling Stock Attributes (The MVA Consultancy, circa 1986, p3)

Rolling Stock Attributes	Value
Heating	0.97%
Seating	0.47%
Information	1.17%
Ride quality	1.11%
Appearance	0.59%
Movement	0.19%
Layout	0.64%
Facilities	0.84%
Total	5.98%

(iii) The prices vary considerably for each attribute level, with some items, such as escalators, costing more than the medium investment budget. Such disparities in price encourage the spending of unused budget points on items of low priority.

(iv) The analysis of the priority evaluator results are too simplistic. Using the proportion of total expenditure allocated to an attribute as a reflection of that attribute's worth ignores the constraint imposed by the unequal pricing structure and the budget limitations.

(v) When assessing the station quality variable in the SP experiment, respondents considered their own optimum budget allocations for the priority evaluator. This means that their valuation of the average station packages, if actually applied to their station, could only be indirectly estimated. The 80% value modification represents a very approximate estimate.

This approach was also used in studies of British Rail rolling stock improvements (The MVA Consultancy, 1986), with the advantage that respondents could be interviewed while travelling, so enabling them to visualise more easily the items under investigation. In response to the limitations of the earlier analysis, the approach was developed for the study of rolling stock improvements in an internal paper (The MVA consultancy, circa 1986.) This suggested that threshold effects may be identified by the proportion of individuals moving at least one level from a particular starting point. Maxima are also represented by the proportion of total expenditure allocated to an attribute, from a particular starting point. It was recommended that these two features should therefore be employed together in the estimation of a fare value for each attribute level. This would be done by using the geometric mean weighted by the number of individuals moving from each level, multiplied by the number of respondents and the value of the investment established in an accompanying SP experiment.

With this approach, information is now provided on the value of each attribute level and percentage fare equivalents for each attribute can then be calculated. This procedure was carried out on a comparison between travellers on standard and refurbished coaches. The difference between the each group's starting levels implied the utility of refurbishment. The resulting value was 5.98% of fare and a breakdown by rolling stock attribute is given in Table 7.18. The analytical approach used here gave a better estimate of priorities than the proportions of total expenditure alone, because it perceives prices as boundary values. It takes account of the numbers choosing to purchase a level (the price is expressed as a percentage of total expenditure) and therefore measures the way valuations change across each level.

Later applications of the priority evaluator method by the MVA consultancy (Copley, Bouma and de Graaf, 1987; Copley and Bates, 1988) used a similar approach to those described above, with a number of important modifications. As before, large-scale attitudinal surveys were carried out, at two stations belonging to Dutch Railways. In the follow-up home interviews, a combination of the priority evaluator method followed by an SP experiment was again used, but changes were introduced into the former survey instrument. These alterations comprised the following:

(i) The attributes were presented on separate boards, which the respondents set out in order of preference. This removed any bias pertaining to the ordering of the variables.

(ii) Respondents were first asked to allocate from base levels a budget equivalent to their present situation. They then allocated budgets of 15 additional points and then a further 10 points. At each stage they had the opportunity to sacrifice all or part of a previously purchased attribute level in order to release points for expenditure elsewhere.

Table 7.18 Values in Minutes and Percentage of Fare Established for Each Station Attribute at Sittard (Copley et al, 1987)

Attribute	Level	Value in Minutes	Value as a Percentage of Fare		
			0.12	0.18	0.26 *
General Appearance	Level 2	2.3	1.7	2.5	3.6
	Level 3	7.5	6.3	8.0	11.6
	Level 4	23.8	17.0	25.4	35.7
Facilities Around Station	Level 2	3.9	2.8	4.1	6.0
	Level 3	12.8	9.1	13.6	19.7
	Level 4	16.4	11.7	18.0	25.4
Access to Platforms	Level 2	5.9	4.2	6.3	9.1
	Level 3	13.0	9.3	13.9	20.0
	Level 4	27.4	19.5	29.3	42.3
Waiting Facilities	Level 2	3.2	2.3	3.4	4.9
	Level 3	6.4	4.6	6.9	10.0
	Level 4	12.8	9.1	13.7	19.8
Information	Level 2	1.4	1.0	1.5	2.1
	Level 3	3.7	2.7	4.0	5.8
	Level 4	12.8	5.2	7.8	11.2

* Alternative values of time in Dutch florins per minute

(iii) The earlier requirement to spend as much of the budget as possible was removed. Price variation was also less extreme so that together, these changes might reduce expenditure of 'left-over' points on unimportant items.

(iv) A travel time variable was included on the priority evaluator game-board as well as in the SP experiment.

(v) Random utility theory, applied within a discrete choice model, was used to analyse the responses.

Having respondents allocate budgets in relation to an approximation of the present situation places them more clearly within their current travel experiences, but it is likely that a degree of inertia may result from respondents beginning from their previous allocation. Removing any requirement to spend all the budget avoids misleading expenditure of surplus budget on less important items. The separation of the attributes onto individual boards approaches a problem not previously identified in studies that use a conventional priority evaluator game-board. It is perhaps likely that the arrangement of the scales is important, as respondents may scan the board from one side to the other and unconsciously establish a hierarchy associated with the order in which they identify the attributes. The exercise of ordering the attribute boards prior to the exercise also facilitates their familiarity with the variables and the layout.

Although a 'station quality' variable was presented in an accompanying SP experiment which included fare and journey time, it was considered valuable to also include the journey time variable in the priority evaluator, so that direct relative valuations of facilities against time could be calculated. While the inclusion of such a variable has its attractions, it may be seen that such a development presents a potential problem that Copley and Bates (1988, p1) previously identified regarding the mixing of secondary variables with primary variables in SP experiments. It is a possibility that the latter will dominate the choices.

In defence of this development, it can be argued that as the priority evaluator method allows many more levels of the time variable to be presented, respondents can achieve certain minimum thresholds, after which point secondary items become more important. Careful pricing and incrementing of the time variable will be critical in adjusting the options to the point where secondary items receive sufficient attention from the respondents. This particular topic is an important issue and one, it would seem, that is not considered in reports of the above survey.

The most significant innovation introduced by this priority evaluator application is the analytical approach. It will be seen that in other studies, the analytical approaches are compromised by the constraints imposed by the price structure and budget used. In response to the limitations of these previous approaches, the MVA consultancy present a useful advance by the application of discrete choice theory to the analysis.

It will be seen that in choosing a particular combination of variable levels, a respondent rejects every other possible combination. On a priority evaluator game-board, the total number of possible combinations is the sum of all the variables, defined by the number levels each possesses. Thus a game-board with two three-level variables and one four-level variable presents 36 (3 x 3 x 4) combinations. Of course, not all these combinations will be available to respondents, as the budget available will constrain their choices. This much has already been established in previous studies. The step taken by the MVA consultancy was to represent the utility of a combination as a function of the constituent variable levels. The coefficients can then be estimated using maximum likelihood procedures within a standard multinomial logit model.

The model parameters established with this method were based on analysis of the responses of 159 individuals. From these results, the value of items in terms of time units could be estimated by dividing each model coefficient by the coefficient for travel time. The resulting values, together with their conversion to monetary units over a range of fare/time exchange rates, are shown in Table 7.19. From these values for each variable level, the overall value of a package of improvements can be calculated, according to the number of respondents choosing each level.

In the SP experiment which followed the priority evaluator exercise at each interview, 'station quality' was offered as a two-level variable (present level and improvement) with time and fare at three levels. Unlike the earlier applications, the higher level of the 'station quality' variable was equivalent to an improvement package already planned by NS, rather than the respondents' individual budget allocations.

The valuations of travel time derived through the application of an 'exploded' logit model are also presented in Table 7.19. The model was expanded down to the fifth ranked choice, from a total of nine ranked options. The value of time was considered to be very high, approximately twice the expected figure, so that the value of the station improvements would be more in the region of 11% of fare if the NS established value of time was used. No explanation for this high value of time is offered.

The use of discrete choice theory in the analysis of the priority evaluator is a procedure already established in the field of SP experiments. In its application to the priority evaluator method, it represents an analysis of comparable sophistication to those applied to such experiments.

Table 7.19 Coefficients and Values Established in a Stated Preference Experiment Administered at Sittard Station (Copley et al, 1987)

Variable	Coefficient*	t-Statistic**
Time (mins)	-0.217	-17
Cost (fl)	-0.419	-17
Facilities	-1.738	-17
Value of Time (fl/mins)	0.13-0.26	16
Value of Improvements (% of fare paid)	11.05-22.09	14

*The coefficients are negative because they relate to loss of utility as time and fares are increased and the improvements option is foregone.

**t-statistics have not been adjusted to take account of "exploded" observations

Discrete choice theory advances the analysis of the priority evaluator method in the following ways:

(i) It uses a robust modelling procedure (maximum likelihood estimation) which yields statistics indicating the reliability of the coefficients derived (for example, t-statistic and rho bar squared);

(ii) It allows the responses of individuals facing different choice sets to be combined, if the full choice set open to each respondent is identified.

Regarding the second item above, it was found that each respondent was presented with a very large choice set. This imposed a burdensome computational task. As a result, the concept of 'dominance' was applied to eliminate alternatives considered unlikely to be chosen. One alternative was considered to dominate another if at least one of its elements was higher while all the rest were the same or greater. Note that these assumptions are necessary only when respondents are not required to spend all their budget. In a situation where all the budget must be allocated, lower value options are automatically eliminated. The application of the dominance criteria necessitates the rejection of responses from those individuals who chose dominated alternatives.

The robustness of the logit model estimated from the responses to the priority evaluator is greatly improved by the removal of dominated alternatives. Nevertheless, the model only uses the respondents' first choices, compared to the multiple responses from conjoint analysis experiments.

This limitation has to be weighed against the advantages previously associated with the priority evaluator method:

> "For any group of respondents that have faced the same priority evaluator board, the variables making up a given combination are constant: the variation is derived from the different choice sets and, of course, the chosen combination. The lack of variation in the underlying variables tends to make the estimation relatively less efficient, compared with standard SP problems where the efficiency of co-efficient estimates for a given sample size is very high. For this reason, larger sample sizes are required with the priority evaluator. The compensation comes from the greater clarity of the task and the ability to deal with more descriptive material than can easily be accommodated in a typical SP study."
>
> Copley and Bates (1988, p6)

It may be seen, from the research undertaken by the MVA Consultancy, that the usefulness of the priority evaluator method has been significantly enhanced, particularly in the advancement of the analytical approach. In these studies, it has been presented as an exercise more suitable than conjoint analysis SP experiments in the measurement of the relative values of secondary choice factors. This is contrast to the earlier series of studies for London Underground, though it has been recognised that the early London Transport work had some deficiencies not reproduced in these later BR studies.

Rather than reject conjoint analysis approaches, the MVA studies suggest that the priority evaluator method and conjoint analysis should be given different roles. The former allows quantification of qualitative variables; this allows valuations of any combination of the factors in an improvement package. The latter allows the valuation of a planned improvement package in the context of the overall journey.

Thus a two-step process is established, in which the preferences established by the priority evaluator method indicate optimum improvement strategies, which may then be assessed in relation to primary factors through the application of SP experiments.

7.4 Previous Applications of SP Approaches to the Valuation of Passenger Facilities: Overview

This chapter has reviewed the application of SP techniques over a number of studies for London Underground, British Rail and some continental operators. It has shown how conjoint measurement (ranking conjoint analysis) was introduced in preference to early applications of the priority evaluator method. In later studies, this in turn was rejected in favour of a more advanced version of the priority evaluator method, used in conjunction with simpler conjoint measurement experiments.

Comparison of Valuations

Table 7.20 to Table 7.22 summarise some of the values for passenger facilities obtained from the studies examined above. Note that there are difficulties in making direct comparisons between the studies because of the different rail systems involved, the different definitions of the facilities and the different survey techniques employed. Despite this, the following may be noted:

Valuations of individual attributes

(i) there is a strong contrast between the train facility values derived from conjoint measurement alone and the priority evaluator with conjoint measurement, which are much lower;

(ii) London Underground and DSB valuations are of similar magnitudes, but Bristol Parkway's are lower and the NS valuations higher (all these studies include fare or time directly with individual attributes.)

Valuations of overall improvement packages

(iii) it appears that individual attribute valuations are inappropriate: valuations of single overall improvement variables produce much smaller and more closely positioned values.

Table 7.20 Summary of Valuations Obtained From a Range of Studies: On-Train Passenger Facilities and Related Features

All valuations expressed as percentages of fare paid.

Individual improvement	London Underground[1]	London Underground[2]	Inter-City[3]	Inter City[4]	DSB (inter-city)[5]
Well ventilated carriage	12	-	-	-	-
Clean carriage	11	-	-	-	-
Well regulated heating	10	-	-	-	-
Smooth ride	10	-	-	-	-
Uncrowded carriage	9	-	-	-	-
Comfortable seats	9	-	-	-	-
Quiet carriage	9	-	-	-	-
Information (delays)	-	10	-	-	-
Less noisy	-	8	-	-	-
Reliable	-	4	-	-	-
Information (direction)	-	2	-	-	-
Low seat density	-	-	8.5	-	-
On-train Catering	-	-	1.1	-	-
Heating	-	-	-	1	-
Seat quality	-	-	-	0.5	-
Information	-	-	-	1	-
Ride quality	-	-	-	1	-
Appearance	-	-	-	0.5	-
Layout	-	-	-	0.5	-
"Facilities"	-	-	-	1	-
Improved toilet	-	-	-	-	16.5
Bikes allowed on train	-	-	-	-	14.5
Drinks & cold meals	-	-	-	-	14
Family carriages	-	-	-	-	13
Adjustable seating	-	-	-	-	11
Improved environment	-	-	-	-	10
Only half seats occupied	-	-	-	-	6.5
More cleaning	-	-	-	-	5
2+1 seating	-	-	-	-	3
Open plan coaches	-	-	-	-	-6
All seats reserved	-	-	-	-	-7

1 London Transport (1984)
2 Steer, Davies & Gleave Ltd (1984d)
3 Steer, Davies & Gleave Ltd (1984c)
4 The MVA Consultancy (1986)
5 Steer, Davies & Gleave Ltd (1986a)

Table 7.21 Summary of Valuations Obtained From a Range of Studies: Individual Station Passenger Facilities and Related Features

All valuations expressed as percentages of fare paid.

Individual improvements	London Underground[1]	Inter-City[2]	DSB (inter-city)[3]	Network South-East[4]	NS (urban)[5]
Clean station	14	-	-	-	-
Modernised station	6	-	-	-	-
(Clean/Modernised)	28	-	-	-	-
Arrival information	10	-	-	-	-
Information (delays)	10	-	-	-	-
Ticket offices	6	-	-	-	-
Faster ticket purchase	4	-	-	-	-
Escalators to platform	-	7	-	-	-
All-weather shelter	-	5	-	-	-
Lift to platform	-	6	-	-	-
No of Ticket Windows	-	3	-	-	-
Enclosed waiting room	-	3	-	-	-
Trolleys available	-	3	-	-	-
Ramp to platform	-	3	-	-	-
(Trolleys and ramp)	-	7	-	-	-
Travel centre	-	3	-	-	-
TV screen information	-	3	-	-	-
2 x seats on platform	-	2	-	-	-
Direct access to trains	-	2	-	-	-
Separate cafeteria	-	2	-	-	-
Telephones	-	2	-	-	-
Toilets on platform	-	1	-	-	-
Improved toilet	-	-	17	-	-
Bikes allowed on train	-	-	15	-	-
Drinks & cold meals	-	-	14	-	-
Family carriages	-	-	13	-	-
Adjustable seating	-	-	11	-	-
Improved environment	-	-	10	-	-
Only half seats occupied	-	-	7	-	-
More cleaning	-	-	5	-	-
2+1 seating	-	-	3	-	-
All seats reserved	-	-	-7	-	-
Open plan coaches	-	-	-6	-	-
Appearance	-	-	-	0.5	-
Ticketing	-	-	-	1	-
Information	-	-	-	0.5	-
Toilets	-	-	-	0.5	-
Shelter	-	-	-	1	-
Movement	-	-	-	1	-
Purchasing	-	-	-	0.5	-
Access	-	-	-	-	29
Appearance	-	-	-	-	25
Facilities	-	-	-	-	18
Waiting facilities	-	-	-	-	14
Information	-	-	-	-	8

1 SDG, 1984d; 2 SDG, 1985; 3 SDG, 1986a; 4 MVA, 1985d; 5 Copley & Bates, 1988

Table 7.22 Summary of Valuations Obtained From a Range of Studies: Overall Train and Station Improvements

All valuations expressed as percentages of fare paid. Unless a single variable was used to express an overall package of improvements, values are sums of individual attribute values (marked "s").

Overall improvements	% Fare Value
Trains	
London Underground[1] (s)	24
DSB (Denmark inter-city)[2] (s)	94 (improvements only)
InterCity[3]	6
Network SouthEast[4]	4
Stations	
London Underground[1] (s)	58
DSB (Denmark inter-city)[2] (s)	115
InterCity[5] (s)	47
NS (Netherlands)[7] (s)	18-94
London Underground[6]	14-29
InterCity[8]	2
InterCity[5]	8
InterCity[9]	4-5
Network SouthEast[10]	5
NS (Netherlands)[7]	11-22

1 Steer, Davies & Gleave Ltd (1984d)
2 Steer, Davies & Gleave Ltd (1986a)
3 The MVA Consultancy (1986)
4 SDG Research (1989)
5 Steer, Davies & Gleave Ltd (1985)
6 Maw, J; Muir, RM & Hendy, P (1976)
7 Copley and Bates (1988)
8 Steer, Davies & Gleave Ltd (1984c)
9 Steer, Davies & Gleave Ltd (1986b)
10 The MVA Consultancy (1986d)

Problems With Previous Applications of SP Techniques

Concerning issues relating to the application of SP methods to this area, the main criticism of the use of conjoint measurement was the implausibly high values obtained, not only when totalled across all improvements but also for some individual improvements. One reason put forward for this was the inclusion of only one primary variable (fare) in an exercise. Some evidence for this appears to have been provided in the study of Bristol Parkway and in later studies by the MVA Consultancy. In these instances, variables representing packages of improvements are given values against fare and other primary variables which are much lower than the totals implied by the summing of individual improvement-fare trade-offs.

The inclusion of only one primary variable may contribute to the high values observed for individual facilities, but it may be argued that a stronger effect is likely to exist in the presence of thresholds on travellers' willingness to pay for improvements. This is recognised in the report of the Bristol Parkway study but not enlarged upon. It is implicit in the later MVA approaches by the omission of a fare variable at the priority evaluator stage and the derivation of individual attribute values (by calculating proportional values from the values of a total investment package in a later conjoint measurement exercise.) Only in later applications of the priority evaluator is a quantitative variable (time) included. This suggests a recognition that there is a distinction to be made between the value of a facility introduced on its own and the value of a facility as part of a wider package of improvements.

Regarding the specific SP survey instruments examined above, there still remains some concern relating to the derivation of plausible monetary values for facilities. This is for two reasons:

(i) the processes of ranking alternatives or spending an investment budget are far removed from real life travel choice situations;

(ii) there are technical limitations in both these approaches, particularly in the analysis of the results: rankings can "force" a weighting to be obtained for an unimportant variable because they present choice situations in the absence of a full range of other alternatives, while the values from budget allocations are very much conditioned by the price structure/budget combination, for which suitable design guidelines have not been established.

These problems were to give the impetus for the new development in the valuation of passenger facilities, outlined in chapter eight: the use of discrete choice SP experiments in conjunction with conjoint measurement exercises.

8 THE EVALUATION OF IMPROVEMENTS TO PASSENGER FACILITIES ON LONDON'S UNDERGROUND: BACKGROUND RESEARCH AND SURVEY DESIGN

8.1 Introduction

These next three chapters report on the development and application by the author of stated preference (SP) techniques to the evaluation of investments in rolling stock and station facilities. The work reported here was financed as a consultancy project commissioned by London Underground Ltd (LUL). The study was directed by the objective of producing monetary valuations of qualitative service improvements which could be applied in a demand forecasting model. A major issue was the general opinion among LUL planners that the monetary values obtained from previous SP studies were generally implausibly high. A principal aim in each case was to establish a more theoretically robust approach than those used hitherto, with the reasonable expectation that lower monetary valuations would result.

8.1.2 Study Context

At the beginning of 1990, London Underground Ltd (LUL) identified the Northern Line section of its network as a priority case for investment in new signalling, tunnel improvements, new rolling stock and some station upgrading. It was intended that some £500 million would be made available for these improvements, which would be implemented before the end of the century.

A substantial proportion of the intended expenditure would be available for new or improved rolling stock, but it was recognised that certain priorities had to be set in the early planning stages with regard to the type of facilities and quality of environments that could be provided in new carriages. The Northern Line Modernization Team (NLMT), an internal body created specifically for the project, decided that market research should be carried out before any major commitments were made concerning the design of new rolling stock. The value of alternative improvement strategies would be assessed in terms of their impact on passengers' perceptions of the quality of service delivered by the new rolling stock and the willingness of passengers to pay higher fares for such improvements.

An SP survey was considered the most suitable approach because of the difficulty of obtaining revealed preferences relating to qualitative service improvements and the intention to evaluate some facilities which had not as yet been introduced anywhere on the Underground network.

LUL released a project brief to competitive tender. The author wrote a study proposal on behalf of Steer Davies Gleave and the work was subsequently awarded to this company. The author was assigned to manage the market research, with direct responsibility for the design and analysis of the SP surveys. The analysis presented here, and reported in the next two chapters, represents a subtantial amount of additional work to that carried out for the original study.

In the project proposal submitted by Steer Davies Gleave, it was argued that conjoint analysis was the most appropriate SP technique to use. The main alternative was considered to be the priority evaluator method, which had been applied to rolling stock and station facilities research in other studies (see chapter seven.) The preference for a conjoint analysis approach over the priority evaluator method was for the following reasons:

(i) the priority evaluator requires respondents to behave as 'quasi-planners' rather than consumers;

(ii) the priority evaluator requires the order of the improvements to rolling stock and station attributes to be determined in advance of the surveys;

(iii) the analytical techniques available for the priority evaluator are less theoretically robust than those developed for conjoint analysis approaches, with the inability to measure interaction effects;

(iv) sophisticated software was available for the design of conjoint analysis SP experiments, allowing the surveys to be carried out on portable computers.

8.1.3 The Study Objectives

The project brief provided by LUL required the following:

> "In order to provide a realistic assessment of the relative merits of the options it will be necessary to evaluate relative passenger responses to the proposals. Furthermore, in most cases a monetary equivalent of the passenger's willingness to pay through higher fares for given improvements will be needed."

A wide range of improvements were being considered. Most were related to characteristics of rolling stock, except for one item relating to station environments. The attributes under consideration by LUL were:

(i) air conditioning/ventilation/heating improvements;
(ii) improved quality of ride (through suspension improvements and/or track repairs and renewals;)
(iii) draught relief at stations;
(iv) articulation and inter-car gangways;
(v) train cleanliness and newness;
(vi) noise.

The first concern arising from this list was that item (iii), relating to station environment, would not sit easily in the same SP exercise as the other variables, as it related to the transfer and interchange elements of Underground journeys and not the train journey proper. It was therefore decided that a separate station attributes SP exercise should be developed, in which additional items could be included if desired. Discussions with the NIMT resulted in three additional attributes being investigated: escalator reliability; train exterior colouring and the removal of external graffiti on trains. The exterior quality of trains would be experienced while travellers waited at their stations, so that these train attributes appeared appropriate for inclusion in the additional station SP exercise.

Information facilities were not included in the project brief, as this was regarded by the NLMT as an issue likely to be considered separately from the current Northern Line proposals (a network-wide strategy for information improvements was currently being implemented independently from the Northern proposals.) However, subsequent discussion with the NLMT led to information being added to the list of attributes. This decision was made on the basis that the results from the study could have a wider application than the Northern Line proposals (for example, a new scheme was being considered for the Jubilee Line) and information facilities were likely to be important components of future projects.

The monetary valuations derived from the SP surveys would be used by LUL as components in their demand forecasting models. The models are used to predict:

(i) the potential for recouping the cost of introducing new passenger facilities through increased fares;

(ii) the potential for increasing demand for Underground services through the provision of new facilities.

These models use elasticities derived from aggregate time series demand data. While they are based on real travel data, they are incapable of fine measurements relating to marginal items such as passenger facilities, because of the aggregate nature of the data. For this reason, the results of SP studies act as important enhancements to these models, but are subject to some of the problems relating to issues of realism and scaling discussed in the context of forecasting in chapter six. The way in which the SP results would be applied had some influence on the design of the SP studies and this is discussed in a later section.

Discussions with the NLMT identified a number of travellers' characteristics on which the survey sample could be segmented, which included the time of day a journey is made, geographical area (central/outer London), journey purpose, ticket type, frequency of travel on the Underground, sex, age and income. The study budget only allowed for six hundred interviews, given that face-to-face interviewing was considered essential (a lot of graphical and other explanatory material would be needed to describe the qualitative service improvements and the use of portable computers was the preferred method of data collection.) With target of some 100 individuals in each segment (75-100 respondents representing a reliable number for modelling purposes), the number of possible segments was limited.

Earlier passenger surveys undertaken by LUL suggested that time of travel was correlated with journey purpose, which in turn had strong correlations with sex, age and income. The peak and off-peak periods of the day, as defined in LUL's pricing policies and service operations, were therefore taken as the main form of segmentation. Within these two categories, distinctions would then be made on the basis of ticket type, which could be grouped as period tickets, One Day Travelcards (off-peak only) and ordinary single/return tickets. The buying decisions associated with each of these types of ticket were seen to be quite different (relating to very different prices and periods of travel.) Time of travel and ticket type therefore produced five segments (One Day Travelcards were not valid for the peak). Each of these corresponded to the main segments used in the NLMT's forecasting models. It was decided that geographical location was of less importance, because it was not a variable specifically used in the forecasting model, but efforts should be made to ensure that a representative spread of respondents over the Northern Line should be achieved.

It was considered prudent to carry out exploratory research in advance of the SP survey. This involved two preceding stages:

(i) a literature analysis of previous London Underground studies of passenger facilities, some of which used SP techniques and have been discussed in chapter seven;

(ii) group discussions with Northern Line users, in which the issues of particular interest to the study could be examined in depth.

The information obtained from this initial research would provide:

(i) a framework for understanding the way passengers perceive rolling stock attributes and prioritise between them and

(ii) guidance on the way the attributes of interest should be defined and presented in the SP surveys.

Once the questionnaire was designed, a pilot survey would be conducted to test its efficiency, before embarking on the main survey fieldwork.

The study was commissioned on 7 March 1990, the results to be submitted in a final report at the end of July 1990. This timescale was considered practical but inevitably there was some limitation on the time available to develop and test the research method. In later sections, criticism is made of a number of shortcomings identified in the research design. Some of the weaknesses of the final SP survey design might have been identified before the main fieldwork if more time and resources had been available for testing it.

8.2 Findings from the Preliminary Research

A general discussion of Underground users' attitudes towards passenger facilities was provided in chapter two, in which passenger facilities were seen to be given low ratings in terms of the quality of service. These results relate to the whole Underground network, so that attitudes towards facilities on the Northern Line are only indirectly inferred. Given that the Northern Line has some of the oldest rolling stock and station fabric on the network, it is unlikely that travellers' perceptions of the quality of service delivered by passenger facilities would be higher than the network averages.

Over the years, various reports have been commissioned by London Underground relating specifically to individual service attributes relevant to the improvements examined in this study. These are referenced in the sections below, which identify the main issues relating to different service characteristics. Findings from the group discussions and earlier SP studies also contribute to these sections.

Four group discussions were undertaken in April, two at the North end of the northern Line (Finchley) and two in the South (Wimbledon.) At each location, one group was composed of peak-time travellers and the other of off-peak travellers. The participants were recruited at local Underground stations and paid to attend the group discussions. These divisions aimed for a good spread of respondents and reflected the assumption that preferences were likely to alter between peak and off-peak travellers. Experienced group moderators were used and the author was able to observe these group discussions at first hand.

Each group discussion began with a general airing of views on the Underground and the Northern Line in particular. These tended to focus on the perceived lack of or inaccuracy of information about services, overcrowding, fear of attacks and a sense of decline related to the Northern Line, despite some awareness of proposed improvements. The discussions were then guided into a construction of the various elements of the train services and stations. In most of the groups, some of these arose spontaneously, either in the context of the problems of passenger security or in trying to identify the factors contributing to the overall sense of discomfort expressed by many of the respondents.

Ventilation/Heating/Air Conditioning

In-train climatic environment is a complex variable which does not easily lend itself to graphical presentation or precise verbal description. Climatic environment is the product of a number of factors: temperature, humidity and air speed. Even in engineering terms, this attribute is difficult to quantify because a number of different approaches to measurement can be taken, none of which are fully compatible with each other (British Rail Research, 1990.) The objective at this stage, for all the attributes being examined, was to establish definitions which would be meaningful to both the NIMT and the respondents to the SP survey.

The quality of in-train climatic environment is closely related to crowding levels, and this was recognised by respondents in the peak users group discussions. Conditions were recognised as generally uncomfortable, effects such as condensation on windows and unpleasant odours illustrating poor quality ventilation. Such perceptions of poor quality are significant in the light of the importance given to ventilation in Table 2.2 in chapter two. It was appreciated that conditions were more tolerable when the train was moving, though windows that could be opened had the disadvantage of allowing dirt to be blown in from the tunnels. Delays in tunnels were therefore made more frustrating because of the deterioration in air quality that they produced.

One factor which respondents identified as an influence on attitudes to this issue is the time of year when travellers are asked to consider it. The main SP fieldwork was to be undertaken in June, which could have had the effect of heightening the importance of ventilation in respondents' minds.

Ride Quality

A detailed study of ride quality on Underground trains was undertaken by SDG Research in 1987b (report no 386; also London Underground Ltd, 1987.) The study sought to identify perceptions of ride quality and investigate its relationship with passengers' activities while aboard their train. The study concluded that:

> "Passengers rating ride as unfavourable found it jerky, bumpy and rough. Roughness or harshness appeared to be the major cause of dissatisfaction. Passengers reporting an unfavourable ride also viewed the train as a difficult place to undertake activities (reading, writing, etc) whilst travelling. However, whether or not a passenger actually entered into activities was not particularly dependent on his perception of ride."
>
> (SDG Research, 1987)

This was an important finding in relation to this study, because it had been considered that quality of ride could be defined in terms of the ease of undertaking different activities. This was still relevant, but the earlier research suggested that passengers can perceive and evaluate quality of ride characteristics as an attribute in its own right. The present study therefore benefited from further exploration of the way individuals describe ride quality and how these should be used to describe different rolling stock options (eg 'bumpy'; 'rough'.)

In the group discussions, most respondents were conscious of varying levels of quality of ride. They were able to describe different types of motion along different sections of track and, significantly, at alternative levels of crowding. This last factor appeared important because it determined the ease with which passengers could obtain a seat or access to support rails. Trains were also perceived to sway more noticeably when a carriage was virtually empty. The Northern Line was viewed as 'bumpier' than most other lines, though the quality of rolling stock was only seen as one factor: speed and track levels were also mentioned as possible sources of an uneven ride. The NLMT expected that the new rolling stock would allow smoother braking and acceleration than at present, but respondents in the group discussions did not perceive current changes of speed to be a problem. Only unexpected braking caused problems. A minority mentioned activities as indicators of ride quality, the ability to read a newspaper being the most commonly cited.

Articulation and Inter-Car Gangways

No market research appeared to have been carried out previously on the subject of access between carriages. This was a result of articulated rolling stock, with gangways running from one carriage to the next, being a fairly new concept for the Underground (the Docklands Light Railway carriages are the only current examples.) Also, articulated rolling stock would be beneficial chiefly to the operators. This is because it would reduce the number of bogies required by each carriage and therefore the running costs. Nevertheless, inter-car gangways could offer some benefits to passengers, by increasing standing room and allowing movement away from crowded parts of the train.

The lack of previous market research meant that the group discussions would be the main guide to defining inter-car gangways in the SP surveys. In the group discussions, the issue of interconnecting doors on trains was not raised spontaneously, but passenger security was. It was in this context that the value of inter-car gangways seemed to be mainly evaluated by respondents. Some viewed them positively, as a means by which to escape from threatening behaviour or to be near other passengers and the guard late at night (this was particularly important to women.) Others viewed inter-car gangways negatively, because they would enable muggers to move more efficiently along the train (so-called 'steaming', already facilitated by use of the existing emergency doors between carriages.) Some respondents did perceive potential benefits in terms of being able to move away from crowded areas, though the advantages were thought to be limited.

Train Cleanliness and Newness

No specific market research has been carried out on cleanliness and newness of rolling stock, though cleanliness has often been included in general attitude studies. In chapter two, Table 2.2, cleanliness was seen to be ranked above most other attributes of Underground services. While the quality of general cleanliness was perceived by passengers to be less unsatisfactory than most other attributes (Table 2.5, chapter two), it was still rated poorly overall.

A study of passengers' attitudes towards different rolling stock designs for the Central Line (SDG Research 1987a), in which three working prototype carriages were used, showed that 'newness' appeared to be perceived as an attribute in its own right, though it was related to a number of other attributes. Newness implied better reliability, ride quality and most of all, better standards of cleanliness. This last relationship was noted in an earlier study (THP, 1983, p21) where it was stated that 'modernity tended to subsume perceptions of cleanliness.'

In the group discussions, the link between cleanliness and newness was reaffirmed. Standards of cleanliness were regarded as poor on the Northern Line and this was blamed largely on the age and design of rolling stock (eg the ribbed flooring was seen as a major dirt trap.) For these reasons, refurbishment of existing carriages was not looked on favourably, though it was recognised that very regular and thorough cleaning would greatly improve the interiors.

Off-peak travellers perceived standards of cleanliness the most negatively, partly because of the litter that would be left after the peak period and partly, it seemed, because litter and graffiti were more noticeable in uncrowded carriages. Respondents generally described low standards of cleanliness chiefly in terms of dirt and oil rubbing off onto hands and clothing. Comparisons were made spontaneously with other Underground lines, most of which were seen as offering better standards than Northern Line trains. District and Circle Line trains were considered the cleanest, though this may reflect the fact that these lines operate larger and airier rolling stock through 'cut-and-cover' tunnels as opposed to smaller tube tunnels as on the Northern Line.

Noise

Noise nuisance experienced on Underground trains was previously the subject of a detailed study (London Underground Ltd, 1987), of which some of the results have been discussed in chapter seven. It was found that passenger discomfort from noise was only partly the result of actual noise levels: the type and variability of noise could also be a source of annoyance. Table 8.1 summarises passengers ratings of different noise sources, in terms of the annoyance they caused. Some noise sources, such as passengers' stereos, are beyond the control of the Underground operators, while others, such as the noise of braking, are directly related to rolling stock design. The biggest difference between the ratings for the Northern Line and the Underground overall is that the public address system causes more annoyance on the former, relative to other sources of noise annoyance. Of all the Underground lines, the Northern Line was rated the second worst for noise annoyance.

Table 8.1 Passenger Ratings of Noise Annoyance on the Underground, by Source of Noise

Item	All Underground	Northern Line Only
Other peoples' stereos	2.62	2.66 (2)
Noise of train braking	2.75	2.58 (1)
Rattles and bangs from inside train	2.79	2.90 (5)
General roar of train	2.81	2.75 (3)
Rattles and bangs from outside the train	2.88	2.95 (9)
Squeaks from inside the train	2.88	2.90 (5)
Noise from children	2.96	2.94 (8)
Public address system	3.08	2.82 (4)
Noise of train starting	3.16	2.90 (5)
Doors opening/closing	3.21	2.97 (10)
Other peoples' conservation	3.31	3.17 (11)

The London Underground study of noise annoyance also related passenger ratings with scientific observations of noise. A close correlation between the two was observed on all lines except the Bakerloo and Jubilee Lines, for reasons that are not clear. This would suggest that the operator's and passengers' understanding of noise levels is likely to be more closely related than with some other variables, which are experienced more subjectively. In describing noise annoyance, passengers often related the effects to activities they would like to carry out while using an Underground train. Conversation was not seen as practical even when noise annoyance was minimal, which suggests that a certain level of train noise is acceptable to passengers (this issue was raised in the group discussions.) The ability to relax, read and think all improved with reductions in noise annoyance and this suggested that some reference to passengers' ability to undertake activities should be included in the present SP survey to convey different degrees of noise annoyance.

In the group discussions, general background noise from the train was perceived as unavoidable by many respondents, and was therefore not cited as a particularly high priority for improvement. Irregular and varying types of noise were seen as more intrusive, described as sudden thumps and screeches. This seemed to confirm some of the findings from the earlier Noise Annoyance study, where 'rattles', 'bangs', 'squeaks' and sounds from braking were in the top half of the list of intrusions.

Information

The study was concerned with only a few specific methods of information presentation (electronic displays; a public address system) and the content of the information they might provide (eg the position of the train on the line; announcements of station stops.) The earlier study of passenger attitudes towards alternative Central Line designs (SDG Research, 1987) included the examination of new on-train visual displays and announcements. The visual displays were viewed positively by respondents but concern was expressed over their usefulness in crowded conditions, when they would be difficult to see. A majority of respondents (63%) preferred all station stops to be announced during the journey, but these were largely infrequent travellers. In preceding group discussions, regular peak time travellers thought that regular announcements of station stops could become irritating.

The 1983 Rail Priority Study (London Transport, 1984) included on-train information as one item among a number of attributes. The previous chapter showed that this study observed that a new visual display device, giving direction of travel and location, was given a low value (less than one pence fare increase) when compared to such attributes as improved cleaning, which had a value of seven pence. On-train announcements, in contrast, were given a substantial value (five pence).

These values do not indicate the intrinsic values of information facilities: only the value of improvements to the present situation. These different values may indicate that consistent information such as the location of stations was satisfactory, but irregular information, such as delays or service cancellations, would benefit from improvement. In the group discussions undertaken for this study, electronic displays were welcomed but doubts were expressed over their reliability and likely resistance to vandalism. As in earlier studies, regular peak travellers did not see the need for the announcement of station stops, yet valued the provision of irregular information. There was a desire that the information content of such information should be improved, mainly in terms of its accuracy and the clarity of the announcements.

One passenger facility not investigated in previous research was a direct communication link to the train driver. It was felt that such a system would only be feasible for emergencies, as few drivers were considered likely to be co-operative if they were used for general enquiries from the public. As a security measure, a telephone link was supported strongly.

Station and Train Exterior Improvements

Of all the previous Underground studies, only the earliest stated preference work (Maw, Muir and Hendry, 1976) investigated the value of draught reduction at stations and improvements to train exteriors. Draught reductions were given a medium value but the painting of carriage doors only received a moderate valuation. Escalators were seen to be valued highly as replacements for stairs and lifts, but there is no indication in any previous research of the value of improvements to the reliability of escalators.

In the group discussions, the topic of draughts was not brought up spontaneously by respondents. On being asked about them, respondents suggested that they offered some benefits as well as problems. Draughts announced the imminent arrival of a train and helped to reduce the stuffiness of the air at stations. Strong draughts up escalator shafts were more of a problem, but in general draught reduction was not seen as a major priority on the Underground. Graffiti on trains, particularly on the exteriors, was criticised severely, but many considered that the Underground authorities were 'trying their best'. Painted surfaces could discourage vandalism, it was suggested, although maintenance costs would rise. While welcomed as a moderate improvement, exterior painting was not considered a priority.

Conclusions from the Preliminary Research

The background research provided guidance on the way the SP surveys could use visual and verbal descriptions to convey qualitative improvements to passenger facilities. The findings from earlier studies would provide useful comparisons with those from the present study, though differences in market research techniques and other factors would in some cases limit such comparisons.

The group discussions allowed the issues of interest to be examined in depth, but caution has to be used in the interpretation of the findings. The respondents are encouraged to consider issues in more detail than might be practical in an SP survey and their stated opinions are the result of interaction with other group members. Skilled moderators are able to condense the highly qualitative information into useful summaries, but some detail is obviously lost and the reported findings inevitably subjected to the moderators' own interpretation. Hence the advantage of the individual responsible for the SP survey design being able to observe some of these discussions as they take place. It intended that the findings from the group discussions should be used to aid the interpretation of the SP data, but their chief contribution would be in providing guidance on the design of the descriptive material used in the surveys.

The preliminary research suggested that most passengers could envisage the many different improvements that could be introduced to rolling stock and stations, as well as the principal of paying higher fares to finance them. Many already appeared well informed as to the financial and operational constraints that required the operators to prioritise between alternative improvement strategies. Passengers already perceived varying degrees of service quality across the network and had identified their own priorities for improvements. The most important issue related to these attitudes towards improvements to qualitative attributes appeared to be the context in which they were introduced. Such improvements, while welcomed by passengers, were seen to be secondary to such issues as better train reliability, reductions in crowding and improvements to passenger security. If these factors remained deficient, the benefits of lesser qualitative improvements would be reduced.

8.3 Design of the Stated Preference Survey

8.3.1 Theoretical Issues

As the previous chapter demonstrated, SP techniques had been used in a number of earlier studies for London Underground. The opinion among the NLMT was that SP techniques could provide plausible measurements of the relative strength of passengers' preferences towards qualitative service attributes, but implausibly high monetary equivalent values. The fare variable was seen to have less impact on responses to SP exercises than would be expected in real choice situations.

From the discussion of response error in chapter six, it is clear that the way in which one designs and presents SP exercises is likely to affect the realism with which respondents assess such variables as fare. The primary concerns related to the design of the more recent SP studies carried out for London Underground were:

(i) only ranking exercises were used, with the following problems:

(a) respondents assessed 'within-mode' situations, in which the attributes would be presented as choices between competing versions of the same rail service, not in the context of real travel decisions already familiar to travellers, such as the mode choice or the decision whether or not to make the journey at all,

(b) rankings do not give a clear indication of behavioural responses to changes in passenger facilities, in that a traveller expresses a hierarchical set of preferences but does not indicate if the changes will make him or her more or less likely to use the rail service or pay more for it;

(ii) practical limitations on the number of options that could be presented to respondents required the use of fractional factorial designs which prohibited the investigation of most interaction effects.

The latter weakness was potentially of some significance, because of the way some respondents might approach the situation of trading fare for service improvements. For example, an SP exercise, using rankings and a fractional factorial design, might infer the following benefits attached by a respondent to three individual improvements:

(i) improvement A = five 'utils'
(ii) improvement B = seven 'utils'
(iii) improvement C = three 'utils'

The term 'util' is an abstract term implying a unit of utility. It corresponds to a relative measure of preference that might be implied by the values of the coefficients derived from a model of SP responses. In the above example, without a measure of any interaction effects, it could be assumed that the total benefit of these three improvements would be equivalent to 15 'utils'. Yet an important constraint could be operating to undermine this assumption: there may be some duplication of benefits derived from the different improvements, so that the introduction of one improvement may have little impact if another improvement introduced at the same time provides most of the same benefits.

Extending the example above, the following interaction effects might exist, such that:

(i) improvement A & improvement B = combined benefit of 10 'utils'

(ii) improvement A & improvement C = combined benefit of five 'utils'

(iii) improvement B & improvement C = combined benefit of 10 'utils'

(iv) improvement A, improvement B & improvement C = combined benefit of 10 'utils'

In this example, improvement B substitutes some of the benefit of improvement A in interaction (i), so that the sum benefit of the two improvements is reduced by two 'utils' from 12 'utils' to 10 'utils'. In interaction (ii), improvement A wholly substitutes the benefit of improvement C, so that the combined benefit is still only five 'utils' (the value of improvement A alone.) In interaction (iii), there is no substitution between improvements B and C, so that the combined benefit of 10 'utils' is the same as the sum of the separate benefits (seven 'utils' + three 'utils'.) When all three improvements are introduced, the benefits of improvement C are negated by the presence of improvement A, which in turn has some of its additional benefit subsumed by improvement B. The final benefit when all three are introduced (interaction (iv)) is considerably less than the sum of the individual benefits. In this case, an SP design that would not allow the measurement of interactions would lead the researcher to over-estimate the benefits of policies which involved more than one improvement.

