The Effects of Stated Preference Design on Bias in Responses

By

Hui Lu

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Institute for Transport Studies

June, 2007

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to pay special thanks to my supervisors, Dr Tony Fowkes and Prof. Mark Wardman, for their help, encouragement and supervision throughout this research. I would also like to thank GMPTE and ITS, University of Leeds for funding support of the SP survey, and to Northern Rail Ltd for permission to survey on their stations. Particularly, I would like to thank Mr. Ian Palmer and Mr. Steve Magner from GMPTE, and Miss Pat Beijer from Northern Rail Ltd. for their help with the SP survey.

Thanks are also due to Dr. Gerard Whelan and Dr. Richard Batley for help in getting started using the ALOGIT and GAUSS program and discrete choice models. I would also like to thank Mr. Caussade and Prof. Ortuzar for sharing the code for heteroskedastic logit model analysis. Thanks are due to Prof. Bonsall for the discussion of decision behaviour and Dr. Nicolas Ibanez for the discussion of choice models. Special thanks to Jamie, Clare, Sunny and many friends in ITS who kindly helped me and took the time to proofread and comment on the draft chapters.

I have greatly benefited from the teaching experience offered by demonstration in transport courses for Master and Undergraduate students, and am grateful to Dr Paul Firmin, Dr Susan Grant-Muller, Mr. Frank Montgomery, Dr Haibo Chen, Dr Nick Marler and Dr. Greg Marsden for offering this opportunity.

I also wish to thank my friend, Jiao, Hedi, Kaushali and Na who gave warm encouragement throughout my difficult time. I wish to thank all those people who have supported and encouraged during this study.

I am deeply indebted to my husband Bo Peng, who has given me encouragement and full loving support, and has been patient throughout the long difficult time. Finally, I do not have the appropriate words to express my deep thanks to my parents and sister, for loving support and continuous encouragement throughout the research and absence abroad. Anything was not possible without their help.

Abstract

Stated Preference (SP) methods have been used extensively in transport research and elsewhere both for demand forecasting purposes and to value the importance attached to different product features and travel attributes. Alongside the broader acceptance and wider application of SP methods, some practitioners (Bates, 1998; Ampt et al., 2000; Wardman and Shires, 2001) have argued for greater openness in discussing what they see as significant concerns surrounding SP. The present study is motivated by the desire to analyse and reduce biases in the SP application, specifically addressing the issue of the strategic biasing of SP responses.

The review of biases observed in the previous SP applications explored the sources of bias, which can be categorized as unrealistic design, incentive to strategic bias and task complexity effects. Amongst these, the issues of design/scenarios specification and task complexity have received a considerable amount of attention. On the other hand, and despite serious concerns in the early literature, the strategic biasing of responses tends to have been overlooked in recent times, particularly within the SP methodology. This study is motivated by the desire to investigate the incentives for respondents to bias their answer in the SP survey and methods to amend the bias.

This study reviewed and summarised concerns surrounding the extent to which the SP responses to hypothetical questions reliably reflect individuals' true preferences when there is an incentive to bias responses. The discussion was illustrated with examples from research in transport field, environment science and marketing.

In an empirical demonstration using data obtained from 1222 respondents (10885 preference observations) on the valuation of the improved rolling stock in Greater Manchester, UK, this study presented results for different designs. Based on the review of studies on rolling stock in recent years, a suite of SP experiments were designed to investigate the effects of different designs on responses. Two factors were introduced into the experiment, a 'cheap-talk' script and 'adding more attributes to mask the research aim', to amend incentives to bias. In the experiment, post-questionnaire questions on respondents' perception of experiments were introduced. More specifically, respondents' perceptions of the task load, familiarity of experiment alternatives together with their perceptions of the attribute change were added to probe the decision making process and the impact of perception on the decision making.

Standard logit models were used to demonstrate the overall effects of variables for the whole sample. The segmentation model, based on the incremental factors, was used to identify

respondents' taste variations. The heteroskedastic multinomial logit (HMNL) model was used to incorporate the impact of design factors, respondents' characteristics and perceptions into the scale parameter, which were unable to be captured by the standard logit model.