The opportunity to investigate interactions is inevitably limited by the need to keep the number of options assessed by respondents to a practical minimum and therefore to use fractional factorial designs. This problem was recognised from the beginning and was one of the reasons why qualitative research was carried out in advance of the stated preference surveys. The group discussions would aim to establish, among other issues, the degree of substitution between different passenger facilities. This information would be used to establish variables that could be considered to be independent from one another. Where a close association was identified, interactions for these variables could be incorporated into the experimental design.

Another issue arises when the researcher attempts to derive a monetary value for each improvement and a value of all improvements introduced together. From chapter six, it can be seen that the ratios of the coefficients of the improvements to the coefficient of fare implies the rate of trade-off and therefore the marginal monetary value of the improvements. This is theoretically robust, providing the respondents are consistently compensatory in their choice behaviour and no interactions apply (see above.) In fact, the trade off between improvements and fare may not be consistent over all values of fare changes. Respondents' willingness to pay for improvements may be subject to a constraint in the form of thresholds.

Continuing with the above example, it might be seen that a respondent is willing to pay one pence extra fare for each 'util' of benefit derived from each improvement. The implied monetary values for improvements A, B and C are five pence, seven pence and three pence respectively. When all improvements are introduced together, and interactions/substitutions are taken into account, this might lead the researcher to assume that the monetary value of all three improvements together will be 10 pence (= 10 'utils'.) In the example used above, a respondent may be willing to exchange up to eight pence fare increase for improvements, but no more. Thus a cut-off point is imposed above which the benefits of additional improvements will not be measured in terms of a passenger's willingness to pay for them. This illustrates an important distinction between the relative benefits attached to different improvements when fare (or other quantitative service variables) are not included and the relative values when fare increases are included.

In an extreme development of the example used above, the individual might pay eight pence higher fare regardless of the type or number of improvements made. In terms of monetary equivalent values, no distinction can be made between improvements A, B or C, yet the respondent may in fact obtain more than twice the benefit from improvement B (seven 'utils') as from improvement C (three 'utils'.)

From this discussion, it may be concluded that limitations in the design of previous SP studies had the potential to give misleading results. Potential sources for error could be identified in the nature of the choices offered to respondents ('within-mode' ranking exercises require judgements unrelated to real choice constraints) and the ability of the experimental design to yield sufficient data (interactions and thresholds.) The objective in the design of the SP experiments for the current LUL study was to take into account these issues and develop an approach which improved the theoretical robustness of the results.

One significant advantage which this study had over previous studies for LUL was the availability of portable computers and a suitable software package ('Alastair', developed by Steer Davies Gleave) to administer the SP surveys. As well as the practical advantages such as improved data processing and faster interviews, a computer-based approach allowed greater flexibility in the design of the SP experiments.

Another important issue relating to the design of the SP experiments was how the results would be applied by the NLMT in their demand forecasting models, to determine the financial benefit (through higher fares and/or increased demand) that might be derived from introducing the improvements. The method for applying the SP results, as adopted in the London Underground demand forecasting models, is fairly simplistic: the introduction of new passenger facilities is taken to have the same impact on demand as the fare value implied by their monetary value from the SP studies. Thus, an improvement with a value of 5% of fare is taken to increase demand by the same degree as a 5% fare decrease.

This heightens the importance of the context in which the SP scenarios are presented to travellers. 'Within-mode' SP exercises, such as those used in previous London Underground studies, obtain direct trade-offs between fare and qualitative service improvements, but cannot directly infer the implications for the demand for rail services. SP exercises which consider new facilities in the context of wider issues such as mode choice will yield more information on the relationship between qualitative service attributes and the demand for rail services.

The issue of applying established demand elasticities to the monetary values derived from SP studies also relates to the type of people who are interviewed. Previous London Underground studies were only carried out with current Underground users, on the basis of cost: potential users of the Underground are only a small proportion of the travelling public in greater London and their recruitment in a survey is therefore very costly. For the same reason of cost, the present study would only interview current Northern Line users. In terms of relating the trade-offs between fares and qualitative service attributes to travel demand, this has important implications for the use of the SP results. Current users, who have some experience of the qualitative attributes being investigated are likely to have a different set of preferences and values attached to these items than potential or infrequent users of the Underground who currently travel some other way or do not travel at all.

Thus, a demand forecasting model which applies demand elasticities derived for all travellers to SP values obtained only from surveys of current Underground users has the potential to be quite inaccurate. In this respect the current application of the London Underground models has some weaknesses: a sounder approach would be to calculate the effects of qualitative service improvements on current users alone, and assume that their effect on non-users of the Underground is negligible. This may under-estimate the impact of passenger facilities on the total demand for Underground services, but represents a more consistent use of the SP results. Whatever the limitations imposed by LUL's demand forecasting approach, the SP survey had to conform to its requirements and reflect practical constraints. The value of the research reported here has to be evaluated in this context.

8.3.2 Definitions of the Qualitative Service Improvements

The first concern in the design of the SP experiments was to define levels for each of the attributes which would have a clear meaning to respondents and the NLMT, and would be easily distinguishable from one another. They also had to be limited in number to keep the size of the experimental designs manageable.

The group discussions gave useful indications of the way in which facilities could be described and combined to represent levels of each of the attributes. Some base levels needed to be defined as well as the proposed improvements. These could actually be perceived by some travellers as below the present service levels, but would be interpreted by the NLMT as the minimum level of service that could be provided.

It was decided that no attribute should have more than three levels and each exercise presented to respondents should not contain more than five variables. This was due to the belief that this approached the limit in the number of items which an individual could realistically assess at any one time and also because the size of the computer screens made more than this number of items impractical to describe. It was decided that train cleanliness and newness should be treated as separate variables, because it was possible that new carriages could be provided but poorly maintained, or old/refurbushed carriages provided with a good level of maintenance. The group discussions suggested that cleanliness and newness were nevertheless closely associated in the minds of travellers and that interaction effects should be investigated at least for these two items (as in earlier studies, the opportunity to investigate interaction effects remained limited for practical reasons.)

The levels finally developed for the SP experiments are shown in Table 8.2. A detailed description of what each of these levels represented is given in Appendix 2, which contains the show cards and photographs given to respondents.

As a practical limit of five variables was imposed on any one SP exercise, it was necessary to divide the rolling stock attributes into two groups. A third group of attributes, relating to stations and train exteriors, would be presented in a separate SP exercise. The groups of rolling stock attributes were defined in terms of their association with each other as suggested by the group discussions.

Table 8.2 Attribute Levels Defined for the SP Experiments

Cleanliness:	base - "Dirty/Vandalised train"; clean 1 - "Dirty train"; Clean 2 - "Clean train";
Newness:	base - "Old train"; New 1 - "Refurbished train"; New 2 - "Brand new train";
Air Quality:	base - "Poor ventilation"; Air 1 - "Forced air ventillation"; Air 2 - "Air conditioning";
Noise:	base - "Very noisy train"; Noise 1 - "Fairly noisy train"; Noise 2 - "Fairly quiet train";
Ride Quality:	base - "Bumpy train"; Ride 1 - "Fairly smooth train"; Ride 2 - "Very smooth train";
Information:	base - "Minimum information"; Info 1 - "Indicator boards"; Info 2 - "Indicators & Announcements";
Gangways:	base - "No gangways"; Gangway 1 - "Gangways"; Gangway 2 - "Gangways & phone link";
Interactions:	Clean 1 New 1 - "Clean and brand new train"; Noise 2 Ride 2 - "Fairly quiet and very smooth train".

Draughts:	base - "Draughty station"; no draughts "Station not draughty";
Exteriors:	base - "Trains have plain exteriors"; Exteriors - "Trains have coloured exteriors";
Escalators:	base - "Unreliable escalators"; Escalators - "Reliable escalators";
Graffiti:	base - "Trains have graffiti on the outside"; No graffiti - "Trains have no graffiti;

Cleanliness and newness have already been identified as close associates and it seemed appropriate to include with these air quality, as all three attributes related to passengers' perceptions of the carriage environmental. Noise levels would be closely associated with improvements to ride quality, as both would result from better suspension and tracks, and it appeared that some travellers perceived such an association. Information and gangways did not have any obvious connection with the other attributes.

The rolling stock attributes were divided into two groups:

(i) cleanliness, newness, air quality, gangways, fare;
(ii) noise, ride quality, information, gangways, fare.

Gangways were included in both groups because it was believed that least was known about this attribute and it should therefore have the benefit of being evaluated with all other attributes. In the station exercise, the items examined were: reduction of draughts, improved escalator reliability, painted train exteriors and reduced graffiti.

8.3.3 Development of the Experimental Designs

The main limitation on the design of the SP experiments was the need to keep the number of options to a minimum, preferably a single Figure. As in earlier studies, the use of full factorial designs was impractical, even though these would be the only way of ensuring that all interaction effects could be measured. A full factorial design for the rolling stock attributes alone would require 2,187 options (3^7.) Even if blocked full factorial designs were used, it would then have been necessary to obtain an extremely large sample of respondents, because many blocks would be required (over 240 if each block was kept to 9 options), each requiring a statistically reliable number of respondents (some 20-30 per block.) This would require a sample of some 5000-7000 for each segment: this was obviously far in excess of the sample of 600 dictated by the budget available for the project.

The separate station attributes exercise, with only two-level variables (except for fare, which would have three levels) would produce a much smaller factorial design, with 48 options (2^4*3^1), but even this would require six blocks, with 120-180 respondents in each segment. This was close to being achievable in a sample of 600 respondents, though five segments were to be identified in advance and the samples assigned to each block would be at the low end of the target of 20-30 individuals.

With these limitations in mind, an alternative approach was developed. It lacked the ideal qualities of a full factorial design, but as will be argued, it had the potential to overcome some of the limitations identified in earlier studies. It required that each respondent would be presented with two types of experimental design.

The first design would be a 'within-mode' exercise, in which improvements to the various rolling stock attributes would varied independently, together with a fare variable. Fractional designs would be used and the rolling stock improvements divided into groups, each containing a common fare variable (this approach was used in the study for Danish National Railways in Steer, Davies and Gleave Ltd, 1986.) In this respect, the first SP experiment would be similar to earlier approaches, though the presentation was to differ.

The second design would be a 'between-mode' exercise. Respondents would have a choice between their present Underground journey or a journey by their best alternative. As well as fare, journey time and service frequency would be included in the exercise. Improvements to rolling stock attributes, instead of varying as separate variables as in the first exercise, would now be presented as components of an 'overall quality of service' variable. Only the attributes of the Underground journey would vary, with fare, journey time and service frequency becoming worse and passenger facilities improving. Thus respondents would be persuaded to choose their alternative mode through the increased costs and deterioration of their Underground service, but this would be off-set to some degree by better quality passenger facilities.

The two-stage approach was considered to offer the following attractions:

(i) the first exercise would use an approach already developed in earlier studies, using established procedures and enabling the direct comparison of results with previous research;

(ii) the first exercise would obtain measurements of the relative preferences of respondents to each individual improvement;

(iii) the second exercise would place the assessment of rolling stock improvements in the more familiar context of mode choice, with other variables than fare for respondents to consider;

(iv) the second exercise would obtain monetary values for discrete packages of variables, in which interaction and threshold effects would be implicit.

This approach was suggested by two separate strands of SP research reported in the literature:

(i) The station improvement studies undertaken for Network SouthEast and later for Dutch National Railways (Copley & Bates, 1988; Copley, Bouma & de Graaf, 1987; MVA, 1985d) used conjoint measurement SP techniques to scale relative values obtained from applications of the priority evaluator method. This approach recognised the importance of different choice exercises and contexts as factors affecting the valuations obtained. The main weaknesses of this approach are chiefly related to the shortcomings of the priority evaluator and the inconsistency of using the two SP techniques together.

(ii) Studies related to mode choice and general travel demand forecasting (Ben-Akiva & Morikawa, 1990; Bradley & Kroes, 1990a, 1990b) showed that individuals demonstrated different rates of trade-off between journey attributes in different choice contexts. This was most apparent in comparisons between real choices and stated preferences, but also between 'within-mode' and 'between mode' SP exercises and between judgmental (ranking) and discrete choice SP exercises.

From the previous section in this chapter, it will be seen that three groups of attributes were to be assessed in separate SP exercises: two sets of rolling stock attributes and one set of station attributes. Using the two-stage approach described above, this would produce six separate SP exercises: far too many to introduce in one interview. Some strategy was therefore required which kept the number of SP exercises in each interview to a practical minimum, while allowing each group of attributes to be assessed by a representative sample of respondents.

The problem could be simplified by identifying the station/train exterior attributes exercise as less important than the rolling stock exercises, because it only contained one attribute originally of interest to the NIMT (draught reduction).

This led to the following strategy:

(i) Respondents would be presented with two 'within-mode' ranking exercises, randomly selected so that they with an equal probability of being presented with either:

(a) two rolling stock exercises or

(b) one rolling stock exercise and the station/train exterior exercise.

In the latter case, each of the two rolling stock exercises had an equal probability of being presented.

(ii) Respondents would be presented with one 'between-mode' fixed choice exercise, containing the rolling stock attributes that were presented in the preceding exercises. The rolling stock attributes from the first exercise (in either iia or iib above) would be used.

This approach meant that the station/train exterior attributes were never presented in the second stage and appeared less frequently than the rolling stock attributes in the first stage. The reason for this decision lay in the concern that the fixed-choice exercises, which required as many respondents in each segment as possible, should concentrate on the attributes of most interest. Any scaling effect on the monetary values for rolling stock attributes derived from the first SP exercises would be assumed the same for station/train exterior attributes. Subsequent analysis and interpretation of the results suggested that this assumption was simplistic, as it could be argued that station/exterior attributes would be subject to greater scaling effects, simply because they affect the passenger for a much shorter part of the total journey than do rolling stock attributes and are therefore likely to be more peripheral to the mode choice process. Nevertheless, it allowed the data collection resources to be concentrated principally upon the rolling stock attributes, which were the main concern of the study.

The random selection between the three types of 'within-mode' exercise meant that from a sample of 600 respondents, some 450 would be presented with each of the rolling stock exercise and some 300 with the station/train exterior improvements exercise.

In the 'within-mode' exercises, two fractional factorial designs were used. These were developed using the SPEED experiment editor (Hague Consulting Group, 1989.) The two 'within-mode' rolling stock exercises used the same design, the only variation being the verbal descriptions of the variables. The design for the rolling stock exercises was divided into blocks, to keep the number of options within a practical limit. This first design was chosen to allow at least one interaction effect to be investigated and this is illustrated in Table 8.3.

Table 8.3 Experimental Design For the Rolling Stock Exercise, Divided into Four Blocks

BLOCK 1

FARE	CLEANESS	NEWNESS	AIR	GANGWAYS
2	3	3	1	3
2	2	1	2	1
3	3	3	2	2
3	2	1	3	3
1	3	3	3	1
1	1	2	2	3
1	2	1	1	2

BLOCK 2

FARE	CLEANESS	NEWNESS	AIR	GANGWAYS
2	1	2	3	2
2	1	1	1	3
3	1	2	1	1
3	2	2	2	2
3	2	3	1	1
1	3	1	2	3
1	1	1	3	1

BLOCK 3

FARE	CLEANESS	NEWNESS	AIR	GANGWAYS
2	3	2	2	1
2	2	2	1	3
3	3	2	3	3
3	3	1	1	1
3	1	1	2	2
1	1	3	1	2
1	2	3	2	3

BLOCK 4

FARE	CLEANESS	NEWNESS	AIR	GANGWAYS
2	1	3	2	1
3	1	3	3	3
2	2	3	3	2
1	2	2	3	1
2	3	1	3	2
1	3	2	1	2

In the second SP exercise, cleanliness was replaced by noise, newness by quality of ride and quality of air by information.

A simpler design could be used for the station/train exteriors exercise, because each item had only two levels, with the exception of fare. This second design is shown in Table 8.4. Because of the secondary importance of this exercise, a larger design which would have enabled interactions to be investigated was not used (a smaller design shortened the length of the interview.)

In order that the design might be divided into blocks, an important assumption had to be made concerning the degree of homogeneity that was likely to exist in respondents' choice behaviour. Ideally, a full or large fractional design should be divided into smaller fractional designs, so that the data can be analysed at the level of individuals. However, when a design cannot be broken into smaller fractional designs (ie the design used here was the smallest possible for five variables and at least one interaction effect), further divisions produce sub-sections of the design which cannot be analysed on their own. This is because it is very likely that they will not produce data with sufficient variation or freedom from collinearity between variables to support analysis at the level of the individual.

With this approach, the researcher was committed to the analysis of responses from groups of respondents only. Given that the preferred method of analysis was the logit model, for reasons discussed in chapter six, this was not considered to represent too great a deficiency, especially when weighed against the advantages gained through the simplification of the SP exercises when presented to respondents. The logit model requires more data than can be practically provided by a single individual, although inter-personal taste variation must be kept as low as possible (Fowkes and Wardman, 1988.) To ensure this, some segmentation of the sample is desirable. The segments were determined by London Underground to make the findings compatible with their forecasting models. This in turn determined the sampling quotas. Nevertheless, it was considered that these segments represented the most distinctive differences between Underground users, determined by the time of day on which they travelled and the ticket used.

Table 8.4 Experimental Design For the Station Exercise

FARE	DRAUGHTS	EXTERIORS	ESCALATORS	GRAFFITI
2	2	1	2	1
2	1	2	1	2
3	2	1	1	2
3	1	2	2	1
1	2	2	2	2
3	2	2	1	1
3	1	1	2	2
1	1	1	1	1

The attribute definitions and fractional design for the 'between-mode' SP exercises are shown in Table 8.5 and Table 8.6 respectively. Only four variables were included, but two - fare and time - were of four levels. This was to allow the analysis of any non-linearities that might exist in the relationship between the qualitative service improvements and these continuous variables. A non-linear relationship could have been detected with only three levels for fare and time, but it was considered that the extra level would improve the accuracy with which threshold effects are detected.

The fare levels used in the SP exercises, together with the other continuous variables in the 'between-mode' exercise, were chosen to encompass the likely ranges of valuations that would occur, while stimulating a sufficient variation in responses. This is an approximate procedure, but can be guided by reference to previous research, such valuations for qualitative service improvements (ranking exercise) and values of travel time (fixed design exercise.) The upper fare level of 30% in the first SP exercises was considered sufficient to cover most likely valuations, given the findings from earlier research. In the second exercises, a higher level of 40% was used to ensure that a enough respondents would switch modes to produce adequate data for the models.

Table 8.5 Attribute Definitions For the 'Between Mode' SP Exercise

Rolling Stock Quality

1	All base levels from 'within-mode SP
2	All intermediate levels from 'within-mode' SP
3	All maximum levels from 'within-mode' SP

Fare

1	Current
2	+5%
3	+15%
4	+40%

Journey time

1	Current
2	+5%
3	+15%
4	+40%

Service headway

1	Current
2	+5%
3	+15%
4	+40%

Table 8.6 Experimental Design For the 'Between Mode' SP Exercise

BLOCK 1

UFREQ	STOCK	UCOST	UTIME
2	1	1	3
1	1	2	4
3	1	1	1
3	2	1	4
1	2	4	1
3	3	4	3
2	2	3	2
1	1	3	3

BLOCK 2

UFREQ	STOCK	UCOST	UTIME
2	3	2	1
3	2	2	3
3	1	2	2
1	3	1	2
2	1	4	4
3	1	4	2
3	1	3	1
3	3	3	4

8.3.4 Presentation of the Stated Preference Exercises

The principal of adaptive ranking used in the 'Alastair' interviewing software has been described in chapter five. Before commencing each of the 'within-mode' SP exercises, the respondent would be asked to rank the improvements and fare variables in order of the importance they attached to each. This information would be used by the program to identify the pairs of options to be shown first. If fare was ranked highest, for example, pairs showing different fare levels in the two options would not be shown immediately, as it could be assumed that the cheaper option would dominate. Only in later pairs, when trade-offs between the other attributes had been explored, would options with different prices be selected.

The 'Alastair' package does not record the initial rankings given by respondents, but a measure of the consistency of responses to the SP exercises was implied by the number of pairs which had to be displayed for the program to achieve a full set of rankings.

Figure 8.1 shows the screen layouts used to present the three types of 'within-mode' exercises. Note that in addition to the scales indicating preference for either option, or indifference between them, a 'choose neither' response was introduced. This is because some options could conceivably present a set of passenger facilities perceived to be inferior to the present situation, but more importantly, fare levels could often be higher than the present situation.

In the 'between-mode' exercise, respondents always faced a choice between journey by Underground or by their best alternative. Figure 8.2 demonstrates the layout for this exercise. The values upon which both journey descriptions are based were obtained from information gained earlier on in the questionnaire.

Figure 8.1 Examples of Screen Layouts From the 'Within-Mode' SP Exercises

First 'Within Mode' SP Game:

WHICH OF THESE ALTERNATIVE NORTHERN LINE TRAINS WOULD YOU CHOOSE ?

OPTION 1	or	OPTION 2
This train is....		This train is....
▮ BRAND NEW		▮ BRAND NEW
▮ CLEAN		DIRTY & VANDALISED
It has....		It has....
POOR VENTILATION		▒ "FORCED AIR" VENTILATION
▮ GANGWAYS BETWEEN CARRIAGES & PHONE LINK TO DRIVER		NO GANGWAYS BETWEEN CARRIAGES
* The FARE is: £ 550.00		* The FARE is: £ 550.00

Definitely option 1	Possibly option 1	Cannot decide	Possibly option 2	Definitely option 2	Neither of these

Second 'Within Mode' SP Game:

WHICH OF THESE ALTERNATIVE NORTHERN LINE TRAINS WOULD YOU CHOOSE ?

OPTION 1	or	OPTION 2
This train is....		This train is....
▒ FAIRLY SMOOTH		▮ VERY SMOOTH
▮ VERY QUIET		▒ FAIRLY QUIET
It has....		It has....
▒ INDICATOR BOARD		▮ INDICATOR & ANNOUNCEMENTS
NO GANGWAYS BETWEEN CARRIAGES		▒ GANGWAYS BETWEEN THE CARRIAGES
* The FARE is: £ 2.60		* The FARE is: £ 2.60

Definitely option 1	Possibly option 1	Cannot decide	Possibly option 2	Definitely option 2	Neither of these

Station SP Game:

WHICH OF THESE ALTERNATIVE SITUATIONS WOULD YOU CHOOSE ?

OPTION 1	or	OPTION 2
The station is....		This station is ....
▮ NOT DRAUGHTY		▒ DRAUGHTY
The trains have....		The trains have....
▮ PAINTED EXTERIORS ░ GRAFFITI ON THE OUTSIDE		▒ PLAIN EXTERIORS ▮ NO GRAFFITI
The escalators are....		The escalators are....
░ UNRELIABLE		▮ RELIABLE
* The FARE is: £ 3.38		* The FARE is: £ 3.38

Definitely option 1	Possibly option 1	Cannot decide	Possibly option 2	Definitely option 2	Neither of these

Figure 8.2 Example of Screen Layout From the 'Between-Mode' SP Exercises

Final 'Between Mode' SP Game:

WHICH OF THESE ALTERNATIVES WOULD YOU CHOOSE ?

The journey is from: Woodside Park
to: Richmond

OPTION 1	or	OPTION 2
THE JOURNEY BY UNDERGROUND The trains are.... Noisy, bumpy with minimum information and NO gangways between carriages The fare is: £ 2.40 Tube journey time: 52 mins Other journey time: 13 mins The trains run every 7 mins		THE ALTERNATIVE By car (Driver) Journey time is: 60 mins The cost is: £ 1.90

Definitely option 1	Possibly option 1	Cannot decide	Possibly option 2	Definitely option 2	Neither of these

8.3.5 Pilot Survey

In addition to the SP designs, conventional survey questions were also developed. These were to provide information to be used directly in the SP exercises (eg fare) and to assist in the analysis of the SP responses (segmenting variables and related attitudinal questions.) The final structure of the questionnaire was composed of six sections, as follows:

(i) questions obtaining details of the respondent's most recent journey on the Northern Line, which would be used in the SP exercises (particularly the final 'between-mode' exercise);

(ii) questions measuring the attitudes and perceptions of the respondent towards different rolling stock attributes on the Northern Line;

(iii) a 'within-mode' adaptive ranking SP exercise (one of two), in which the respondent was asked to trade between different rolling stock attribute levels and fare;

(iv) a 'within-mode' adaptive ranking SP exercise in which the respondent was asked to trade either between more rolling stock attributes and fare or between external train and selected station attributes and fare;

(v) a 'between-mode' fixed design SP exercise in which the respondent was asked to trade qualitative attributes against wider-ranging mode choice factors;

(vi) questions obtaining details of respondent's socio-economic characteristics.

The questionnaire was tested in a pilot survey before the main fieldwork was carried out. A listing of the final questionnaire is given in Appendix 2. The main aims of the pilot survey were:

(i) to establish the likely interviewing rates that would be achieved;

(ii) to train the survey work force.

(iii) to test the SP exercises, specifically in terms of:

(a) the clarity of presentation;

(b) the practicality of using three exercises;

(c) the efficiency of the designs;

(iv) to test the wording of the questions.

It was not considered practical to administer the survey aboard trains or at stations because neither offered suitable environments. Home interviews would be very expensive to administer, because the sampling method would be inefficient (requiring either pre-recruitment of travellers or a wide number of household contacts, of which only a small proportion would be in scope.) The preferred method was therefore to carry out the interviews in 'hall test' situations. The computers would be set up in hired rooms close to Underground stations. The interviewers would recruit travellers from around the stations and bring them back to the rooms to be interviewed. As an incentive for respondents to participate, they were invited to take part in a free prize draw which offered a cash prize.

This method had the following advantages over other methods:

(i) it was cost effective;

(ii) it allowed close supervision of the interviewers;

(iii) it provided a comfortable environment for respondents;

(iv) it provided the opportunity to use larger visual display materials than would normally be practical in mobile interviewing conditions.

Two disadvantages were identified with this approach:

(i) it was likely to discriminate against travellers who could not spare the time to be interviewed;

(ii) the approach had not been used in previous studies by Steer Davies Gleave and the interviewers would have had limited experience of working in hall tests.

The first disadvantage would apply to many commuters, so it was considered essential that suitable quotas should be set in the main fieldwork to ensure the adequate representation of such travellers. Quotas were not set in the pilot survey to allow any inherent bias of the sampling method towards certain groups of travellers to be identified. The second disadvantage was a minor one, given that 'hall tests' are an established market research technique (see chapter five.) Nevertheless, this emphasised the importance of a thorough pilot survey.

The pilot survey was carried out at two locations: Tooting Bec (9 May 1990) and Goodge Street (10 May 1990.) The survey period was from 8.00 am to 7.00 pm each day, split into two six hour shifts that overlapped for an hour between 1.00 pm and 2.00 pm. A team of four interviewers worked each shift, supervised by the author. The intention was to obtain 40-50 interviews at each site. In total, 89 complete interviews were carried out, 44 at Goodge Street and 45 at Tooting Bec. Two more interviews had to be terminated due to respondents' time constraints.

Concerning the interviewing rates, recruitment in the peak period (before 9.00 am) was found to be very difficult, with interview rates of about one per hour per interviewer for the first two hours. This improved to about two per hour per interviewer in the off-peak, when it was found that a number of morning peak period travellers could be recruited at lunch time (assisted by the overlapping interviewer shifts.) Despite this last point, peak travellers were still found to be under-represented: 38% of the sample were period Travelcard users (of whom the majority are peak travellers), compared with 57% of travellers interviewed in LUL's large scale passenger monitoring surveys on the Northern Line. As the sample proportion of 38% indicated the proportion of Travelcard users among the population to be between the range of about 28% to 48% at the 95% level of confidence, the more reliable passenger monitor Figure can be seen to be well outside this range. This confirmed the need for suitable quotas in the main fieldwork. The aim would not be to obtain proportions similar to the passenger monitor, but to ensure adequate numbers of respondents from both the peak and off-peak periods and their associated ticket types.

The majority of respondents found the verbal and visual descriptions of the qualitative service improvements easy to understand. As a result it was not considered necessary to modify the presentation materials. In the SP exercises, most respondents again appeared to understand the definitions of the attribute levels and the idea of trading between alternative packages of improvements. The interviews took an average of 20 minutes to complete, which did not appear to impose too much upon respondents' time.

The pilot survey provided enough data to show how respondents assessed the SP options and to allow logit models of that response behaviour to be developed. This would confirm whether plausible responses and models could be estimated from the SP designs used. It was not expected that the models would be of particularly good quality, because the small sample size (89 respondents) prohibited segmentation. For example, 29 respondents used ordinary (single/return) tickets, 26 used One Day travelcards, 31 used period travelcards and three used a pass. None of these key segments represented a large enough sub-sample for which separate models could be estimated. Despite this limitation, it was believed that models developed for the whole pilot sample could still be assessed for internal consistency and robustness.

A summary of the pilot analysis is given in Appendix 3. The estimates from the models suggested that plausible results could be obtained from the 'within-mode' SP exercises. The limited sample size meant that the models were unreliable for the 'between-mode' exercises, except at the most aggregate level of analysis. At this level, a scaling effect on the monetary value of the qualitative service improvements could be observed, which provided support for the two-stage approach.

A number of respondents showed considerable inconsistency in their responses when taking part in the 'within-mode' exercises. In the pilot, the 27 options in each design were divided into three blocks of nine options each. A respondent who was entirely consistent with his or her initial rankings of the attributes would be expected to complete each exercise over about eight or nine pairs. A number respondents exceeded this figure, one responding to as many as 27 pairs of options before the SP exercise was completed. This variation in the number of responses per individual suggested the importance of weighting the data in the main survey analysis, so that each individual would have the same impact on the estimation of the models coefficients.

When the respondents were observed making their responses, a few suggested that the task was difficult because the variations in the options were often minimal and sometimes hard to detect, even though the descriptions of each attribute level appeared clear enough. This problem could not be easily rectified for the main survey, as it appeared to be a feature of the research objective itself: small variations in the quality of passenger facilities were the subject of the study. Despite this, it was considered that smaller blocks of options would at least lighten the burden of the task for respondents in the main survey and as a result, the 27 options were divided into four blocks, three with seven options and one with six options. This would have the effect of reducing the total amount of data obtained from the survey, but not to a major extent.

With a similar aim of simplifying the task for respondents in the station exercise, the eight options were reduced to seven by removing the one with all attributes at their most attractive levels. This was done on the basis that respondents would choose this dominant option and the information was therefore of little value, when compared to the advantages in reducing the length of the exercise. In the event, this action had serious consequences for the quality of the data relating to the station attributes, which are discussed in the next chapter.
The 'between-mode' choice exercises appeared to be well understood by respondents, but a high proportion did not switch to their alternative mode at any point in the exercise. The sensitivity of the models that can be developed from mode choice data may be expected to improve with the amount of mode switching that can be observed. It was therefore considered that the amount by which the fare and journey time variables increased should be larger for the main survey. Both were therefore increased to levels of +15% and +40% over the present values to encourage more mode switching. The only drawback with these values was that they would now be slightly different from the fare levels used in the 'within-mode' exercise (+10% and +30%), which did not appear in need of being increased. It was considered that these differences would not be large enough to greatly distort comparisions between the monetary valuations derived from each exercise.

8.4 Conclusions

This chapter has described the steps taken in the design of an SP survey instrument for measuring the monetary values attached to Underground service quality improvements. It has described in detail the process by which a qualitative understanding of passenger preferences could be obtained in advance of designing the SP questionnaire. Particularly, it has sought to emphasise the important contribution such a process played in defining the attribute levels to be presented in the SP experiments and the advantages it might offer in aiding the interpretation of models fitted to the SP data.

The theoretical concerns relating to previous applications of SP techniques were discussed and the case made for using a combination of 'within-mode' ranking SP exercises and 'between-mode' choice exercises. The intention was that the two approaches should complement one another, the first enabling trade-offs between detailed service improvements to be observed, the second providing a more realistic context for estimating monetary values of service quality improvements.

The practical considerations relating to the design of the SP exercises and the administration of the survey have also been reported, discussing how certain compromises in the design of the experiments (separate exercises; fractional designs; blocks) were necessary to cope with the number of attributes and levels to be incorporated. The SP exercises (and the whole questionniare) were tested in a pilot survey but not, it is worth noting, with simulated data, this procedure not being well-established at this time. Weaknesses in the final SP designs that might have been identified with simulations are discussed at the end of chapter ten, in the light of what was learned from the analysis of the SP data.

9: THE EVALUATION OF IMPROVEMENTS TO PASSENGER FACILITIES ON LONDON'S UNDERGROUND: ANALYSIS OF GENERAL SAMPLE CHARACTERISTICS AND 'WITHIN MODE' SP DATA

9.1 Introduction

This and the next chapter report on the analysis of data obtained from the Northern Line survey, the design of which was reported in the previous chapter. The analytical procedures that were used are fully described, together with a number of strategies adopted with the aim of improving the statistical models that were fitted to the data. The aim of the analysis was:

(i) to derive quantitative measures of the survey respondents' preferences towards different service quality improvements and to express them in units of percentage fare equivalents;

(ii) to observe the stability of these results under different treatments of the data and

(iii) to assess the comparative performance of the two types of SP exercise, particularly in their ability to yield plausible monetary values for the service improvements.

In a concluding section to the next chapter, a comparison is made between the findings of the study and those from previous research, both for the Underground and for rail services generally. Discussion is made concerning the strengths and weaknesses of the survey design and analytical procedures used, with recommendations for the modification and development of this approach in future studies.

9.2 Main Fieldwork

In addition to the two sites used in the pilot survey, a further two were selected for the main survey. The dates of the survey and the four locations were:

(i) Goodge Street (29 May - 1 June, 1990);
(ii) Golders Green (29 May - 1 June, 1990);
(iii) Tooting Bec (4 - 6 June, 1990);
(iv) London Bridge (4 - 6 June, 1990).

These locations were chosen to ensure a reasonable geographical spread of interviews across the Northern line and because suitable hired rooms were available very close to these stations. Two teams of interviewers were used with the same day shift arrangements as in the pilot survey. Surveying did not take place at weekends, as the NLMT's forecasting models only required information on weekday travel. The objective was to obtain 150 interviews at each site, spread across the five main segments defined by the time of the respondent's trip and the ticket he or she used. The final number of usable interviews obtained was 589, the composition of which is shown in Table 9.1.

The total number was slightly below the target of 600 interviews because of the difficulty experienced in recruiting period Travelcard users who had made an off-peak journey. In fact, a total of 632 complete interviews were achieved, but this included 43 individuals who travelled using special passes, such as those issued to London Underground or British Rail employees. The interviewers had not been instructed to screen such individuals out and as such a degree of the survey resources had been wasted on collecting irrelevant information.

The minimum targets were set to ensure a good geographical spread in each segment, as well as adequate numbers. The different characteristics of the people using each station was bound to affect the ease with which some of the targets could be met, but where the target failed to be reached, the deficiency was only in terms of a few individuals.

Table 9.1 Distribution of Sample Across Pre-defined Segments

Location	Peak Period T'card	Peak Ordinary	Off-Peak Period T'card	One Day T'card	Off-Peak Ordinary	Total
Goodge St	46	21	21	35	24	147
Minimum	25	25	25	25	25	125
Tooting Bec	34	30	24	34	21	143
Minimum	25	25	25	25	25	125
Golders Grn	30	35	24	41	25	155
Minimum	25	25	25	25	25	125
London Bdge	38	28	20	29	29	144
Minimum	25	25	25	25	25	125
TOTAL	148	114	89	139	99	589
Minimum	100	100	100	100	100	500

9.3 Analysis of the Non-SP Information

9.3.1 Travel Characteristics

Table 9.2 summarises the main journey purposes reported by respondents and the frequency with which they used the Underground over the preceding four weeks. As might be expected, work and business trips are the majority in the peak, non-work trips are more prevalent in the off-peak. The high proportions of journeys to work and 'other' for the off-peak Period Travelcard group might suggest some people returning from work and points to a weakness in the design of this question.

Though the categories of journey purpose are compatible with those used in previous Underground surveys, additional categories would have been useful: namely, 'regular journey from work place'. Respondents for whom this category would have been appropriate should have chosen 'other', but the high proportions in the ' regular journey to work' category may suggest that some respondents on journeys from work chose this category. Those who specified 'other' were not asked to give another purpose, so the proportion of journeys from work in this category is not known. The substantial proportions stating 'other' for non-period ticket user groups may be composed primarily of tourism and education trips.

From the responses relating to frequency of Underground use, most respondents had used an Underground service in the last week, though sizeable proportions of one-day Travelcard and ordinary ticket users had travelled less frequently. Regarding the particular origin station and destination station visited on the recent journey to which the interview related, frequency of Underground use declines noticeably for all categories except peak Period Travelcard users. The implication of this observation for respondents' likely valuations of service improvements is that peak Period Travelcard users are more likely to consider them in the context of regular use compared to other categories of traveller.

Table 9.2 Journey Purpose and Experience of Using the Underground

Main Journey Purpose

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Regular journey to work	75%	42%	29%	10%	9%
Firm's business	6%	17%	11%	12%	16%
Visiting friends/relatives	8%	14%	12%	16%	19%
Shopping	0%	6%	12%	16%	14%
Leisure	3%	7%	12%	20%	24%
Other	8%	15%	25%	26%	18%
Total	100%	100%	100%	100%	100%

How often Travelled on the Underground in the Last Four Weeks

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Every day	47%	27%	52%	17%	12%
5-6 days a week	36%	18%	24%	12%	8%
2-4 days a week	11%	34%	13%	33%	46%
Once a week	2%	10%	7%	16%	17%
Less than once a week	4%	12%	4%	20%	17%
This was the first time	0%	0%	0%	2%	1%
Total	100%	100%	100%	100%	100%

How often Travelled Between Origin Station and Destination Station in the Last Four Weeks

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Every day	24%	12%	11%	7%	2%
5-6 days a week	42%	19%	13%	5%	4%
2-4 days a week	13%	23%	21%	17%	21%
Once a week	3%	14%	11%	12%	18%
Less than once a week	10%	15%	18%	26%	19%
This was the first time	8%	18%	26%	34%	36%
Total	100%	100%	100%	100%	100%

Table 9.3 gives details of the tickets purchased and the time constraints acting on respondents when they made their journeys. The average costs for respondents' tickets include Underground tickets that are also valid for British Rail services. The higher values associated with such tickets inflate the averages and lead to some relatively large outliers. Despite this, the coefficients of variation are not large.

The values for Period Travelcard users are estimated unit costs for a single journey, calculated using respondents' reported use of the Underground and the duration of the ticket. The costs for one-day Travelcards have been divided by three, to represent the approximate number of journeys made with these tickets. All these values are therefore approximate. Among ordinary tickets, returns are also represented as single costs (ie halved.) It should also be remembered that all values are as reported by respondents and are therefore likely to be subject to further inaccuracy.

From a comparison of the average values, it can be seen that Period Travelcard Users, though having to consider the cost of their tickets aggregated over many more days than for ordinary ticket users, enjoy a lower unit cost per journey. A comparison with unit values estimated from Underground profile data shows one-day Travelcard and ordinary ticket users to have similar reported values. In contrast, Period Travelcard users report much higher values. The reason for this is likely to be found in the question used to record frequency of Underground travel. The most frequent category which respondents could choose was 'use everyday'. Those travellers making more than one return trip per day would therefore have a lower frequency recorded than their real travel behaviour. This weakness in the questionnaire is likely to account for much of the discrepancies.

Regarding who paid for respondents' tickets, most bore the cost themselves or had it paid by a household member. Thus the majority would be trading off between fare and service improvements on the basis that any implied costs would have to be paid for by themselves.

Table 9.3 Payment for Ticket and Journey Constraints

Average Cost of Tickets

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Average cost	£0.82	£1.14	£0.67	£0.87	£0.93
Coefficient of variation	59%	77%	52%	43%	42%

Who Paid for Ticket?

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Respondent	92%	93%	86%	95%	90%
Household member	2%	2%	5%	2%	2%
Employer	5%	4%	5%	3%	8%
Other	1%	1%	5%	1%	0%
Total	100%	100%	100%	100%	100%

Average Group Numbers

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Average group size	1.15	1.29	1.38	1.68	1.67
Coefficient of variation	18%	30%	103%	224%	113%

Was Respondent Travelling for a Certain time? (Commuters Excluded)

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Yes	71%	53%	30%	44%	33%
No	29%	47%	70%	56%	67%
Total	100%	100%	100%	100%	100%

How Late Could He or She be? (All Respondents)

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Vital to arrive on time	41%	37%	21%	24%	14%
Up to 10 mins late	22%	19%	23%	15%	12%
Up to 20 mins late	37%	45%	56%	61%	73%
Total	100%	100%	100%	100%	100%

Time constraints on respondents did not appear particularly severe. Although a number of Non-commuters stated that they were aiming for a certain time (it was assumed most commuters were aiming for a certain time, so these were not asked this question), less than half the peak travellers and less than a quarter of the off-peak travellers stated that it was vital that they arrive on time. Over a third of the peak travellers and over one half of the off-peak travellers said they could be up to 20 minutes late. These observations suggest that respondents might not be very sensitive to limited increases in travel time.

Tables 9.4 and 9.5 summarise key stages of respondents' journeys, beginning with the access modes they used and the lengths of their access journeys. Off-peak Travelcard users stand out with the highest proportion of pedestrian access and correspondingly low use of car and BR train. Average access time is noticeably less for this group and as a proportion of total travel time is also the lowest. Off-peak Travelcard users have much higher average access times (over twice the length of time) and a greater dispersion of reported values, despite having a similar profile of access modes to other groups.

Once at the station, all respondents have a similar average estimated waiting time. Most reported either a very short time indeed (users of termini who walked onto a waiting train reported waiting time as zero, even though it is possible that the train did not leave for a few minutes longer) or tended to round to five or ten minutes. From the estimates of service frequencies, most respondents perceived a headway of ten minutes or less. Peak travellers rightly observe higher frequencies.

A large proportion of travellers in each segment changed Underground trains at least once during their journey and in most of these cases an Underground line other than the Northern Line was used. This suggests that a number of people could make an immediate comparison between their experience of the Northern Line and that of other Lines.

Table 9.4 Journey Characteristics (1)

Access mode to Underground

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Car	5%	5%	0%	4%	5%
Bus	13%	9%	13%	13%	9%
BR Train	13%	11%	6%	12%	12%
Walked all the way	69%	75%	79%	71%	71%
Other	0%	0%	2%	0%	2%
Total	100%	100%	100%	100%	100%

Average waiting times

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Average wait time	4.21	4.62	4.92	4.83	4.57
Standard deviation	3.27	4.60	4.33	3.82	3.74
Coefficient of variation	78%	100%	88%	79%	82%

Service Frequency

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
One every 5 minutes	64%	59%	52%	53%	50%
One every 10 minutes	23%	20%	29%	24%	28%
One every 15 minutes	4%	2%	2%	5%	9%
One every 20 minutes	1%	0%	4%	2%	0%
Infrequently	2%	3%	1%	3%	2%
Don't know	5%	16%	12%	13%	11%
Total	100%	100%	100%	100%	100%

Table 9.5 Journey Characteristics (2)

Interchange

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Yes	34%	45%	48%	57%	38%
No	66%	55%	52%	43%	62%
Total	100%	100%	100%	100%	100%

Used Other Underground Lines?

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Yes	28%	39%	40%	50%	31%
No	72%	61%	60%	50%	69%
Total	100%	100%	100%	100%	100%

Train delayed?

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Yes	37%	33%	31%	27%	19%
No	63%	67%	69%	73%	81%
Total	100%	100%	100%	100%	100%

Length of Time Train Delayed

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Average Delay	2.10	2.41	1.71	1.27	1.12
Coefficient of variation	197%	229%	210%	250%	264%

Egress Mode

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Car	2%	2%	0%	0%	3%
Bus	4%	3%	5%	3%	6%
BR Train	4%	1%	5%	3%	1%
Walked all the way	90%	94%	90%	92%	89%
Other	0%	0%	0%	2%	1%
Total	100%	100%	100%	100%	100%

A factor likely to have some influence on travellers' valuations of such facilities as air conditioning was delays in tunnels. The group discussions highlighted this perception among travellers. As Table 9.5 shows, up to a third or more travellers perceived some delay. The average delay for all travellers is fairly small (one to two minutes), but the large coefficients of variation identify some large delays for some individuals.

Finally, it can be seen that most respondents walked from their destination station. As the proportion is noticeably lower than for access modes, there is some suggestion of bias towards a particular leg of travellers' journeys, probably the outward one.

In Table 9.6, three items relating to the experience of travelling on the Underground train are compared across the key segments. Concerning position on the train, Period Travelcard users were more likely to travel in an end carriage than other ticket users. This may perhaps reflect a slightly higher proportion of these travellers using a termini at the beginning of their Northern Line journey or their greater propensity to seek out end carriages, if their regular travel experience has shown these to be less crowded.

As might be expected, perceptions of crowding are higher during the peak and lower in the off-peak, though Period Travelcard users in both time periods are more likely to identify at least some crowding. A cross-tabulation of position on train against perceptions of crowding suggests that travellers in end carriages might be less likely to perceive crowding than elsewhere, though this distinction, through the use of the chi-squared test, has not been found to be significantly different from a random distribution. A chi-squared value of only 1.24 was obtained: with two degrees of freedom (from a two by three matrix of end carriage/other against crowded/partly crowded/not crowded), a figure of 5.99 or more would be required for the relationship to be significant at the 95% level of confidence. As with crowding, seat availability is correspondingly poor in the peak and good in the off-peak.

Table 9.6 Summary of Experiences on Train

Position on Train

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
End carriage	19%	12%	17%	13%	12%
Towards front or back	33%	30%	43%	29%	41%
Towards middle	47%	56%	36%	53%	43%
Don't know	1%	2%	5%	5%	3%
Total	100%	100%	100%	100%	100%

Did Train Become Crowded?

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Yes: throughout journey	47%	39%	15%	21%	17%
Yes: some of the time	26%	25%	31%	22%	19%
No	27%	37%	54%	57%	64%
Total	100%	100%	100%	100%	100%

Position v Crowding

	Crowded Throughout	Crowded Some of the Time	Not Crowded	Total
End Carriage	25%	29%	46%	100%
Towards front/back	32%	23%	44%	100%
Towards middle	28%	25%	47%	100%
Don't know	31%	6%	63%	100%
Total	29%	25%	46%	100%

Seat Availability

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Seat for whole journey	38%	53%	73%	72%	73%
Seat for part of journey	23%	18%	17%	14%	13%
Seat available, but prefered to stand	4%	2%	5%	5%	2%
No seat available	34%	27%	6%	9%	11%
Total	100%	100%	100%	100%	100%

9.3.2 Perceptions of Service Quality

The absolute benefit that travellers would derive from rolling stock improvements would depend on the current levels of service quality to which they could be compared. Before being presented with the SP exercises, each respondent was therefore asked to identify which of the three levels for each service attribute being considered most nearly represented their perception of the current level of service. Newness, gangway and information improvements were not presented, as the intermediate and maximum levels of these improvements clearly represented additional facilities to those currently offered.

Table 9.7 summarises perceptions for the other four attributes. Regarding cleanliness, it can be seen that the majority in each segment perceived their train to be dirty, but few identified vandalism. Off-peak Period Travelcard users are the most united in their perception of cleanliness (three quarters identified dirty trains), off-peak ordinary ticket users are the most diverse. It is not obvious why this difference should occur.

Few respondents considered the air in the train to be fresh and the majority in each segment identified it as 'stuffy'. This may reflect the time of year in which the survey was carried out (May/June), when the weather was generally warm for most of the time. Stuffy air conditions were more frequently identified in the peak, perhaps reflecting the higher perceptions of crowding among these travellers.

Concerning levels of noise in the train, few describe conditions as quiet. The distribution of responses is similar across most segments, with the majority settling for the middle level. Only peak Period Travelcard users show a distinctly different set of perceptions, a higher proportion choosing the worst level of noise. Peak Period Travelcard users are also the most critical of the quality of ride, in marked contrast to off-peak Travelcard users, over two thirds of which consider the ride to be 'fairly smooth'. This may reflect the fact that a much higher proportion of the former group had to stand during their journeys (see Table 9.6.)

Table 9.7 Summary of Perceived Service Quality

Cleanliness

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Dirty, vandalised train	11%	12%	8%	13%	16%
Dirty train	63%	60%	76%	65%	58%
Clean train	26%	28%	15%	22%	27%
Total	100%	100%	100%	100%	100%

Quality of the Air

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Fairly Stuffy	65%	70%	58%	63%	59%
Neither Stuffy nor Fresh	30%	23%	36%	34%	36%
Fresh	5%	7%	6%	3%	6%
Total	100%	100%	100%	100%	100%

Noise Level

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Noisy Train	49%	39%	38%	34%	38%
Fairly Noisy Train	44%	50%	56%	58%	53%
Very Quiet Train	7%	11%	6%	8%	9%
Total	100%	100%	100%	100%	100%

Quality of Ride

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Bumpy Train	49%	42%	29%	43%	42%
Fairly Smooth Train	47%	55%	68%	51%	52%
Very Smooth Train	4%	3%	4%	6%	6%
Total	100%	100%	100%	100%	100%

Perceptions relating to station service attributes were not investigated to keep the length of the questionnaire to a minimum. Though current perceptions of these attributes would no doubt affect the absolute value of improvements, the omission of perceptual questions for these items reflected the lower degree of importance placed upon them, relative to the objective of designing an efficient questionnaire.

9.3.3 Socio-Economic Characteristics

Questions were asked concerning the occupational status of respondents, approximate personal income and age. Their gender was also recorded. Table 9.8 summarises these characteristics for each ticket/time segment. In the peak, most respondents worked full-time, reflecting the predominance of commuting and business trips in this period. Even in the off-peak, about half the respondents worked full-time. The only other group of any significant size are students in higher education. Most respondents were willing to give an approximation of their personal annual income. Predictably, incomes are higher for peak travellers, reflecting the higher proportions of full-time workers.

In keeping with the distributions of occupational status in each segment, the average age of peak travellers was higher than off-peak travellers. The proportion of people over 60 years of age was very small, even among off-peak travellers. Finally, the majority of all travellers in each segment were male, though among peak ordinary ticket users a balance was almost achieved.

These socio-economic characteristics allow for some validation of the sample against larger-scale surveys of the Northern Line. The Northern Line Profile Survey was considered the most reliable source of information and comparisons with it are made in Table 9.9. Only aggregate market profiles were made available. To make a comparison between the sample and this profile realistic, the averages across the ticket/time segments were weighted using market share information provided by the NLMT. These proportions are shown at the top of the table.

Table 9.8 Summary of Socio-Economic Characteristics

Main Occupation

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Working full-time	80%	64%	49%	47%	57%
Working part-time	4%	17%	7%	9%	10%
Housewife/husband	1%	3%	0%	4%	4%
Retired	0%	1%	1%	2%	6%
At school	1%	2%	2%	4%	0%
Student (HE)	13%	9%	32%	25%	14%
Other	1%	4%	8%	9%	9%
Total	100%	100%	100%	100%	100%

Personal Income

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Less than £10K	32%	37%	50%	50%	46%
£10K - £20K	45%	39%	30%	29%	23%
£20K - £30K	12%	16%	6%	9%	10%
£30K - £40K	4%	1%	4%	1%	9%
More than £40K	4%	2%	5%	3%	7%
Not disclosed	4%	6%	6%	9%	6%
Total	100%	100%	100%	100%	100%

Age Group

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Less than 16 years	1%	2%	1%	2%	0%
16 - 24 years	36%	33%	46%	51%	37%
25 - 34 years	34%	31%	29%	26%	39%
35 - 44 years	19%	20%	13%	12%	11%
45 - 59 years	10%	11%	8%	8%	10%
60 - 64 years	0%	4%	2%	1%	2%
65 - 74 years	0%	0%	0%	0%	0%
75 years or more	0%	0%	0%	0%	1%
Total	100%	100%	100%	100%	100%

Gender

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Female	33%	48%	43%	44%	36%
Male	67%	52%	57%	56%	64%
Total	100%	100%	100%	100%	100%

Table 9.9 Proportions in Sample Against Proportions Observed by London Underground Ltd

	Peak T'card	Peak Ordinary	Off-Peak T'card	One-Day T'card	Off-Peak Ordinary
Proportions in sample	25.1%	19.4%	15.1%	23.6%	16.8%
Proportions from LUL	39.9%	14.5%	17.2%	17.7%	10.8%

Main Occupation	**Sample Average**	**Weighted Average**	**Northern Line Profile**
Working full-time	61%	65%	71%
Working part-time	9%	8%	7%
Housewife/husband	2%	2%	3%
Retired	2%	1%	4%
At school/Student	20%	19%	12%
Other	6%	5%	3%
Total	100%	100%	100%

Age Group	**Sample Average**	**Weighted Average**	**Northern Line Profile**
Less than 16 years	1%	1%	1%
16 - 24 years	41%	39%	39%
25 - 34 years	32%	33%	30%
35 - 44 years	15%	16%	16%
45 - 59 years	9%	9%	10%
60 years or more	2%	1%	4%
Total	100%	100%	100%

Gender	**Sample Average**	**Weighted Average**	**Northern Line Profile**
Female	40%	38%	39%
Male	60%	62%	61%
Total	100%	100%	100%

It can be seen that most occupation categories in the sample are comparable to the profile survey, with the exception of students. This bias probably had two sources:

(i) students were perhaps more likely to have time to participate in 'hall-test' interviews;

(ii) one of the sites, Goodge Street, was situated fairly near University College, London.

For the other characteristics, age and gender, a very close match is achieved.

9.4 Preparation of the SP Data for Analysis

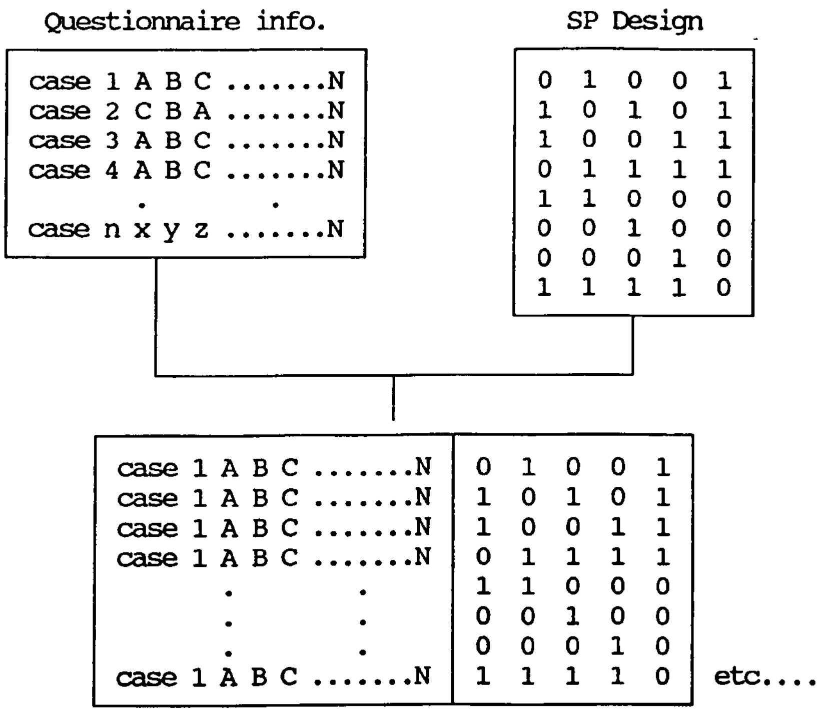
The survey data was stored by the 'Alastair' software on two types of DBASE III database files:

(i) REPLIES.DBF - one file containing the responses of all individuals to the non-SP questions

(ii) GAME*.DBF - a set of files, each containing responses to a block of SP options (in the case of the 'between-mode' exercises) or pairs of SP options (in the case of 'within-mode' exercises.)

Having eliminated incomplete or practice interviews from the REPLIES.DBF, the next step was to merge the multiple responses from the GAME*.DBF files with the single questionnaire records from REPLIES.DBF and the experimental design information stored in the DESIGN*.DBF files. An illustration of the process is given in Figure 9.1. Though the GAME*.DBF and DESIGN*.DBF files contained the main information needed for use in the logit models, questions from the REPLIES.DBF such as time of travel and ticket used would be required for segmenting the data.

Figure 9.1: Diagram of the Merging Process For Data Obtained From the Computerised Interviews



The information from the questionnaire is stored on a single dbase record per respondent. This has to be merged in the above way with the codes for each option in the SP design, so that the model can be estimated across each SP observation, not simply each individual traveller.

The files were merged using unique information on the date and time of the interviews to match the appropriate records. A simple program written by the author in the CLIPPER database language was used to carry out this process. This would later be developed into a more general application for use in subsequent SP studies.

Although the 589 interviews all represented completed interviews, 15 did not contain three completed SP exercises, because it was possible for interviewers to 'escape' from an SP exercise during an interview. This could be detected by the interviewing software and a note of it stored in the REPLIES.DBF file. Terminations of SP exercises sometimes occurred if the respondent was in a hurry. It was decided that only interviews that had all three exercises completed should be used. This is because an incomplete exercise called into question the reliability of the responses to the other complete exercises that might have been obtained in the interview (eg a respondent in a hurry may have given particularly unreliable responses to the completed exercises.)

9.5 Analysis of the First SP Exercises (Adaptive Rankings)

9.5.1 Initial Observations

As with all adaptive SP approaches, the algorithm produced sets of choice situations that would differ across individuals. Though each respondent undertaking each SP exercise would assess options from the same design, the combination of these options would vary according to the pattern of each individual's responses. The more inconsistent the individuals' choice behaviour in relation to his or her initial rankings of the attributes, the more diffuse the pattern of responses. This additional error in the data was not likely to be systematic, but would weaken the models to be estimated.

Table 9.10 and Table 9.11 show the number of pairs assessed by respondents in the 'within-mode' exercises. It can be seen that a number of respondents assessed more than the five to eight pairs that would be associated with responses that were consistent with their initial rankings of the attributes. Clearly, the rejection of respondents assessing more than eight pairs would have led to a substantial reduction in the size of the samples. There is also the argument that some of these inconsistencies reflected the way some respondents are likely to make choices between competing alternatives, so that the rejection of such individuals might be seen as artificially improving the data to conform to the assumptions of rational economic decision making that underlie the statistical analysis. As further paragraphs will show, the rejection of specific illogical responses (as opposed to internally inconsistent responses) improves the quality of the models, while maintaining broadly similar fare values for each improvement. An important result of the different numbers of responses per individual would be the requirement to weight the data in the analysis so that each respondent had the same representation.

Before developing the logit models from the data, it was advisable to gain some information at a simpler level, using 'naive' methods. It was hoped that this would provide some general guidance on the substance of the data before commencing the more complex analysis. Because the design was adaptive, with each respondent responding to different pairs of options, the scope for obtaining preference weightings for each attribute from simple choice frequency procedures was limited.

Some indication of the pattern of responses obtained from the SP exercises could be gained from simply observing the frequency with which a particular option was chosen from all the pairs in which it appeared. One clear observation from such an analysis was that most responses were to be found at the extreme ends of the response scales (eg 'definitely choose option 1'). This would justify the collapsing of the 'definitely' and 'probably' categories into single choice categories to simplify the data. The frequencies from this analysis are summarised in Tables 9.12, 9.13 and 9.14 for the rolling stock and station/train exterior exercises respectively.

Table 9.10 Numbers of Pairs Assessed by Respondents in the Rolling Stock Exercises

Exercise 1 improvements

Pairs	Peak Tcd	Peak Ody	Off Peak Tcd	Off Peak Tcd	Off Peak Ody	Total	Cum. %
5	6%	5%	8%	4%	8%	6%	6%
6	14%	16%	24%	16%	26%	18%	25%
7	19%	19%	15%	23%	16%	19%	44%
8	24%	16%	17%	18%	11%	18%	62%
9	7%	11%	10%	4%	15%	9%	70%
10	14%	15%	8%	11%	8%	12%	82%
11	4%	1%	15%	10%	6%	7%	89%
12	6%	7%	0%	6%	2%	5%	94%
13	6%	4%	0%	6%	3%	4%	98%
14	0%	4%	2%	1%	3%	2%	100%
16	0%	1%	0%	0%	2%	1%	100%
Total	100%	100%	100%	100%	100%	100%	
N	108	74	59	97	62	400	

Exercise 2 improvements

Pairs	Peak Tcd	Peak Ody	Off Peak Tcd	Off Peak Tcd	Off Peak Ody	Total	Cum. %
5	6%	3%	11%	6%	6%	6%	6%
6	25%	29%	18%	26%	23%	25%	31%
7	19%	15%	29%	12%	28%	19%	50%
8	18%	23%	16%	22%	13%	19%	69%
9	11%	13%	7%	13%	8%	11%	80%
10	8%	9%	13%	5%	4%	7%	87%
11	6%	4%	2%	5%	8%	5%	92%
12	2%	0%	0%	4%	6%	2%	94%
13	1%	1%	2%	2%	0%	1%	96%
14	2%	1%	2%	1%	0%	1%	97%
15	2%	0%	0%	2%	2%	1%	98%
16	3%	0%	0%	0%	4%	1%	99%
17	0%	0%	0%	2%	0%	0%	100%
18	0%	1%	0%	0%	0%	0%	100%
Total	100%	100%	100%	100%	100%	100%	
N	120	78	55	100	53	406	

Table 9.11 Numbers of Pairs Assessed by Respondents in the Station/Train Exteriors Exercise

Pairs	Peak Tcd	Peak Cdy	Off Peak Tcd	Off Peak Tcd	Off Peak Cdy	Total	Cum. %
6	33%	39%	43%	25%	23%	32%	32%
7	11%	11%	15%	30%	21%	18%	50%
8	13%	13%	13%	15%	14%	13%	63%
9	9%	7%	13%	10%	0%	8%	71%
10	13%	7%	3%	3%	12%	8%	79%
11	4%	4%	8%	7%	14%	7%	85%
12	9%	6%	0%	3%	12%	6%	91%
13	3%	2%	0%	3%	0%	2%	93%
14	1%	4%	0%	0%	2%	1%	95%
15	3%	2%	5%	0%	0%	2%	97%
16	0%	2%	0%	3%	0%	1%	98%
17	0%	2%	3%	2%	2%	1%	99%
18	0%	2%	0%	0%	0%	0%	100%
19	1%	0%	0%	0%	0%	0%	100%
Total	100%	100%	100%	100%	100%	100%	
N	70	54	40	61	43	268	

Table 9.12 Frequencies With Which SP Options Were Selected in the First "Within-Mode" Rolling Stock Exercise

Fare Increase	Cleanliness	Newness	Air	Gangways	Peak Tcd	Peak Cdy	Off Peak Tcd	One Day Tcd	Off Peak Cdy
0%	Good	Poor	Good	Medium	95%	82%	70%	78%	84%
10%	Good	Poor	Medium	Good	68%	84%	72%	76%	78%
0%	Good	Medium	Medium	Poor	48%	71%	70%	73%	63%
0%	Medium	Good	Good	Medium	72%	53%	73%	60%	57%
10%	Good	Good	Good	Poor	66%	57%	50%	65%	62%
30%	Good	Good	Medium	Medium	48%	68%	54%	60%	54%
0%	Good	Good	Poor	Good	59%	57%	60%	55%	52%
30%	Good	Medium	Good	Good	39%	58%	50%	57%	57%
10%	Medium	Good	Medium	Good	57%	54%	47%	50%	53%
0%	Poor	Medium	Good	Medium	51%	41%	58%	38%	60%
10%	Good	Medium	Poor	Medium	51%	63%	47%	44%	37%
10%	Poor	Medium	Medium	Good	47%	41%	51%	43%	51%
30%	Poor	Good	Good	Medium	42%	42%	40%	44%	37%
10%	Medium	Medium	Good	Poor	44%	36%	34%	29%	43%
10%	Poor	Poor	Good	Poor	42%	42%	32%	33%	27%
0%	Poor	Poor	Poor	Good	37%	34%	33%	27%	37%
0%	Medium	Medium	Poor	Good	31%	34%	31%	42%	28%
30%	Medium	Poor	Good	Good	28%	44%	33%	28%	27%
30%	Good	Poor	Poor	Poor	20%	41%	24%	31%	42%
0%	Poor	Good	Medium	Poor	36%	30%	34%	28%	25%
10%	Poor	Good	Poor	Medium	35%	25%	29%	27%	36%
0%	Medium	Poor	Medium	Poor	25%	32%	42%	22%	25%
30%	Poor	Poor	Medium	Medium	14%	42%	15%	34%	41%
30%	Medium	Medium	Medium	Good	14%	25%	22%	30%	38%
10%	Medium	Poor	Poor	Medium	25%	14%	19%	19%	29%
30%	Medium	Good	Poor	Poor	16%	12%	10%	7%	17%
30%	Poor	Medium	Poor	Poor	2%	0%	0%	8%	0%

Table 9.13 Frequencies With Which SP Options Were Selected in the Second "Within-Mode" Rolling Stock Exercise

Fare Increase	Noise	Ride Quality	Information	Gangways	Peak Tcd	Peak Cdy	Off Peak Tcd	One Day Tcd	Off Peak Cdy
0%	Medium	Good	Good	Medium	70%	71%	77%	76%	73%
0%	Medium	Medium	Poor	Good	63%	68%	60%	67%	68%
10%	Medium	Good	Medium	Good	62%	63%	61%	64%	65%
0%	Good	Poor	Good	Medium	77%	72%	47%	64%	53%
0%	Good	Good	Poor	Good	80%	61%	56%	65%	43%
10%	Good	Poor	Medium	Good	48%	58%	64%	58%	56%
10%	Good	Good	Good	Poor	44%	63%	51%	50%	61%
0%	Good	Medium	Medium	Poor	56%	42%	60%	53%	46%
0%	Poor	Medium	Good	Medium	51%	45%	46%	53%	56%
0%	Medium	Poor	Medium	Poor	55%	54%	50%	46%	45%
10%	Poor	Poor	Good	Poor	43%	57%	38%	46%	60%
10%	Good	Medium	Poor	Medium	51%	52%	39%	44%	52%
30%	Good	Good	Medium	Medium	46%	46%	39%	53%	51%
30%	Poor	Poor	Medium	Medium	46%	43%	44%	46%	44%
10%	Poor	Medium	Medium	Good	42%	43%	36%	41%	53%
30%	Medium	Good	Poor	Poor	34%	40%	39%	41%	57%
10%	Medium	Medium	Good	Poor	42%	40%	44%	35%	34%
10%	Medium	Poor	Poor	Medium	42%	35%	34%	32%	37%
10%	Poor	Good	Poor	Medium	40%	29%	44%	27%	31%
0%	Poor	Good	Medium	Poor	25%	42%	43%	35%	19%
30%	Medium	Poor	Good	Good	28%	33%	20%	34%	42%
30%	Good	Poor	Poor	Poor	29%	35%	35%	27%	28%
30%	Medium	Medium	Medium	Good	24%	22%	30%	29%	37%
0%	Poor	Poor	Poor	Good	33%	28%	18%	25%	34%
30%	Good	Medium	Good	Good	10%	16%	19%	14%	20%
30%	Poor	Medium	Poor	Poor	12%	15%	25%	0%	26%
30%	Poor	Good	Good	Medium	10%	4%	14%	0%	6%

Table 9.14 Frequencies With Which SP Options Were Selected in the Station/Train Exterior Exercise

Fare	Draughts	Exteriors	Escalators	Graffiti	Peak Tcd	Peak Ody	Off-Peak Tcd	One Day Tcd	Off-Peak Ody
Poor	Good	Poor	Poor	Good	53%	54%	45%	44%	58%
Medium	Poor	Good	Poor	Good	42%	47%	31%	47%	49%
Poor	Good	Good	Poor	Poor	41%	40%	39%	39%	43%
Medium	Good	Poor	Good	Poor	25%	36%	28%	32%	33%
Poor	Poor	Good	Good	Poor	38%	24%	40%	18%	30%
Good	Poor	Poor	Poor	Poor	22%	18%	12%	28%	22%
Poor	Poor	Poor	Good	Good	0%	0%	0%	0%	0%

Examination of these tables offers the following observations:

Rolling Stock Exercises

(i) The increases in fare levels do not dominate the responses; some options with the highest fare increases still have sizeable choice frequencies.

(ii) There is considerable variation between the choice frequencies of the different segments. This would suggest some difference in respondents preferences, though random error due to the small sample sizes must also have an effect.

(iii) Inconsistencies in the responses are suggested by the counter-intuitive ordering of some of the options in terms of choice frequency. The most notable illustration of this is in the last two cases in the second 'within-mode' exercise (Table 9.12), where the option at the bottom of the list is ranked below an obviously inferior option. This indication of inconsistency is only approximate, as inferior options would have appeared in fewer pairs than superior options due to the action of the adaptive ranking algorithm. This suggests that the frequencies of choice for the former options will be less reliable than for the latter.

Station/Train Exteriors Exercises

(iv) The fare and escalator attribute appear to have the least impact on responses, while reductions in draughts appear to be most important, particularly off-peak ordinary ticket users.

Having subjected the data to a preliminary analysis using a 'naive' choice frequency approach, the next stage is to consider a more sophisticated method of modelling the responses in relation to the different attribute levels. Such an approach needs to relate the discrete choices of the respondents (the dependent variable) to the combined variation of service improvements and fare (independent variables) in a theoretically sound and statistically robust way.

The data could allow the researcher to construct rankings for some of the respondents across the options in the design and model the rankings as the dependent variable using an appropriate method such a MONANOVA or exploded logit. This approach was rejected on the following grounds:

(i) for many respondents, the data did not contain enough information to construct rankings, because of inconsistencies in their responses (ie leading to high numbers of pairs;)

(ii) such rankings, if constructed, would to some degree be 'manufactured'.

The last point highlights the fact that the rankings would not have been specified directly by respondents. They would be dependent on certain assumptions about dominance which might have been appropriate for enabling the 'Alastair' software to select competitive pairs of options during the interviews, but not for the development of the final models.

The preferred approach was therefore one that would be consistent with the format in which the data was presented. As respondents considered pairs of options, binary logit models could be estimated, relating directly to the paired choices offered to respondents. The discrete choice categories would represent the collapsed 'probably choose'/'definitely choose' scale indicating a preference for the option on the left hand side of the computer screen or the right hand side. All other responses would be omitted. In respect of this last point, an alternative formulation would have been multinomial logit models, in which the additional responses of 'cannot choose' and 'choose neither' could have been included as additional, separate response categories.

This approach was rejected on the grounds that these extra categories gave little additional information on the importance of the passenger facilities and, in the case of the rolling stock exercises, were rarely chosen. A larger proportion of responses in the station facilities exercise were given to the 'neither' option and the significance of this is discussed in the later section on the analysis of that exercise.

9.5.2 Statistical Analysis of the Rolling Stock Exercises

As already noted, changes to the station attributes exercise created some problems, so that this part of the analysis is treated separately from that of the two rolling stock exercises. The next sections are therefore concerned only with the analysis of the rolling stock attributes; a section on the station attributes follows afterwards.

For the analysis of the rolling stock exercises, the binary logit models would have the following form:

$$P_r = \frac{1}{1 + \exp^{(U_l - U_r)}}$$

where: P_r = Probability of choosing the right hand option
U_l = Utility of left hand option
U_r = Utility of right hand option

To enable non-linearities to be detected, fare was represented by dummy variables for each level and interactions between the appropriate facilities were included, such that:

$$\begin{aligned} U_l = \; & a_1 + a_2{*}Fare(10\%) + a_3{*}Fare(30\%) + a_4{*}att2(2) + a_5{*}att2(3) \\ & + a_6{*}att3(2) + a_7{*}att3(3) + a_8{*}att4(2) + a_9{*}att4(3) \\ & + a_{10}{*}att5(2) + a_{11}{*}att5(3) + a_{12}{*}att2(2){*}att3(2) \\ & + a_{13}{*}att2(2){*}att3(3) + a_{14}{*}att2(3){*}att3(2) \\ & + a_{15}{*}att2(3){*}att3(3) \end{aligned}$$

$$\begin{aligned} U_r = \; & a_2{*}Fare(10\%) + a_3{*}Fare(30\%) + a_4{*}att2(2) + a_5{*}att2(3) \\ & + a_6{*}att3(2) + a_7{*}att3(3) + a_8{*}att4(2) + a_9{*}att4(3) \\ & + a_{10}{*}att5(2) + a_{11}{*}att5(3) + a_{12}{*}att2(2){*}att3(2) \\ & + a_{13}{*}att2(2){*}att3(3) + a_{14}{*}att2(3){*}att3(2) \\ & + a_{15}{*}att2(3){*}att3(3) \end{aligned}$$

where: Fare(X%) = fare level, indicating percentage increase;
att1..5 = qualitative service improvements;
$a_1 \ldots a_{15}$ = coefficients

In this scheme, the fare levels and attributes of the left hand options presented on the computer screen are used in U_l, those of the right hand options in U_r. The numbers in brackets indicate the particular levels of fare and each attribute, relative to the base levels of zero fare increase and the lowest standards of each attribute. In the station/train exteriors exercise, interactions could not be investigated under the design used, so that coefficients a_{12} to a_{15} were not applicable. Note that the coefficients are 'generic', in that they are not specific to either the left or right options. This is because the 'within-mode' exercises compare variations of the same items: sets of rolling stock or station/train exterior attributes. Only the constant a_1 has been assigned (arbitrarily) to one option to detect any bias respondents might have towards one side of the computer screen. It would be hoped that the constant would therefore be small and non-significant.

The representation of fare changes as percentages in the models appeared the most appropriate approach, in that it was consistent with the stated preference design (everyone received 10% and 30% increases) and would give monetary valuations in units consistent with the NIMT's forecasting models. As the following discussion shows, differences in the make-up of each of the samples responding to the two rolling stock exercises, together with their different methods of payment, suggested that the results should also be interpreted in relation to actual monetary fare values.

The ALOGIT modelling software (Hague Consulting Group, 1989) was used to estimate the models. The results of the first set of models estimated for each 'within-mode' SP exercise and each time-of-travel/ticket type segment are summarised in Table 9.15 and Table 9.16. In these model runs, all respondents not eliminated by the procedures described earlier were included in the models. These tables show the following:

(i) Model coefficients - labelled in the 'Attribute' column and listed under the 'Coeff' heading;

(ii) Significance - indicated under the 'Sig' heading, in which a question mark shows a coefficient is not significantly different from zero at the 95% level of confidence and an exclamation mark indicates non-significance even at the 90% level of confidence;

Table 9.15 Summary of Weighted Logit Models Estimated From the First "Within-Mode" Rolling Stock Exercise: All Responses

Mode Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Constant	0.00	0.0 !	0.13	0.9 !	-0.05	-0.3 !	0.08	0.6 !	0.06	0.4 !
10% fare	-0.37	-1.9 ?	-0.44	-1.9 ?	-0.73	-2.6	-0.26	-1.3 !	-0.13	-0.6 !
30% fare	-1.52	-5.7	-0.92	-3.4	-2.32	-5.4	-0.90	-3.6	-0.80	-2.7
Clean1	-0.79	-1.1 !	-1.10	-1.3 !	-0.63	-0.6 !	-2.21	-2.8	-1.62	-2.2
Clean2	-0.76	-2.6	-1.59	-3.8	-1.01	-2.3	-1.25	-4.0	-1.16	-3.0
New1	-0.29	-0.9 !	0.00	0.0 !	0.17	0.3 !	-0.07	-0.2 !	-0.39	-1.0 !
New2	0.40	0.8 !	0.10	0.2 !	0.18	0.3 !	0.04	0.1 !	0.75	1.3 !
Air1	-0.70	-3.6	-0.23	-1.0 !	-0.33	-1.2 !	-0.22	-1.2 !	-0.35	-1.5 !
Air2	-1.09	-4.4	-1.32	-4.5	-1.60	-4.3	-0.96	-3.9	-1.18	-4.3
Gangway1	0.09	0.5 !	-0.24	-1.0 !	-0.04	-0.1 !	-0.10	-0.5 !	0.00	0.0 !
Gangway2	-0.63	-3.1	-0.72	-2.9	-0.80	-2.5	-0.68	-3.4	-0.64	-2.5
c1n1	0.04	0.0 !	-0.27	-0.2 !	-0.65	-0.4 !	0.96	0.8 !	1.18	1.1 !
c1n2	-0.82	-1.0 !	-0.68	-0.7 !	-1.14	-1.0 !	0.27	0.3 !	-0.38	-0.4 !
c2n1	-0.01	0.0 !	-0.13	-0.2 !	-0.43	-0.6 !	-0.32	-0.6 !	0.48	0.8 !
c2n2	-1.32	-1.3 !	0.15	0.1 !	-0.84	-0.5 !	-1.91	-1.8 ?	-1.51	-1.3 !
Obs	700		561		382		694		509	
Sample	98		73		65		98		69	
Rho bar sq	0.18		0.21		0.27		0.23		0.17	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
Clean1	0%	0%	0%	0%	0%	0%	100%	100%	100%	100%
New1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Air1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Gangway1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Clean2	31%	31%	44%	44%	30%	30%	43%	43%	39%	39%
New2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Air2	44%	44%	36%	36%	47%	47%	33%	33%	40%	40%
Gangway2	25%	25%	20%	20%	23%	23%	24%	24%	22%	22%

Table 9.16 Summary of Weighted Logit Models Estimated From the Second "Within-Mode" Rolling Stock Exercise: All Responses

Model Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Constant	0.10	0.8 !	-0.06	-0.4 !	0.04	0.2 !	0.01	0.1 !	0.07	0.5 !
10% fare	-0.95	-4.2	-0.54	-2.5	-0.93	-3.2	-0.81	-3.9	-0.27	-1.1 !
30% fare	-1.89	-6.9	-1.67	-5.3	-2.34	-5.4	-1.86	-6.6	-0.65	-2.2
Noise1	-0.43	-0.9 !	-0.35	-0.6 !	0.60	0.9 !	0.23	0.4 !	-0.15	-0.3 !
Noise2	-0.99	-3.1	-1.13	-3.3	-0.93	-2.4	-1.24	-4.0	-0.96	-2.3
Ride1	-0.14	-0.4 !	-0.76	-1.8 ?	0.23	0.5 !	-0.14	-0.4 !	0.03	0.1 !
Ride2	-0.62	-1.4 !	-0.88	-1.8 ?	-0.46	-0.8 !	-1.10	-2.4	-0.57	-1.1 !
Info1	-0.20	-0.9 !	-0.07	-0.3 !	0.09	0.3 !	-0.01	0.0 !	-0.03	-0.1 !
Info2	-0.16	-0.6 !	-0.81	-3.0	-0.89	-2.7	-1.04	-4.1	-0.73	-2.5
Gangway1	-0.30	-1.2 !	-0.26	-0.9 !	-0.09	-0.3 !	-0.35	-1.3 !	-0.16	-0.6 !
Gangway2	-1.04	-3.5	-0.85	-2.6	-0.56	-1.5 !	-1.04	-3.6	-0.68	-2.0
n1r1	-0.01	0.0 !	0.57	0.8 !	-0.96	-1.1 !	-0.44	-0.7 !	0.24	0.4 !
n1r2	-0.26	-0.5 !	-0.36	-0.6 !	-1.32	-1.8 ?	-0.47	-0.9 !	0.10	0.2 !
n2r1	0.08	0.2 !	0.24	0.4 !	-0.49	-0.7 !	0.04	0.1 !	0.23	0.4 !
n2r2	-1.17	-1.7 ?	-0.12	-0.2 !	-0.65	-0.7 !	-0.41	-0.5 !	0.39	0.4 !
Obs	810		570		394		730		461	
Sample	110		78		60		100		64	
Rho bar sq	0.21		0.19		0.21		0.23		0.12	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
L of C:	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
Noise1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ride1	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%
Info1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Gangway1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Noise2	49%	49%	31%	40%	51%	51%	28%	28%	41%	41%
Ride2	0%	0%	24%	0%	0%	0%	25%	25%	0%	0%
Info2	0%	0%	22%	29%	49%	49%	24%	24%	31%	31%
Gangway2	51%	51%	23%	30%	0%	0%	24%	24%	29%	29%

(iii) Data quantities - summarised by the number of observations and the number of individuals (labelled 'Sample') providing these observations;

(iv) Goodness of fit - indicated by the rho bar statistic, which is with respect to zero coefficients/constants;

(v) Interactions - these coefficients represent every combination of levels for the two attributes in each exercise for which interactions can be measured; 'c1n1' refers to 'clean level 1 * new level 1' and so on.

Beneath the summaries of each model is a sub-table indicating the relative impact of each attribute on the total benefit derived by passengers when all intermediate levels of the attributes are introduced and when all maximum levels are introduced. These groupings relate to the packages presented in the succeeding 'between-mode' exercises. The values are calculated by dividing each coefficient by the total of the coefficients in each block of four attribute levels. Two levels of confidence are used (indicated by the label 'L of C'). Those coefficients not significantly different from zero at these levels are treated as zero.

The data used in all the 'within-mode' choice models reported here was weighted two ways: firstly to account for imbalances in the random distribution of the design blocks and secondly to take account of the varying number of responses per individual. Table 9.17 shows the distribution of the design blocks in each rolling stock exercise. It can be seen that in some cases the imbalances were quite noticeable, suggesting that weighting was warranted.

Table 9.17 Proportions of Respondents Being Presented With Each Design 'Block' in Each "Within-Mode" Rolling Stock Exercise: All Responses

	Peak T'card	Peak Ordinary	Off Peak T'card	Off Peak One Day T'card	Off Peak Ordinary
Exercise 1					
Block 1	18	25	25	20	30
Block 2	36	23	18	15	26
Block 3	30	34	38	51	27
Block 4	15	18	19	14	16
Total	**100**	**100**	**100**	**100**	**100**
Exercise 2					
Block 1	42	28	27	25	35
Block 2	23	26	24	31	24
Block 3	20	24	32	28	22
Block 4	15	22	17	16	19
Total	**100**	**100**	**100**	**100**	**100**

In the second case, an average number of responses per individual was calculated for each model segment. The observations in each model were then weighted according to this average. For example, the average number of responses for peak Travelcard users in the first within-mode exercise (Table 9.15) was 7.14 (ie 700 observations/98 individuals.) Thus, each observation for an individual submitting a total of 12 responses received a weight of 0.595 (ie 7.14/12); each observation for one submitting six responses received a weight of 1.19 (ie 7.14/6.)

The t-Statistics have been adjusted to take account of repeated observations in the data, using the approach recommended by Bradley and Kroes (1990). This approach suggests that the sample sizes implicit in the standard errors of a model estimated from SP data should be adjusted downwards, because they represent repeated observations (ie the actual number of respondents is considerably lower than the number of observations, which are grouped around each individual.) To weight the sample size to the equivalent of one observation to each respondent would be too severe, so a compromise is suggested in the form of a halfway figure. For example, nine observations per person suggest that the adjusted standard error $(\sigma * \sqrt{n}/2) = 3\sigma/2$.

Tables 9.15 and 9.16 exhibit the following general characteristics:

(i) The goodness-of-fit is not strong for some models (rho bars are in most cases close to 0.20 in value), being particularly weak for off-peak ordinary ticket users in the second 'within-mode' exercise.

(ii) All the model constants are weak and not significantly different from zero at the 95% level of confidence. This is to be hoped for, given that it does not indicate a bias towards one side of the choice pairs.

(iii) many intermediate levels of rolling stock improvements are not significantly different from zero.

(iv) While the 30% fare coefficients are all significantly different from zero, all but one of the 10% fare coefficients in the first within-mode models and one in the second within-mode models are not significantly different from zero. The importance of the fare levels relates to the importance of the other attributes, which appears greater in the first exercise than in the second.

(v) No interaction coefficients are significantly different from zero at the 95% level of confidence. Only three are significant at the 90% level of confidence.

The sub-tables showing the relative importance for each attribute level allow some comparisons across the five segments. In the first exercise cleanliness and air are the only attributes of any importance at the intermediate levels, and these only for some groups. In the second exercise, the intermediate levels are valued even less, though ride quality is of value to peak ordinary ticket users. A broader spread of values can be seen at the maximum levels, though newness remains unimportant to all travellers, as does ride quality for most.

The only anomalies in the model results can be observed for the off-peak Travelcard and ordinary ticket users, relating to the cleanliness attributes. Here the intermediate improvement is valued more highly than the maximum improvement. For the one-day Travelcard users, the difference in value is not significant at the 95% level of confidence. Some explanation for these anomalies may be found in the strong interaction effects for these groups, though none are statistically significantly different from zero at the 95% level of confidence. If these were added to the weight for maximum improvements to cleanliness, the result would exceed that for the intermediate improvement.

These comparisons of the relative importance of each attribute level within each exercise can be extended to comparisons across the two exercises by standardizing the values against the fare variable. Tables 9.18 and 9.19 show the coefficients for the improvements in the previous tables divided by the fare variables.

Table 9.18 Summary of Weighted Fare Valuations Estimated From the First "Within-Mode" Rolling Stock Exercise: All Responses

Fare Values (%)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
10% fare	n/a	0.7 !	n/a	1.4 !	n/a	1.0 !	n/a	0.9 !	n/a	0.5 !
30% fare	1.4 !	n/a	0.7 !	n/a	1.0 !	n/a	1.2 !	n/a	2.0 !	n/a
Clean1	22 !	16 !	25 !	36 !	9 !	8 !	86 !	74	122 !	61 ?
Clean2	21 !	15	36 ?	52	14 ?	13	48 !	42	87 !	43
New1	8 !	6 !	0 !	0 !	-2 !	-2 !	3 !	2 !	29 !	15 !
New2	-11 !	-8 !	-2 !	-3 !	-2 !	-2 !	-2 !	-1 !	-56 !	-28 !
Air1	19 ?	14	5 !	7 !	5 !	4 !	9 !	7 !	26 !	13 !
Air2	30 ?	22	30 ?	43	22	21	37 !	32	89 !	44
Gangway1	-3 !	-2 !	5 !	8 !	0 !	0 !	4 !	3 !	0 !	0 !
Gangway2	17 !	12	16 !	23	11 ?	10	26 !	23	48 !	24 ?
c1n1	-1 !	-1 !	6 !	9 !	9 !	8 !	-37 !	-32 !	-89 !	-45 !
c1n2	22 !	16 !	16 !	22 !	15 !	15 !	-11 !	-9 !	28 !	14 !
c2n1	0 !	0 !	3 !	4 !	6 !	6 !	12 !	11 !	-36 !	-18 !
c2n2	36 !	26 !	-3 !	-5 !	11 !	11 !	74 !	64 !	114 !	57 !