This study found that the cheap-talk script decreased the valuation of the improved rolling stock by 20% on average, through increasing respondents' sensitivity to the cost attribute in the SP survey. However, this impact was not significant at the 5% significance level. This indicates that the warning message will help individuals to amend the incentive to strategic bias in the SP experiment; however bias may remain in our study.

This study did not detect significant impact of the complex design on the valuation of the improved rolling stock, although task complexity effects were detected where a large error variance was found in the complex SP design.

Individuals' perceptions have significant impacts on the valuation and model estimation precision. Individuals' familiarity with alternatives in the experiment increased the value of the improved rolling stock and improved the estimation precision. Individuals' perceptions of potential price increase have an impact on the valuation and estimation precision. The more likely respondents perceived the potential price increase, the fewer preferences were given to the improved rolling stock and respondents were observed to be more consistent in their choice making.

In brief, this study suggests that incentives to strategic bias exist in the SP experiment due to its hypothetical nature. Warning message such as a CT script is helpful to amend individuals' incentive to strategic bias. Attention should be made to the complexity of the experiment, as respondents are subjected to certain cognitive ability. In the SP analysis, individuals' perceptions can be incorporated into the model analysis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

This paper considers whether Stated Preference (SP) experiments may be prone, in certain circumstances, to strategic bias by respondents who guess the purpose of the exercise and believe that they will not in practice be required to pay the amounts they say they would be willing to pay. The present study is motivated by the desire to analyse and reduce biases in the SP application, specifically addressing the issue of the strategic biasing of SP responses.

1.2 Research Background

1.2.1 Introduction

Stated Preference (hereafter, SP) methods include a variety of ways to elicit individuals' preferences in addition to the possibility of estimating willingness to pay (WTP) for improving specific attributes. For this reason, SP methods have been used extensively in research, especially in transport, market research and health economics (Louviere et al., 2000).

In UK, SP techniques have already proved to be useful tools for travel demand analysis and for valuing attributes such as time savings (Fowkes, Nash and Whiteing, 1985; Wardman 1987; MVA/ITS/TSU 1987; Hague Consulting Group et al., 1999; Mackie, Wardman and Fowkes, 2003). Other recent applications of SP techniques include valuation of accidents (Ortuzar and Rizzi, 2001), atmospheric pollution (Ortuzar and Rodriguez, 2002), environmental science (Adamowicz et al., 1998), urban design (Cooper, Ryley and Smyth, 2001) and evaluation of aircraft noise (Wardman and Bristow, 2004).

Louviere (1988, p.114) states that "*...there is considerable evidence to support the conclusion that appropriately designed, implemented and analysed conjoint studies can predict the real behaviour of real individuals in real markets*". Compared to other methods, SP method is "*a reasonably accurate guide to true underlying preference*" (Wardman, 1988, p.89). Louviere and Swait (1996) state that there is growing body of evidence to suggest SP choice process can be very similar in real and hypothetical markets.

Although SP methods have been increasingly applied in transportation research, "*their gradual acceptance in the transportation research community has not taken place without criticism*" (Arentze et al., 2003, p.229). From the beginning of SP applications, there were concerns about the reliability and validity. A basic question is how much faith we can put on individuals actually doing what they stated they would do when the case arises (Ampt et al., 2000; Carson et al., 2000 and Wardman, 2003).

The inconsistency between the SP survey result and the reality imply the existence of factors (errors) that affect the validity and reliability of SP results from individuals' responses, namely biases. The impact and reduction of biases in SP method remain an issue in research.

The present study is motivated by the desire to analyse and reduce biases in the SP application. This research will seek to identify some sources of biases that occurred in the SP, and investigate the incentives and reasons of biases, once obtained, to use them for the optimisation of SP design and models.

1.2.2 Gaps in existing research concerned with biases in SP practice

Research gaps are found regarding to biases in the SP practice.

Gap1: Existence of strategic bias in SP studies

Research needs to be done to investigate reasons and effects of biases in SP method. Only if the sources of biases are known, researchers can reduce or eliminate them in the design and modelling stage and improve the reliability of SP method.