Package Fare values (90% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
Medium	19%	14%	0%	0%	0%	0%	0%	74%	0%	61%
Maximum	30%	49%	66%	118%	47%	44%	0%	96%	0%	112%

Package Fare values (95% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
Medium	0%	14%	0%	0%	0%	0%	0%	74%	0%	0%
Maximum	0%	49%	0%	118%	22%	44%	0%	96%	0%	88%

Table 9.19 Summary of Weighted Fare Valuations Estimated From the Second "Within-Mode" Rolling Stock Exercise: All Responses

Fare Values (%)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
10% fare	n/a	1.5 !	n/a	1.0 !	n/a	1.2 !	n/a	1.3 !	n/a	1.2 !
30% fare	0.7	n/a	1.0 !	n/a	0.8 !	n/a	0.8 ?	n/a	0.8 !	n/a
Noise1	4 !	7 !	6 !	6 !	-6 !	-8 !	-3 !	-4 !	6 !	7 !
Noise2	10	16	21	20	10 ?	12	15	20	36 !	45 !
Ride1	1 !	2 !	14 !	14 ?	-2 !	-3 !	2 !	2 !	-1 !	-1 !
Ride2	7 !	10 !	16 !	16 ?	5 !	6 !	14	18	21 !	26 !
Info1	2 !	3 !	1 !	1 !	-1 !	-1 !	0 !	0 !	1 !	1 !
Info2	2 !	3 !	15 ?	15	10	11	13	17	27 !	34 ?
Gangway1	3 !	5 !	5 !	5 !	1 !	1 !	4 !	6 !	6 !	8 !
Gangway2	11	17	16 ?	15	6 !	7 !	13	17	26 !	32 !
n1r1	0 !	0 !	-11 !	-10 !	10 !	12 !	5 !	7 !	-9 !	-11 !
n1r2	3 !	4 !	7 !	7 !	14 !	17 ?	6 !	8 !	-4 !	-5 !
n2r1	-1 !	-1 !	-4 !	-4 !	5 !	6 !	-1 !	-1 !	-9 !	-11 !
n2r2	12 !	19 ?	2 !	2 !	7 !	8 !	5 !	7 !	-15 !	-18 !

Package Fare values (90% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
Medium	0%	0%	0%	14%	0%	0%	0%	0%	0%	0%
Maximum	21%	32%	51%	66%	20%	32%	55%	72%	0%	34%

Package Fare values (95% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
Medium	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Maximum	21%	32%	21%	50%	10%	23%	55%	72%	0%	0%

Tables 9.18 and 9.19 contain the following:

(i) Fare Values - calculated as the ratio of each attribute coefficient to 10% and 30% fare coefficients that have been standardised for a percentage cost, inferring the rate of trade-off between each rolling stock improvement and fare. Those values not significant at the 95% level of confidence are shown with a question mark and those not significant even at the 90% level of confidence are shown with an exclamation mark (for a discussion of the method used to determine the significance of ratios of coefficients, see Appendix four.)

(ii) Package Fare Values - the packages of improvements introduced in the succeeding between-mode exercises and defined in the previous tables are given a total fare value by summing the fare values of each attribute level. As before, non-significant values are treated as zero. Two sets of totals are provided for the 90% and 95% levels of confidence.

The most notable observation is that there is a wide variation in the relationship of the 10% fare coefficient with that of the 30% fare level. Not only are there major differences between the segments, but also between the two SP exercises, within each segment. The largest variations between the exercises are for peak Travelcard users and off-peak ordinary ticket users, where the relationships are reversed. It is important to recognise that the ratios of the two fare coefficients are almost all non-significant even at the 90% level of confidence. The very wide confidence limits therefore provide little support that these variations in values are representative of actual responses to fare changes.

The low significance of the intermediate levels of the fare variable is also a problem for the monetary values to be inferred for the rolling stock attributes, as it undermines the confidence that can be placed in them. The majority of valuations are not significantly different from zero at the 95% level of confidence, though some of the maximum attribute levels, when matched against the 30% fare coefficient, are significant (cleanliness is significant at both levels against 30% fare, for all segments.)

The fare valuations in many cases appear implausibly high, as was expected from within-mode exercises. The total fare values, as indicated by the sub-tables, accentuate this, even when all non-significant values are treated as zero. In some cases, travellers appear to be willing to pay more than twice their current fare to obtain rolling stock improvements.

The weaknesses in these first models suggest that there would be some merit in exploring strategies for improving the degree of confidence that can be placed in these findings. These weaknesses are indicated by the weak goodness-of-fit (low rho bar statistics), low t-Statistics and the volatility of the fare coefficients across the two SP exercises.

The following developments of the modelling approach were therefore considered:

(i) Respondents seen to be illogical in their responses (or, more correctly, responding in a way that was inconsistent with utility maximisation theory) could be removed from the data set.

(ii) The level of data aggregation could be increased, by one or both of the following strategies:

   (i) aggregating the segments (eg merging peak and off-peak by ticket type);

   (ii) combining the two SP exercises.

These approaches are discussed below.

Removing 'Illogical' Responses

Responses inconsistent with utility maximisation theory could be identified in pairs where one option obviously dominated the other. Such pairs were displayed by the 'Alastair' software on occasions when respondents were inconsistent in relation to their initial rankings of the attributes or assumptions of dominance were violated. The proportions of illogical responses identified in this way are shown in Table 9.20.

It would be expected that the removal of individuals with such 'illogical' responses would enhance the quality of the models. Nevertheless, an important weakness, already discussed, is the likelihood that only some of these responses might be the result of respondents not understanding the exercises, becoming fatigued or being subject to other response errors identified previously in chapter six. Some of these responses might accurately reflect the way some individuals interpret real choices and make real decisions (see above.) Nevertheless, the hypothetical and, to some extent, marginal nature of the choices offered would suggest that the majority of the data removed would be the product of error due to the limitations of SP surveys and not individuals' normal choice processes.

A second set of models was estimated from the within-mode SP data, in which all the records of individuals displaying one or more illogical response were rejected. This was considered preferable to the rejection of only the single identified illogical responses, on the basis that the other pairs for which an illogical response could not be identified could still be the result of economically non-rational behaviour.

Tables 9.21 and 9.22 summarise the results of the models fitted to data from which illogical responses have been removed. In most cases, the rho bar figures are higher than in the preceding tables, though not dramatically so. The number of coefficients that are significant at the 95% level of confidence also increases, despite the reduced sample sizes (the square roots of which determine reciprocally the size of the t-statistic for each coefficient.)

Table 9.20 Proportions of Illogical Responses to the "Within-Mode" Rolling Stock Exercises by Segment

Exercise	Peak T'card	Peak Ordinary	Off Peak T'card	One Day T'card	Off Peak Ordinary
1	13%	19%	12%	9%	11%
2	12%	12%	14%	16%	10%

Table 9.21 Summary of Weighted Logit Models Estimated From the First "Within-Mode" Rolling Stock Exercise: Illogical Responses not Included

Model Results

Attribute	Peak T'card Coeff	t	Peak Ordinary Coeff	t	Off-Peak T'card Coeff	t	Off-Peak One Day T'card Coeff	t	Off-Peak Ordinary Coeff	t
Constant	-0.04	-0.3 !	0.05	0.3 !	-0.05	-0.3 !	0.03	0.2 !	0.11	0.7 !
10% fare	-0.46	-2.2	-0.62	-2.3	-0.73	-2.5	-0.50	-2.1	-0.20	-0.8 !
30% fare	-1.60	-5.4	-1.19	-3.6	-2.32	-5.3	-1.25	-4.2	-0.97	-3.0
Clean1	-0.80	-1.1 !	-0.70	-0.8 !	-0.63	-0.6 !	-2.44	-2.7	-1.13	-1.4 !
Clean2	-0.92	-2.8	-1.70	-3.5	-1.01	-2.3	-1.67	-4.2	-1.21	-3.0
New1	-0.21	-0.6 !	0.02	0.0 !	0.17	0.3 !	-0.49	-0.9 !	-0.35	-0.8 !
New2	0.15	0.3 !	-0.15	-0.3 !	0.18	0.3 !	-0.20	-0.4 !	0.61	1.0 !
Air1	-0.75	-3.6	-0.28	-1.1 !	-0.33	-1.2 !	-0.33	-1.5 !	-0.28	-1.1 !
Air2	-1.19	-4.2	-1.45	-4.2	-1.60	-4.2	-1.31	-4.4	-1.21	-4.0
Gangway1	-0.09	-0.4 !	-0.34	-1.2 !	-0.04	-0.1 !	-0.31	-1.3 !	-0.17	-0.7 !
Gangway2	-0.73	-3.2	-0.94	-3.2	-0.80	-2.4	-1.07	-4.3	-0.84	-3.0
c1n1	-0.16	-0.1 !	-0.62	-0.5 !	-0.65	-0.4 !	1.31	1.0 !	0.64	0.5 !
c1n2	-0.72	-0.8 !	-0.95	-0.9 !	-1.14	-1.0 !	0.19	0.2 !	-0.80	-0.8 !
c2n1	-0.37	-0.7 !	-0.22	-0.3 !	-0.43	-0.6 !	-0.25	-0.4 !	0.30	0.5 !
c2n2	-1.01	-0.9 !	0.61	0.5 !	-0.84	-0.5 !	-1.52	-1.3 !	-0.76	-0.6 !
Obs	611		456		382		629		452	
Sample	83		59		54		85		61	
Rho bar Sq	0.19		0.22		0.27		0.29		0.17	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

L of C:	Peak T'card 90%	95%	Peak Ordinary 90%	95%	Off-Peak T'card 90%	95%	Off-Peak One Day T'card 90%	95%	Off-Peak Ordinary 90%	95%
Clean1	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%
New1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Air1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Gangway1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Clean2	33%	33%	42%	42%	30%	30%	41%	41%	37%	37%
New2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Air2	42%	42%	35%	35%	47%	47%	32%	32%	37%	37%
Gangway2	26%	26%	23%	23%	23%	23%	26%	26%	26%	26%

Table 9.22 Summary of Weighted Logit Models Estimated From the Second "Within-Mode" Rolling Stock Exercise: Illogical Responses not Included

Model Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**
Constant	0.18	1.3 !	-0.01	-0.1 !	0.07	0.4 !	0.01	0.1 !	-0.03	-0.2 !
10% fare	-0.91	-3.6	-0.63	-2.6	-0.89	-2.8	-1.02	-4.2	-0.61	-2.1
30% fare	-1.94	-6.3	-1.83	-5.1	-2.36	-4.9	-2.13	-6.3	-1.39	-3.6
Noise1	-0.39	-0.7 !	-0.62	-1.0 !	0.51	0.7 !	0.00	0.0 !	-0.48	-0.8 !
Noise2	-1.00	-2.7	-1.55	-3.6	-0.95	-2.1	-1.44	-3.9	-1.06	-2.2
Ride1	-0.17	-0.4 !	-0.86	-1.7 ?	0.16	0.3 !	-0.11	-0.3 !	-0.19	-0.4 !
Ride2	-0.80	-1.5 !	-1.34	-2.4	-0.69	-1.1 !	-1.12	-2.2	-1.18	-1.9 ?
Info1	-0.19	-0.7 !	-0.07	-0.2 !	0.12	0.3 !	-0.09	-0.3 !	-0.13	-0.4 !
Info2	-0.43	-1.5 !	-0.93	-2.8	-0.88	-2.4	-1.02	-3.2	-1.06	-3.0
Gangway1	-0.41	-1.4 !	-0.43	-1.3 !	-0.37	-0.9 !	-0.55	-1.7 ?	-0.32	-0.9 !
Gangway2	-1.19	-3.4	-1.01	-2.5	-0.86	-2.1	-1.35	-3.6	-1.24	-2.8
n1r1	-0.04	-0.1 !	0.54	0.7 !	-0.93	-1.0 !	-0.49	-0.6 !	0.13	0.2 !
n1r2	-0.19	-0.3 !	-0.02	0.0 !	-1.06	-1.3 !	-0.34	-0.5 !	0.43	0.6 !
n2r1	0.00	0.0 !	0.37	0.5 !	-0.34	-0.4 !	0.14	0.2 !	-0.03	0.0 !
n2r2	-1.13	-1.4 !	0.25	0.3 !	-0.68	-0.7 !	-1.01	-1.1 !	0.43	0.4 !
Obs	711		502		337		616		414	
Sample	94		67		50		83		56	
Rho bar sq	0.23		0.23		0.24		0.27		0.19	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
L of C:	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**
Noise1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ride1	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%
Info1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Gangway1	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%
Noise2	46%	46%	32%	32%	35%	35%	29%	29%	23%	32%
Ride2	0%	0%	28%	28%	0%	0%	23%	23%	26%	0%
Info2	0%	0%	19%	19%	33%	33%	21%	21%	23%	32%
Gangway2	54%	54%	21%	21%	32%	32%	27%	27%	27%	37%

The improvement in the significance of some coefficients is particularly noticeable for the 10% fare level in the first exercise, where all but one are now significant at 95% level of confidence. The t-statistics for some rolling stock attributes improve also, though many still remain non-significant. The degree of improvement is not large in most cases, but it is clear that the coefficients for the intermediate levels of a number of attributes benefit most from this procedure. A notable exception, however, is the intermediate level for cleanliness among off-peak ordinary ticket users, which becomes non-significant.

With the improved significance of some coefficients, the proportional values of each item within each 'package' of improvements become more comparable between peak and off-peak travellers within each category of ticket type (one-day Travelcard users show some similarities with off-peak ordinary ticket users.)

Concerning the fare valuations of the rolling stock attributes, tables 9.23 and 9.24 show similar results to the earlier ones, but again with some improvement in the significance of the estimates. Overall, the numerical values do not differ greatly from the earlier values, though the ratios between the fare coefficients change considerably for some segments.

Increasing the Level of Data Aggregation

Two strategies were identified for increasing the level of data aggregation. The simplest was to group the time/ticket segments into broader categories. The second was to maintain the previous segments but combine the data from the two stated preference exercises. With regard to the first approach, it did not seem practical to group period and ordinary ticket types together, because of the different purchasing decisions and unit costs. Instead, peak and off-peak were combined in two ticket categories: period Travelcards and other tickets (ordinaries and one-day Travelcards.) These groupings seemed appropriate for aggregation because the earlier models implied some similarities between the constituent segments.

Table 9.23 Summary of Weighted Fare Valuations Estimated From the First "Within-Mode" Rolling Stock Exercise: Illogical Responses Not Included

Fare Values (%)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
10% fare	n/a	0.9 !	n/a	1.6 !	n/a	1.0 !	n/a	1.2 !	n/a	0.6 !
30% fare	1.1 !	n/a	0.6 !	n/a	1.0 !	n/a	0.8 !	n/a	1.6 !	n/a
Clean1	17 !	15 !	11 !	18 !	9 !	8 !	49 ?	59	56 !	35 !
Clean2	20 ?	17	28 ?	43	14 ?	13	33 ?	40	60 !	37
New1	4 !	4 !	0 !	0 !	-2 !	-2 !	10 !	12 !	17 !	11 !
New2	-3 !	-3 !	3 !	4 !	-2 !	-2 !	4 !	5 !	-30 !	-19 !
Air1	16 ?	14	5 !	7 !	5 !	4 !	7 !	8 !	14 !	9 !
Air2	26 ?	22	23	37	22	21	26 ?	32	60 !	37
Gangway1	2 !	2 !	6 !	9 !	0 !	0 !	6 !	8 !	8 !	5 !
Gangway2	16 ?	14	15 ?	24	11 ?	10	21 ?	26	42 !	26
c1n1	3 !	3 !	10 !	16 !	9 !	8 !	-26 !	-31 !	-31 !	-20 !
c1n2	16 !	14 !	15 !	24 !	15 !	15 !	-4 !	-4 !	40 !	25 !
c2n1	8 !	7 !	4 !	6 !	6 !	6 !	5 !	6 !	-15 !	-9 !
c2n2	22 !	19 !	-10 !	-15 !	11 !	11 !	30 !	37 !	37 !	23 !

Package Fare values (90% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
Medium	16%	14%	0%	0%	0%	0%	49%	59%	0%	0%
Maximum	61%	53%	66%	103%	47%	44%	81%	97%	0%	101%

Package Fare values (95% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
Medium	0%	14%	0%	0%	0%	0%	0%	59%	0%	0%
Maximum	0%	53%	23%	103%	22%	44%	0%	97%	0%	101%

Table 9.24 Summary of Weighted Fare Valuations Estimated From the Second "Within-Mode" Rolling Stock Exercise: Illogical Responses not Included

Fare Values (%)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
10% fare	n/a	1.4 !	n/a	1.0 !	n/a	1.1 !	n/a	1.4 !	n/a	1.3 !
30% fare	0.7 ?	n/a	1.0 !	n/a	0.9 !	n/a	0.7 ?	n/a	0.7 !	n/a
Noise1	4 !	6 !	10 !	10 !	-6 !	-6 !	0 !	0 !	8 !	10 !
Noise2	11	15	25	25	11 ?	12 ?	14	20	17 !	23 ?
Ride1	2 !	3 !	14 !	14 !	-2 !	-2 !	1 !	2 !	3 !	4 !
Ride2	9 !	12 !	21 ?	22	8 !	9 !	11 ?	16	19 !	26 ?
Info1	2 !	3 !	1 !	1 !	-1 !	-1 !	1 !	1 !	2 !	3 !
Info2	5 !	7 !	15 ?	15	10 ?	11	10	14	17 ?	23
Gangway1	5 !	6 !	7 !	7 !	4 !	5 !	5 !	8 ?	5 !	7 !
Gangway2	13	18	16 ?	17	10 ?	11 ?	13	19	20 ?	27
n1r1	0 !	1 !	-9 !	-9 !	10 !	12 !	5 !	7 !	-2 !	-3 !
n1r2	2 !	3 !	0 !	0 !	12 !	13 !	3 !	5 !	-7 !	-9 !
n2r1	0 !	0 !	-6 !	-6 !	4 !	4 !	-1 !	-2 !	0 !	1 !
n2r2	12 !	17 !	-4 !	-4 !	8 !	9 !	10 !	14 !	-7 !	-9 !

Package Fare values (90% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
Medium	0%	0%	0%	0%	0%	0%	0%	8%	0%	0%
Maximum	24%	34%	77%	79%	30%	34%	48%	70%	38%	98%

Package Fare values (95% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
Medium	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Maximum	24%	34%	25%	79%	0%	11%	37%	70%	0%	50%

This aggregation of the segments was not of practical use for application to the NIMT's forecasting model, but served to indicate how the reliability of the results might improve with sample size. As the results would not be used for forecasting, the combination of ticket holders in each group was not weighted to record actual market shares. The models estimated for the two aggregated groups are shown in Table 9.25. Economically irrational responses were removed in the light of the earlier moderate improvements to the models, and the same weighting procedures were applied to adjust for variable numbers of responses per individual. Each model for the aggregated segments shows an acceptable value for the rho bar statistic and the significance of some coefficients has been greatly enhanced due to the larger sample sizes. This improves the reliability of the fare valuations.

In Table 9.26, the differences in the ratios between the two fare levels are less extreme than previously seen, tending to a linear relationship in the first exercise but remaining distinctly non-linear in the second exercise. The greater significance of the fare valuation for some items (eg the top level of ride quality) generally increases the overall values of the packages of items, when compared to what would be obtained if the values from the earlier models were simply averaged across the relevant segments. In this respect the benefits of aggregation are most noticeable.

From this aggregation exercise, it may be argued that aggregation improves significantly the statistical confidence that may be placed in the results, particularly the fare valuations, but not so dramatically as to compensate for the loss of sensitivity to the characteristics of the market as identified by less aggregate segmentation.

The second approach to increasing the level of data aggregation was to combine the data from the two stated preference exercises while maintaining the previous segments. The effect of this would be to produce generic values for the fare and gangway attributes and separate values for all the other attributes. The coefficients for fare and gangways would be based on approximately twice the sample sizes as before.

Table 9.25 Summary of Weighted Logit Models Estimated From the First and Second "Within-Mode" Rolling Stock Exercises: Segments Aggregated and Illogical Responses not Included

Model Results

Attribute	Period T'card Coeff	t	Ordinary & One Day T'card Coeff	t	Attribute	Period T'card Coeff	t	Ordinary & One Day T'card Coeff	t
Constant	-0.04	-0.4 !	0.07	0.9 !	Constant	0.13	1.2 !	0.00	0.0 !
10% fare	-0.55	-3.2	-0.42	-2.9	10% fare	-0.87	-4.5	-0.74	-5.2
30% fare	-1.84	-7.6	-1.08	-6.1	30% fare	-2.04	-8.0	-1.73	-8.7
Clean1	-0.81	-1.3 !	-1.43	-2.9	Noise1	-0.06	-0.1 !	-0.38	-1.1 !
Clean2	-0.98	-3.7	-1.49	-6.3	Noise2	-0.96	-3.4	-1.32	-5.6
New1	-0.08	-0.3 !	-0.29	-1.1 !	Ride1	-0.08	-0.2 !	-0.37	-1.4 !
New2	0.19	0.5 !	0.07	0.2 !	Ride2	-0.75	-1.9 ?	-1.12	-3.5
Air1	-0.61	-3.7	-0.28	-2.1	Info1	-0.06	-0.3 !	-0.12	-0.8 !
Air2	-1.34	-6.0	-1.29	-7.3	Info2	-0.60	-2.7	-0.96	-5.1
Gangway1	-0.07	-0.4 !	-0.26	-1.8 ?	Gangway1	-0.38	-1.7 ?	-0.38	-2.0
Gangway2	-0.78	-4.2	-0.94	-6.1	Gangway2	-1.04	-4.0	-1.18	-5.2
c1n1	-0.23	-0.3 !	0.46	0.6 !	n1r1	-0.30	-0.6 !	0.10	0.2 !
c1n2	-0.78	-1.1 !	-0.49	-0.9 !	n1r2	-0.49	-1.0 !	-0.01	0.0 !
c2n1	-0.31	-0.7 !	-0.07	-0.2 !	n2r1	-0.10	-0.2 !	0.16	0.4 !
c2n2	-1.04	-1.2 !	-0.56	-0.8 !	n2r2	-0.89	-1.5 !	-0.15	-0.3 !
Obs	993		1,537			1,048		1,532	
Sample	137		205			144		206	
Rho bar sq	0.22		0.23			0.23		0.22	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

L of C:	Period T'card 90%	95%	Ordinary & One Day T'card 90%	95%		Period T'card 90%	95%	Ordinary & One Day T'card 90%	95%
Clean1	0%	0%	73%	84%	Noise1	0%	0%	0%	0%
New1	0%	0%	0%	0%	Ride1	0%	0%	0%	0%
Air1	100%	100%	14%	16%	Info1	0%	0%	0%	0%
Gangway1	0%	0%	13%	0%	Gangway1	100%	0%	100%	100%
Clean2	31%	31%	40%	40%	Noise2	29%	37%	29%	29%
New2	0%	0%	0%	0%	Ride2	22%	0%	24%	24%
Air2	43%	43%	35%	35%	Info2	18%	23%	21%	21%
Gangway2	25%	25%	25%	25%	Gangway2	31%	40%	26%	26%

Table 9.26 Summary of Weighted Fare Valuations Estimated From the First and Second "Within-Mode" Rolling Stock Exercises: Segments Aggregated and Illogical Responses not Included

Fare Values (%)

	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
Fare:	10%	30%	10%	30%		10%	30%	10%	30%
10% fare	n/a	0.9 !	n/a	1.2 !	10% fare	n/a	1.3 !	n/a	1.3 !
30% fare	1.1	n/a	0.9 ?	n/a	30% fare	0.8	n/a	0.8	n/a
Clean1	15 !	13 !	34	40	Noise1	1 !	1 !	5 !	7 !
Clean2	18	16	36	41	Noise2	11	14	18	23
New1	1 !	1 !	7 !	8 !	Ride1	1 !	1 !	5 !	6 !
New2	-4 !	-3 !	-2 !	-2 !	Ride2	9 ?	11 ?	15	19
Air1	11	10	7 ?	8	Info1	1 !	1 !	2 !	2 !
Air2	25	22	31	36	Info2	7	9	13	17
Gangway1	1 !	1 !	6 !	7 ?	Gangway1	4 !	6 !	5 ?	7 ?
Gangway2	14	13	23	26	Gangway2	12	15	16	20
c1n1	4 !	4 !	-11 !	-13 !	n1r1	3 !	4 !	-1 !	-2 !
c1n2	14 !	13 !	12 !	14 !	n1n2	6 !	7 !	0 !	0 !
c2n1	6 !	5 !	2 !	2 !	n2r1	1 !	1 !	-2 !	-3 !
c2n2	19 !	17 !	13 !	15 !	n2r2	10 !	13 !	2 !	3 !

Package Fare values (90% level of confidence)

	Period T'card		Ordinary & One Day T'card		Period T'card		Ordinary & One Day T'card	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%
Medium	11%	10%	41%	54%	0%	0%	5%	7%
Maximum	57%	51%	89%	103%	38%	49%	62%	79%

Package Fare values (95% level of confidence)

	Period T'card		Ordinary & One Day T'card		Period T'card		Ordinary & One Day T'card	
Fare:	10%	30%	10%	30%	10%	30%	10%	30%
Medium	11%	10%	34%	47%	0%	0%	0%	0%
Maximum	57%	51%	89%	103%	30%	38%	62%	79%

As the number of respondents in each data set differed within each segment, the responses were weighted to ensure an equal balance. This was in addition to the weighting method used in the preceding models. The weighting ensured that the combined segment sample sizes were reproduced. As before, economically irrational responses were also omitted.

Table 9.27 summarises the model results for the combined data sets, using the original segments. The rho bar statistic for each segment is no worse than observed previously, with that for off-peak ordinary ticket users being the weakest, as before. The proportional values of each attribute level when placed into the relevant packages have changed little from those previously observed, with the exception of a more significant gangway coefficient altering the proportions for one-day Travelcard users.

Table 9.28 summarises the fare valuations for this group of models. Some non-linear relationships still exist, but even at this level of aggregation most of the fare to fare ratios are still non-significant. Nevertheless, this may represent evidence for a non-linear effect for fare among ordinary ticket users, against a linear effect for Travelcard users. As for the remaining attributes, the majority of intermediate levels remain weak and non-significant, as do the interaction terms. The maximum levels yield generally high fare values, particularly against the 30% fare shift among ordinary ticket users. Note that the anomaly identified earlier concerning off-peak Travelcard users' valuations of cleanliness improvements still persists, though the difference between intermediate and maximum improvements are not as pronounced. The package values are in some cases noticeably higher than in previous models: the maximum(2) package for peak ordinary ticket users has a value of 91% of fare compared to 79% from the earlier separate analysis of the second exercise.

Table 9.27 Summary of Weighted Logit Models Estimated From Combined "Within-Mode" Rolling Stock Exercises: Illogical Responses not Included

Model Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Constant	0.06	0.7 !	0.03	0.2 !	0.02	0.1 !	0.01	0.1 !	0.04	0.3 !
10% fare	-0.66	-4.1	-0.64	-3.6	-0.85	-4.0	-0.77	-4.6	-0.38	-2.0
30% fare	-1.78	-8.4	-1.48	-6.2	-2.37	-7.4	-1.70	-7.6	-1.18	-4.8
Clean1	-0.60	-0.8 !	-0.77	-0.9 !	-0.29	-0.3 !	-2.28	-2.6	-1.17	-1.6 !
Clean2	-1.04	-3.3	-1.84	-4.0	-1.11	-2.5	-1.84	-4.8	-1.27	-3.1
New1	-0.10	-0.3 !	0.02	0.0 !	0.29	0.5 !	-0.41	-0.8 !	-0.34	-0.8 !
New2	0.14	0.3 !	-0.15	-0.3 !	0.07	0.1 !	-0.36	-0.8 !	0.64	1.0 !
Air1	-0.75	-3.7	-0.30	-1.2 !	-0.28	-1.0 !	-0.31	-1.4 !	-0.30	-1.2 !
Air2	-1.30	-4.9	-1.54	-4.7	-1.60	-4.4	-1.43	-4.7	-1.30	-4.3
Noise1	-0.21	-0.4 !	-0.71	-1.2 !	0.48	0.7 !	0.04	0.1 !	-0.32	-0.6 !
Noise2	-0.93	-2.5	-1.44	-3.5	-0.90	-2.1	-1.34	-3.8	-1.07	-2.3
Ride1	-0.05	-0.1 !	-0.78	-1.6 !	0.20	0.4 !	-0.07	-0.2 !	-0.12	-0.3 !
Ride2	-0.66	-1.4 !	-1.15	-2.2	-0.56	-0.9 !	-1.01	-2.2	-1.11	-2.1
Info1	-0.27	-1.2 !	-0.07	-0.3 !	0.03	0.1 !	-0.11	-0.5 !	-0.18	-0.6 !
Info2	-0.52	-1.9 ?	-0.87	-2.7	-0.92	-2.6	-0.96	-3.4	-1.09	-3.3
Gangway1	-0.19	-1.1 !	-0.33	-1.6 !	-0.19	-0.8 !	-0.45	-2.3	-0.22	-1.1 !
Gangway2	-0.84	-4.7	-1.03	-4.4	-0.78	-3.1	-1.17	-5.6	-0.95	-4.2
c1n1	-0.53	-0.5 !	-0.63	-0.5 !	-1.24	-0.8 !	0.89	0.7 !	0.66	0.6 !
c1n2	-0.98	-1.2 !	-1.01	-1.0 !	-1.45	-1.3 !	-0.14	-0.1 !	-0.86	-0.9 !
c2n1	-0.42	-0.8 !	-0.29	-0.4 !	-0.51	-0.7 !	-0.46	-0.7 !	0.26	0.4 !
c2n2	-0.71	-0.7 !	0.65	0.5 !	-0.28	-0.2 !	-1.12	-0.9 !	-0.77	-0.6 !
n1r1	-0.19	-0.3 !	0.71	0.9 !	-0.89	-1.0 !	-0.46	-0.7 !	0.02	0.0 !
n1r2	-0.46	-0.8 !	0.06	0.1 !	-1.18	-1.6 !	-0.32	-0.6 !	0.31	0.5 !
n2r1	-0.27	-0.5 !	0.39	0.6 !	-0.53	-0.7 !	0.11	0.2 !	-0.06	-0.1 !
n2r2	-0.80	-1.1 !	-0.03	0.0 !	-0.74	-0.8 !	-0.75	-0.9 !	0.74	0.8 !
Obs	1,322		958		719		1,245		866	
Sample	177		126		104		168		117	
Rho bar Sq	0.21		0.23		0.26		0.28		0.18	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Table 9.27 (Contd) Summary of Weighted Logit Models Estimated From Combined "Within-Mode" Rolling Stock Exercises: Illogical Responses not Included - Part Two

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
L of C:	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
Clean1	0%	0%	0%	0%	0%	0%	84%	84%	0%	0%
New1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Air1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Gangway1	0%	0%	0%	0%	0%	0%	16%	16%	0%	0%
Clean2	33%	33%	42%	42%	32%	32%	41%	41%	36%	36%
New2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Air2	41%	41%	35%	35%	46%	46%	32%	32%	37%	37%
Gangway2	26%	26%	23%	23%	22%	22%	26%	26%	27%	27%
Noise1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ride1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Info1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Gangway1	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%
Noise2	41%	53%	32%	32%	35%	35%	30%	30%	25%	25%
Ride2	0%	0%	25%	25%	0%	0%	23%	23%	26%	26%
Info2	23%	0%	19%	19%	35%	35%	21%	21%	26%	26%
Gangway2	37%	47%	23%	23%	30%	30%	26%	26%	23%	23%

Table 9.28 Summary of Weighted Fare Valuations Estimated From Combined "Within-Mode" Rolling Stock Exercises: Illogical Responses not Included

Fare Values (%)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
10% fare	n/a	1.1 !	n/a	1.3 !	n/a	1.1 !	n/a	1.4 !	n/a	1.0 !
30% fare	0.9	n/a	0.8 ?	n/a	0.9	n/a	0.7	n/a	1.0 !	n/a
Clean1	9 !	10 !	12 !	16 !	3 !	4 !	30	40	31 !	30 !
Clean2	16	18	29	37	13	14	24	33	34 ?	32
New1	2 !	2 !	0 !	0 !	-3 !	-4 !	5 !	7 !	9 !	9 !
New2	-2 !	-2 !	2 !	3 !	-1 !	-1 !	5 !	6 !	-17 !	-16 !
Air1	11	13	5 !	6 !	3 !	4 !	4 !	5 !	8 !	8 !
Air2	20	22	24	31	19	20	19	25	34 ?	33
Noise1	3 !	4 !	11 !	14 !	-6 !	-6 !	0 !	-1 !	9 !	8 !
Noise2	14	16	23	29	11 ?	11	17	24	28 !	27
Ride1	1 !	1 !	12 !	16 !	-2 !	-2 !	1 !	1 !	3 !	3 !
Ride2	10 !	11 !	18 ?	23	7 !	7 !	13	18	29 !	28 ?
Info1	4 !	5 !	1 !	1 !	0 !	0 !	1 !	2 !	5 !	5 !
Info2	8 ?	9 ?	14	18	11	12	13	17	29 ?	28
Gangway1	3 !	3 !	5 !	7 !	2 !	2 !	6	8	6 !	6 !
Gangway2	13	14	16	21	9	10	15	21	25 ?	24
c1n1	8 !	9 !	10 !	13 !	15 !	16 !	-12 !	-16 !	-17 !	-17 !
c1n2	15 !	17 !	16 !	20 !	17 !	18 !	2 !	2 !	23 !	22 !
c2n1	6 !	7 !	5 !	6 !	6 !	6 !	6 !	8 !	-7 !	-7 !
c2n2	11 !	12 !	-10 !	-13 !	3 !	4 !	15 !	20 !	20 !	20 !
n1r1	3 !	3 !	-11 !	-14 !	10 !	11 !	6 !	8 !	-1 !	-1 !
n1r2	7 !	8 !	-1 !	-1 !	14 !	15 !	4 !	6 !	-8 !	-8 !
n2r1	4 !	5 !	-6 !	-8 !	6 !	7 !	-1 !	-2 !	1 !	1 !
n2r2	12 !	14 !	1 !	1 !	9 !	9 !	10 !	13 !	-20 !	-19 !

Package Fare values (95% level of confidence)

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Fare:	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**	**10%**	**30%**
Medium1	11%	13%	0%	0%	0%	0%	36%	48%	0%	0%
Maximum1	48%	54%	69%	90%	41%	44%	58%	78%	0%	89%
Medium2	0%	0%	0%	0%	0%	0%	6%	8%	0%	0%
Maximum2	27%	30%	53%	91%	20%	33%	58%	79%	0%	79%

Finally, Table 9.29 shows models from the combined data sets with just the two ticket segments: Period Travelcards and ordinaries/ one-day Travelcards. This represents what may be considered to be the highest level of aggregation that may be applied to the data, given that a single model for all the data would contain very different fare variables (although all values are expressed in percentage changes, actual values are very different depending on whether a period or other ticket has been used.) The significance of the parameters improves slightly with the increase in segment sizes while the rho bar statistics remain small but acceptable. The proportional values (Table 9.29) and fare valuations (Table 9.30) show that intermediate levels are generally poorly valued, while at the maximum levels, the values are more evenly spread across the improvements.

Conclusions to the Analysis of the 'Within-Mode' Rolling Stock Exercises

The preceding section has reported on a range of analytical strategies based around the application of logit models to the adaptive ranking data derived from the 'within-mode' rolling stock SP exercises. It has demonstrated how the quality of the models may be enhanced by different treatments, such as the removal of economically irrational responses and the aggregation of the data.

Table 9.31 shows the average valuations for each improvement over all the different treatments, firstly for those analyses which included the peak/off-peak split then secondly for all the analyses, the data aggregated according to ticket types alone. This allows a general overview of the results obtained from the 'within-mode' ranking exercises:

(i) for intermediate improvements, significant valuations can only be obtained for cleanliness, air quality and gangways, and these only for some segments;

(ii) maximum improvements mostly receive a significant valuation, the only exception being newness;

Table 9.29 Summary of Weighted Logit Models Estimated From Combined "Within-Mode" Rolling Stock Exercises: Segments Aggregated and Illogical Responses not Included

Model Results

	Period T'card		Ordinary & One Day T'card	
Attribute	**Coeff**	**t**	**Coeff**	**t**
Constant	0.04	0.6 !	0.04	0.6 !
10% fare	-0.71	-5.6	-0.58	-5.9
30% fare	-1.96	-11.2	-1.41	-10.7
Clean1	-0.56	-1.0 !	-1.40	-3.1
Clean2	-1.08	-4.3	-1.60	-6.9
New1	0.04	0.1 !	-0.25	-0.9 !
New2	0.14	0.3 !	0.04	0.1 !
Air1	-0.58	-3.6	-0.28	-2.1
Air2	-1.40	-6.7	-1.40	-7.9
Noise1	0.03	0.1 !	-0.35	-1.1 !
Noise2	-0.90	-3.2	-1.26	-5.5
Ride1	0.00	0.0 !	-0.31	-1.2 !
Ride2	-0.62	-1.6 ?	-1.01	-3.6
Info1	-0.13	-0.7 !	-0.13	-0.9 !
Info2	-0.64	-3.0	-0.93	-5.3
Gangway1	-0.20	-1.5 !	-0.31	-2.7
Gangway2	-0.83	-5.7	-1.04	-8.3
c1n1	-0.67	-0.8 !	0.31	0.5 !
c1n2	-1.05	-1.6 !	-0.65	-1.2 !
c2n1	-0.37	-0.9 !	-0.20	-0.5 !
c2n2	-0.64	-0.8 !	-0.41	-0.6 !
n1r1	-0.39	-0.8 !	0.12	0.3 !
n1r2	-0.69	-1.6 !	0.00	0.0 !
n2r1	-0.31	-0.7 !	0.16	0.5 !
n2r2	-0.76	-1.4 !	-0.05	-0.1 !
Obs	2,041		3,069	
Sample	281		311	
Rho bar Sq	0.22		0.22	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Table 9.29 (Contd) Summary of Weighted Logit Models Estimated From Combined "Within-Mode" Rolling Stock Exercises: Segments Aggregated and Illogical Responses not Included - Part Two

Proportional Values

	Period T'card		Ordinary & One Day T'card	
L of C:	**90%**	**95%**	**90%**	**95%**
Clean1	0%	0%	70%	70%
New1	0%	0%	0%	0%
Air1	100%	100%	14%	14%
Gangway1	0%	0%	15%	15%
Clean2	33%	33%	40%	40%
New2	0%	0%	0%	0%
Air2	42%	42%	35%	35%
Gangway2	25%	25%	26%	26%
Noise1	0%	0%	0%	0%
Ride1	0%	0%	0%	0%
Info1	0%	0%	0%	0%
Gangway1	0%	0%	100%	100%
Noise2	30%	38%	30%	30%
Ride2	21%	0%	24%	24%
Info2	21%	27%	22%	22%
Gangway2	28%	35%	25%	25%

Table 9.30 Summary of Weighted Fare Valuations Estimated From Combined "Within-Mode" Rolling Stock Exercise: Segments Aggregated and Illogical Responses not Included

Fare Values (%)

	Period T'card		Ordinary & One Day T'card	
Fare:	10%	30%	10%	30%
10% fare	n/a	1.1 !	n/a	1.2 !
30% fare	0.9	n/a	0.8	n/a
Clean1	8 !	9 !	24	30
Clean2	15	17	28	34
New1	-1 !	-1 !	4 !	5 !
New2	-2 !	-2 !	-1 !	-1 !
Air1	8	9	5	6
Air2	20	22	24	30
Noise1	0 !	0 !	6 !	7 !
Noise2	13	14	22	27
Ride1	0 !	0 !	5 !	7 !
Ride2	9 !	9 !	17	22
Info1	2 !	2 !	2 !	3 !
Info2	9	10	16	20
Gangway1	3 !	3 !	5	6
Gangway2	12	13	18	22
c1n1	9 !	10 !	-5 !	-7 !
c1n2	15 !	16 !	11 !	14 !
c2n1	5 !	6 !	3 !	4 !
c2n2	9 !	10 !	7 !	9 !
n1r1	5 !	6 !	-2 !	-3 !
n1r2	10 !	11 !	0 !	0 !
n2r1	4 !	5 !	-3 !	-3 !
n2r2	11 !	12 !	1 !	1 !

Package Fare values

	Period T'card		Ordinary & One Day T'card	
Fare:	10%	30%	10%	30%
(90% level of confidence)				
Medium1	8%	9%	29%	36%
Maximum1	48%	52%	89%	110%
Medium2	0%	0%	5%	6%
Maximum2	34%	36%	73%	90%
(95% level of confidence)				
Medium1	8%	9%	34%	42%
Maximum1	47%	51%	70%	86%
Medium2	0%	0%	5%	6%
Maximum2	34%	36%	73%	90%

Table 9.31 Summary of Weighted Fare Valuations From "Within-Mode" Rolling Stock Exercises: Averages Over All Analyses

Average Fare Valuations Over First Three Approaches (95% L of C)

Intermediate Levels	Fare	Clean-liness	New-ness	Air Qua'ty	Noise	Ride Qua'ty	Info	Gang-ways
Peak TC	10%	0	0	4	0	0	0	0
Off-Peak TC	10%	0	0	0	0	0	0	0
Peak Ordinary	10%	0	0	0	0	0	0	0
One Day TC	10%	10	0	0	0	0	0	1
Off-Peak Ordy	10%	0	0	0	0	0	0	0
Peak TC	30%	0	0	13	0	0	0	0
Off-Peak TC	30%	0	0	0	0	0	0	0
Peak Ordinary	30%	0	0	0	0	0	0	0
One Day TC	30%	57	0	0	0	0	0	2
Off-Peak Ordy	30%	0	0	0	0	0	0	0

Maximum Levels	Fare	Clean-liness	New-ness	Air Qua'ty	Noise	Ride Qua'ty	Info	Gang-ways
Peak TC	10%	5	0	7	12	0	0	7
Off-Peak TC	10%	4	0	21	0	0	7	2
Peak Ordinary	10%	10	0	16	23	0	5	3
One Day TC	10%	8	0	6	16	9	12	8
Off-Peak Ordy	10%	0	0	0	0	0	0	0
Peak TC	30%	17	0	22	16	0	0	15
Off-Peak TC	30%	13	0	21	8	0	11	6
Peak Ordinary	30%	44	0	37	25	15	16	20
One Day TC	30%	38	0	30	21	17	16	21
Off-Peak Ordy	30%	38	0	38	9	0	17	15

Average Fare Valuations Over All Five Approaches (95% L of C)

Intermediate Levels	Fare	Clean-liness	New-ness	Air Qua'ty	Noise	Ride Qua'ty	Info	Gang-ways
Travelcards	10%	0	0	5	0	0	0	0
Ordinary/OD TC	10%	14	0	1	0	0	0	0
Travelcards	30%	0	0	9	0	0	0	0
Ordinary/OD TC	30%	27	0	3	0	0	0	1

Maximum Levels	Fare	Clean-liness	New-ness	Air Qua'ty	Noise	Ride Qua'ty	Info	Gang-ways
Travelcards	10%	10	0	16	9	0	5	0
Ordinary/OD TC	10%	16	0	15	14	5	8	7
Travelcards	30%	16	0	21	16	3	7	0
Ordinary/OD TC	30%	38	0	33	19	15	18	12

(iii) significant valuations are more likely to be obtained against 30% changes in fare (which have more significant coefficients), with a higher value than those significant valuations observed for a 10% change.

Concerning the stability of the results across the different treatments, Table 9.32 provides a basis for comparison between the improvements by summarising the coefficients of variation (average values over standard deviation across the treatments) for each item. Generally, the valuations derived against 30% fare changes are the most stable.

9.5.3 Statistical Analysis of the Station Attribute Exercise

As in the analysis of the rolling stock exercises, the binary logit models would have the following form:

$$P_r = \frac{1}{1 + \exp^{(Ul-Ur)}}$$

where: P_r = Probability of choosing the right hand option
U_l = Utility of left hand option
U_r = Utility of right hand option

Interactions could not be examined under the design used, so main effects only were included as components of the utility functions:

$$U_l = a_1 + a_2*Fare(10\%) + a_3*Fare(30\%) + a_4*att2(2) + a_5*att3(2) + a_6*att4(2) + a_7*att5(2)$$

$$U_l = a_1 + a_2*Fare(10\%) + a_3*Fare(30\%) + a_4*att2(2) + a_5*att3(2) + a_6*att4(2) + a_7*att5(2)$$

where: Fare(X%) = fare level, indicating percentage increase;
att1..5 = qualitative service improvements;
$a_1 \ldots a_7$ = coefficients

Table 9.32 Summary of Variations in Fare Valuations From "Within-Mode" Rolling Stock Exercises: Variations Over All Analyses

Coefficients of Variation For Fare Valuations (Average / Std Deviation)

Over First Three Approaches

Intermediate Levels	**Fare**	**Clean-liness**	**New-ness**	**Air Qua'ty**	**Noise**	**Ride Qua'ty**	**Info**	**Gang-ways**
Peak TC	10%	-	-	173%	-	-	-	-
Off-Peak TC	10%	-	-	-	-	-	-	-
Peak Ordinary	10%	-	-	-	-	-	-	-
One Day TC	10%	173%	-	-	-	-	-	224%
Off-Peak Ordy	10%	-	-	-	-	-	-	-
Peak TC	30%	-	-	5%	-	-	-	-
Off-Peak TC	30%	-	-	-	-	-	-	-
Peak Ordinary	30%	-	-	-	-	-	-	-
One Day TC	30%	29%	-	-	-	-	-	224%
Off-Peak Ordy	30%	-	-	-	-	-	-	-

Maximum Levels	**Fare**	**Clean-liness**	**New-ness**	**Air Qua'ty**	**Noise**	**Ride Qua'ty**	**Info**	**Gang-ways**
Peak TC	10%	-	-	-	-	-	-	-
Off-Peak TC	10%	173%	-	173%	17%	-	-	92%
Peak Ordinary	10%	173%	-	8%	-	-	87%	224%
One Day TC	10%	173%	-	87%	8%	-	173%	224%
Off-Peak Ordy	10%	173%	-	173%	11%	87%	14%	92%
Peak TC	30%	-	-	-	-	-	-	-
Off-Peak TC	30%	8%	-	2%	1%	-	-	16%
Peak Ordinary	30%	4%	-	1%	87%	-	2%	91%
One Day TC	30%	17%	-	16%	18%	87%	10%	19%
Off-Peak Ordy	30%	13%	-	12%	9%	7%	9%	16%

Over Five Approaches

Intermediate Levels	**Fare**	**Clean-liness**	**New-ness**	**Air Qua'ty**	**Noise**	**Ride Qua'ty**	**Info**	**Gang-ways**
Travelcards	10%	-	-	95%	-	-	-	300%
Ordinary/OD TC	10%	107%	-	224%	-	-	-	287%
Travelcards	30%	-	-	10%	-	-	-	282%
Ordinary/OD TC	30%	37%	-	139%	-	-	-	205%

Maximum Levels	**Fare**	**Clean-liness**	**New-ness**	**Air Qua'ty**	**Noise**	**Ride Qua'ty**	**Info**	**Gang-ways**
Travelcards	10%	92%	-	44%	27%	-	74%	80%
Ordinary/OD TC	10%	98%	-	82%	13%	118%	44%	98%
Travelcards	30%	6%	-	1%	21%	224%	70%	41%
Ordinary/OD TC	30%	14%	-	13%	45%	36%	21%	43%

The first attempt to estimate this model failed due to perfect correlation between the coefficients (not, it should be noted, between the independent variables themselves.) This compared dramatically with the very low correlations observed throughout the analysis of the rolling stock exercises and indicated that the properties of the original experimental design had been completely undermined by the removal of the best option (where all the levels were 'good'.) This failure of the model was the case for all five segments.

As no dominant option was present in the set of options, there was no scope for removing economically irrational choices, though it is doubtful that such a strategy would solve the serious problem of correlation. The loss of orthogonality from the removal of the dominant option does not suggest a reason for the model failures either, as it was the coefficients and not the independent variables that were correlated. Nevertheless, with a view to obtaining a workable model, the re-insertion of this option might have some benefit. The most likely explanation appears to be that there has been some loss in the degrees of freedom in the data, relative to the number of coefficients being estimated.

The following strategies were therefore considered:

(i) The use of a more parsimonious model formulation. It was possible that much of the correlation could be attributable to a single coefficient correlated with all the others. The reduction of model parameters is an established way of trying to deal with correlation in revealed preference models.

(ii) The addition of artificial observations to the data, representing the effect of including the missing option from the design, in which all attributes were at their best levels. It would have to be assumed that such an option would dominate in all pairs and the t-Statistics derived from the models would have to be scaled appropriately.

With respect to the first strategy, a series of models were run in which each coefficient was removed one at a time. All the models continued to fail, with the exception of those in which one of the fare coefficients was omitted. Tables 9.33 and 9.34 show the effects of removing the 30% fare coefficient and the 10% fare coefficient respectively. In the first models, fare is consistently the wrong sign, as are most of the other coefficients, with the exception of escalators. In the second models, fare is the correct sign, as are most of the the other coefficients. Those coefficients that are of the wrong sign in the second models are non-significant, with the exception of the exteriors coefficient for off-peak Travelcard users. Contrary to the impression given by the initial observation of choice frequencies, escalators appear to be by far the most important station attribute. Correlations for the first models were in some cases still high, especially between the two fare coefficients, but in these second models, all correlations were low.

The first strategy for obtaining workable models appeared to work, in terms of producing internally consistent models. In the second strategy (the re-inserting of the missing option), each individual was deemed to have assessed 15 additional pairs of choices, representing the introduction of two new options: one where all attributes were at their best (the missing option) and one when all were at their worst. The 15 pairs represented every combination of these artificial options with the seven options actually included in the design, and with each other. Despite these additions, the model with all coefficients failed. This suggests that the data suffered from additional problems, in that correlations between the coefficients were the stronger influence.

Models with only one or other of the fare coefficients produced the results shown in tables 9.35 and 9.36. The models in the first table have improved on Table 9.33, insofar as more coefficients are of the correct sign. This is true of the fare coefficient, but in most cases it is weak and non-significant. Note the higher rho bar statistics compared to the previous models, indicating the effect of many deterministic choices in the data.

Table 9.33 Summary of Weighted Logit Models Estimated From the "Within-Mode" Station Facilities Exercise: 30% Fare Coefficient Removed

Model Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Constant	0.04	0.3 !	-0.02	-0.1 !	0.50	2.7	0.10	0.6 !	0.09	0.5 !
10% fare	0.84	3.3	1.06	3.6	1.22	4.1	0.61	2.3	0.78	2.7
30% fare	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Draughts	0.61	2.3	0.00	0.0 !	0.53	1.9 ?	-0.21	-0.8 !	0.18	0.6 !
Exteriors	0.68	2.9	0.57	2.2	1.00	3.7	0.16	0.7 !	0.71	2.6
Escalators	-0.64	-3.1	-1.65	-5.7	-0.60	-2.5	-1.27	-5.4	-1.42	-5.1
Graffiti	0.32	1.5 !	-0.07	-0.3 !	0.43	1.8 ?	-0.07	-0.3 !	0.00	0.0 !
Obs	443		351		262		385		361	
Sample	57		48		41		52		46	
Rho bar Sq	0.10		0.25		0.21		0.16		0.19	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
L of C:	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
Draughts	47%	176%	0%	0%	57%	0%	0%	0%	0%	0%
Exteriors	53%	224%	100%	374%	108%	403%	0%	0%	80%	297%
Escalators	-50%	-242%	-288%	-988%	-64%	-265%	-774%	-3309%	-159%	-571%
Graffiti	0%	0%	0%	0%	46%	0%	0%	0%	0%	0%

Fare Values (%)

	Peak T'card	Peak Ordinary	Off-Peak T'card	Off-Peak One Day T'card	Off-Peak Ordinary
Fare:	10%	10%	10%	10%	10%
Draughts	7 ?	0 !	4 ?	-3 !	2 !
Exteriors	8	5 ?	8	3 !	9 ?
Escalators	-8	-16	-5	-21	-18
Graffiti	4 !	-1 !	3 !	-1 !	0 !

Table 9.34 Summary of Weighted Logit Models Estimated From the "Within-Mode" Station Facilities Exercise: 10% Fare Coefficient Removed

Model Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Constant	0.04	0.3 !	-0.02	-0.1 !	0.52	2.7	0.10	0.6 !	0.09	0.5 !
10% fare	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
30% fare	-0.84	-3.3	0.00	0.0 !	-1.53	-4.6	-0.61	-2.3	-0.78	-2.7
Draughts	0.19	0.8 !	-0.53	-2.0	0.02	0.1 !	-0.52	-2.2	-0.21	-0.8 !
Exteriors	0.26	1.3 !	0.04	0.2 !	0.47	2.1	-0.14	-0.7 !	0.32	1.3 !
Escalators	-1.06	-4.6	-2.19	-6.5	-1.35	-4.6	-1.57	-5.8	-1.81	-5.6
Graffiti	-0.10	-0.5 !	-0.60	-2.4	-0.22	-0.9 !	-0.37	-1.7 ?	-0.39	-1.6 !
Obs	443		351		262		385		361	
Sample	57		48		41		52		46	
Rho bar Sq	0.10		0.25		0.18		0.16		0.19	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
L of C:	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
Draughts	0%	0%	0%	19%	0%	0%	0%	28%	0%	0%
Exteriors	0%	0%	0%	0%	0%	-85%	0%	0%	0%	0%
Escalators	0%	100%	0%	60%	0%	185%	0%	72%	0%	100%
Graffiti	0%	0%	0%	22%	0%	0%	0%	0%	0%	0%

Fare Values (%)

	Peak T'card	Peak Ordinary	Off-Peak T'card	Off-Peak One Day T'card	Off-Peak Ordinary
Fare:	30%	30%	30%	30%	30%
Draughts	-7 !	0 !	0 !	25 !	8 !
Exteriors	-9 !	0 !	-9 ?	7 !	-12 !
Escalators	38	0 !	27	77	69
Graffiti	4 !	0 !	4 !	18 !	15 !

Table 9.35 Summary of Weighted Logit Models Estimated From the "Within-Mode" Station Facilities Exercise: Artificial Data Attached and 30% Fare Coefficient Removed

Model Results

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
Attribute	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**
Constant	0.09	0.4 !	0.02	0.1 !	0.11	0.4 !	0.20	0.8 !	0.08	0.3 !
10% fare	-0.59	-1.6 !	-0.22	-0.5 !	-0.25	-0.6 !	-0.54	-1.3 !	-0.33	-0.8 !
30% fare	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Draughts	-1.31	-4.2	-1.40	-3.8	-1.23	-3.3	-1.71	-4.6	-1.38	-3.8
Exteriors	-1.16	-4.2	-1.21	-3.7	-1.12	-3.4	-1.42	-4.5	-0.99	-3.2
Escalators	-1.78	-5.2	-2.31	-4.9	-1.84	-4.2	-2.23	-5.3	-2.25	-5.1
Graffiti	-1.20	-4.3	-1.48	-4.2	-1.36	-3.9	-1.31	-4.0	-1.24	-3.8
Obs	1,339		1,093		878		1,267		1,047	
Sample	57		48		41		52		46	
Rho bar Sq	0.57		0.63		1.73		0.66		0.58	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One Day T'card		Off-Peak Ordinary	
L of C:	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**
Draughts	31%	31%	28%	28%	29%	29%	32%	32%	30%	30%
Exteriors	27%	27%	25%	25%	27%	27%	26%	26%	22%	22%
Escalators	42%	42%	47%	47%	44%	44%	42%	42%	49%	49%
Graffiti	28%	28%	30%	30%	32%	32%	24%	24%	27%	27%

Fare Values (%)

	Peak T'card	Peak Ordinary	Off-Peak T'card	Off-Peak One Day T'card	Off-Peak Ordinary
Fare:	**10%**	**10%**	**10%**	**10%**	**10%**
Draughts	22 !	64 !	49 !	32 !	42 !
Exteriors	20 !	56 !	45 !	26 !	30 !
Escalators	30 !	106 !	74 !	41 !	68 !
Graffiti	20 !	68 !	54 !	24 !	38 !

Table 9.36 Summary of Weighted Logit Models Estimated From the "Within-Mode" Station Facilities Exercise: Artificial Data Attached and 10% Fare Coefficient Removed

Model Results

Attribute	Peak T'card Coeff	t		Peak Ordinary Coeff	t		Off-Peak T'card Coeff	t		Off-Peak One Day T'card Coeff	t		Off-Peak Ordinary Coeff	t	
Constant	0.05	0.2	!	-0.20	-0.7	!	0.32	1.0	!	0.10	0.4	!	0.11	0.4	!
10% fare	N/A	N/A		N/A	N/A		N/A	N/A		N/A	N/A		N/A	N/A	
30% fare	-1.71	-4.4		-1.68	-3.7		-2.43	-4.1		-1.42	-3.4		-1.66	-3.9	
Draughts	-0.86	-2.3		-1.11	-2.6		-0.78	-1.5	!	-1.43	-3.4		-1.00	-2.4	
Exteriors	-0.77	-2.7		-0.77	-2.2		-0.50	-1.3	!	-1.04	-3.3		-0.58	-1.8	?
Escalators	-2.26	-5.7		-2.79	-5.2		-2.42	-4.3		-2.61	-5.6		-2.83	-5.5	
Graffiti	-1.13	-3.9		-1.52	-4.1		-1.38	-3.4		-1.24	-3.7		-1.34	-3.8	
Obs	1,339			1,093			878			1,267			1,047		
Sample	57			48			41			52			46		
Rho bar Sq	0.63			0.69			0.69			0.69			0.65		

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Proportional Values

L of C:	Peak T'card 90%	95%	Peak Ordinary 90%	95%	Off-Peak T'card 90%	95%	Off-Peak One Day T'card 90%	95%	Off-Peak Ordinary 90%	95%
Draughts	6%	6%	8%	8%	0%	0%	9%	9%	9%	9%
Exteriors	5%	5%	5%	5%	0%	0%	6%	6%	5%	0%
Escalators	16%	16%	20%	20%	31%	31%	16%	16%	24%	24%
Graffiti	8%	8%	11%	11%	18%	18%	8%	8%	12%	12%

Fare Values (%)

Fare:	Peak T'card 30%	Peak Ordinary 30%	Off-Peak T'card 30%	Off-Peak One Day T'card 30%	Off-Peak Ordinary 30%
Draughts	15	20	10 !	30	18
Exteriors	14	14 ?	6 !	22	11 !
Escalators	40	50	30	55	51
Graffiti	20	27	17	26	24

The rho bar statistic for off-peak Travelcard ticket users indicates that the maximum likelihood estimate was actually worse when the model finally converged than at the beginning! For this segment particularly, twenty iterations were carried out before the convergence criterion was satisfied (the normal number is in the range of three to eight iterations), suggesting that the resulting model is not a reliable fit. In contrast, the second models show plausible results for all segments, though the rho bar statistics remain large as a result of the presence of the deterministic artificial choices.

As observed above, the initial 'naive' examination of the responses to the station facilities exercise (see Table 9.14) suggested that escalators and fare were unimportant. The models estimated above contradict this. As already noted, the 'naive' analysis does not take account of how frequency of choice is related to the options rejected as well as chosen, and in this case some importance must be attached to the high proportion of responses in the 'neither option' category. A total of 15% of responses fell into this category.

Table 9.37 breaks down the proportion of 'neither' responses for each option and shows three important features:

(i) Respondents are sensitive to the 'poor' level of fare (+30%), choosing neither option in up to a quarter of the responses.

(ii) Respondents are noticeably less sensitive to the 'medium' level of fare.

(iii) Of the four station improvements, escalators have the most important influence on whether the 'neither option' response is chosen. For example, in the option on the third line, which offers improved escalators, a lower proportion of 'neither' responses is obtained than for the option on the fourth line, where improved escalators are not offered, even though the latter option is less expensive and offers better train exteriors.

Table 9.37 Frequencies With Which the "Neither" Response Was Selected in the Station/Train Exterior Exercise

Fare	Draughts	Exteriors	Escalators	Graffiti	% stating "Neither"
Medium	Good	Poor	Good	Poor	5%
Good	Poor	Poor	Poor	Poor	13%
Poor	Poor	Poor	Good	Good	13%
Medium	Poor	Good	Poor	Good	15%
Poor	Poor	Good	Good	Poor	15%
Poor	Good	Poor	Poor	Good	23%
Poor	Good	Good	Poor	Poor	26%

This supports the importance given to escalators in the models above. As can be seen from Table 9.38, the package fare values are high and in many cases exceed those obtained for rolling stock attributes. As already discussed, the station attributes would not be presented in the 'between-mode' exercises and these values would be scaled using those factors derived for rolling stock attributes. The drawbacks of this approach are discussed towards the end of the next chapter.

9.6 Summary

This chapter has provided a summary of the sample characteristics, followed by a detailed account of the analysis of the 'within-mode' and station attributes SP exercises. The intention has been to report, step by step, each procedure undertaken in the analysis. This will allow future researchers to understand the processes involved in the analysis of these types of SP techniques and make an informed assessment for themselves of the validity of the results.

Sample Characteristics (Sections 9.2 to 9.4)

The sample appeared representative when compared with other profile information on Northern Line users, though period Travelcards were slightly under-represented and students over-represented. A number of respondent characteristics were identified as likely to have an influence on the service attribute values that would be derived from the SP data. These were: frequency of travel, use of other Underground lines, fare paid, perceptions of current travel conditions (delays, crowding, seat availability) and perceptions of service quality attributes (air quality, noise, etc). Some of these would be referred to in the later interpretaion of the SP model results.

Table 9.38 Summary of Weighted Fare Valuations Estimated From the "Within-Mode" Station Facilities Exercise: Artificial Data Attached and 10% Fare Coefficient Removed

Package Fare values (90% level of confidence)

	Peak T'card	Peak Ordinary	Off-Peak T'card	Off-Peak One Day T'card	Off-Peak Ordinary
Fare:	**30%**	**30%**	**30%**	**30%**	**30%**
All items	88%	110%	47%	133%	93%

Package Fare values (95% level of confidence)

All Items	88%	97%	47%	133%	93%

Analysis of SP data (Sections 9.4 and 9.5)

In the preparation of the SP data for analysis the first stage was to remove incomplete and "test" interviews and merge the responses to the SP experiments with the information from the non-SP parts of the questionnaire. This was followed by the application of weights to each observation, to adjust for the different numbers of responses given by each individual in each adaptive ranking SP exercise. Through this process, each individual would be equally represented. As the analysis progessed, it was also found necessary to weight the data so that each design "block" used in the SP experiment was also equally represented.

Before statistical models were estimated from the data, a simpler analysis of the responses was conducted, which allowed some understanding of respondents' priorities in advance of the more complex analysis. Respondents were clearly trading between service attributes and fare, though there appeared evidence of inconsistent choice behaviour.

Valuation of Rolling Stock and Station Attributes

Binary logit models were calibrated on the SP data. Over the course of a number of treatments, it was found that the plausibility and statistical robustness of the results was noticeably enhanced not only by the weighting of the data, but also by the removal of respondents making economically non-rational responses, by aggregation of the two rolling stock SP data sets and by aggregation across the ticket/time of day segments.

Intermediate improvements were not generally valued very highly, though cleanliness and air quality were given consistently strong values at both intermediate and maximum improvements, across all segments. Ride quality, noise reduction and information links with the driver were more valued (at the maximum levels) by travellers in the off-peak periods than in the peak. Newness and inter-car gangways received little value and no strong interaction effects were observed between the attributes.

The station attributes SP exercise proved harder to analyse, despite it being a simpler exercise. Initial models were statistically weak and produced implausible parameter values. This was first thought to reflect omission of a dominant option in the SP experimental design. However, the insertion of artificial observations to correct this did not greatly improve matters. The main improvements came from removing one of the fare variables, specifically the weaker 10% fare change. The reason for this is that fare dominated the choices, leading to a large number of "choose neither" responses. Once plausible model results were finally obtained, all the station service improvements were seen to have some value, but escalators were the most important.

After a range of developments in the treatment of the data had been applied, it is argued that robust valuations of improvements were obtainable from these SP exercises. Nevertheless, it would appear that the results that emerge prior to the analysis of the "between-mode" SP exercises are, when expressed in terms of equivalent percentage fare values, implausibly high.

10: THE EVALUATION OF IMPROVEMENTS TO PASSENGER FACILITIES ON LONDON'S UNDERGROUND: ANALYSIS OF 'BETWEEN MODE' SP EXERCISES

10.1 Initial Observations

In the 'between-mode' SP exercises, respondents were asked to consider choices between their current Underground journey and their best alternative mode of transport. Throughout each exercise, the characteristics of the alternative mode remained as the respondent described them. Only the attributes of the Underground service were altered. In a more comprehensive mode choice study, variations to both modes would have been of interest, as would interviews with non-users of the Underground. For this study, the concern was only to see how valuations of rolling stock improvements might be obtained in a realistic mode choice context, as opposed to the more hypothetical circumstances of the 'within-mode' exercises. The packages of facilities presented in each SP exercise corresponded to the groupings of intermediate and maximum levels of rolling stock improvements presented to them in the 'within-mode' exercises. The improvements from the first 'within-mode' rolling stock exercise to be presented to them were used. To avoid respondents concentrating exclusively on the facilities/fare trade-offs, journey times and service frequency were also altered.

Table 10.1 shows the proportion of choices in favour of the Underground over the sixteen options in each SP design (each respondent was only presented with half of these.) Cost is clearly an important variable in both exercises, whereas travel time is not. Improved facilities and better service frequency appear to have some value, but most noticeably when fare is +15% or less. The ordering of the options on the basis of choice frequencies is very similar for the two SP exercises, with the exception of one or two options. These variations suggest that fare is more important in the second exercise (where rolling stock improvements correspond to those presented in the second of the two 'within-mode' rolling stock SP exercises) and service frequencies and rolling stock improvements are less important.

Table 10.1 Proportions of Respondents Loyal to London Underground in the "Between-Mode" SP Exercises (by Time/Ticket Segments)

Exercise 1 Improvements — **Segments (% Choosing LUL)**

Headway	Stock	Cost	Time	Peak Tcd	Peak Ody	Off Peak Tcd	One Day Tcd	Off Peak Ody	Average
Good	Good	+0%	+5%	90	86	92	90	80	88
Medium	Good	+5%	+0%	86	87	73	87	84	83
Medium	Poor	+0%	+15%	72	72	65	45	62	63
Poor	Medium	+5%	+15%	55	65	50	67	61	60
Poor	Good	+15%	+40%	57	74	50	60	46	57
Medium	Medium	+15%	+5%	65	62	53	49	57	57
Poor	Medium	+0%	+40%	62	52	67	50	43	55
Poor	Poor	+0%	+0%	67	61	53	37	34	50
Poor	Poor	+5%	+5%	46	55	37	47	46	46
Poor	Good	+40%	+15%	57	53	39	37	41	45
Good	Poor	+5%	+40%	63	48	41	21	48	44
Good	Poor	+15%	+15%	59	52	36	37	29	43
Poor	Poor	+15%	+0%	41	55	30	42	31	40
Good	Medium	+40%	+0%	44	46	22	27	34	35
Poor	Poor	+40%	+5%	27	28	10	33	25	25
Medium	Poor	+40%	+40%	30	29	10	22	27	23

Exercise 2 Improvements

Headway	Stock	Cost	Time	Peak Tcd	Peak Ody	Off Peak Tcd	One Day Tcd	Off Peak Ody	Average
Good	Good	+0%	+5%	83	80	80	90	91	85
Medium	Good	+5%	+0%	77	80	71	69	87	77
Medium	Poor	+0%	+15%	80	59	64	69	70	68
Poor	Medium	+0%	+40%	74	59	55	76	66	66
Poor	Medium	+5%	+15%	75	76	48	71	57	65
Medium	Medium	+15%	+5%	69	69	37	64	72	62
Good	Poor	+5%	+40%	58	39	52	59	70	56
Poor	Poor	+0%	+0%	76	36	43	66	57	56
Poor	Good	+15%	+40%	55	65	35	54	57	53
Good	Poor	+15%	+15%	58	37	47	62	60	53
Poor	Poor	+5%	+5%	51	44	39	47	58	48
Poor	Good	+40%	+15%	30	42	32	48	59	42
Good	Medium	+40%	+0%	30	44	25	39	66	41
Poor	Poor	+15%	+0%	47	44	25	42	45	41
Poor	Poor	+40%	+5%	15	25	25	16	38	24
Medium	Poor	+40%	+40%	21	28	18	13	38	24

An important observation is that a fifth of respondents remain with the Underground even when the options are at their worst and a further tenth choose the alternative mode even when the best options are presented. This demonstrates the presence of 'non-traders' (individuals who stay with the same alternative, Underground or competing mode, throughout the exercise.) With regard to those who chose the alternative mode at the best Underground option, the implication is that even a 5% increase in journey time is enough to persuade them to leave the Underground. This contrasts with the impression that most travellers are insensitive to the time increases. Another possibility is that some of these respondents are in the habit of using different modes on different days for reasons not investigated in the SP exercises, or had used the Underground on the day of the survey for exceptional reasons.

10.2 Statistical Analysis

As noted for the 'within-mode' exercises, the data would need to be weighted to account for imbalances in the distribution of the design blocks. Table 10.2 shows the division of the two blocks in each SP exercise for each segment. While the variation is generally small, the difference between the proportions presented with each block is noticeably large for off-peak ordinary ticket users in the first exercise. The second weighting strategy adopted for the 'within-mode' exercises, namely the use of an average number of responses per person, was not relevant here, because the design was not adaptive and all respondents completed eight choice options.

As before, logit models would be estimated from the data to infer importance weightings for each attribute in the SP exercises. As fare non-linearities were of interest, separate coefficients were used for each level of fare change. Single coefficients were used for service headway times and travel time because these attributes were of less interest. As discrete variables, the two levels of rolling stock improvements were given separate coefficients.

Table 10.2 Proportions of Respondents Being Presented With Each Design 'Block' in Each "Between-Mode" Rolling Stock Exercise: All Responses

	Peak T'card	Peak Ordinary	Off Peak T'card	Off Peak One Day T'card	Off Peak Ordinary
Exercise 1					
Block 1:	56.7%	48.4%	55.0%	50.9%	37.2%
Block 2:	43.3%	51.6%	45.0%	49.1%	62.8%
Exercise 2					
Block 1:	52.4%	55.9%	50.6%	45.0%	51.6%
Block 2:	47.6%	44.1%	49.4%	55.0%	49.4%

The form of the logit models was as follows:

$$P_l = \frac{1}{1 + \exp^{(Ua-Ul)}}$$

where: P_l = Probability of choosing Underground
U_l = Utility of Underground
U_a = Utility of alternative mode

Interactions could not be examined under the design used, so main effects only were included as components of the utility functions:

$$U_l = a_1 + a_2{*}Fare(5\%) + a_3{*}Fare(15\%) + a_4{*}Fare(40\%) + a_5{*}stock1 + a_6{*}stock2 + a_7{*}headway + a^8{*}journey\ time$$

$$U_r = 0$$

where:

Fare(X%) = fare level, indicating percentage increase;

stock1 = intermediate package of rolling stock improvements;

stock2 = maximum package of rolling stock improvements;

headway = continuous value of headway, in minute increases;

j. time = continuous value of journey time, in percentage increases;

$a_1 \ldots a_8$ = coefficients.

In this model, the basic bias of travellers for or against the Underground service would be indicated by a positive or negative value for the constant a^1. Headway was coded in minutes because it did not progress in consistent percentage values like the other continuous variables. Respondents had the opportunity to reject both options (ie choose a third mode or not travel), on the basis that they should not be restricted only to their best alternative, but could have the opportunity to change their minds and go by some other way, if Underground was rejected. This opportunity to choose a third option was taken up in less than 10% of all cases and such responses were omitted from the models. The standard errors of the coefficient estimates were scaled as in the earlier models to adjust for repeated observations.

The same segments were used as for the 'within-mode' models and the results of the first set of model runs are summarised in Table 10.3 and Table 10.4. The format of these tables, as it is for subsequent tables relating to models derived from the 'between-mode' SP data, is as follows:

(i) Model coefficients, t-Statistics, rho bar statistics and observations/sample sizes for each segment (see the 'within-mode' section for a description of these items.) In these models, rolling stock improvements were coded in such a way that an improvement would have a positive coefficient (the opposite convention had been used in the 'within-mode' exercises.) All other attributes would be expected to be negative, with the exception of the constant, which could be negative or positive.

(ii) Fare values, derived from the division of each coefficient by one of the fare coefficients. For clarity, only values against the 15% and 40% increases are shown, as in most cases the 5% fare increase coefficients are weak and non-significant.

(iii) Comparisons with the 'within-mode' package values, significant at 90% level of confidence (unless otherwise specified.) These are taken from the models from which illogical responses were removed. The term 'medium' refers to packages of intermediate levels of improvement, the term 'maximum' to packages of the highest levels of improvement. Each one has a value relating to the earlier 10% and 30% fare changes. Below these are shown scaling factors which relate these figures to the 15% and 40% fare values for the matching stock coefficients derived from the 'between-mode' exercise. A question mark occurs where a significant value has been obtained from the 'between-mode' models but not from the 'within-mode' models (ie a division of zero.)

Table 10.3 Summary of Logit Models Estimated From the First "Between-Mode" Rolling Stock Exercise: All Responses

Model Results

Attribute	Peak T'card Coeff	t	Peak Ordinary Coeff	t	Off Peak T'card Coeff	t	Off Peak One Day T'card Coeff	t	Off Peak Ordinary Coeff	t
Constant	1.53	3.7	1.09	2.4	0.84	1.6 !	1.04	2.4	1.02	1.8 ?
Fare5	-0.53	-1.4 !	-0.26	-0.6 !	-0.86	-1.7 ?	-0.21	-0.5 !	0.20	0.4 !
Fare15	-0.71	-1.8 ?	-0.44	-1.0 !	-1.28	-2.6	-0.55	-1.3 !	-0.73	-1.4 !
Fare40	-1.53	-3.9	-1.35	-2.9	-2.47	-4.3	-1.41	-3.2	-1.20	-2.2
Stock1	0.28	0.9 !	0.37	1.0 !	0.75	1.7 ?	0.68	2.0	0.54	1.2 !
Stock2	1.24	3.4	1.27	3.0	1.43	3.1	1.72	4.2	1.29	2.6
Time	-0.01	-0.9 !	-0.01	-1.4 !	-0.01	-1.0 !	-0.02	-2.1	-0.01	-0.8 !
Headway	-0.05	-3.1	-0.02	-1.5 !	-0.02	-0.8 !	-0.05	-3.1	-0.06	-3.0
Obs	531		368		302		440		269	
Sample	68		48		40		60		34	
Rho bar sq	0.13		0.11		0.16		0.14		0.13	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

Fare:	Peak T'card 15%	40%	Peak Ordinary 15%	40%	Off Peak T'card 15%	40%	Off-Peak One Day T'card 15%	40%	Off Peak Ordinary 15%	40%
Constant	-32 !	-40	-37 !	-32 ?	-10 !	-14 !	-29 !	-30 ?	-21 !	-34 !
Fare5	2.2 ?	2.7	1.8 !	1.5	2.0	2.8	1.2 !	1.2	-0.8 !	-1.3 ?
Fare15	n/a	1.2	n/a	0.9 ?	n/a	1.4	n/a	1.0	n/a	1.6 ?
Fare40	0.8 !	n/a	1.2 !	n/a	0.7 !	n/a	1.0 !	n/a	0.6 !	n/a
Stock1	-6 !	-7 !	-13 !	-11 !	-9 !	-12 !	-19 !	-19 ?	-11 !	-18 !
Stock2	-26 !	-32	-44 !	-38	-17	-23	-47 !	-49	-27 !	-43 ?
Time	0 !	0 !	1 !	0 !	0 !	0 !	1 !	1 ?	0 !	0 !
Headway	1 !	1	1 !	1 !	0 !	0 !	1 !	1	1 !	2 ?
Within Mode Values										
Medium	16	14	0	0	0	0	49	59	0	0
Maximum	61	53	66	103	47	44	81	97	0	101
Mdm Ratio	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00
Max Ratio	0.00	0.61	0.00	0.36	0.36	0.52	0.00	0.50	0.00	0.43

Totals use figures significant at the 90% level of confidence.

Table 10.4 Summary of Logit Models Estimated From the Second "Between-Mode" Rolling Stock Exercise: All Responses

Model Results

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
Attribute	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**
Constant	2.32	5.1	0.96	2.2	0.84	1.8 ?	1.88	4.4	2.00	4.3
Fare5	-0.78	-1.8 ?	0.00	0.0 !	-0.50	-1.1 !	-0.83	-2.0	-0.25	-0.6 !
Fare15	-1.11	-2.6	-0.27	-0.6 !	-1.20	-2.6	-1.12	-2.8	-0.70	-1.6 !
Fare40	-2.70	-5.9	-1.26	-2.8	-1.77	-3.7	-2.34	-5.5	-1.17	-2.6
Stock1	0.67	1.9 ?	1.09	2.9	0.12	0.3 !	0.94	2.8	0.65	1.8 ?
Stock2	0.70	1.9 ?	1.25	3.1	0.80	2.0	1.01	2.9	1.00	2.5
Time	-0.01	-1.1 !	-0.01	-0.9 !	0.00	-0.2 !	-0.01	-0.9 !	-0.01	-1.1 !
Headway	-0.07	-4.1	-0.06	-3.9	-0.02	-1.3 !	-0.06	-3.9	-0.08	-4.9
Obs	502		401		348		538		448	
Sample	64		52		46		69		57	
Rho bar sq	0.22		0.15		0.10		0.18		0.17	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off-Peak One Day T'card		Off Peak Ordinary	
Fare:	**15%**	**40%**	**15%**	**40%**	**15%**	**40%**	**15%**	**40%**	**15%**	**40%**
Constant	-31	-34	-53 !	-31 ?	-11 !	-19 !	-25	-32	-43 !	-69
Fare5	2.1	2.3	0.0 !	0.0 !	1.3	2.3	2.2	2.8	1.1 !	1.7
Fare15	n/a	1.1	n/a	0.6 !	n/a	1.8	n/a	1.3	n/a	1.6
Fare40	0.9 ?	n/a	1.7 !	n/a	0.6 !	n/a	0.8 ?	n/a	0.6 !	n/a
Stock1	-9 !	-10 ?	-60 !	-35	-2 !	-3 !	-13	-16	-14 !	-22 !
Stock2	-9 !	-10 ?	-69 !	-40	-10 !	-18 ?	-14	-17	-21 !	-34 ?
Time	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 !
Headway	1	1	3 !	2	0 !	0 !	1	1	2 !	3
Within Mode Values										
Medium	0	0	0	0	0	0	0	8	0	0
Maximum	24	34	77	79	30	34	48	70	38	98
Mdm Ratio	0.00	*	0.00	*	0.00	0.00	*	2.07	0.00	0.00
Max Ratio	0.00	0.31	0.00	0.50	0.00	0.52	0.28	0.25	0.00	0.35

Totals use figures significant at the 90% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

The impression from the first two summary tables is that the initial model runs have produced poor results, in terms of the following:

(i) The rho bar statistics are weak in almost all cases.

(ii) Few coefficients are significant even at the 90% level of confidence.

It is important to note that most coefficients are of the expected sign and that the sample sizes in almost all cases are unacceptably small, given the general guidance figure of 75 or more respondents per segment. In some cases, it might be reasonable to expect weak and non-significant coefficients (eg for the lowest fare change), but as these results stand it does not appear that reliable conclusions can be drawn from them.

Despite these problems, there are some early indications of different constraints acting on different groups of respondents and a scaling effect on the valuations of the rolling stock improvements. The fare values tables (also in Tables 10.3 and 10.4) show that some segments have a stronger attachment to the Underground than others (indicated by the size and sign of the constant.) Large positive constants (which have a negative value when divided by fare) can be observed consistently across the two exercises for all travellers, though in some cases they are non-significant.

Concerning a scaling effect on the value of the rolling stock improvements, the fare tables also show that when fare values significant at the 90% level are used, the corresponding values from the 'within-mode' exercises are considerably higher. The scaling effects on these 'within-mode' values are in the order of 0.3 to 0.6. The only exceptions to this observation are the results for the package of intermediate improvements introduced in the second SP exercise to peak travellers and off-peak one-day Travelcard users. Here, the values are actually higher in the 'between-mode' exercise than in the 'within-mode' exercise, where the latter are sometimes not significantly different from zero at the 90% level of confidence.

These latter observations highlight a major difference in the values obtained from the second 'within-mode' exercise and the associated 'between-mode' exercise. In the models derived from the former exercise, the intermediate levels for each item of rolling stock improvement were all weak and nearly all non-significant. In contrast, the packages of intermediate improvements in the latter exercise are stronger and in most cases almost the same value as for the package of maximum improvements. It would appear that in these instances the values of the packages of these particular attributes remain fairly consistent, while the values of the individual components within the package can differ markedly. This contrasts with the attributes in the first 'within-mode' and 'between-mode' exercises, where the variation between intermediate and maximum levels is more consistent.

The progression of unit values for each level of fare increase are in most cases non-linear. Nevertheless, the degree of variability of the fare/fare ratios over the two exercises, together with the low levels of confidence that can be attached to them suggest that no firm conclusions can be drawn about the nature of the fare variable at this stage. It is worth noting that in a number of instances the 5% fare level, though often non-significant, has a greater weight per unit than either of the higher levels.

Given the weaknesses of these first models, a number of strategies could be adopted to attempt to improve them. Two strategies used for the 'within-mode' exercises could be adopted here also, together with a third approach relevant only to 'fixed choice' SP exercises (as opposed to ranking/adaptive ranking exercises.) These were:

(i) The aggregation of segments by ticket type.

(ii) The combination of the two exercises in a common model.

(iii) The removal of 'non-traders'.

The latter approach refers to the removal of those respondents who throughout the exercise chose the same alternative every time. The results of each of these approaches is discussed below. A fourth approach, used with the 'within-mode' mode exercises, would be to remove economically irrational responses. For these exercises, this was rejected on the following basis:

(i) Such responses are difficult to identify - such individuals could only be identified when, for example, an individual chose the Underground when headway was poor, rolling stock was poor, fare was +40% and time +5%, but chose his or her alternative mode when headway was good, rolling stock was good, fare was +0% and time +5%. Such obvious inconsistencies are rare, while few other dominating options with more subtle differences exist in the design.

(ii) In the context of mode choice, some of the error in responses may be a faithful reflection of the error and uncertainty that affects real mode choices and the removal of economically irrational responses may reduce the realism of the data.

Increasing the Level of Data Aggregation

As noted for the earlier analysis of the 'within-mode' exercises, aggregation of the segments had no direct relevance to the NLMT forecasting model but would allow more representative results to be obtained from the data. Table 10.5 summarises the results for models fitted to segments aggregated by ticket type. The sample sizes are now all in excess of 100 respondents and the benefit can be seen in terms of the larger t-Statistics for a number of coefficients. The larger sample sizes invest more confidence in the conclusions that may now be drawn.

Table 10.5 Summary of Logit Models Estimated From Both "Between-Mode" Rolling Stock Exercises: Segments Aggregated

Model Results

	First Exercise					Second Exercise			
	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
Attribute	Coeff	t	Coeff	t		Coeff	t	Coeff	t
Constant	1.28	4.0	1.04	3.8	Constant	1.68	5.2	1.63	6.5
Fare5	-0.65	-2.1	-0.10	-0.4 !	Fare5	-0.64	-2.1	-0.40	-1.6 !
Fare15	-0.91	-3.0	-0.53	-2.0	Fare15	-1.12	-3.7	-0.73	-3.0
Fare40	-1.82	-5.8	-1.32	-4.8	Fare40	-2.27	-6.9	-1.62	-6.5
Stock1	0.41	1.6 ?	0.53	2.4	Stock1	0.41	1.6 !	0.87	4.2
Stock2	1.24	4.5	1.46	5.8	Stock2	0.73	2.7	1.05	4.8
Time	-0.01	-1.3 !	-0.02	-2.6	Time	-0.01	-1.0 !	-0.01	-1.4 !
Headway	-0.04	-3.1	-0.04	-4.5	Headway	-0.05	-4.1	-0.06	-7.4
Obs	833		1,077		Obs	850		1,387	
Sample	108		142		Sample	110		178	
Rho bar sq	0.12		0.12		Rho bar sq	0.14		0.16	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
Fare:	15%	40%	15%	40%		15%	40%	15%	40%
Constant	-21	-28	-30 ?	-32	Constant	-22	-30	-34	-40
Fare5	2.1	2.8	0.6 !	0.6	Fare5	1.7	2.2	1.6	2.0
Fare15	n/a	1.3	n/a	1.1	Fare15	n/a	1.3	n/a	1.2
Fare40	0.7 ?	n/a	0.9 !	n/a	Fare40	0.8	n/a	0.8 ?	n/a
Stock1	-7 !	-9 !	-15 !	-16	Stock1	-6 !	-7 !	-18	-21
Stock2	-20	-27	-42 ?	-44	Stock2	-10	-13	-22	-26
Time	0 !	0 !	0 !	0	Time	0 !	0 !	0 !	0 !
Headway	1	1	1 ?	1	Headway	1	1	1	2
Within Mode Values									
Medium	11	10	41	54	Medium	0	0	5	7
Maximum	57	51	89	103	Maximum	38	49	62	79
Mdm Ratio	0.00	0.00	0.00	0.29	Mdm Ratio	0.00	0.00	3.52	3.28
Max Ratio	0.36	0.54	0.47	0.43	Max Ratio	0.25	0.26	0.35	0.33

Totals use figures significant at the 95% level of confidence

In these models, the positive and significant constants indicate strong bias in favour of the Underground service and are fairly consistent over the two exercises. It would at first appear that ordinary and one-day Travelcard users have a stronger allegiance to the Underground, due to the higher fare values for the constants. However, the fare coefficients are weaker for these groups and the difference in the constants is not so large.

All the fare levels are significant at the 90% level of confidence, with the exception of the coefficient for the 5% fare level for ordinary/one-day Travelcard users in the both exercises (the 5% fare level in the second exercise only just fails the 90% threshold.) The 'stock' coefficients representing levels of improved passenger facilities are significant at the 90% level of confidence, with the exception of period ticket users in the second exercise, though this is only just non-significant.

The headway coefficients are strongly significant, while travel time remains small and non-significant for three of the four groups. As before, the rho bar statistics remain weak. Using the fare value table to comment on this group of models, the following observations can be made:

(i) The weightings for fare only appear to progress in a linear manner for ordinary/one-day Travelcard users in the first exercise, and that only above the 5% fare level. For the other groups, the unit value of fare still reduces noticeably as fare increases.

(ii) The values for the packages of rolling stock improvements vary in line with those observed in the 'within-mode' exercises, being subjected to scaling effects in the region of 0.3 to 0.5 at the levels of maximum improvement. A notable exception, observed for the preceding models, is the package of intermediate improvements for ordinary/one-day Travelcard users in the second exercise, for which the scaling factors are much larger than parity.