There are several reasons why individuals' responses to hypothetical questions might not reflect their true preferences (Bonsall, 1986). This study will present a discussion of sources of biases observed from previous SP application (see section 2.4). Amongst these, the issues of design/scenarios specification and task complexity have received a considerable amount of attention (Bradley and Daly, 1994). On the other hand, and despite serious concerns in the early literature, the strategic biasing of responses tends to have been overlooked in recent times, particularly within the SP methodology (Wardman and Bristow, 2005).

Empirical evidence in transport has found the existence of strategic bias, which will be summarised in detail in the literature review (section 2.5.4). This study will examine the existence and consequence of strategic bias in the SP studies. The literature review suggests two methods for further testing on amending individuals' incentives to bias, which are cheap-talk script (CT) and adding more attributes to mask the research aim.

Cheap-talk (hereafter, CT) is a warning message that explicitly discusses the bias that occurs in the previous studies and a reminder for the budget constraints in hypothetical experiments. CT has been initially introduced in the Contingent Valuation (CV) studies (Cummings and Taylor, 1999) to test the existence of hypothetical bias (respondents overestimate the WTP of the product) and eliminate the bias. It is found that a properly designed CT script can effectively amend respondents' incentive to overestimate the valuation. Recently, it is being introduced into SP methods in the food and environmental science (Carlsson et al., 2005; List et al., 2006), however its impact and reliability still needs further examination. We will introduce this method to the present study. There is no evidence of applying the CT script in the transport related topics and consequently the present study will be innovative in this way.

The second method is motivated by Wardman and Bristow (2003)'s successful empirical evidence on eliminating the strategic bias in the evaluation of aircraft noise. Their study suggested that where the objective of the SP exercise is obvious, especially where the issue is contentious, the strategic bias is likely to occur. 'Introducing more attributes', they found that respondents would less be likely to be able to perceive the aim of the SP study, hence respondents are less likely to strategically bias their answers in the SP study.

We will discuss these two methods in detail in the literature review in chapter 2. The present study will examine the impacts of these two methods on amending individuals' incentive to strategic bias and on respondents' choice making.

Gap 2: Impacts of SP design on responses – complexity

Efforts need to be made to explore the impacts of SP design on responses. Related literature in economics and behaviour decision theory has convincingly illustrated how changes in task environment result in changes in decision-making, which will be provided in section 2.7.2. This also has been supported by large empirical demonstration in the past SP studies (DeShazo and Fermo, 2001), which will be presented in section 2.7.3.

As stated above, we will introduce two methods, namely a CT script and adding more attributes to mask the research aim into this study to examine their impact on amending individuals' incentives to strategic bias. However, the addition of information (the CT script and masking research aim by introducing more attributes) may run the risk of changing the way people make their decision. How respondents cope with this information in the choice making process needs further exploration. Will these two methods introduce different kind of biases (or less consistent choice), such as task complexity effects to SP responses? This study will explore the impacts of SP design (i.e. more information in the SP choice) on respondents' choice making processes.

Gap 3: Impacts of respondents' perceptions on their choice making

SP methods are dedicated to analysing individuals' choice behaviour. It is important to know what factors affect their decision making and impacts of those factors on SP estimation. As stated by McFadden (2001, p.31): *"The potentially important role of perceptions, ranging from* *classical psychophysical perceptions of attributes, through psychological shaping of perceptions to reduce dissonance, to mental accounting for times and costs, remains largely unexplored in empirical research on economic choice. Finally, the feedback from the empirical study of choice behaviour to the economic theory of the consumer has begun, through behaviour and experimental economics, but is still in its adolescence."*

Research needs to be done to investigate the impacts of individuals' perceptions on their choice making. This study will investigate the influences of individuals' perceptions of SP game on their choice making behaviour and will explore the relationship between the perception and existence of bias in SP responses.

1.3 Objectives and Methodology

1.3.1 Research objectives

Objective 1: *to identify incentives to strategic bias in SP surveys and investigate how incentives to bias vary across different circumstances and their consequences; and how to amend them in SP design.*

Objective 2: *to analyse the effects of SP design on the biases in SP responses.*

The results of this research may have implications for questionnaire design and the interpretation of SP results, improve reliability of SP methods and obtain more accurate analysis of transport behaviour.