(iii) The very weak value of travel time appears to be the result of the sample being a choice-based one. When asked in the interview why they had chosen the Underground over other modes, the most frequent reason was faster journey time. It would appear that for many respondents, increases of even 40% still resulted in favourable travel times for the Underground.

The last point is confirmed by average reported Underground journeys of 25 minutes against estimated alternative average journey times of 53 minutes. In addition, over 50% stated that they could afford to arrive up to twenty minutes later than their actual arrival time.

The second approach to aggregation was to combine the data from the two 'between-mode' exercises while maintaining the earlier segments. To ensure that both sets of data had an equal influence on the model results, appropriate weights were used. The models would have a similar structure to the previous models, with the addition of extra 'stock' coefficients to represent separately the packages from each exercise. The coefficients stock11 and stock12 refer to the intermediate and maximum levels of improvement from the first exercise, stock21 and stock22 the same from the second exercise.

The model results are summarised in Table 10.6. Each segment now represents a more reliable sample size and as a result the t-Statistics improve appreciably for most coefficients, when compared with the first 'between-mode' models. In the accompanying fare value table, there are non-linear progressions of the unit value of each fare level for most groups. Period Travelcard users place a high unit value on the 5% fare change, but this declines with larger fare changes, as observed in the previous aggregate models. One-day Travelcard users show a similar trend, but ordinary ticket users behave quite differently. For these respondents, the 5% fare change is less important, particularly in the off-peak, while the larger fare changes vary quite markedly between time periods. Peak ordinary ticket users place a higher unit value on the 40% fare increase, while off-peak ordinary ticket users do the opposite.

Table 10.6 Summary of Logit Models Estimated From Combined "Between-Mode" Rolling Stock Exercises: All Responses

Model Results

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
Attribute	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**	**Coeff**	**t**
Constant	1.89	6.2	0.98	3.2	0.82	2.4	1.40	4.6	1.52	4.3
Fare5	-0.64	-2.2	-0.09	-0.3 !	-0.64	-1.9 ?	-0.49	-1.7 ?	-0.01	0.0 !
Fare15	-0.89	-3.1	-0.31	-1.0 !	-1.22	-3.6	-0.80	-2.8	-0.70	-2.1
Fare40	-2.05	-6.8	-1.27	-3.9	-2.05	-5.6	-1.84	-6.0	-1.15	-3.4
Stock11	0.26	0.9 !	0.62	1.7 ?	0.58	1.4 !	0.42	1.5 !	0.17	0.6 !
Stock21	1.27	3.7	1.59	3.7	1.26	2.9	1.39	4.2	0.76	2.5
Stock12	0.67	2.0	0.82	2.6	0.24	0.7 !	1.22	3.5	1.14	2.5
Stock22	0.66	2.0	1.00	3.0	0.88	2.6	1.30	3.5	1.57	3.2
Time	-0.01	-1.5 !	-0.01	-1.4 !	-0.01	-0.7 !	-0.01	-2.0	-0.01	-1.3 !
Headway	-0.06	-5.0	-0.05	-4.1	-0.02	-1.5 !	-0.05	-4.9	-0.07	-5.5
Obs	1,033		769		650		978		717	
Sample	132		100		86		129		91	
Rho bar sq	0.16		0.12		0.12		0.15		0.15	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off-Peak One Day T'card		Off Peak Ordinary	
Fare:	**15%**	**40%**	**15%**	**40%**	**15%**	**40%**	**15%**	**40%**	**15%**	**40%**
Constant	-32	-37	-47 !	-31	-10	-16 !	-26	-30	-33 ?	-53 !
Fare5	2.1	2.5	0.9 !	0.6 ?	1.6	2.5	1.8	2.1	0.0 !	0.1 !
Fare15	n/a	1.2	n/a	0.7	n/a	1.6	n/a	1.2	n/a	1.6
Fare40	0.9	n/a	1.5 !	n/a	0.6 ?	n/a	0.9 ?	n/a	0.6 !	n/a
Stock11	-4 !	-5 !	-30 !	-20 !	-7 !	-11 !	-8 !	-9 !	-4 !	-6 !
Stock21	-21	-25	-76 !	-50	-15	-25	-26	-30	-16 !	-26
Stock12	-11 ?	-13 ?	-39 !	-26	-3 !	-5 !	-23	-26	-24 !	-39
Stock22	-11 ?	-13 ?	-48 !	-32	-11	-17	-24	-28	-34 ?	-54
Time	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 ?	0 !	0 !
Headway	1	1	2 !	1	0 !	0 !	1	1	2 ?	3

Table 10.7 compares the significant valuations of the stock coefficients with those obtained from the 'within-mode' exercises. In the first set of comparisons, fare values significant at the 90% level of confidence are compared with those from the models derived separately from each 'between-mode' exercise. This lower level of confidence is used to take account of the smaller sample sizes in these earlier models. Scaling factors largely in the range of 0.3 to 0.5 can again be observed, together with the factor of over 3.0 for the intermediate package in the one-day Travelcard group (similarly large factors for peak travellers are also implied by the asterisks, which indicate a significant value from the 'between-mode' exercises compared against a non-significant value from the 'within-mode' exercises.)

In the second set of comparisons in Table 10.7, a higher level of confidence (95%) is used, to reflect the larger sample sizes of the combined 'within-mode' exercises. Less significant values are to be seen for some groups (especially off-peak ordinary ticket users) but for those scaling factors that can still be calculated, the values are generally similar to those observed in the first set of comparisons.

One final aggregation is to combine the two SP exercises and aggregate the segments by ticket type. A summary of models fitted to this aggregated data is given in Table 10.8. Earlier characteristics are repeated with greater significance, with a declining unit value for fare as the levels increase, weak values for the stock11 variable, a coefficient for stock21 that is almost as large as stock22 (for ordinary/Travelcard users) and very weak time coefficients.

Comparisons between the package values from Table 10.8 and corresponding values from the 'within-mode' exercises are made in Table 10.9, and show similar scaling values to those seen earlier. It is worth noting the scaling values for the period users' valuations of the Medium1 package (corresponding to Stock11), which are higher than the general range of 0.3 to 0.5 observed for other packages (values for period users' Medium1 package were previously non- significant.)

Table 10.7 Comparisons of Package Fare Values Estimated From Combined "Between-Mode" Rolling Stock Exercises With Comparable "Within-Mode" Exercises: All Responses

Within Mode Values (Separate Exercises)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off-Peak One Day T'card		Off Peak Ordinary	
Fare:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Medium1	16	14	0	0	0	0	49	59	0	0
Maximum1	61	53	66	103	47	44	81	97	0	101
Medium2	0	0	0	0	0	0	0	8	0	0
Maximum2	24	34	77	79	30	34	48	70	38	98
Mdm1 Ratio	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max1 Ratio	0.35	0.46	0.00	0.49	0.33	0.55	0.32	0.31	0.00	0.26
Mdm2 Ratio	*	*	0.00	*	0.00	0.00	*	3.41	0.00	*
Max2 Ratio	0.46	0.38	0.00	0.40	0.36	0.50	0.51	0.41	0.89	0.55

Totals use figures significant at the 90% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Within Mode Values (Combined Exercises)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off-Peak One Day T'card		Off Peak Ordinary	
Fare:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Medium1	11	13	0	0	0	0	36	48	0	0
Maximum1	48	54	69	90	41	44	58	78	0	89
Medium2	0	0	0	0	0	0	6	8	0	0
Maximum2	27	30	53	91	20	33	58	79	0	79
Mdm1 Ratio	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max1 Ratio	0.44	0.46	0.00	0.56	0.38	0.56	0.45	0.39	0.00	0.29
Mdm2 Ratio	0.00	0.00	0.00	*	0.00	0.00	3.93	3.35	0.00	*
Max2 Ratio	0.41	0.43	0.00	0.35	0.54	0.52	0.42	0.36	0.00	0.69

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Table 10.8 Summary of Logit Models Estimated From Combined "Between-Mode" Rolling Stock Exercises: Segments Aggregated

Model Results

All Responses

Attribute	Period T'card Coeff	t	Ordinary & One Day T'card Coeff	t
Constant	1.47	6.5	1.35	7.3
Fare5	-0.64	-2.9	-0.26	-1.4 !
Fare15	-1.02	-4.7	-0.63	-3.5
Fare40	-2.04	-8.9	-1.48	-8.0
Stock11	0.36	1.6 !	0.41	1.9 ?
Stock21	1.19	4.7	1.41	5.6
Stock12	0.47	2.0 ?	0.90	5.1
Stock22	0.78	3.1	1.08	5.7
Time	-0.01	-1.6 !	-0.01	-2.6
Headway	-0.04	-5.0	-0.06	-8.6
Obs	1,683		2,464	
Sample	228		320	
Rho bar sq	0.13		0.14	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

Fare:	Period T'card 15%	40%	Ordinary & One Day T'card 15%	40%
Constant	-22	-29	-32	-37
Fare5	1.9	2.5	1.3	1.4
Fare15	n/a	1.3	n/a	1.1
Fare40	0.7	n/a	0.9	n/a
Stock11	-5 !	-7 !	-10 ?	-11 ?
Stock21	-18	-23	-34	-38
Stock12	-7 ?	-9 ?	-21	-24
Stock22	-12	-15	-26	-29
Time	0 !	0 !	0	0
Headway	1	1	1	2

Table 10.9 Comparisons of Package Fare Values Estimated From Combined "Between-Mode" Rolling Stock Exercises With Comparable "Within-Mode" Exercises: Segments Aggregated

Within Mode Values (Separate Exercises)

	Period T'card		Ordinary & One Day T'card	
Fare:	15%	40%	15%	40%
Medium1	11	10	34	47
Maximum1	57	51	89	103
Medium2	0	0	0	7
Maximum2	30	38	62	79
Mdm1 Ratio	0.00	0.00	0.28	0.23
Max1 Ratio	0.31	0.46	0.38	0.37
Mdm2 Ratio	0.00	0.00	*	3.72
Max2 Ratio	0.39	0.40	0.42	0.37

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Within Mode Values (Combined Exercises)

	Period T'card		Ordinary & One Day T'card	
Fare:	15%	40%	15%	40%
Medium1	8	9	34	42
Maximum1	47	51	70	86
Medium2	0	0	5	6
Maximum2	34	36	73	90
Mdm1 Ratio	0.65	0.80	0.28	0.26
Max1 Ratio	0.37	0.46	0.48	0.44
Mdm2 Ratio	0.00	0.00	4.09	3.73
Max2 Ratio	0.34	0.42	0.35	0.32

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Removal of 'Non-Traders'

Because the sample was choice-based, it was likely that the Underground would remain attractive to a number of respondents, regardless of the fare, time and headway increases that were introduced. The large difference between average Underground journey times and those of the alternative modes has already been discussed as a principle reason for such 'non-trading' in the SP exercises. The presence of these 'non-traders' reduces the quality of the models that have been fitted to the SP data. This is because they assume compensatory choice behaviour explains all the SP responses, when for such 'non-traders' there is obviously no compensatory decision-making taking place. Wardman (1988 p78) suggests that a number of possible reasons may exists to explain why an individual does not switch modes, some of which are due to common survey problems (not understanding the nature of the task required, justification of existing travel behaviour; strategies aimed to influence policy decisions) and some to genuine forms of decision making (non-compensatory, 'lexicographic' and habit inertia.)

Table 10.10 summarises the proportions of 'non-traders' in each SP exercise and their allegiance to the Underground or their alternative mode. This table may be compared with Table 10.1, which earlier showed a breakdown of those proportions of respondents choosing the Underground in each option of the SP exercises. It can be seen that the majority of those not choosing the Underground even when at its most attractive and those choosing the Underground even at its most unattractive were non-traders.

Returning to Table 10.10, there are some differences between the segments, but more noticeably between SP exercises, within segments. Off-peak Travelcard users contain higher proportions of non-traders in the second SP exercise than in the first; off-peak ordinary ticket users show a balance of non-traders in the first SP exercise but a pronounced imbalance in the second SP exercise.

Table 10.10 Proportions of Respondents Being Presented With Each Design 'Block' in Each "Between-Mode" Rolling Stock Exercise: All Responses

All Responses Included (See Table 10.)

	Peak T'card	Peak Ordinary	Off Peak T'card	Off Peak One Day T'card	Off Peak Ordinary
Exercise 1					
Block 1:	56.7%	48.4%	55.0%	50.9%	37.2%
Block 2:	43.3%	51.6%	45.0%	49.1%	62.8%
Exercise 2					
Block 1:	52.4%	55.9%	50.6%	45.0%	51.6%
Block 2:	47.6%	44.1%	49.4%	55.0%	49.4%

Non-Trading Responses Removed

	Peak T'card	Peak Ordinary	Off Peak T'card	Off Peak One Day T'card	Off Peak Ordinary
Exercise 1					
Block 1:	66.5%	55.9%	59.0%	53.2%	39.4%
Block 2:	33.5%	44.1%	41.0%	46.8%	61.6%
Exercise 2					
Block 1:	52.6%	64.3%	56.8%	53.1%	64.9%
Block 2:	47.4%	35.7%	43.2%	46.9%	35.1%

As the only difference between the exercises is the type of rolling stock improvements presented, a superficial conclusion might be that the number of 'non-traders' is influenced by these items. However, a more convincing explanation for these variations may be found in the types of alternative modes identified by 'non-traders'. Table 10.11 shows that for the two segments showing the greatest variations in the proportions of 'non-traders', between the SP exercises, the differences in the types of alternative modes identified are also most pronounced. In each case, there is greater diversity of alternatives in the second exercise.

The sizeable mismatch among off-peak ordinary ticket users, between those 'non-traders' allied to the Underground and those allied to an alternative mode, could be a result of the high proportions of respondents with a non-vehicular alternative (walking or not travelling.) Further evidence of differences in the alternatives corresponding to varying amounts of 'non-traders' in each SP exercise is provided by Table 10.12. Here, journey cost and time differences between the Underground and the alternative modes contrast most between SP exercises in the two segments for which the most variation in the proportions of 'non-traders' was observed between the SP exercises. Off-peak ordinary ticket users have, on average, much less attractive modes in the second SP exercise compared with the first, corresponding with the much higher allegiance to the Underground in this case. Less obvious is why off-Peak Travelcard users have more 'non-traders' allied to their alternative modes in the second exercise than in the first, given that proportional differences in the characteristics are fairly similar. In absolute terms, the time difference is appreciably lower in the second exercise than in the first, but this observation is also true for other segments, for which variations in 'non-traders' between exercises are not so apparent.

These effects of variations in the alternative modes and their perceived characteristics relative to the Underground suggests that there is some advantage in aggregating the two SP exercises. In addition to increased sample sizes in each segment, as already observed, aggregation also reduces the likely impacts of imbalances of 'non-traders' on the model results.

Table 10.11 Non-Traders As Proportions of Each Segment, by Mode Allegiance and SP Exercise

Exercise	Mode Allegiance	Peak T'card	Peak Ordinary	Off-Peak T'card	Off-Peak One-day T'card	Off-Peak Ordinary
One	Underground	29%	29%	8%	20%	21%
	Alternative	10%	13%	8%	10%	21%
	(Sample)	(68)	(48)	(40)	(60)	(36)
Two	Underground	20%	23%	16%	23%	38%
	Alternative	11%	17%	20%	10%	5%
	(Sample)	(64)	(52)	(46)	(69)	(57)

Table 10.12 Type of Alternative Mode by SP Exercise

	Peak T'card		Peak Ordinary		Off-Peak T'card		Off-Peak One-Day T'card		Off-Peak Ordinary		TOTAL	
Alternative	**Ex 1**	**Ex 2**	**Ex 1**	**Ex 2**	**Ex 1**	**Ex 2**	**Ex 1**	**Ex 2**	**Ex 1**	**Ex 2**	**Ex 1**	**Ex 2**
Car Driver	10	11	13	25	0	7	5	9	6	7	7	11
Car Pas'ger	0	3	2	0	0	2	7	4	6	4	3	3
Bus	64	57	69	58	90	69	60	57	62	43	68	56
BR Train	7	14	10	0	0	9	10	7	0	4	6	7
Taxi	4	2	2	6	5	2	10	6	15	16	7	6
Walk	9	11	2	8	0	7	3	7	6	14	4	9
Not Travel	6	3	2	4	5	4	5	10	6	13	5	7
Total	100	100	100	100	100	100	100	100	100	100	100	100
Sample	(68)	(64)	(48)	(52)	(40)	(46)	(60)	(69)	(36)	(57)	(250)	(288)

Another issue to consider is whether the proportion of 'non-traders' varies across the two blocks used in the SP design. Table 10.13 compares the proportions for segments with 'non-traders' removed with those for all responses (reproduced for comparison from Table 10.2.) In general, more non-traders could be observed for the second blocks than for the first, reflecting the characteristics of the eight choice situations in each. Weighting of the data on this basis would therefore play an even more important role in the analysis of data from which 'non-traders' have been removed.

Evidence from other studies suggests that the effect of non-traders on the models is chiefly to weaken their goodness-of-fit and reduce the statistical significance of the coefficients. In cases where non-traders are largely allied to one mode (in this case chiefly the Underground), they are also likely to have a major influence on the size and direction of any mode-specific constants. The relative values of other coefficients would not be expected to be greatly affected, as these are derived directly from information on mode-switching (which does not occur, by definition, among non-traders.) Given the potential for improved models, the previous model runs were all repeated with the exclusion of 'non-traders'. The results can be seen over tables 10.14 to 10.20.

Table 10.13 Perceived Journey Characteristics of Underground Services and Alternative Modes by SP Exercise

Segment	SP Exercise	Underground Fare	Underground Time	Alternative Mode Cost	Alternative Mode Time	Differences (Absolute) Cost	Differences (Absolute) Time	(Percentage) Cost	(Percentage) Time
Peak	Ex 1	£3.21	29	£3.10	61	£0.11	-33	3%	-114%
Travelcard	Ex 2	£4.58	29	£3.76	55	£0.82	-26	18%	-90%
Peak Ordinary	Ex 1	£1.34	29	£1.46	48	-£0.13	-19	-10%	-66%
	Ex 2	£1.66	26	£2.10	51	-£0.44	-25	-27%	-96%
Off-Peak	Ex 1	£2.50	28	£2.79	60	-£0.30	-32	-12%	-114%
Travelcard	Ex 2	£4.31	27	£3.65	48	£0.65	-22	15%	-81%
Off-Peak One-Day	Ex 1	£2.40	29	£2.52	59	-£0.13	-30	-5%	-103%
Travelcard	Ex 2	£2.81	31	£2.83	52	-£0.02	-21	-1%	-68%
Off-Peak	Ex 1	£1.19	23	£1.58	43	-£0.39	-20	-33%	-87%
Ordinary	Ex 2	£1.16	21	£3.19	44	-£2.02	-23	-174%	-110%

Table 10.14 Summary of Logit Models Estimated From the First "Between-Mode" Rolling Stock Exercise: Non-trading Responses Not Included

Model Results

Attribute	Peak T'card Coeff	t	Peak Ordinary Coeff	t	Off Peak T'card Coeff	t	Off Peak One Day T'card Coeff	t	Off Peak Ordinary Coeff	t
Constant	0.98	1.8 ?	0.62	1.1 !	1.51	2.5	0.74	1.4 !	0.74	1.0 !
Fare5	-0.65	-1.3 !	-0.26	-0.4 !	-1.16	-2.0	-0.56	-1.1 !	0.27	0.4 !
Fare15	-0.79	-1.6 !	-0.53	-0.9 !	-1.65	-2.9	-0.92	-1.8 ?	-1.34	-1.8 ?
Fare40	-2.35	-4.2	-2.09	-3.2	-2.93	-4.4	-2.41	-3.9	-2.34	-2.7
Stock1	0.31	0.8 !	0.45	0.9 !	0.89	1.8 ?	1.32	3.0	1.10	1.8 ?
Stock2	1.87	3.8	1.91	3.1	1.53	2.9	2.45	4.4	2.33	2.9
Time	-0.01	-1.1 !	-0.02	-1.5 !	-0.01	-0.9 !	-0.03	-2.2	-0.03	-1.5 !
Headway	-0.03	-1.1 !	0.00	0.0 !	-0.04	-2.0	-0.04	-1.8 ?	-0.03	-1.3 !
Obs	325		222		261		308		160	
Sample	41		28		34		40		20	
Rho bar sq	0.17		0.16		0.20		0.22		0.23	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

Fare:	Peak T'card 15%	40%	Peak Ordinary 15%	40%	Off Peak T'card 15%	40%	Off-Peak One Day T'card 15%	40%	Off Peak Ordinary 15%	40%
Constant	-19 !	-17 ?	-18 !	-12 !	-14 ?	-21	-12 !	-12 !	-8 !	-13 !
Fare5	2.5 !	2.2	1.5 !	1.0	2.1	3.2	1.8 !	1.9	-0.6 !	-0.9
Fare15	n/a	0.9	n/a	0.7 ?	n/a	1.5	n/a	1.0	n/a	1.5
Fare40	1.1 !	n/a	1.5 !	n/a	0.7 !	n/a	1.0 !	n/a	0.7 !	n/a
Stock1	-6 !	-5 !	-13 !	-9 !	-8 !	-12 ?	-21 !	-22	-12 !	-19 !
Stock2	-35 !	-32	-54 !	-37	-14	-21	-40 ?	-41	-26 !	-40
Time	0 !	0 !	1 !	0 !	0 !	0 !	0 !	0 ?	0 !	0 !
Headway	1 !	0 !	0 !	0 !	0 ?	1 ?	1 !	1 !	0 !	1 !

Within Mode Values

Medium	16	14	0	0	0	0	49	59	0	0
Maximum	61	53	66	103	47	44	81	97	0	101
Mdm Ratio	0.00	0.00	0.00	0.00	0.00	*	0.00	0.37	0.00	0.00
Max Ratio	0.00	0.60	0.00	0.35	0.30	0.47	0.49	0.42	0.00	0.40

Totals use figures significant at the 90% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Table 10.15 Summary of Logit Models Estimated From the Second "Between-Mode" Rolling Stock Exercise: Non-trading Responses Not Included

Model Results

Attribute	Peak T'card Coeff	Peak T'card t	Peak Ordinary Coeff	Peak Ordinary t	Off Peak T'card Coeff	Off Peak T'card t	Off Peak One Day T'card Coeff	Off Peak One Day T'card t	Off Peak Ordinary Coeff	Off Peak Ordinary t
Constant	2.38	4.0	0.26	0.5 !	1.93	2.6	2.15	3.8	0.48	0.8 !
Fare5	-0.96	-1.8 ?	-0.17	-0.3 !	-1.20	-2.0 ?	-1.23	-2.3	0.23	0.4 !
Fare15	-1.54	-2.8	-0.22	-0.4 !	-2.08	-3.3	-1.63	-3.1	-0.50	-0.9 !
Fare40	-4.51	-6.0	-2.06	-3.3	-3.00	-4.3	-3.45	-5.7	-1.28	-2.2
Stock1	1.09	2.3	2.01	3.7	0.32	0.6 !	1.23	2.8	0.91	1.9 ?
Stock2	1.27	2.3	1.64	3.1	0.92	1.7 ?	1.23	2.7	1.34	2.7
Time	-0.02	-1.5 !	-0.01	-0.9 !	-0.01	-0.4 !	-0.01	-1.0 !	-0.01	-1.1 !
Headway	-0.07	-2.8	-0.03	-1.2 !	-0.05	-1.6 !	-0.06	-3.3	-0.05	-2.2
Obs	344		244		229		367		251	
Sample	44		31		30		46		33	
Rho bar sq	0.33		0.20		0.20		0.25		0.14	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

Fare:	Peak T'card 15%	Peak T'card 40%	Peak Ordinary 15%	Peak Ordinary 40%	Off Peak T'card 15%	Off Peak T'card 40%	Off-Peak One Day T'card 15%	Off-Peak One Day T'card 40%	Off Peak Ordinary 15%	Off Peak Ordinary 40%
Constant	-23	-21	-18 !	-5 !	-14	-26	-20	-25	-14 !	-15 !
Fare5	1.9	1.7	2.4 !	0.7 ?	1.7	3.2	2.3	2.9	-1.4 !	-1.4 ?
Fare15	n/a	0.9	n/a	0.3 !	n/a	1.9	n/a	1.3	n/a	1.1 ?
Fare40	1.1 ?	n/a	3.5 !	n/a	0.5 !	n/a	0.8 ?	n/a	1.0 !	n/a
Stock1	-11 ?	-10	-139 !	-39	-2 !	-4 !	-11	-14	-27 !	-28 !
Stock2	-12 ?	-11	-113 !	-32	-7 !	-12 !	-11	-14	-40 !	-42 ?
Time	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 !	0 !
Headway	1	-1	2 !	0 !	0 !	1 !	1	1	2 !	2 !

Within Mode Values

	Peak T'card 15%	Peak T'card 40%	Peak Ordinary 15%	Peak Ordinary 40%	Off Peak T'card 15%	Off Peak T'card 40%	Off-Peak One Day T'card 15%	Off-Peak One Day T'card 40%	Off Peak Ordinary 15%	Off Peak Ordinary 40%
Medium	0	0	0	0	0	0	0	8	0	0
Maximum	24	34	77	79	30	34	48	70	38	98
Mdm Ratio	0.00	*	0.00	*	0.00	0.00	*	1.83	0.00	0.00
Max Ratio	0.52	0.33	0.00	0.40	0.00	0.00	0.23	0.20	0.00	0.43

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Table 10.16 Summary of Logit Models Estimated From Combined "Between-Mode" Rolling Stock Exercises: Non-trading Responses Not Included

Model Results

Attribute	Peak T'card Coeff	t	Peak Ordinary Coeff	t	Off Peak T'card Coeff	t	Off Peak One Day T'card Coeff	t	Off Peak Ordinary Coeff	t
Constant	1.70	4.3	0.35	0.9 !	1.68	3.6	1.26	3.4	0.55	1.2 !
Fare5	-0.80	-2.2	-0.18	-0.4 !	-1.18	-2.8	-0.78	-2.2	0.25	0.6 !
Fare15	-1.17	-3.2	-0.32	-0.8 !	-1.84	-4.3	-1.16	-3.3	-0.89	-2.1
Fare40	-3.34	-7.5	-2.02	-4.4	-2.96	-6.2	-2.78	-6.6	-1.75	-3.6
Stock11	0.22	0.5 !	0.58	1.4 !	0.75	1.9 ?	0.84	2.5	0.74	2.1
Stock21	2.11	3.9	1.97	3.8	1.35	3.1	1.86	4.5	1.53	3.7
Stock12	1.05	2.7	1.83	3.5	0.47	0.9 !	1.64	3.7	1.20	2.0
Stock22	1.13	2.7	1.48	2.9	1.11	2.0	1.67	3.5	1.82	2.7
Time	-0.02	-1.8 ?	-0.01	-1.3 !	-0.01	-0.9 !	-0.02	-2.1	-0.02	-1.8 ?
Headway	-0.05	-2.8	-0.01	-1.0 !	-0.05	-2.6	-0.05	-3.4	-0.04	-2.4
Obs	669		466		490		675		411	
Sample	85		59		64		86		53	
Rho bar sq	0.24		0.17		0.20		0.22		0.18	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

Fare:	Peak T'card 15%	40%	Peak Ordinary 15%	40%	Off Peak T'card 15%	40%	Off-Peak One Day T'card 15%	40%	Off Peak Ordinary 15%	40%
Constant	-22	-20	-17 !	-7 !	-14	-23	-16	-18	-9 !	-13 !
Fare5	2.1	1.9	1.7 !	0.7	1.9	3.2	2.0	2.2	-0.8 !	-1.1
Fare15	n/a	0.9	n/a	0.4 ?	n/a	1.7	n/a	1.1	n/a	1.4
Fare40	1.1	n/a	2.4 !	n/a	0.6	n/a	0.9	n/a	0.7 !	n/a
Stock11	-3 !	-3 !	-27 !	-11 !	-6 ?	-10 ?	-11	-12	-12 !	-17 ?
Stock21	-27	-25	-92 !	-39	-11	-18	-24	-27	-26 ?	-35
Stock12	-13	-13	-86 !	-36	-4 !	-6 !	-21	-24	-20 !	-27 ?
Stock22	-15	-14	-69 !	-29	-9 ?	-15 ?	-22	-24	-31 !	-42
Time	0 !	0 ?	0 !	0 !	0 !	0 !	0 ?	0	0 !	0 !
Headway	1	1	1 !	0 !	0	1	1	1	1 !	1

Table 10.17 Comparisons of Package Fare Values Estimated From Combined "Between-Mode" Rolling Stock Exercises With Comparable "Within-Mode" Exercises: All Responses

Within Mode Values (Separate Exercises)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off-Peak One Day T'card		Off Peak Ordinary	
Fare:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Medium1	16	14	0	0	0	0	49	59	0	0
Maximum1	61	53	66	103	47	44	81	97	0	101
Medium2	0	0	0	0	0	0	0	8	0	0
Maximum2	24	34	77	79	30	34	48	70	38	98
Mdm1 Ratio	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.21	0.00	0.00
Max1 Ratio	0.44	0.47	0.00	0.38	0.24	0.41	0.30	0.28	0.00	0.35
Mdm2 Ratio	*	*	0.00	*	0.00	0.00	*	3.03	0.00	0.00
Max2 Ratio	0.60	0.40	0.00	0.37	0.30	0.44	0.45	0.35	0.00	0.43

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Within Mode Values (Combined Exercises)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off-Peak One Day T'card		Off Peak Ordinary	
Fare:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Medium1	11	13	0	0	0	0	36	48	0	0
Maximum1	48	54	69	90	41	44	58	78	0	89
Medium2	0	0	0	0	0	0	6	8	0	0
Maximum2	27	30	53	91	20	33	58	79	0	79
Mdm1 Ratio	0.25	0.21	0.00	0.00	0.00	0.00	0.30	0.25	0.00	0.00
Max1 Ratio	0.56	0.47	1.33	0.44	0.27	0.41	0.41	0.34	0.00	0.39
Mdm2 Ratio	*	*	0.00	*	0.00	0.00	3.62	2.98	0.00	0.00
Max2 Ratio	0.54	0.45	1.32	0.32	0.45	0.46	0.37	0.30	0.00	0.53

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Table 10.18 Summary of Logit Models Estimated From Both "Between-Mode" Rolling Stock Exercises: Segments Aggregated; Non-trading Responses Not Included

	First Exercise					Second Exercise			
	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
	Coeff	t	Coeff	t	Attribute	Coeff	t	Coeff	t
Constant	1.25	3.1	0.69	2.0	Constant	2.16	4.8	1.14	3.6
Fare5	-0.89	-2.4	-0.24	-0.7 !	Fare5	-1.09	-2.7	-0.49	-1.6 !
Fare15	-1.20	-3.2	-0.82	-2.4	Fare15	-1.76	-4.4	-0.90	-3.0
Fare40	-2.62	-6.2	-2.22	-5.6	Fare40	-3.78	-7.6	-2.35	-6.9
Stock1	0.56	1.8 ?	0.95	3.3	Stock1	0.72	2.1	1.25	4.7
Stock2	1.69	4.7	2.20	6.1	Stock2	1.07	2.8	1.31	4.7
Time	-0.01	-1.4 !	-0.03	-3.0	Time	-0.01	-1.3 !	-0.01	-1.8 ?
Headway	-0.03	-2.2	-0.03	-2.0	Headway	-0.06	-3.0	-0.05	-4.2
	586		690		Obs	573		862	
	75		88		Sample	74		110	
	0.18		0.19		Rho bar sq	0.26		0.18	

! = not significant at 90% level of confidence

? = not significant at 95% level of confidence

Fare Values (%)

	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
Fare:	15%	40%	15%	40%		15%	40%	15%	40%
Constant	-16	-19	-13 !	-12 ?	Constant	-18	-23	-19	-19
Fare5	2.2	2.7	0.9 !	0.9	Fare5	1.9	2.3	1.6	1.7
Fare15	n/a	1.2	n/a	1.0	Fare15	n/a	1.2	n/a	1.0
Fare40	0.8 ?	n/a	1.0 !	n/a	Fare40	0.8	n/a	1.0 ?	n/a
Stock1	-7 !	-9 ?	-17	-17	Stock1	-6 ?	-8	-21	-21
Stock2	-21	-26	-40	-40	Stock2	-9	-11	-22	-22
Time	0 !	0 !	0 ?	0	Time	0 !	0 !	0 !	0 ?
Headway	0 ?	1	0 !	0 ?	Headway	0	1	1	1

Within Mode Values

Medium	11	10	41	54	Medium	0	0	5	7
Maximum	57	51	89	103	Maximum	38	49	62	79
Mdm Ratio	0.00	0.87	0.43	0.31	Mdm Ratio	0.00	*	4.14	3.27
Max Ratio	0.37	0.51	0.45	0.38	Max Ratio	0.24	0.23	0.36	0.28

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Table 10.19 Summary of Logit Models Estimated From Combined "Between-Mode" Rolling Stock Exercises: Segments Aggregated; Non-trading Responses Not Included

Model Results

	Period T'card		Ordinary & One Day T'card	
Attribute	Coeff	t	Coeff	t
Constant	1.67	5.6	0.84	3.7
Fare5	-0.99	-3.6	-0.35	-1.5 !
Fare15	-1.47	-5.4	-0.82	-3.6
Fare40	-3.15	-9.7	-2.24	-8.7
Stock11	0.48	1.7 ?	0.77	3.2
Stock21	1.70	4.9	1.94	6.7
Stock12	0.80	2.6	1.38	5.3
Stock22	1.10	3.3	1.47	5.3
Time	-0.01	-2.0	-0.01	-3.1
Headway	-0.04	-3.7	-0.04	-4.4
Obs	1,159		1,552	
Sample	149		198	
Rho bar sq	0.21		0.18	

! = not significant at 90% level of confidence
? = not significant at 95% level of confidence

Fare Values (%)

	Period T'card		Ordinary & One Day T'card	
Fare:	15%	40%	15%	40%
Constant	-17	-21	-15	-15
Fare5	2.0	2.5	1.3	1.2
Fare15	n/a	1.2	n/a	1.0
Fare40	0.8	n/a	1.0	n/a
Stock11	-5 !	-6 !	-14	-14
Stock21	-17	-22	-35	-35
Stock12	-8	-10	-25	-25
Stock22	-11	-14	-27	-26
Time	0 ?	0 ?	0	0
Headway	0	1	1	1

Table 10.20 Comparisons of Package Fare Values Estimated From Combined "Between-Mode" Rolling Stock Exercises With Comparable "Within-Mode" Exercises: Segments Aggregated; Non-trading Respondents Not Included

Within Mode Values (Separate Exercises)

	Period T'card		Ordinary & One Day T'card	
Fare:	**15%**	**40%**	**15%**	**40%**
Medium1	11	10	34	47
Maximum1	57	51	89	103
Medium2	0	0	0	7
Maximum2	30	38	62	79
Mdm1 Ratio	0.00	0.00	0.41	0.29
Max1 Ratio	0.31	0.43	0.40	0.34
Mdm2 Ratio	*	*	*	3.77
Max2 Ratio	0.38	0.37	0.44	0.33

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

Within Mode Values (Combined Exercises)

	Period T'card		Ordinary & One Day T'card	
Fare:	**15%**	**40%**	**15%**	**40%**
Medium1	8	9	34	42
Maximum1	47	51	70	86
Medium2	0	0	5	6
Maximum2	34	36	73	90
Mdm1 Ratio	0.60	0.68	0.41	0.33
Max1 Ratio	0.37	0.42	0.51	0.40
Mdm2 Ratio	*	*	4.81	3.78
Max2 Ratio	0.33	0.38	0.37	0.29

Totals use figures significant at the 95% level of confidence. A "*" indicates that a significant value for "stock" in the "between-mode" exercise is divided by a non-significant package value from the "within-mode" exercise.

The preceding seven tables, when compared to those shown earlier, exhibit the following important characteristics:

(i) In most cases, the absolute sizes of the coefficients increase, indicating a reduction in the amount of error in the models.

(ii) The rho bar statistics improve to more acceptable levels.

(iii) Despite smaller sample sizes, many coefficients become statistically more significant.

(iv) When expressed in terms of equivalent fare values, the model constants are smaller in some cases than before, demonstrating that the majority of 'non-traders' are allied to the Underground. An exception to this may be seen for off-peak Travelcard users, whose constants are seen to increase. This is consistent with the observation above that in the first SP exercise, off-peak Travelcard 'non-traders' allied to their alternative modes were in the same proportion as those allied to the Underground; in the second exercise, they formed a higher proportion of the sample.

(v) Though for some groups the fare values for the rolling stock improvements remain similar (peak Travelcard users especially), for others, the values have changed quite noticeably. The most important change is the occurrence of an anomaly not present in the earlier models. This may be observed in the result for peak ordinary ticket users. The intermediate package of rolling stock improvements from the second exercise exhibits a higher fare value than the package of maximum improvements. Given that the latter package subsumes the intermediate package, this would appear counter-intuitive. However, the difference between the coefficients is not significantly different from zero, even at the 90% level of confidence.

The last finding is of most interest, because it seems to contradict the assumption derived from other studies and from a previous understanding of the modelling process, that the removal of non-traders would not have a major impact on the relative values of the coefficients. Low correlations between the estimates were observed for all the models throughout, so that these variations cannot be explained by the effects of correlation. Another possible explanation is that as each individual was only presented with half of the fractional experimental design in each exercise, the loss of more 'non-traders' from one of these blocks than from the other could have led to a slight distortion of the values put upon the 'stock' variables. Though weightings were applied to compensate for imbalances in the mixture of design blocks, it is likely that imprecision in this process could allow slight variations in values such as those observed here.

10.3 Combining the Results of the Two SP Exercises

The previous chapter reported on the analysis of the 'within-mode' adaptive ranking exercises, producing for each service quality improvement an implied monetary value and a utility weighting relative to other attributes that would be introduced as packages in the succeeding 'between-mode' choice exercises. Variations in these values were observed for different treatments of the data (removal of economically irrational responses; greater aggregation.) Section 10.2 reported the analysis of the 'between-mode' choice exercises, also deriving an implied monetary value for packages of service quality improvements. These values were also observed to vary under different treatments of the data (greater aggregation; removal of 'non-traders'.) The main finding from this section was that the monetary values obtained for the packages of improvements were in most cases considerably less than the sum of the values for the package components obtained from the 'within-mode' exercises.

It was argued in the previous chapter that the 'between-mode' exercises had the potential to produce more realistic (ie lower) monetary values than the 'within-mode' exercises, because of the more realistic choice context and the likelihood of spending thresholds being identified in relation to more than one improvement being introduced at any one time. Of course, a 'between-mode' exercise could not be regarded as a substitute for a 'within-mode' exercise, because the former would not be able to present the detailed combinations of attribute variations that were possible in the latter, while keeping the respondent's task manageable. Thus, the two exercises were intended to compliment one another: the 'within-mode' exercise deriving relative values for individual attribute improvements, the 'between-mode' exercise deriving monetary values within a realistic mode choice context.

To combine the results of the two exercises and obtain the most plausible monetary valuations for individual attributes, the simplest approach would be to apply the fare scaling factors derived from section 10.2 to the monetary valuations from chapter nine. However, this would distort the relative weights attached to the service attribute improvements. It will be remembered that some significant coefficients attached to each level of service attribute improvement in the 'within-mode' exercise did not produce a *significant* value when related to the fare coefficients from the same exercise (due to a weaker fare coefficient and the loss of significance inevitable when estimating a ratio of coefficients.) The more appropriate method would therefore be to apply the relative weights implied by the coefficients for each improvement in the 'within-mode' exercise (ignoring the ratio against fare) directly to the implied monetary values for the packages of improvements obtained from the 'between-mode' exercise. This approach was therefore adopted.

Tables 10.21 to 10.26 summarise the final values calculated for each rolling stock improvement. It will be seen that the totals for each set of improvements match the values established for each package of improvements presented in the 'between-mode' exercises.

Table 10.21 Summary of Final Fare Valuations: Cleanliness; Newness; Air Quality and Gangways

Proportional Fare Values (90% Level of Confidence)

All Responses

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
L of C:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Clean1	0	0	0	0	0	0	0	19	0	0
New1	0	0	0	0	0	0	0	0	0	0
Air1	0	0	0	0	0	0	0	0	0	0
Gangway1	0	0	0	0	0	0	0	0	0	0
Total	0	0	0	0	0	0	0	19	0	0
Clean2	0	11	0	16	5	7	0	20	0	16
New2	0	0	0	0	0	0	0	0	0	0
Air2	0	14	0	13	8	11	0	16	0	16
Gangway2	0	8	0	9	4	5	0	13	0	11
Total	0	32	0	38	17	23	0	49	0	43

Non-Traders Removed

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
L of C:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Clean1	0	0	0	0	0	0	0	22	0	0
New1	0	0	0	0	0	0	0	0	0	0
Air1	0	0	0	0	0	0	0	0	0	0
Gangway1	0	0	0	0	0	0	0	0	0	0
Total	0	0	0	0	0	0	0	22	0	0
Clean2	0	10	0	15	4	6	16	17	0	15
New2	0	0	0	0	0	0	0	0	0	0
Air2	0	13	0	13	7	10	13	13	0	15
Gangway2	0	8	0	8	3	5	10	11	0	10
Total	0	32	0	37	14	21	40	41	0	40

Totals are from the "Between-mode" exercises, upon which the values for individual improvements are based. The columns will not always sum to the total figure due to rounding.

Table 10.22 Summary of Final Fare Valuations: Noise; Quality of Ride; Information and Gangways

Proportional Fare Values (90% Level of Confidence)

All Responses

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
L of C:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Noise1	0	*	0	0	0	0	0	0	0	0
Ride1	0	*	0	35	0	0	0	0	0	0
Info1	0	*	0	0	0	0	0	0	0	0
Gangway1	0	*	0	0	0	0	13	16	0	0
Total	0	10	0	35	0	0	13	16	0	0
Noise2	0	5	0	13	0	6	4	5	0	8
Ride2	0	0	0	11	0	0	3	4	0	9
Info2	0	0	0	8	0	6	3	4	0	8
Gangway2	0	6	0	8	0	6	4	5	0	9
Total	0	10	0	40	0	18	14	17	0	34

Non-Traders Removed

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
L of C:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Noise1	0	*	0	0	0	0	0	0	0	0
Ride1	0	*	0	39	0	0	0	0	0	0
Info1	0	*	0	0	0	0	0	0	0	0
Gangway1	0	*	0	0	0	0	11	14	0	0
Total	0	10	0	39	0	0	11	14	0	0
Noise2	6	5	0	10	0	0	3	4	0	10
Ride2	0	0	0	9	0	0	3	3	0	11
Info2	0	0	0	6	0	0	2	3	0	10
Gangway2	7	6	0	7	0	0	3	4	0	11
Total	12	11	0	32	0	0	11	14	0	42

Totals are from the "Between-mode" exercises, upon which the values for individual improvements are based. The columns will not always sum to the total figure due to rounding.

Table 10.23 Summary of Final Fare Valuations: All Attributes, Segments Aggregated, Separate SP Exercises

Proportional Fare Values (95% Level of Confidence)

All Responses

	First Exercise					Second Exercise			
	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
L of C:	15%	40%	15%	40%		15%	40%	15%	40%
Clean1	0	0	0	13	Noise1	0	0	0	0
New1	0	0	0	0	Ride1	0	0	0	0
Air1	0	0	0	3	Info1	0	0	0	0
Gangway1	0	0	0	0	Gangway1	0	0	18	21
Total	0	0	0	16		0	0	18	21
Clean2	6	9	0	18	Noise2	4	5	6	7
New2	0	0	0	0	Ride2	0	0	5	6
Air2	9	12	0	15	Info2	2	3	5	5
Gangway2	5	7	0	11	Gangway2	4	5	6	7
Total	20	27	0	44		10	13	22	26

Non-Traders Not Included

	First Exercise					Second Exercise			
	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
L of C:	15%	40%	15%	40%		15%	40%	15%	40%
Clean1	0	0	15	14	Noise1	0	*	0	0
New1	0	0	0	0	Ride1	0	*	0	0
Air1	0	0	3	3	Info1	0	*	0	0
Gangway1	0	0	0	0	Gangway1	0	*	21	21
Total	0	0	17	17		0	8	21	21
Clean2	7	8	16	16	Noise2	3	4	6	6
New2	0	0	0	0	Ride2	0	0	5	5
Air2	9	11	14	14	Info2	2	3	5	5
Gangway2	5	7	10	10	Gangway2	4	5	6	6
Total	21	26	40	40		9	11	22	22

Totals are from the "Between-mode" exercises, upon which the values for individual improvements are based. The columns will not always sum to the total figure due to rounding.

Table 10.24 Summary of Final Fare Valuations: All Attributes; SP Exercises Combined

Proportional Fare Values (95% Level of Confidence)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
L of C:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Clean1	0	0	0	0	0	0	0	0	0	0
New1	0	0	0	0	0	0	0	0	0	0
Air1	0	0	0	0	0	0	0	0	0	0
Gangway1	0	0	0	0	0	0	0	0	0	0
Total	0	0	0	0	0	0	0	0	0	0
Clean2	7	8	0	21	5	8	11	13	0	9
New2	0	0	0	0	0	0	0	0	0	0
Air2	9	10	0	18	7	11	8	10	0	10
Gangway2	6	7	0	12	3	5	7	8	0	7
Total	21	25	0	50	15	25	26	30	0	26
Noise1	0	0	0	*	0	0	0	0	0	*
Ride1	0	0	0	*	0	0	0	0	0	*
Info1	0	0	0	*	0	0	0	0	0	*
Gangway1	0	0	0	*	0	0	23	26	0	*
Total	0	0	0	26	0	0	23	26	0	39
Noise2	0	0	0	10	4	6	7	8	0	14
Ride2	0	0	0	8	0	0	6	6	0	14
Info2	0	0	0	6	4	6	5	6	0	14
Gangway2	0	0	0	7	3	5	6	7	0	12
Total	0	0	0	32	11	17	24	28	0	54

Totals are from the "Between-mode" exercises, upon which the values for individual improvements are based. The columns will not always sum to the total figure due to rounding.

Table 10.25 Summary of Final Fare Valuations: All Attributes; Segments Aggregated; SP Exercises Combined; Non-Traders Not Included

Proportional Fare Values (95% Level of Confidence)

	Peak T'card		Peak Ordinary		Off Peak T'card		Off Peak One Day T'card		Off Peak Ordinary	
L of C:	15%	40%	15%	40%	15%	40%	15%	40%	15%	40%
Clean1	0	0	0	0	0	0	9	10	0	0
New1	0	0	0	0	0	0	0	0	0	0
Air1	0	0	0	0	0	0	0	0	0	0
Gangway1	0	0	0	0	0	0	2	2	0	0
Total	0	0	0	0	0	0	11	12	0	0
Clean2	9	8	0	16	3	6	10	11	0	13
New2	0	0	0	0	0	0	0	0	0	0
Air2	11	10	0	14	5	8	8	9	0	13
Gangway2	7	7	0	9	2	4	6	7	0	9
Total	27	25	0	39	11	18	24	27	0	35
Noise1	*	*	0	*	0	0	0	0	0	0
Ride1	*	*	0	*	0	0	0	0	0	0
Info1	*	*	0	*	0	0	0	0	0	0
Gangway1	*	*	0	*	0	0	21	24	0	0
Total	13	13	0	36	0	0	21	24	0	0
Noise2	6	6	0	9	0	0	6	7	0	11
Ride2	0	0	0	7	0	0	5	5	0	11
Info2	3	3	0	6	0	0	5	5	0	11
Gangway2	5	5	0	7	0	0	6	6	0	9
Total	15	14	0	29	0	0	22	24	0	42

Totals are from the "Between-mode" exercises, upon which the values for individual improvements are based. The columns will not always sum to the total figure due to rounding.

Table 10.26 Summary of Final Fare Valuations: All Attributes; Segments Aggregated; SP Exercises Combined

Proportional Fare Values (95% Level of Confidence)

All Responses					Non-traders Not Included				
	Period T'card		Ordinary & One Day T'card			Period T'card		Ordinary & One Day T'card	
L of C:	15%	40%	15%	40%	L of C:	15%	40%	15%	40%
Clean1	0	0	0	0	Clean1	0	0	10	10
New1	0	0	0	0	New1	0	0	0	0
Air1	0	0	0	0	Air1	0	0	2	2
Gangway1	0	0	0	0	Gangway1	0	0	2	2
Total	0	0	0	0	Total	0	0	14	14
Clean2	6	8	13	15	Clean2	6	7	14	14
New2	0	0	0	0	New2	0	0	0	0
Air2	7	10	12	13	Air2	7	9	12	12
Gangway2	4	6	9	10	Gangway2	4	5	9	9
Total	18	23	34	38	Total	17	22	35	35
Noise1	0	0	0	0	Noise1	0	0	0	0
Ride1	0	0	0	0	Ride1	0	0	0	0
Info1	0	0	0	0	Info1	0	0	0	0
Gangway1	0	0	21	24	Gangway1	0	0	25	25
Total	0	0	21	24	Total	0	0	25	25
Noise2	4	6	8	9	Noise2	4	5	8	8
Ride2	0	0	6	7	Ride2	0	0	6	6
Info2	3	4	6	6	Info2	3	4	6	6
Gangway2	4	5	6	7	Gangway2	4	5	7	6
Total	12	15	26	29	Total	11	14	27	26

Totals are from the "Between-mode" exercises, upon which the values for individual improvements are based. The columns will not always sum to the total figure due to rounding.

From these seven tables a number of important anomalies can be identified:

(i) In some cases, only one improvement in the intermediate range obtained a significant coefficient in the 'within-mode' exercises, so that all the value measured in the corresponding 'between-mode' has been assigned to it. The improvement for which this occurs most frequently is 'gangway1'. When compared against the values obtained for maximum improvements, the assigned value is lower because significant coefficients were obtained for more improvements in the 'within-mode' exercise.

(ii) For a limited number of intermediate improvements, no significant coefficients were obtained from the 'within-mode' exercise, yet a significant value is obtained from the 'between-mode' exercise. The most noticeable occurrence for this is in Table 10.25, among peak ordinary ticket users with non-traders removed and the SP exercises aggregated (the higher package value for intermediate improvements, compared to that for the maximum improvements, was noted in the previous section.)

One interpretation that could be placed upon these observations is that in some cases the exercises have detected respondents' willingness to pay for even the smallest improvement, but a reluctance to increase the amount when more are added. Another interpretation would be that facilities given little weighting in the 'within-mode' exercise give some benefit when introduced all together: that is, the instances where all the value from the 'between-mode' exercise has been assigned to one attribute produce an over-estimate of the value of that one attribute. This is supported by the occasions when a value obtained from the 'between-mode' exercise cannot be assigned to any attributes because none were significant in the 'within-mode' exercise.

Table 10.27 shows a comparison of the study findings with comparable estimates from earlier studies. The figures used are weighted combinations of values from the segments used in Tables 9.28 (chapter nine) and 10.25 previously. These have been used to represent values for all London Underground passengers that are comparable with those derived from the earlier studies. These results are from models in which inconsistent responses ('within-mode' exercise) and non-traders ('between-mode' exercise) have been removed, as they represent the most robust estimates. Only the maximum improvements have been included, as these appear most consistent with the improvements investigated in the other studies. The weights used to combine the segments were derived from contemporary origin-destination surveys conducted on the Northern Line by LUL. It can be seen that the values from the 'within-mode' exercises (Table 9.28) are of the same magnitude as those from earlier studies and generally slightly higher. In contrast, the values from the 'between-mode' exercises (Table 10.25) are somewhat lower, in the region of about half the earlier values.

10.4 Chapter Summary and Conclusions to the Analysis of Improvements to the Northern Line

As in the previous chapter, this chapter has presented the main stages involved in the analysis of the "between-mode" SP exercises with the intention of providing a detailed guide of how such an analysis may be conducted.

For this SP data set, weighting was less critical as it applied only to the design "blocks" and not the number of choices per individual (all respondents faced the same number of SP choice situations.) A preliminary non-statistical analysis of the responses provided some initial indication of how respondents prioritised between the service quality improvements and changes to other journey attributes. Binary logit models were again estimated (Underground v best alternative mode/ not travel) and overall valuations for packages of service improvements derived.

Table 10.27 Comparisons of Fare Valuations with Previous Underground Studies

Values are percentages of fare

Improvement (Maximum)	Table 9.28	Table 10.25	London Transport (1984)	Steer, Davies & Gleave (1984d)
Clean2	24	10	16	-
New2	0	0	-	-
Air2	25	10	18	-
Gangway2	17	7	-	-
Noise2	20	6	12	8
Ride2	7	3	14	-
Info2	11	5	-	10

Weights used to derive total sample values:

Ticket	Peak	Off-Peak
Ordinary	14.5%	10.8%
One-day T'card		17.7%
Period T'card	39.9%	17.2%

As before, aggregation of the data (across the two SP exercises particularly) was found to improve the performance of the models. Also, the removal of "non-traders" (those who always chose the Underground or always chose their alternative modes) was found to be beneficial.

This and the previous chapter have between them described the analysis of a study of passenger valuations towards improvements to London Underground's Northern Line services. They have reported on the estimation of a range of models fitted to the 'within-mode' and 'between-mode' SP exercises, and the main conclusions that may be drawn are given below.

'Within-Mode' Exercises

It was demonstrated that simple tabulations of responses to each SP option could provide useful indications of respondent preferences before embarking with more sophisticated analytical approaches. These could confirm that there were inconsistencies in the responses of some individuals, that some trading between the attributes appeared to be taking place and that different priorities appeared to exist for different segments.

Compared to many previous applications of conjoint analysis in studies of passenger facilities, this study used modelling procedures that yielded measures of statistical significance and goodness-of-fit. It was possible to calculate t-Statistics not only for model coefficients but also fare valuations implied by the ratios of coefficients. For the estimation of the logit models, some adjustment was necessary both for repeated observations and for the unequal numbers of observations per respondent due to the adaptive ranking algorithm.

Rolling Stock Exercises

It was shown that generally plausible results could be obtained from the first logit models to be estimated from the 'within-mode' data, though the goodness-of-fit was weak in a number of cases. Two succeeding strategies, the removal of illogical responses and the aggregation of data across the SP exercises and segments, demonstrated that the quality of the models could be improved appreciably.

It was shown that respondents gave a higher weighting per unit to the 10% changes in fare compared to the 30% changes, though the latter were the more strongly significant. The progression of the weights attached to fare changes were not all linear, though became closer to linear as the data was aggregated. Where the SP exercises were analysed separately, the relative weights for each fare change could be seen to be quite different in some cases between comparable groups of respondents. The same could be seen for the gangways attribute, which was universal to both rolling stock exercises. Only when data were aggregated did the values for this attribute, like fare, become more consistent. This suggested that even with sample sizes approaching 100 respondents, the estimates can vary considerably across different groups. The most stable estimates across the different data manipulation exercises were those for maximum level improvements at values calculated against a 30% fare change.

Regarding the conclusions that may be drawn concerning respondents' preferences to improvements in passenger facilities, these may be summarised as follows:

(i) ordinary ticket and one-day Travelcard users placed higher percentage fare values on improvements than period Travelcard users;

(ii) few intermediate improvements are given any importance, while most maximum improvements do, with the exception of newness;

(iii) no significant interaction effects were measured, even at the highest levels of data aggregation;

Cleanliness, air quality and noise reductions were the main priorities for passengers, followed by the ability to contact the driver in an emergency ('gangway2'), on-train information and quality of ride. Newness was never valued, suggesting that all the benefits from a new train are to be derived in improvements to the other attributes. Only one clear anomaly was identified in the valuations derived from these exercises, and that was in the way one-day Travelcard users appeared to place a higher weight on the shift from a dirty, vandalised train to a dirty one than to a clean one. The difference in valuations was significant over all the models, though lessened with the removal of economically irrational responses, and may have resulted from a non-significant but sizeable interaction effect with newness.

Station Exercise

This exercise appeared to suffer from weaknesses in the design (namely the removal of the dominant option in the choice set) and a large proportion of responses where both options in the paired choices were rejected. In this respect, tests with simulated data could probably have provided a warning of such problems (the pilot survey used a design with all options present, which appeared to work efficiently.)

It still proved possible to derive plausible estimates for the values of improvements by omitting the 10% fare variable from the model and the insertion of artificial data to adjust for the missing option in the design, but the reliability of the results cannot be regarded as strong.

'Between-Mode' Exercise

Simple tabulations of the responses to this exercise showed that quality of rolling stock could have an impact on mode choices and that a sizeable proportion of respondents were 'non-traders'. Many of these were allied to the Underground, but a number preferred their alternative throughout. Even in the best Underground option, journey time was increased, suggesting that these non-traders were sensitive even to the smallest time change. It is possible that some of these respondents were 'voting' against any worsening of the Underground service (ie policy response bias) or they may simply have reflected a marginal mode choice situation.

Each respondent considered the same number of choice situations, so that no adjustment was necessary as with the 'within-mode' exercises, but the unequal distribution of the design 'blocks' in some cases did require a weighting procedure to correct the imbalance. For the initial logit models fitted to the data, their goodness-of-fit was seen to be very poor, with resulting fare valuations that were generally non-significant. Aggregation of the data improved the significance of the valuations, but the removal of non-traders was to have a noticeable positive effect on the fit of the models and a reduction in the size of the error term (implied by the increased magnitude of the coefficients.)

In most cases, the values for packages of improvements were considerably lower than those derived by summing the results of the 'within-mode' valuations, implying scaling factors in the region of 0.3 to 0.5 at the levels of maximum improvement. For some segments, a package of facilities in which no individual improvements with a significant valuation in the 'within-mode' exercise received a significant valuation in the 'between-mode' exercise. This occurred for the intermediate levels of noise, air, information and gangways group of facilities. This may suggest that some relatively unimportant improvements, if grouped with others, can be given a sizeable value. In others, the opposite occurred, when improvements given a value in the first exercise were not valued in the second. This occurred for the intermediate level of air quality for peak Travelcard users and for intermediate levels of cleanliness and gangways for all off-peak Travelcard users. In these instances, a variable with some value initially did not register as part of a package of improvements set against other journey attributes.

Use of the Results

The object of the research was to produce values for improvements to passenger facilities that could be used as measures of generalised cost in LUL's forecasting models. It was required that a single value should be obtained for each facility examined in the study, which when added with the value of other facilities would not overstate the final values of such investment. This creates a problem concerning which results should be used from the two types of SP exercise. Those values derived from the 'within-mode' exercises could be taken to represent the value of introducing each facility on its own, while those from the 'between-mode' exercise represent the values of facilities introduced all together.

From the discussions that have taken place in this and previous chapters, there is a view that even as measures of the value of individual facilities, the results from the 'within mode' exercises will not be reliable. This is because they are derived from an SP exercise which requires abstract judgements, set against only one primary journey attribute (fare.) In contrast, the valuations from the 'between-mode' exercises are derived from a familiar mode choice context, with a number of primary journey attributes included. In some cases, the value of a single improvement derived from the 'within-mode' exercise is greater than the value for the corresponding group of improvements in the second exercise. This could suggest some over-statement of the value of individual attributes.

The approach taken in section 10.2, in which the proportional values of improvements in the 'within-mode' exercise are used to calculate monetary values from the results of the 'between-mode' exercise, produces final values that may be considered conservative when used to evaluate schemes with less improvements than those introduced in the 'between-mode' exercises. That is to say, the final value derived for each improvement is the value when all the other related improvements are introduced. It could under-state the value of the improvement if introduced with smaller packages of improvements or on its own.

The values obtained in this study are of course a product of the specific decisions made on how they were to be presented to respondents in the SP exercises. It is hoped that these decisions may be considered reasonable given that they have drawn from past research and considerable exploratory work (though, as has already been noted, testing of the experimental designs with simulated data would have increased confidence in their efficiency and possibly highlighted the difficulties with the 'within-mode' station exercise.) The aim has been to produce results suitable for the IUL forecasting models and it is suggested that those produced here are sufficient for the task. Nevertheless, there remain two uncertainties regarding these values:

(i) it is not clear how the final values would have been affected if all seven on-train facilities had been introduced in one package, as opposed to two separate ones - it is possible that lower values could have resulted due to threshold effects on respondents' willingness to pay for them;

(ii) to use the station facility values for forecasting, these can be scaled by a similar factor to that derived for the fare values for on-train improvements (in the order of 0.3 to 0.5) - the assumption, though, is that station facilities would assume the same level of importance as on-train facilities in a mode choice exercise.

Directions for Further Research

The data collected here is obviously capable of further interrogation, though it is argued that sufficient work has been undertaken to satisfy the goals of this study: namely to determine how more plausible valuations of passenger facilities may be derived from the use of a discrete choice SP exercise. It has examined procedures in the analysis that can improve the estimates and has presented figures which are more conservative (and therefore to be considered more plausible) than those obtained previously in studies for London Underground.

Further research that could be undertaken and which would address issues of relevance to topics covered here are:

(i) Analysis with alternative segmentations - although the segments used represent a convenient way of broadly representing a number of journey and traveller characteristics (eg journey purpose, crowding, age and gender), it would be of interest to observe how valuations differ between other groupings (eg crowded versus uncrowded; "AB" socio-economic groups against others.) In doing this, it would be advisable to maintain the division between period Travelcards and other tickets, because the purchasing decisions involved will be so different.

(ii) Alternative modelling procedures - the merging of the results from the two types of SP exercise to produce scaling factors could be more rigorously undertaken by the use of two procedures suggested in Kroes and Bradley (1990, pp14-15):

(a) sequential scaling, in which the parameters estimated for items in a judgemental SP exercise (eg the 'within mode' exercises) are used to fix values in a corresponding discrete choice SP or RP exercise (eg the 'between mode' exercises);

(b) simultaneous scaling, in which judgemental data is used in conjunction with choice data in a single hierarchical logit model.

11: CONCLUSIONS

11.1 Introduction

This chapter puts forwards the principal conclusions that may be drawn from it and suggests directions for future research. Ostensibly the study has been concerned with the valuations travellers attach to passenger facilities at rail stations and aboard trains, but the main focus has been upon the research methods that may be most effectively applied in the measurement of such valuations. A final section presents a discussion on the future for Stated Preference techniques in transport and potentially wider research.

Reference to chapter one will recall the central research question to which the study was to be addressed, namely:

> "How best can the value to passengers of investment in station and on-train passenger facilities be measured, given the requirement to assess it on the basis of the financial returns which they will yield?"

It was asserted that such facilities had an influence on the benefit travellers derive from a rail service and therefore their willingness to pay for and use it. It was also suggested that the standard of passenger facilities on many rail services is perceived to be poor by many travellers, such that the demand for rail services will be negatively affected.

This poor standard, it was argued, is due not only to limited investment funds, but also to limitations in rail operators' evaluation methods, which can often fail to incorporate the value of secondary qualitative improvements. This state of affairs has been improved by the introduction of new research methods more suited to measuring the values of small service attributes, but it was suggested that further development of such methods was possible and desirable.

The study therefore proceeded with the following:

(i) A discussion of how passenger facilities can influence travellers' perceptions of the attractiveness of rail services and an examination of passenger perceptions from market research surveys (chapter two.)

(ii) A review of developments in evaluation procedures and quantitative approaches to understanding travel behaviour (chapter three). Stated preference (SP) techniques were identified as the most suitable for examining the value of secondary attributes and subsequently made subject to detailed investigation (chapters four, five and six.)

(iii) A review of SP studies of the value of passenger facilities (chapter seven.)

(iv) A detailed case study of the use of SP techniques in a study on the London Underground (chapters eight, nine and ten.) A development in conjoint analysis techniques was applied and a range of alternative procedures considered in the calibration of logit models to the data.

The main findings from each of the above elements are summarised below.

11.2 Conclusions

The conclusions that may be drawn from this study are:

11.2.1 Passenger Facilities and the Demand for Rail Services

(i) (a) Passenger facilities are secondary to individuals' travel decision processes, but have an identifiable role in mitigating the disutility of travel time and interchange.

(b) Given that large perceptual costs may be associated with rail travel, any facilities which may reduce these should be considered by operators and planners, especially in view of the generally small capital costs involved.

(ii) (a) Travellers are able to identify the contribution of passenger facilities to the quality of a rail journey and regard some items as very important.

(b) The perception of the quality of such facilities is variable according to different rail systems and in many cases the level of satisfaction is low.

11.2.2 Quantifying the Value of Passenger Facilities

(iii) (a) As the influence of passenger facilities on travellers' propensity to use and pay for rail services is likely to be small in relation to primary journey attributes, it is difficult to measure their contribution to travel demand.

(b) Equally, the qualitative nature of such items does not allow them to be easily represented against planning criteria which are based on monetary values (financial returns, cost benefit analysis.)

(c) Research techniques that can estimate the monetary (and other quantitative) values of passenger facilities are therefore of value in promoting their proper consideration in the investment evaluation process.

(iv) (a) A range of approaches to understanding travellers' preferences and determining their impact on travel choices have been developed, but the most appropriate for valuing passenger facilities are 'stated preference' (SP) techniques.

(b) SP techniques allow secondary and qualitative attributes to be traded by travellers against fare and other primary attributes in conditions closely controlled by the researcher.

(c) The use of sample segmentation and the development of sophisticated statistical procedures for modelling from discrete choice data has strengthened the procedures available for analysing SP data and broadened the range of SP approaches that may be used, to include options that can closely represent real travel choices.

(d) Although the data obtained from SP surveys represent only judgements or stated intentions, there is limited evidence to suggest that the results from such surveys are often consistent with real travel choices.

(v) (a) A number of studies have been carried out using SP techniques to measure the value of improvements to passenger facilities and general rail travel environments. The values obtained for such improvements vary greatly in magnitude, for which the following reasons are suggested:

(i) They are sensitive to the characteristics of the particular rail system and section of that rail system to which the improvements will be introduced.

(ii) They reflect the realism (or otherwise) of the SP choices presented to them. This relates particularly to the way improvements are presented (individually or as packages), the variable against which they are to be traded (one main item against a variety) and the type of responses they are asked to give (ranging from abstract ranking or budget allocation to more familiar and simpler discrete choices.)

(b) In the design of an SP survey, the researcher should consider the value of preparatory work, principally:

(i) Exploratory research, of which previous studies, attitudinal studies and group discussions all offer useful insights into the topic being examined and guidance for the way improvements to facilities may be described to respondents.

(ii) Testing of the SP designs, of which pilot surveys are particularly important, but some work with simulated data may also be of benefit, to test their ability to efficiently return known valuations.

(c) Where interaction effects may be considered likely to exist, efforts should be made to allow them to be included in the analysis. However, it should be remembered that a number of compromises may be necessary in the design to allow interactions to be measured.

(vi) (a) For the analysis of SP data, a range of approaches may be adopted, from simple "naive" techniques to sophisticated probabilistic models (eg logit.) The latter are to be preferred, as they provide measures of goodness-of-fit and tests of statistical significance, all of which serve to indicate the reliability of the parameters they produce.

(b) The researcher should consider the necessity of the following adjustments when analysing SP data:

(i) Compensation for varying numbers of responses per individual, which is a feature for data derived from adaptive ranking SP methods.

(ii) Adjustment for imbalances in the data arising from the use of randomly allocated design 'blocks'.

(iii) Adjustment of the t-Statistics produced from statistical models to reflect repeated observations.

(c) The statistical robustness of models fitted to SP data can be enhanced appreciably by the following strategies:

(i) removal of 'illogical' responses.

(ii) data aggregation, across related segments and SP exercises with common attributes.

(iii) removal of 'non-traders' (discrete choice exercises.)

(vii) It is suggested that the use of discrete choice SP exercises in conjunction with conjoint measurement is a useful approach to the measurement of values attached to passenger facilities. There is merit in developing and testing this approach further, given that it has demonstrated a pronounced scaling effect on the monetary values for passenger facilities when presented in packages within a realistic travel choice context.

11.3 Further Research

This study has contributed to the development of SP techniques as a tool for valuing qualitative rail service attributes, principally through the use of discrete choice SP exercises and adaptive conjoint measurement. It has presented the application of these methods as a response to some of the weaknesses identified in previous studies of this nature. These weaknesses included the use of SP techniques that presented choice situations removed from travellers' real experiences and the application of often relatively unsophisticated analytical procedures. The study has developed ways in which these problems can be effectively addressed.

From the study of the literature and the findings from the case study, it is suggested that this topic could benefit in a number of ways from further research. It has been argued that there are considerable benefits in public transport operators and planners being able to quantify the value of investment in passenger facilities. The development of suitable measurement techniques should therefore be continued. A number of issues need to be addressed, and these are summarised below:

(i) Validation of SP techniques

Because of their secondary nature, the impact of passenger facilities is very difficult to measure from revealed preferences and SP techniques offer a practical alternative. Nevertheless, it could be possible to set up a real life experiment in which the provision of facilities and an appropriate charge for them could be systematically varied, to infer monetary values. An SP survey could be conducted before and/or afterwards and the findings compared. The cost of such an exercise could of course be expensive and require a lengthy timescale.

(ii) Direct Comparisons of Alternative Methods

The study has shown how the researcher must make a number of judgements concerning the design of SP instruments, only some of which can refer to well-tested procedures. For example, a particular method of data collection may recommend itself for practical reasons, though there has been little formal research to suggest how such a method may be more or less efficient than others. Below are listed a number of elements from the process of SP survey design and analysis for which it is considered that further research would be particularly beneficial.

(i) Self-completion and interviewer-administered surveys - how are respondents' understanding of an SP exercise and preferences influenced by these alternative approaches?

(ii) Conventional questionnaires and computer-based interviews - by what degree is the efficiency of the survey instrument enhanced by the greater flexibility of design offered by computer-based interviews? How are individuals' responses affected, if at all, by the different mediums?

For these first two issues, a useful approach would be to use more than one method within the same survey. Comparisons may then be made between the results from each set of data. Other issues may require more explorative work:

(iii) Presentation of attributes: descriptions - how sensitive are valuations to the different ways in which facilities are presented to respondents? There is some very limited evidence to suggest that there is little difference between visual and verbal descriptions, but this should be examined further. If respondents were asked to assess a mock-up of a carriage, for example, would this produce different results from illustrations or verbal descriptions?

(iv) Presentation of attributes: packages - it has been shown that the values derived for packages of facilities are somewhat lower than the sum of individual values measured for those facilities. How sensitive is this scaling effect to the number of items in a package? By what process may individuals pick out certain items in a package and reject others?

(v) Choice contexts - The scaling effect on monetary values of passenger facilities observed in the case study and some preceding studies was considered to be the result of a combination of using packages of facilities, comparing them with a range of primary journey attributes and presenting them in a realistic travel context, such as mode choice. The use of a simpler procedure for respondents (discrete choices compared to ranking or budget allocation) was also expected to enhance the robustness of the results. Further work should be considered to determine the relative effects of these different approaches.

These latter issues could also be investigated by including more than one method of presentation or choice context in the same SP survey, but in-depth qualitative work would also be of use. For example, respondents could be asked to rationalise their decisions as they proceed through an SP exercise and this qualitative information used to formulate appropriate decision rules.

(vi) Analytical procedures - the theoretical robustness of alternative modelling procedures and the sensitivity of the results to different treatments of the data have been discussed at length in this study, but there remains some value in comparing further the efficiency of alternative modelling methods.

11.4 Stated Preference Techniques: Powerful Tools for Behavioural Research

This study has reported on the application of SP techniques to a very specific subject area: passengers preferences for improvements to secondary rail service attributes. However, these research tools clearly offer benefits to other aspects of transport operations and planning and fields outside transportation. Their principal attraction is that they allow the researcher to experiment with aspects of human behaviour, albeit in a hypothetical context, in ways that would not be practical in real circumstances.

The opportunity is provided for the researcher to examine responses to circumstances which do not currently exist. Alternatively, he or she may recreate existing circumstances in which variations of certain variables may be introduced and their effect on behaviour more precisely identified. Items that are qualitative by nature may have equivalent quantitative values attached to them (inferred by observed trade-offs against quantitative variables), so offering the opportunity for comparisons of qualitative and quantitative impacts on a common scale.

These attractions offered by SP techniques are of course only of value if one can accept that responses to hypothetical scenarios are valid representations of real behaviour. The limited validation work that has been carried out suggests some support for this, but there is clearly a distinction to made between techniques that offer completely hypothetical scenarios (in which the method of presenting choices and eliciting responses, as well as the items presented in the experiment, are fairly abstract) and those that offer choices in a context familiar to respondents. The work presented in this study suggests that more plausible results will be obtained from the latter type of survey (in this case, a mode choice exercise based on an individual's most recent journey) than the former type (a "within-mode" adaptive ranking exercise.)

Behavioural researchers have available to them a wide range of tools which extend from unstructured qualitative methods (group discussions, in-depth interviews) through "pseudo-quantitative" attitudinal measures (eg simple ranking and ratings of attributes) to models of observed behaviour (revealed preferences.) SP techniques occupy a position in this range somewhere between attitudinal measures and revealed preference modelling. In some studies they have the role of a sophisticated form of attitudes/perceptions measurement. In others they may provide a direct substitute for revealed preference (RP) data, which is often expensive to collect and difficult to analyse.

This study, with its emphasis on relative valuations of rail attributes rather than forecasts of behaviour, leans more towards the use of SP techniques as sophisticated measures of attitudes and perceptions. In others, such as in the study undertaken by the Department of Transport into the value of travel time (MVA et al, 1987), the emphasis has been on the quantification of factors for forecasting behaviour at a more refined level than that made possible by the use of revealed preference data. Whatever the role for SP techniques in a particular study, their contribution is made the more convincing when they are conducted in association with other, complementary methods.

In this study, group discussions and simple measures of perception provided information which influenced the design of the SP surveys, threw further light on the findings and provided support for the robustness of the results. In studies of a more quantitative nature, the most convincing results are likely to be obtained from models which combine SP and RP data (Bradley and Kroes, 1990a/b.) The case for using SP techniques is therefore strongest when they are viewed as tools complimentary to other established research methods, rather than as superior replacements.

To conclude, SP techniques have opened up a number of new opportunities for researchers to understand more precisely the processes that can influence certain behaviour (such as mode choice.) The numerous studies that have now been conducted illustrate the usefulness of these methods, but their increasingly widespread application, both in the field of transport research and beyond, should still be regarded with a degree of caution. SP techniques require careful design and survey administration and, as the case study reported here has shown, there are a number of pitfalls that may be encountered in their analysis. Such techniques must not be regarded as tools that can be used "off-the-shelf". Each study will pose specific requirements on the design and analysis of the SP technique used and care must be taken to produce an instrument that is suitable for the task in hand. Again, the use of other established research tools applied as complimentary approaches in guiding the survey design, enhancing the analysis and interpreting the findings cannot be too strongly recommended.

BIBLIOGRAPHY

Aaker, D & Day, G (1986) **MARKETING RESEARCH**, Wiley

Acito, F (1977) **AN INVESTIGATION OF SOME DATA COLLECTION ISSUES IN CONJOINT MEASUREMENT**, in Contemporary Marketing Thought, Proceedings of the American Marketing Association

Acito, F & Jain, AK (1980) Evaluation of Conjoint Measurement Results: A Comparison of Methods, **JOURNAL OF MARKETING RESEARCH**, Vol 17, pp106-112

Ampt, E & Patterson, N (1987) **VALIDATION OF RESPONSE & NON-RESPONSE IN MAIL-BACK TRAVEL SURVEYS**, PTRC 15th Summer Meeting, Bath, 7pp

Ampt, E; Bradley, M & Jones, PMJ (1987) **DEVELOPMENT OF AN INTERACTIVE, COMPUTER-ASSISTED STATED PREFERENCE TECHNIQUE TO STUDY BUS PASSENGER PREFERENCES**, PTRC 15th Summer Meeting, Bath, 18pp

Ampt, E; Bradley, MA; Ferguson, D & Jones, PMJ (1986) **TESTS OF AN INTERACTIVE STATED PREFERENCE APPROACH TO STUDY BUS SERVICE PROVISION OPTIONS**, Transport Studies Unit (ref: 326), Oxford

Andersen, PB; Moeller, J & Sheldon, R (1986) **MARKETING DSB (DANISH STATE RAILWAYS) RAIL SERVICES USING A STATED PREFERENCE APPROACH**, PTRC 14th Summer Meeting, pp263-270

Anderson, NH (1971) Integraton Theory and Attitude Change, **PSYCHOLOGICAL REVIEW**, Vol 78, No 3

Anderson, NH (1970) Functional Measurement and Psycho-Physical Judgement, **PSYCHOLOGICAL REVIEW**, Vol 77, No 3

Atkins (1984) Value of Time: a Critical Appraisal, **TRAFFIC ENGINEERING AND CONTROL**

Axhausen, KW & Smith, RL (1985) Bicyclist Link Evaluation: A Stated Preference Approach, **TRANSPORTATION RESEARCH RECORD No 1085**, Transportation Research Board, USA, pp7-15

Bates, JJ (Jan 1988) Editorial, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp8-11

Bates, JJ (Jan 1988) Econometric Issues in Stated preference Analysis, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp59-70

Bates, JJ (1986) Sensitivity to Level of Service: Evidence From Stated Preference Work, **BEHAVIOURAL RESEARCH FOR TRANSPORT POLICY**, VNU Press, Netherlands, pp269-288

Bates, JJ (1985) **SENSITIVITY TO LEVEL OF SERVICE: EVIDENCE FROM STATED PREFERENCE WORK**, International Conference on Travel Behaviour, The Hague, pp269-289

Bates, JJ (1984) **VALUES OF TIME FROM STATED PREFERENCE DATA**, PTRC 12th Summer Meeting

Bates, J; Copley, G (Jan 1990) Transport Planning Applications, Stated Preference- Special Report, **LOCAL TRANSPORT TODAY**, 24 January, 1990, pp10-13

Bates, JJ; Shepherd, NR & Van der Hoorn, ALJM (1989) **INVESTIGATING DRIVER REACTION TO PEAK HOUR SURCHARGES**, PTRC 17th Summer Meeting

Bates, JJ; Ashley, D & Hyman, G (1987) **THE NESTED INCREMENTAL LOGIT MODEL: THEORY & APPLICATION TO MODAL CHOICE**, PTRC 15th Summer Meeting, Bath, pp75-97

Bates, JJ & Roberts, M (1986) **VALUE OF TIME RESEARCH: SUMMARY OF METHODOLOGY & FINDINGS**, PTRC 14th Summer Meeting, 22pp

Bates, JJ & Roberts, M (1983) **RECENT EXPERIENCE WITH MODELS FITTED TO STATED PREFERENCE DATA**, PTRC 11th Summer Meeting, pp61-82

Beesley, ME & Hensher, DA (1987) The Consistency of Choice Using Experimental Design, **JOURNAL OF ECONOMIC PSYCHOLOGY**

Bell, MC (1976) **PASSENGER TRANSPORT INTERCHANGE BETWEEN INLAND SURFACE MODES - A STATE OF THE ART REVIEW**, Direct Report No 10, TORG, University of Newcastle-upon-Tyne, Newcastle, 149pp

Ben-Akiva M & Lerman, S (1985) **DISCRETE CHOICE ANALYSIS: THEORY & APPLICATION TO TRAVEL DEMAND**, MIT Press, Cambridge, Mass.

Ben-Akiva M & Morikawa, T (1990) **ESTIMATION OF TRAVEL DEMAND MODELS FROM MULTIPLE DATA SOURCES**, Paper presented to the 11th International Symposium on Transportation and Traffic Theory, Yokohama

Benjamin, J (1985) **UTILISATION OF ATTITUDINAL MEASUREMENT TECHNIQUES TO ANALYSE DEMAND FOR TRANSPORTATION: METHODS APPLICATIONS & NEW DIRECTIONS**, International Conference on Travel Behaviour, pp386-406

Berthoud, R & Jowell, R (1973) **CREATING A COMMUNITY: A STUDY OF RUNCORN NEW TOWN**, Social & Community Planning Research, London

Bhattacharyya, GK & Johnson, RA (1977) **STATISTICAL CONCEPTS & METHODS**, John Wiley & Sons, Singapore, 639pp

Bonsall, PN (1985) Transfer Price Data - Its Definition, Collection and Use, in Ampt et al (eds) **NEW SURVEY METHODS IN TRANSPORT**, Oxford

Bonsall, PN (1983) **TRANSFER PRICE DATA - ITS USE AND ABUSE**, PTRC 11th Summer Meeting

Bonsall, PN & Kroes, EP (1985) The Use of Sophisticated Stated Preference Techniques in the Air Transport Industry, **ITA MAGAZINE**, Vol 21, pp15-18

Bottom, C & Jones, PMJ (1982) **DON'T FORGET THE DRIVER - A PILOT STUDY OF ATTITUDES TO SCHEDULES & WORKING CONDITIONS**, Transport Studies Unit (ref 187/CP), Oxford, 25pp

Bradley, M (Jan 1988) Realism & Adaption in Designing Hypothetical Travel Choice Concepts, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp121-137

Bradley, M & Gunn, H (1990) **A STATED PREFERENCE ANALYSIS OF VALUES OF TIME IN THE NETHERLANDS**, PTRC 18th Summer Meeting

Bradley, M & Kroes, K (1990a) **FORECASTING ISSUES IN STATED PREFERENCE SURVEY RESEARCH**, paper presented to the 3rd International Conference on Survey Methods in Transportation, Washington DC

Bradley, M & Kroes, K (1990b) **SIMULTANEOUS ANALYSIS OF STATED PREFERENCE AND REVEALED PREFERENCE INFORMATION**, PTRC 18th Summer Meeting

Bradley, M; Mapp,D; Sheldon R & Steer, J (1989) **BUSINESS TRAVEL DECISON MAKING - ITS A COMPLEX BUSINESS**, PTRC 17th Summer Meeting

Bradley, M; Grosvenor, T & Bouma, A (1988) **AN APPLICATION OF COMPUTER-BASED STATED PREFERENCE TO STUDY MODE-SWITCHING IN THE NETHERLANDS**, PTRC 16th Summer Meeting

Bradley, M & Jones, PM (1987) Development and Assessment of an Interactive Computer-Based Technique to Measure Bus Passenger Preferences, **TRANSPORT AND ROAD RESEARCH LABORATORY CONTRACTOR REPORT 58**, Crowthorne 60pp

Bradley, M; Marks, P & Wardman, M (1986) **A SUMMARY OF FOUR STUDIES INTO THE VALUE OF TRAVEL TIME SAVINGS**, PTRC 14th Summer Meeting, pp1-14

Bradley, M & Bovey, PHL (1986) Functional Measurement of Route Choice Behaviour: A Study of Commuting Cyclists in the Netherlands, **BEHAVIOURAL RESEARCH FOR TRANSPORT POLICY**, VNU Science Press, pp289-305

Bradley, M & Bovey, PHL (1985) **A STATED PREFERENCE ANALYSIS OF BICYCLIST ROUTE CHOICE**, PTRC 13th Summer Meeting, pp39-53

Bradley, J & Maw, JR (Sep 1977a) **ESTIMATES FOR BUS STATION IMPROVEMENTS**, London Transport Marketing Technical Note MTN11, London, 6pp

Bradley, J & Maw, JR (Sep 1977b) **BUS IMPROVEMENTS SURVEY**, London Transport Memorandum M52a, London, 5pp

Bradley, J & Maw, JR (Nov 1975) **UNDERGROUND IMPROVEMENTS SURVEY**, London Transport Memorandum M37, London, 10pp

Bradley, J; Maw, JR & Muir, RM (July 1977) **TRAFFIC GENERATION ESTIMATED FOR THE STATION IMPROVEMENT PROGRAMME**, London Transport Marketing Technical Note MTN 10, London, 7pp

British Rail Research, March 1990, **FEASIBILITY OF SALOON AIR CONDITIONING ON FUTURE ROLLING STOCKS**, Report commissioned by London Underground Ltd.

British Rail (Undated) **ANALYSIS OF THE PRIORITY EVALUATOR - INTERNAL REPORT**, BR (Swindon), pp8

Brog, W & Erl, E (1981) **DEVELOPMENT OF A MODEL OF INDIVIDUAL BEHAVIOUR TO EXPLAIN AND FORECAST DAILY ACTIVITY PATTERNS IN URBAN AREAS**, PTRC 9th Meeting

Brookshire, D; Ives, B & Schultse, W (1976) The Valuation of Aesthetic Preferences, **JOURNAL OF ENVIRONMENTAL ECONOMICS & MANAGEMENT**, Vol 3, pp325-346

Carmone, FJ, Green, PE & Arun, KJ (May 1978) Robustness of Conjoint Analysis; Some Monte Carlo Results, **JOURNAL OF MARKETING RESEARCH**, Vol 15 pp300-303

Chapin, FS (1974) **HUMAN ACTIVITY PATTERNS IN THE CITY**, John Wiley & Sons, USA

Chapman, RG (1984) An Approach to Estimating Logit Models of a Single Decision Maker's Choice Behaviour, **ADVANCES IN CONSUMER RESEARCH**, Vol 11, pp656-661

Chapman, RG & Staelin, R (1982) Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model, **JOURNAL OF THE MARKETING RESEARCH SOCIETY**, Vol 19 pp288-301

Copley, G & Bates JJ (1988) **USING THE PRIORITY EVALUATOR TO MEASURE TRAVELLER NEEDS**, Oxford Conference on Travel & Transportation, 14pp

Copley, G; Bouma, A & de Graaf, E (1987) **STATION INVESTMENT: PRIORITIES & VALUES**, "Energy, Transportation & the Environment", Proceedings from the ESOMAR Conference, Montreux, Switzerland, pp21-45

Crittle, FJ & Johnson, LW (Oct 1980) **BASIC LOGIT (BLOGIT) - TECHNICAL MANUAL**, ATM No 9, Australian Road Research Board, Australia, 66pp

Currim, I (1981) Using Segmentation Approaches for Better prediction & Understanding from Consumer Mode Choice Models, **JOURNAL OF MARKETING RESEARCH**, Vol 18 pp301-309

Currim, I (1982) Predictive Testing of Consumer Choice Models not Subject to Independence of Irrelevant Alternatives, **JOURNAL OF MARKETING RESEARCH**, Vol 19 pp208-222

Department of Transport (1991) **GUIDELINES FOR PUBLIC TRANSPORT PROJECT APPRAISAL: APPLICATIONS FOR SECTION 56 GRANT**, Annex A to Circular 3/89

Dobson, R (1975) Towards the Analysis of Attitudinal and Behavioural Responses to Transportation System Characteristics, **TRANSPORTATION**, Vol 4 pp267-290

Donnelly, EP; Howe, SM & DesChamps, JA (1976) **TRADE-OFF ANALYSIS: THEORY & APPLICATIONS TO TRANSPORTATION POLICY ANALYSIS**, Preliminary Research Report 103, Planning Research Unit, New York State Department of Transportation, Albany, New York, USA

Fowkes, T (1991) **RECENT DEVELOPMENTS IN STATED PREFERENCE TECHNIQUES IN TRANSPORT RESEARCH**, PTRC 19th Annual Meeting, Sussex, pp251-263

Fowkes, T & Wardman, M (Jan 1988) The Design of Stated Preference Experiments, With Special Reference to Inter-personal Taste Variations, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp27-44

Fowkes, T & Tweddle, G (1988) **A COMPUTER GUIDED STATED PREFERENCE EXPERIMENT FOR FREIGHT MODE CHOICE**, PTRC 16th Annual Meeting, pp295-305

Gensch, D & Recker, W (1979) The Multinomial Multiattribute Logit Choice Model, **JOURNAL OF MARKETING RESEARCH**, Vol 16 pp14-32

Golob, T and Golob J (1982) **A CLASSIFICATION OF APPROACHES TO TRAVEL BEHAVIOR ANALYSES**, Resource paper for the Conference on Travel Behaviour and Values, Easton Maryland, USA

Golob, T (1980) **TRANSIT RIDERS' ATTITUDES TO BUS AND TRAIN**, Unpublished Research Document.

Goodwin, P (ed) (1991) **FAIR ASSESSMENT FOR TRANSPORT EXPENDITURE**, Transport 2000, London

Gore, C; Kroes, E & Sheldon, R (1988) **HOW DO RAIL PASSENGERS CHOOSE WHEN TO TRAVEL?**, Oxford Conference on Travel & Transportation, 17pp

Green, P (1984) Hybrid Models for Conjoint Analysis: an Expository Review, **JOURNAL OF MARKETING RESEARCH**, Vol 21 pp155-169

Green, P; Carmone, F & Wachspress, D (1977) On the Analysis of Qualitative Data in Marketing Research, **JOURNAL OF MARKETING RESEARCH**, Vol 14 pp52-59

Green, P & Srinivasan, V (1978) Conjoint Analysis in Consumer Research: Issues & Outlook, **JOURNAL OF CONSUMER RESEARCH**, Vol 5 pp103-121

Gunn, HF (1984) **AN ANALYSIS OF TRANSFER PRICE DATA**, PTRC, pp1-13

Hague Consulting Group (1989), **ALOGIT USERS MANUAL**, Netherlands

Hague Consulting Group (1990), **MINT**; (1989) **SPEED**, Netherlands

Hartgen (1981) Behavioural Research in Transport, in **PSYCHOLOGICAL APPROACHES TO HUMAN BEHAVIOUR**, Vol 3, New York

Hamilton, A (July 1976) Birmingham International Station, in **ARCHITECTS JOURNAL**, pp114-125

Harris, N and Godward, W (Jan 1992) **PLANNING PASSENGER RAILWAYS**, Transport Publishing Company, 255pp

Hatch, D & Flack, M (1974) Priority Research Techniques Applied to Public Transport Investment, **JOURNAL OF THE MARKET RESEARCH SOCIETY**, Vol 16 No 1 pp1-11

Hausman, J & Ruud, P (1986) Specifying & Testing Econometric Models for Rank-Ordered Data, **JOURNAL OF ECONOMETRICS**

Height, F & Cresswell, R (1978) **DESIGN FOR PASSENGER TRANSPORT**, Pergamon, Oxford

Hensher, DA (Oct 1985) **BLOGIT - AN ABRIDGED USERS' GUIDE**, Australian Road Research Board, ATM 17, Australia, 88pp

Hensher, DA (1984) Achieving Representativeness of the Observable Component of the Indirect Utility Function in Logit Choice Models: An Experimental Revelation, **JOURNAL OF BUSINESS**, Vol 57 pp265-286

Hensher, DA; Barnard, PO & Truong, TP (Jan 1988) The Role of Stated Preference Methods in Studies of Travel Choice, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp45-58

Hensher, DA & Louviere, JJ (Sep 1983) Identifying Individual Preferences for International Air Fares, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 17 pp225-245

Hensher, DA & Louviere, JJ (1979) Behavioural Intentions as Predictors of Very Specific Behaviour, **TRANSPORTATION**, Vol 8 pp167-182

Hensher, DA & Stopher, PR (1979) **BEHAVIOURAL TRAVEL MODELLING**, Croom Helm, London

Hensher, DA & Truong, TP (Sep 1985) Valuation of Travel Time Savings - A Direct Experimental Approach, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 19 pp237-261

Hoinville, G (Sep 1977) **THE PRIORITY EVALUATOR METHOD**, SCPR Working Paper No 3, Social & Community Planning Research, London, 52pp

Hoinville, G (1973) Evaluating Community Preferences, **JOURNAL OF THE MARKET RESEARCH SOCIETY**, Vol 15 pp1-23

Hoinville, G (1971) Evaluating Community Preferences, **ENVIRONMENT & PLANNING**, London, Vol 3 pp33-50

Hoinville, G & Berthoud, R (1970a) **IDENTIFYING PREFERENCE VALUES: REPORT ON DEVELOPMENT WORK**, Social & Community Planning Research, London

Hoinville, G & Berthoud, R (1970b) **VALUE OF TIME: DEVELOPMENT PROJECT**, Report on Stage 3, Social & Community Planning Research, London

Hoinville, G & Berthoud, R (1970c) **IDENTIFYING & EVALUATING TRADE-OFF PREFERENCES - AN ANALYSIS OF ENVIRONMENTAL/ACCESSIBILITY PRIORITIES**, Social & Community Planning Research, London

Hoinville, G & Courtenay, G (May 1978) **MEASURING CONSUMER PRIORITIES**, Methodological Working Paper No 15, Social & Community Planning Research, London, 59pp

Hoinville, G & Johnson, E (Dec 1971) **THE IMPORTANCE & VALUES COMMUTERS ATTACH TO TIME SAVINGS**, Social & Community Planning Research, London, P.191, 72pp

Holden, D (1992) **DESIGN PROCEDURES FOR STATED PREFERENCE EXPERIMENTS**, UTSG Conference, Newcastle

Holden, D, Fowkes, T & Wardman, M (1992) **AUTOMATIC STATED PREFERENCE DESIGN ALGORITHMS**, PTRC 20th Annual Summer Meeting, Manchester

Huber, GP (June 1974) Multi-attribute Utility Models: a Review of Field & Field-like Studies, **MANAGEMENT SCIENCE**, Vol 20 No 10 pp1393-1401

Holbrook, MB & Moore, WL (1981) Feature Interactions in Consumer Judgements of Verbal Versus Pictorial Presentations, **JOURNAL OF CONSUMER RESEARCH**, Vol 8 No 1 pp103-113

Institution of Mechanical Engineers, 1988, **RAIL VEHICLES FOR PASSENGERS**, International Conference Proceedings.

Jain, AK; Acito, F; Malhotra, NK & Mahajan, V (Aug 1979) A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models, **JOURNAL OF MARKETING RESEARCH**, Vol 16 pp313-322

Jansen, GRM; Bovy, PHL; van Est, JPJM & le Clerq, F (1979) **NEW DEVELOPMENTS IN MODELLING TRAVEL DEMAND & URBAN SYSTEMS**, Saxon House, Farnborough, pp13-36

Johnson, R (1985) **ADAPTIVE CONJOINT ANALYSIS**, Sawtooth Software Inc., Idaho

Johnson, R (1980) Measurement of Consumer Values Using Computer Interviewing Techniques, in **MARKET MEASUREMENT AND ANALYSIS**, MSI, Cambridge, Mass.

Johnston, RH; Bates, JJ; Roberts, M (1989) **A SURVEY OF PEAK SPREADING IN LONDON : METHODOLOGY AND INITIAL RESULTS**, PTRC, 140pp

Jones, PMJ (1985) **RESPONDENT REACTIONS TO THE VOT PILOT SP BUS EXPERIMENT**, Internal Note, Transport Studies Unit, Oxford, 10pp

Jones, PMJ, Dix, M & Carpenter, L (1983) **UNDERSTANDING TRAVEL BEHAVIOUR**, John Wiley

Jones, PMJ (Dec 1982) **SPECIAL TECHNIQUES - lecture notes for a PTRC course on survey research techniques**, Transport Studies Unit (ref: 203), Oxford, 24pp

Jones, PMJ & Bottom, C (Dec 1982) **THE DEVELOPMENT OF TECHNIQUES TO INVESTIGATE & MONITOR DRIVER ATTITUDES**, Transport Studies Unit (ref: 198/PR), Oxford, 144pp

Knight, Blackledge, Brocklebank, Rasbash (Sep 1989) **THE REINTRODUCTION OF SUBURBAN RAIL PASSENGER SERVICES IN STAFFORDSHIRE AND THE WEST MIDLANDS**, PTRC 17th Summer Meeting, 3pp

Kocur, G; Adler, T; Hyman, W & Aunet, B (Sep 1981) **GUIDE TO FORECASTING TRAVEL DEMAND WITH DIRECT UTILITY ASSESSMENT**, Urban Mass Transportation Administration, USA, Rpt No UMTA-NH-11-0001-82-1

Krantz, DH; Tversky, A (1971) Conjoint-Measurement Analysis of Composition Rules in Psychology, **PSYCHOLOGICAL REVIEW**, Vol 78, No.2, pp151-169

Kroes, EP; Sheldon, RJ; Swanson, J (1984) **DEVELOPING CHOICE MODELS USING STATED PREFERENCE RESEARCH**, PTRC 12th PTRC meeting, 11pp

Kroes, EP & Sheldon, RJ (1988) Stated Preference Methods: An Introduction, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1

Kroes, EP & Sheldon, RJ (1985) The Use of Attitude Models & Stated Preference Models in Practical Transport Analysis, **BEHAVIOURAL RESEARCH FOR TRANSPORT POLICY**, 309-324

Kroes, EP; Sheldon, RJ (July, 1984) **STATED PREFERENCE TECHNIQUES IN MEASURING TRAVEL ELASTICITIES**, Tenth Transportation Planning Research Colloquium, Zandvoort, The Netherlands, Dec 1983, 21pp

Kruskal, J (1965) Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data, **JOURNAL OF ROYAL STATISTICAL SOCIETY**, Series B, Vol 27, pp251-263

Leigh, TW, McKay, DB & Summers, JO (1984) Reliability and Validity of Conjont Analysis and Self-Explicated Weights: A Comparison, **JOURNAL OF MARKETING RESEARCH**, Vol 21, pp456-462

Lerman, SR & Louviere, JJ (1978) On the Use of Functional Measurement to Identify the Functional Form of the Utility Expression in Travel Demand Models, **TRANSPORTATION RESEARCH RECORD No 673**, Transportation Research Board, USA, pp78-86

Levin, I; Louviere,A, Schepaski & Norman, K (1983) **EXTERNAL VALIDITY OF TESTS OF LABORATORY STUDIES OF INFORMATION INTEGRATION**, in Organistational Behaviour and Human Performance, Vol 31 pp173-193

London Transport, 1984, **1983 RAIL PRIORITY SURVEY**, Development Director (Railways)

London Underground Ltd, Nov 1987, **UNDERGROUND QUALITY OF RIDE SURVEY**, OR 87/22

London Underground Ltd, Dec 1985, **A SURVEY OF PASSENGER NOISE ANNOYANCE ON THE UNDERGROUND**, OR 85/2

Louviere, JJ (Jan 1988) Conjoint Analysis Modelling of Stated Preferences, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp93-119

Louviere, JJ (1988) **MODELLING INDIVIDUAL DECISIONS: METRIC CONJOINT ANALYSIS - THEORY, METHODS & PRACTICAL APPLICATIONS**, Sage University Press Series, Sage Publications Inc, Newbury park, CA

Louviere, JJ (1983) Integrating Conjoint & Functional Measurement with Discrete Choice Theory: An Experimental Design Approach, **ADVANCES IN CONSUMER RESEARCH**, Vol 10 pp151-156

Louviere, JJ & Hensher, DA (Dec 1983) Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event, **JOURNAL OF CONSUMER RESEARCH**, Vol 10 pp348-361

Louviere, JJ & Hensher, DA (1982) Design & Analysis of Simulated choice or Allocation Experiments in Travel Choice Modelling, **TRANSPORTATION RESEARCH RECORD No 890**, Transportation Research Board, USA, pp11-17

Louviere, JJ; Henley, DH; Woodworth, G; Meyer, RJ; Levin, IP; Stoner, JW; Curry, D & Anderson, DA (1981) Laboratory-Simulation Versus Revealed Preference Methods for Estimating Travel Demand Models, **TRANSPORTATION RESEARCH RECORD No 794**, Transportation Research Board, USA, pp42-51

Louviere, JJ; Schroeder, H; Woodworth, GG & Louviere, CH (1987) Do The Parameters of Choice Models Depend Upon Mode of Stimulus Presentation? Visual versus Verbal Presentation, **ADVANCES IN CONSUMER RESEARCH**, Vol 14

Louviere, JJ & Woodworth, G (Nov 1983) Design & Analysis of Simulated Consumer Choice or Allocation Experiments: an Approach Based on Aggregate Data, **JOURNAL OF MARKETING RESEARCH**, Vol 20 pp350-367

Luce, RD (1978) The Choice Axiom After Twenty Years, **JOURNAL OF MATHEMATICAL PSYCHOLOGY**, Vol 15 pp215-233

Luce, RD (1959) **INDIVIDUAL CHOICE BEHAVIOUR - A THEORETICAL ANALYSIS**, John Wiley & Sons, London, 149pp

Maw, J (July 1977) **PASSENGER EVALUATION OF BUS STATION IMPROVEMENTS**, London Transport Memorandum 62, London, 10pp

Maw, J & Bradley, J (1977) Passenger Evaluation Of Underground Improvements, **GREATER LONDON INTELLIGENCE JOURNAL**, London, No 37 pp28-31

Maw, J; Muir, RM & Hendy, P (Feb 1976) **PASSENGER EVALUATION OF UNDERGROUND IMPROVEMENTS**, London Transport Memorandum M43, London, 10pp

MAS, 1988, **INTERCITY MONITOR**, InterCity

McCullough, J & Best, R (1979), **CONJOINT MEASUREMENT: TEMPORAL STABILITY AND STRUCTURAL RELIABILITY**, Journal of Marketing Research, Vol.16, pp26-31

McFadden, D (1987) The Choice Theory Approach to Market Research, **MARKETING SCIENCE**, Vol5 No 4 pp275-297

McFadden, D (1985) Regression Based Specification Tests for the Multinomial Logit Model, **JOURNAL OF ECONOMETRICS**

Montgomery, DC (1984) **DESIGN & ANALYSIS OF EXPERIMENTS**, John Wiley & Sons, New York, 538pp

Moore, L (1985) **STATED PREFERENCE ANALYSIS: A CRITICAL REVIEW**, The MVA Consultancy, Woking, 40pp

Moore, WL (1980) Level of Aggregation in Conjoint Analysis: an Empirical Comparison, **JOURNAL OF MARKETING RESEARCH**, Vol 17, pp516-523

Morikawa, T (1990) **COMBINED ESTIMATION WITH SP AND RP DATA - CASE STUDY 2**, Chapter from PhD Thesis, MIT

The MVA Consultancy, The Transport Studies Unit & Leeds ITS (1987) **THE VALUE OF TIME**, Department of Transport

The MVA Consultancy (1987) **CITY OF YORK TRAFFIC & PARKING STUDY - PARK & RIDE: SURVEY & ANALYSIS**, Working paper No 1, Woking, 29pp + app

The MVA Consultancy (1986) **INTERNAL NOTE TO BRITISH RAIL (UNTITLED)**, 5pp

The MVA Consultancy (1985a) **INTERCHANGE**, Report Commissioned by British Rail, Woking

The MVA Consultancy (1985b) **STATION INVESTMENT - Report of Stage II**, Report commissioned by British Rail (contract 391), Woking, 50pp

The MVA Consultancy (1985c) **STATION MODERNISATION PRIORITIES & PAYOFFS - PROPOSAL**, Report commissioned by British Rail (no 581), London, 24pp

The MVA Consultancy (1985d) **STATION MODERNISATION PRIORITIES & PAYOFFS - FINAL REPORT**, Report commissioned by British Rail, London

OECD (1980) **TRANSPORT CHOICES FOR URBAN PASSENGERS**, Paris

Olshavsky, RW & Acito, F (1980) An Information Processing Probe into Conjoint Analysis, **DECISION SCIENCES**, vol 2, pp451-470

Ortuzar, J & Garrido, R (1991) **RANK, RATE OR CHOICE? AN EVALUATION OF SP METHODS IN SANTIAGO**, PTRC 19th Meeting

Ortuzar, J & Willumsen, L (1991) **MODELLING TRANSPORT**, John Wiley

Parker, B & Srinivasan, V (1976) A Consumer Preference Approach to the Planning of Rural Primary Health Care Facilities, **OPERATIONS RESEARCH**, No 24, pp991-1025

Pearmain, DB (1992) **THE MEASUREMENT OF USERS' WILLINGNESS TO PAY FOR IMPROVED RAIL FACILITIES**, PTRC 20th Annual Summer Meeting, Manchester

Pearmain, DB (1989) **THE MEASUREMENT OF PASSENGER VALUATIONS TOWARDS RAIL STATION FACILITIES: AN APPLICATION OF THE PRIORITY EVALUATOR METHOD**, UTSG Conference, Edinburgh, 17pp

Pearmain, DB; Kroes, E; Swanson, J & Bradley, M (Jan 1991) **STATED PREFERENCE TECHNIQUES - A GUIDE TO PRACTICE**, Steer Davies Gleave / Hague Consulting Group, Richmond.

Pearmain, DB & Jones TSM (Nov 1985) **TRANSPORT INTERCHANGES - THEORY & PRACTICE (SUMMARY REPORT)**, Oxford Polytechnic Working Paper No 86, Oxford, 109pp

Pendse, D & Wyckoff, JB (1974) Environmental Goods: Determination of Preferences & Trade-off Values, **JOURNAL OF LEISURE RESEARCH**, Vol6 No 1 pp64-76

Polak, J & Jones, P (1991) **ASSESSING TRAVELLER RESPONSES TO ROAD PRICING OPTIONS**, 6_{th} International Conference on Travel Behaviour, Quebec

Polak, J; Jones, P & Stokes, G (1989) **COMPUTER-BASED PERSONAL INTERVIEWS: A PRACTICAL TOOL FOR COMPLEX TRAVEL SURVEYS**, PTRC 17^{th} Meeting, pp141-154

Prideaux, J (1984) **RAIL/ROAD INTERCHANGES**, Journal of the Institution of Civil Engineers, pp12-16

Preston, J & Wardman, M (1988) **DEMAND FORECASTING FOR NEW LOCAL RAIL SERVICES: A CASE STUDY OF A NEW SERVICE BETWEEN LEICESTER AND BURTON-ON-TRENT**, Leeds ITS, Working Paper 260

Rendell Planning (1992) **ENVIRONMENTAL APPRAISAL: A REVIEW OF MONETARY EVALUATION AND OTHER TECHNIQUES**, TRRL, Contractor Report No 290

RDS Ltd, Oct 1989, **NORTHERN LINE PROPOSALS: QUALITATIVE RESEARCH REPORT**, RDS2612

Sawtooth Software, 1989, **GAINING A COMPETITIVE ADVANTAGE THROUGH PC-BASED INTERVIEWING AND ANALYSIS**, Idaho USA, Vol 1

Saunders, L (Jan 1972) **VALUE OF TIME FOR AIR TRAVELLERS ON THE LAND FEEDER JOURNEY TO AN AIRPORT**, Geographical Paper No 13, University of Reading, Reading, 62pp

SDG Research (1989) **SEASON TICKET PRICING**, Richmond, 45pp

SDG Research (1987a) **PROTOTYPE TRAINS RESEARCH: FINAL REPORT**, Report No 480

SDG Research (1987b) **RIDE QUALITY RESEARCH**, Report No 386

SDG Research (1987c) **LONDON UNDERGROUND IMPROVEMENT STUDY**, Report No 371

Segal, J & Todd, R (1985) **ACCESS TO INTERCITY RAIL SERVICES**, PTRC 13th Summer Meeting, pp117-130

Segal, MN (Feb 1982) Reliability of Conjoint Analysis: Contrasting Data Collection Procedures, **JOURNAL OF MARKETING RESEARCH**, Vol 19 pp139-143

Sheldon, R; Bottom, C & Golob, T (1985) **AN EXAMINATION OF THE PRIORITIES THAT PASSENGERS PLACE UPON FEATURES OF THE LONDON UNDERGROUND SYSTEM**, PTRC 13th Summer Meeting, pp143-154

Sheldon, RJ & Steer, JK (1982) **THE USE OF CONJOINT ANALYSIS IN TRANSPORT RESEARCH**, PTRC 10th Summer Meeting, pp145-158

Steer Davies Gleave (1990) **WORKING PAPER ON THE CHESTER LRT SP SURVEYS**, Richmond

Steer Davies Gleave (1989) **FINAL REPORT ON THE LEEDS ADVANCED TRANSIT STUDY**, Richmond

Steer, Davies & Gleave Ltd (July 1987) **RESEARCH TO EVALUATE PASSENGER INVESTMENT PRIORITIES FOR REGIONAL SERVICES**, Richmond, 22pp

Steer, Davies & Gleave Ltd (1986a) **RESEARCH TO EVALUATE PASSENGER INVESTMENT PRIORITIES**, Richmond, 80pp

Steer, Davies & Gleave Ltd (1986b) **OXFORD STATION IMPROVEMENT EVALUATION**, Richmond, 30pp

Steer, Davies & Gleave Ltd (1985) **BRISTOL PARKWAY STATION IMPROVEMENT EVALUATION**, Richmond, 95pp

Steer, Davies & Gleave Ltd (July 1984a) **METHODOLOGY REVIEW: FORECASTING THE EFFECTS OF CHANGES IN ACCESS ON BR**, Richmond, 42pp

Steer, Davies & Gleave Ltd (July 1984b) **STATION CAR PARK INVESTMENT STUDY**, Report for British Railways Board, Richmond, 54pp

Steer, Davies & Gleave Ltd (Feb 1984c) **RESEARCH INTO INTERCITY SERVICE CHARACTERISTICS**, Report for British Railways Board, Richmond, 50pp

Steer, Davies & Gleave Ltd (1984d) **RESEARCH TO EVALUATE PASSENGER PRIORITIES FOR INVESTMENT IN LONDON'S UNDERGROUND**, Report for London Transport, Richmond

Steer, JK & Willumsen, L (1981) **AN INVESTIGATION OF PASSENGER PREFERENCE STRUCTURES**, PTRC 9th Summer Meeting, Warwick

Strauss, RP & Hughes, GD (1976) A New approach to the Demand for Public Goods, **JOURNAL OF PUBLIC ECONOMICS**, Vol 6 pp191-204

Swanson, J, Pearmain, DB, Loughead, K (1992) **STATED PREFERENCE SAMPLE SIZES**, PTRC 20th Annual Meeting, Manchester

THP, May 1983, **PROTOTYPE TUBE STOCK: QUALITATIVE RESEARCH UNDERTAKEN FOR LONDON TRANSPORT**, J384

Timmermans, H; van der Heijden & Westerveld, H (1982) The Identification of Factors Influencing Destination Choice: An Application of the Repertory Grid Methodology, **TRANSPORTATION**, Vol 11, pp189-203

Tipple, G (1977) Design a House Game: the use of a priority evaluator in Zambia, **TOWN PLANNING REVIEW**, Vol 48 pp141-148

Tipple, G (1972) **ARAB HOUSING PREFERENCES: A STUDY AMONG YEMENI ARABS IN ATTERCLIFFE, SHEFFIELD**, MA Thesis, University of Sheffield

Transecon International, Dec 1985, **THE EXTERIOR APPEARANCE OF UNDERGROUND TRAINS**

Transport & Road Research Laboratory (1980) **THE DEMAND FOR PUBLIC TRANSPORT**, Crowthorne

Wardman, M (1987) PhD Dissertation, Leeds University, Leeds

Wardman, M (Jan 1988) A Comparison of Revealed Preference & Stated Preference Models of Travel Behaviour, **JOURNAL OF TRANSPORT ECONOMICS & POLICY**, Vol 22 No 1 pp71-92

Watson, PL (1974) **THE VALUE OF TIME; BEHAVIOURAL MODELS OF MODAL CHOICE**, Lexington, USA

Westwood, D; Lunn, T & Beazley, D (1974) The Trade-off Model & Its Extensions, **JOURNAL OF THE MARKET RESEARCH SOCIETY**, Vol 16 No 3 15pp

Wicks, J & Beswick, M (1986) **THE NATURE OF RAIL TO RAIL INTERCHANGE**, PTRC 14th Summer Meeting, pp11-22

Wildermuth, HK (1976) **TRANSPORT POLICIES FOR CENTRAL PERTH: REVIEW & FORMULATION**, Project No WDG 75/2, Office of Director General of Transport, Perth, Western Australia

Wilson, RL (1962) Livability of the City: Attitudes & Urban Development, Chapter 11, pp359-99 in Chapin, FS & Weiss, SF (eds) (1962) **URBAN GROWTH DYNAMICS IN A REGIONAL CLUSTER OF CITIES**, Chapter 11, pp359-399, John Wiley & Sons, 484pp

Wilson, T (1983) **THE GENERALISED COST OF TRAVEL INVOLVING INTERCHANGE**, PhD Thesis, Newcastle

Wilson, TK & Bell, MC (1984) **DISAGGREGATE MODELLING OF TRAVEL INVOLVING INTERCHANGE**, PTRC 12th Summer Meeting, pp95-107

Wolfe, RA (1988) **A NON-PARAMETRIC ESTIMATION METHOD FOR RANKED UTILITY MODELS**, Oxford Conference on Travel & Transportation, 28pp

APPENDIX 1: DISCRETE CHOICE TRAVEL MODELS

The most commonly used approaches to modelling discrete choices are:

(i) the linear probability model (LPM);

(ii) the logit model;

(iii) the probit model.

The essential characteristics of each of the three approaches can be most simply understood in relation to dichotomous (or binary) choices, rather than multiple choices. Returning to the earlier car versus bus example, the aim in such a case would be to predict a value between 0 and 1, where 0 might represent no probability of choosing a bus (ie complete probability of choosing car) and 1 represents complete probability of choosing bus. A value of 0.5 will therefore indicate indifference between choosing car or bus.

The probability of choosing one option over another will be represented as a function of one or more variables, some of which may be attributes relating to the options, some characteristics of the individuals making the choices and some external variables acting on the individuals and imposing constraints on choices. With reference to the conceptual model of Figure 3.1 (chapter 3), these different types of independent variables relate to features of the transport environment, individuals' socio-economic characteristics and situational constraints.

When considering the mathematical form of the model, the researcher has to make an implicit assumption about the way the variables combine to influence choices. That is, he or she has to construct a suitable utility function, which may be simply linear (the most common approach) or more complex (eg quadratic, multiplicative.) Further decisions then have to be made regarding the way in which individuals relate the utility of one option to the other in order to make a choice.

The basic steps in this modelling approach are therefore:

(i) define a finite set of alternatives from which individuals are observed to choose a preferred alternative;

(ii) define a utility function in terms of:

  (a) the independent variables thought to influence individuals' choice processes and

  (b) the way in which the independent variables are combined to construct the utility attached to each alternative.

(iii) define a model form in which the probability of choosing an alternative is expressed in terms of the utility functions attached to each alternative in the choice set.

Steps two and three are combined in one function in the linear probability model but are defined separately in logit and probit models. Using the car versus bus example (a finite set of two alternatives), the researcher might assume a linear additive utility function, the component attributes of which are the differences in cost and journey time between the two modes.

In such a case, the three model approaches discussed above would take the following forms:

(i) Linear Probability Model:

$$P_{car} = a_0 + a_1*(\text{bus time-car time}) + a_2*(\text{bus cost-car cost}) + e$$

Where: P_{car} = Probability of choosing car

a_0 = Constant

$a_1..a_2$ = parameters

e = random error term

(ii) Binary Logit Model:

$$P_{car} = 1/(1 + \exp^{-U})$$

Where: $U = a_0 + a_1*(\text{bus time-car time}) + a_2*(\text{bus cost-car cost}) + e$

= the utility of car over bus

All other terms are as above

(iii) Probit Model:

$$P_{car} = 1/\sqrt{2\pi} \int_{-\infty}^{U} exp^{-t^2/2} dt$$

Where: $U = a_0 + a_1*(\text{bus time} - \text{car time}) + a_2*(\text{bus cost} - \text{car cost})$

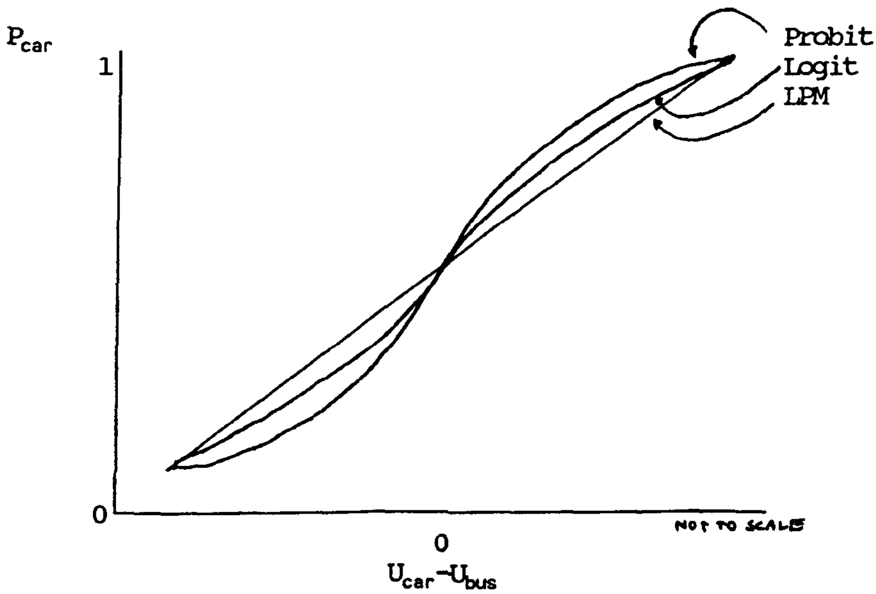
= the utility of car over bus

t = standardised normal variable

All other terms are as above

The simplest way to compare these modelling approaches is to look at the shapes of the curves they produce when the probability of a particular choice is graphed against the utility function. Figure A1.1 illustrates the curves produced from the binary versions of each of the three modelling approaches. In this diagram, the main limitation of the Linear probability model is most clearly seen. As its name implies, the LPM model will produce a linear probability curve, such that the marginal effect on the choice probability of changes to the utilities is constant throughout. In terms of representing travel behaviour, this is not very plausible: it implies, for example, that a 10 pence increase in the cost of bus fares at a point when the utility difference of car and bus is small (ie $P_{car} \approx 0.5$) will induce the same increase in the probability of choosing car as when the utility difference is large and negative or large and positive (eg $P_{car} \approx 0.9$ or $P_{car} \approx 0.1$ respectively.)

Figure A1.1 Illustrations of the Probability Curves Produced by Different Discrete Choice Travel Models



In this respect, logit and probit models produce more plausible non-linear probability curves. In both cases, the S-shaped curves represent the cumulative distribution function of a random variable. The curves ensure that the choice probability approaches zero at an ever decreasing rate as utility diminishes and approaches the value 1 at ever decreasing rates as utility becomes very large. In the context of travel behaviour and the car versus bus example, this is conceptually more attractive. Travellers with a competitive choice between car and bus will be more sensitive to small changes in the attributes of these modes than travellers with a clearly superior car or bus journey.

It is for this reason the logit and probit models are more attractive than the linear probability model, despite their greater computational complexity. It can also be seen from Figure A1.1 that there is little difference between the cumulative distribution functions of logit and probit, the latter offering a slightly steeper gradient around $P^{car} \approx 0.5$ and a faster approach to the axes than the logistic curve. Given that the computation of the logit model is easier than the probit model, and a wider range of suitable software packages are therefore available for it, the logit model is the most commonly used.

The following has been excluded at the request of the university

Appendix 2

APPENDIX 3: NOTE ON THE PILOT STUDY

REPORT ON PILOT SURVEYS CONDUCTED FOR THE NORTHERN LINE PROJECT (CONTRACT NL005; NO.736)

1 Introduction

The pilot surveys were carried out in two locations:

(i) Tooting Bec (9 May 1990)

(ii) Goodge Street (10 May 1990)

A full day's interviewing was carried out at each location, with target sample sizes of 40-50 interviews per site. The pilot survey was a large one because we wished to examine strike rates for recruitment over most of a day in one inner and one outer London location.

The interviews were also quite lengthy, with three stated preference (SP) games in each. We were concerned to see how this might effect the quality of the responses and the completion rates.

2 Questionnaire Design

The interviews were conducted on portable computers. Specialised SP interview software was used (the "Game Generator") to design the computer questionnaire. The broad structure was as follows:

(i) Details of respondents' Northern Line journey;

(ii) "Within mode" SP exercises involving trade-offs between fare, individual train attributes and sometimes individual station attributes (only a sub-sample completed a station attribute SP exercise);

(iii) A "between-mode" SP exercise, in which respondents' traded **packages** of train attributes with Underground fare, time and service frequency. They made choices within a mode choice context (ie. Underground versus their best alternative).

(iv) Details of respondents' socio-economic characteristics.

Four SP exercises were used altogether, but each respondent only undertook three in each interview (this was considered to be the practical limit). Every respondent completed the final "between mode" exercise. For the "within mode" exercises, these were selected randomly. The selection was weighted so that train attributes would receive more coverage, because these were of principal interest.

Therefore, from a total of 100 respondents, each undertaking two "within mode" exercises, the following division of responses would be expected:

(i) First train attribute exercise: 75 respondents

(ii) Second train attribute exercise: 75 respondents

(iii) Station/external train attribute exercise: 50 respondents

3 Survey Administration

Respondents were recruited at or near the Underground stations. The computers were set up in hired rooms nearby. This approach to interviewing is often referred to as "hall tests".

Recruitment was fairly easy during the off-peak periods, but very difficult at peak times. This reflected travellers' greater time constraints at these times of the day. The result was a lower proportion of peak to off-peak travellers. For the main survey, we will aim to recruit peak passengers at station platforms, with a view to inviting them to the hall test in the evening. This may require direct payment as an incentive, although the "prize draw" system worked well for off-peak respondents. A total of 91 passengers were recruited, of which 89 completed the interview.

4 Sample Structure

A summary of respondents' principal Underground journey characteristics is given in **Table 1.** Despite the greater difficulty in recruiting peak travellers, a reasonable number of commuters were interviewed. Note the large proportions who only used the Northern Line during their Underground journey. The considerable difference in journey cost, between period travelcard/pass and other ticket users suggests the importance of modelling the SP responses separately. This is because the absolute values of their valuations will differ in relation to the ticket price.

Table 1: RESPONDENTS' PRINCIPAL UNDERGROUND JOURNEY CHARACTERISTICS (PILOT SURVEY, MAY 1990)

	Survey Location		
	Goodge Street	**Tooting Bec**	**Total**
Purpose:			
Work	23	13	36
Business	3	7	10
Other	18	25	43
Used Other Lines:			
Yes	20	16	36
No	24	29	53
Average Journey Time:	25 mins	28 mins	27 mins
Type of Ticket:			
Travelcard/pass	17	17	34
Other (including one day Travelcard)	27	28	65
Average Cost			
Travelcard/pass	£65.99	£41.60	£53.80
Other (inc. one day Travelcard)	£1.79	£1.91	£1.85
Total:	44	45	89

5 Perceptions of Present Northern Line Services

Table 2 summarises respondents' perceptions of the present level of service given by Northern Line trains. There is a marked difference between the survey sites regarding perceptions of crowding and cleanliness, while the remaining items are generally similar. The quality of air is rated very poorly, while a large proportion found their train only moderately noisy or bumpy.

6 Conclusion

At the time of writing, the data is still being processed for logit analysis, so that statistical models have yet to be developed.

From the pilot study, the following modifications to the survey are recommended:

(i) additional recruitment needs to be carried out in the peak period;

(ii) the within-mode SP exercise need to be shortened if possible. This may necessitate using smaller "blocks" of options, to maintain the present number of variables.

Other than these changes, the survey method proved workable and most respondents were observed to understand the items being considered and to make considered responses to the SP exercises.

Table 2: RESPONDENTS' PERCEPTIONS OF CURRENT NORTHERN LINE CONDITIONS (PILOT SURVEY, MAY 1990)

	Survey Location		
	Goodge Street	Tooting Bec	Total
	(Figures are in percentages)		
Crowding:			
All the time	48	18	33
Some of the time	30	27	28
Not crowded	22	55	39
	(100)	(100)	(100)
Cleanliness:			
Dirty/vandalised	25	13	19
Dirty	41	67	54
Clean	34	20	27
	(100)	(100)	(100)
Noise:			
Noisy	45	51	48
Fairly quiet	45	47	46
Very quiet	10	2	6
	(100)	(100)	(100)
Ride:			
Bumpy	48	36	41
Fairly smooth	48	56	52
Very smooth	4	8	7
	(100)	(100)	(100)
Air:			
Stuffy	75	62	69
Not stuffy or fresh	25	33	29
Fresh	0	5	2
	(100)	(100)	(100)

"WITHIN-MODE" SP EXERCISES (PILOT)

<u>Exercise 1</u>

Attribute	Monetary Valuation (As % of fare)		
	G oodge Street	Tooting Bec	
Dirty Train	+ 51.4%	+ 25.5%	Base = Dirty/Vandalised Train
Clean Train	+ 58.2%	+ 28.7%	
Refurbished Train	+ 3.0%	- 1.0%	Base = Old
Brand New Train	+ 18.0%	+ 2.35%	
"Forced Air"	+ 8.2%	+ 6.1%	Base = Poor Ventilation
Air conditioning	+ 60.4%	+ 26.7%	
Gangways	+ 10.6%	+ 4.1%	Base = No Gangways
Gangways with Phone Link	+ 40.0%	+ 17.3%	

Exercise 2

Attribute	Monetary Valuation (As % of fare)		
	Goodge Street	Tooting Bec	
Fairly Noisy Train	+ 1.81%	+ 5.5%	Base = Noisy Train
Quiet Train	+ 13.5%	+ 27.0%	
Fairly Smooth Train	+ 3.3%	+ 9.8%	Base = Bumpy Train
Very Smooth Train	+ 27.01	+ 22.2%	
Indicator Boards	+ 9.2%	+ 14.1%	Base = Minimum Information
Indicator and Announcements	+ 21.6%	+11.5%	
Gangways	+ 4.4%	- 1.3%	Base = No Gangways
Gangways with Phone Link	+ 28.9%	+ 16.9%	

MONETARY VALUATIONS

	Improvements 1		Improvements 2	
	Medium	**Good**	**Medium**	**Good**
<u>Goodge St</u>				
Travelcard/Pass	£1.75 (2.65%)	£15.76 (23.9%)	Not enough data	
Single/Return One Day TC	£0.02 (1.1%)	£1.06 (59.2%)	£0.61 (34.1%)	£0.98 (54.7%)
<u>Tooting Bec</u>				
Travelcard/Pass	£95.80 (230.3%)	£117.17 (281.7%)	Not enough data	
Single/Return One Day TC	£0.84 (44.0%)	£3.94 (206.3%)	£11.52 (593.8%)	£11.73 (604.6%)

Improvements 1 = newness, cleanliness, air quality, gangways

Improvements 2 = noise, ride quality, information, gangways

APPENDIX 4: FORMULAE FOR THE CALCULATION OF THE VARIANCE OF THE RATIO OF TWO RANDOM VARIABLES

FORMULAE FOR THE VARIANCE OF A FUNCTION OF RANDOM VARIABLES

A1.1 If $\underline{b}$ is a vector of coefficients, with variance-covariance matrix $\underline{\underline{V}}$, we can derive a formula for the variance of a scalar function of $\underline{b}$, say $h(\underline{b})$ as follows.

The variance of any random variable y is given as

$$\text{Var}(y) = E[y - E(y)]^2 = E(y^2) - [E(y)]^2$$

where E denotes expected value.

Suppose $E(\underline{b}) = \underline{\beta}$. We now expand the function $h(\underline{b})$ as a Taylor series around $\underline{\beta}$, as follows:

$$h(\underline{b}) \sim h(\underline{\beta}) + (\underline{b} - \underline{\beta}).\ \underline{\nabla h} + (\underline{b} - \underline{\beta})^T.\ \nabla^2 h.(\underline{b} - \underline{\beta})/2 + \ldots\ldots\ldots\ldots \quad (1)$$

ignoring terms higher than second order in $(\underline{b} - \underline{\beta})$.

A1.2 We wish to obtain $\text{Var}[h(\underline{b})] = E([h(\underline{b})]^2) - [E(h(\underline{b}))]^2$

By substitution, we have

$$E(h(\underline{b})) = h(\underline{\beta}) + E[(\underline{b} - \underline{\beta})^T.\ \nabla^2 h.\ (\underline{b} - \underline{\beta})]/2 + \ldots\ldots \quad (2)$$

and
$$[h(\underline{b})]^2 = [h(\underline{\beta})]^2 + 2h(\underline{\beta}).(\underline{b} - \underline{\beta}).\ \underline{\nabla h}$$
$$+ (\underline{b} - \underline{\beta})^T.\ \underline{\nabla h}^T.\ \underline{\nabla h}.(\underline{b} - \underline{\beta})$$
$$+ h(\underline{\beta}).\ [(\underline{b} - \underline{\beta})^T.\ \nabla^2 h.(\underline{b} - \underline{\beta})] + \ldots\ldots \quad (3)$$

Since $E(\underline{b} - \underline{\beta}) = 0$, we obtain

$$\text{var}[h(\underline{b})] \sim E[\underline{\nabla h}^T.(\underline{b} - \underline{\beta})^T(\underline{b} - \underline{\beta}).\underline{\nabla h}]$$

$$= \underline{\nabla h}^T . V . \underline{\nabla h}$$

$$= \sum_{i}\sum_{j} \left(\frac{\partial h}{\partial b_i}\right)\left(\frac{\partial h}{\partial b_j}\right) V_{ij} \quad (4)$$

A1.3 Two special cases are of interest:

a) To obtain the variance of the difference between two coefficients b_1 and b_2.

We have $h(b_1, b_2) = b_1 - b_2$; $\frac{\partial h}{\partial b_1} = 1$, $\frac{\partial h}{\partial b_2} = -1$

Hence
$$\text{var}(b_1 - b_2) = \text{var}(b_1) + \text{var}(b_2) - 2\,\text{cov}(b_1, b_2) \quad (5)$$

b) To obtain the variance of the ratio between two coefficients b_1 and b_2 (as in estimating values of time).

We have $h(b_1, b_2) = b_1/b_2$; $\frac{\partial h}{\partial b_1} = \frac{1}{b_2}$, $\frac{\partial h}{\partial b_2} = -\frac{b_1}{b_2^2}$

Hence $$\text{var}(b_1/b_2) = \frac{1}{b_2^2}\left[\text{var}(b_1) - 2\frac{b_1}{b_2}\text{cov}(b_1, b_2) + \frac{b_1^2}{b_2^2}\text{var}(b_2)\right] \quad (6)$$

if we write $b_1 = \psi$, $b_2 = \lambda$ and $\text{vot} = \psi / \lambda$

then $$\text{var}(\text{vot}) = \frac{1}{\lambda^2}\left[\text{var}(\psi) - 2\,\text{vot}\,\text{cov}(\psi, \lambda) + \text{vot}^2\,\text{var}(\lambda)\right] \quad (7)$$

All variances of value of time, etc used in the study have been derived using the approach set out in this Appendix.

Source: MVA et al (1987) The Value of Time, Department of Transport, appendix 1