1.3.2 Proposed SP experiment context

The SP experiment context is chosen to be users' valuation of rolling stock. Wardman and Whelan (2001) conducted a meta-analysis based on a large number of published and un-published studies on users' valuation of rail passenger rolling stock. They found the stock values from SP experiment were approximately three times higher than that obtained from the demand analysis using ticket sales data. The reason suspected is the existence of strategic bias in SP responses. When individuals perceive the aim of the SP study is to evaluate a new rolling stock from which they will not have to pay extra, they have the incentive to overestimate the valuation, thus increasing the possibility for the introduction of the new train.

1.3.3 Proposed research hypotheses

Hypotheses are set from the objectives and background ideas in order to guide and shape the boundary of this research.

The first research hypothesis is based on some theoretical evidence from experimental economics and empirical evidence in CV and SP studies (Cumming and Taylor, 1999; Wardman and Whelan, 2001; Wardman and Bristow, 2003). Due to the hypothetical nature of CV and SP studies, strategic biases are observed in lots of applications. A review of the incentive to strategic bias (will be presented in section 2.5) from both theoretical and empirical sides leads to the first research hypothesis:

Hypothesis 1 (H1): *the incentive to strategic bias exists in the SP exercises.*

Respondents will overestimate the utility/valuation of service improvement for which they will not have to pay extra, to increase the likelihood of its introduction. We select users' valuation of rail passenger rolling stock as the SP experiment context in our case study.

To test the existence of incentive to strategic bias, a CT script and adding more attributes to mask the research aim are introduced in the SP experiment.

Hypothesis 2: *The adding of cheap-talk can amend respondents' incentive to strategic bias.*

Hypothesis 3A (H3A): *Masking the research aim (by introducing more attributes) can amend incentive to strategic bias.*

We will examine the impact of adding more attributes to mask the research aim on amending individuals' incentives to bias. However, this method might have the potential drawback of adding task complexity. Task complexity (choice complexity) is defined as the context and format of the SP. A review of task complexity effects will be provided in section 2.7. This leads to the second hypothesis:

Hypothesis 3B (H3B): *An increase in the number of attributes will always increase the variance of error terms.*

In our study, we will test the impact of adding more attributes to the SP experiment on respondents' choice making. Chapter 2 will provide a detailed literature review and explanation of the research hypotheses.

A suite of SP experiments are developed to test the research hypotheses. Two measures are introduced into four SP designs: adding a cheap-talk script and adding more attributes to mask the research aim. By comparing the responses from four different groups, the effects of the CT script and task complexity on the valuation of rolling stock will be investigated.

In summary, this research is aimed at examining some sources of bias in SP responses, more specifically, the incentive to strategic bias and task complexity effects. By conducting a series of

SP experiments, the existence and consequence of biases are examined in the context of users' valuation of rolling stock. Some experimental factors such as cheap-talk and masking research aim are tested to see their impacts on respondents' choice making.

1.4 Outline of Thesis

The thesis is written in the order of the research process. A graphic representation of the thesis's outline is present in Figure 1.1.

The first part, chapters 1 to 4, explains the construction of the study's framework. Chapter 1 presents the objective and research hypotheses, which are set up from the literature review and previous case studies in chapter 2. Chapter 3 provides a review of valuation of rolling stock studies. Chapter 4 introduces the methodology of this research.

The second part of the research involves the design of SP experiments as shown in chapters 5 and 6. The design and development of SP exercises, through two pilot surveys, is described. Chapter 6 also presents the data collection process and the sample characteristics.

The third part, chapters 7 to 9, comprises the analysis of results and conclusions. Chapter 7 establishes a base model for users' valuation of the improved rolling stock. The base model controlled several factors (i.e. income and journey purpose) which cause the variation of valuations to avoid their potential confounding effects. Chapter 8 explores the effects of design factors (the cheap-talk script and the complex design) on SP responses from the base model. Chapter 9 draws together a summary of research objectives and methodology, and the main findings on the effects of SP design on biases in responses. It also provides suggestions for future studies.

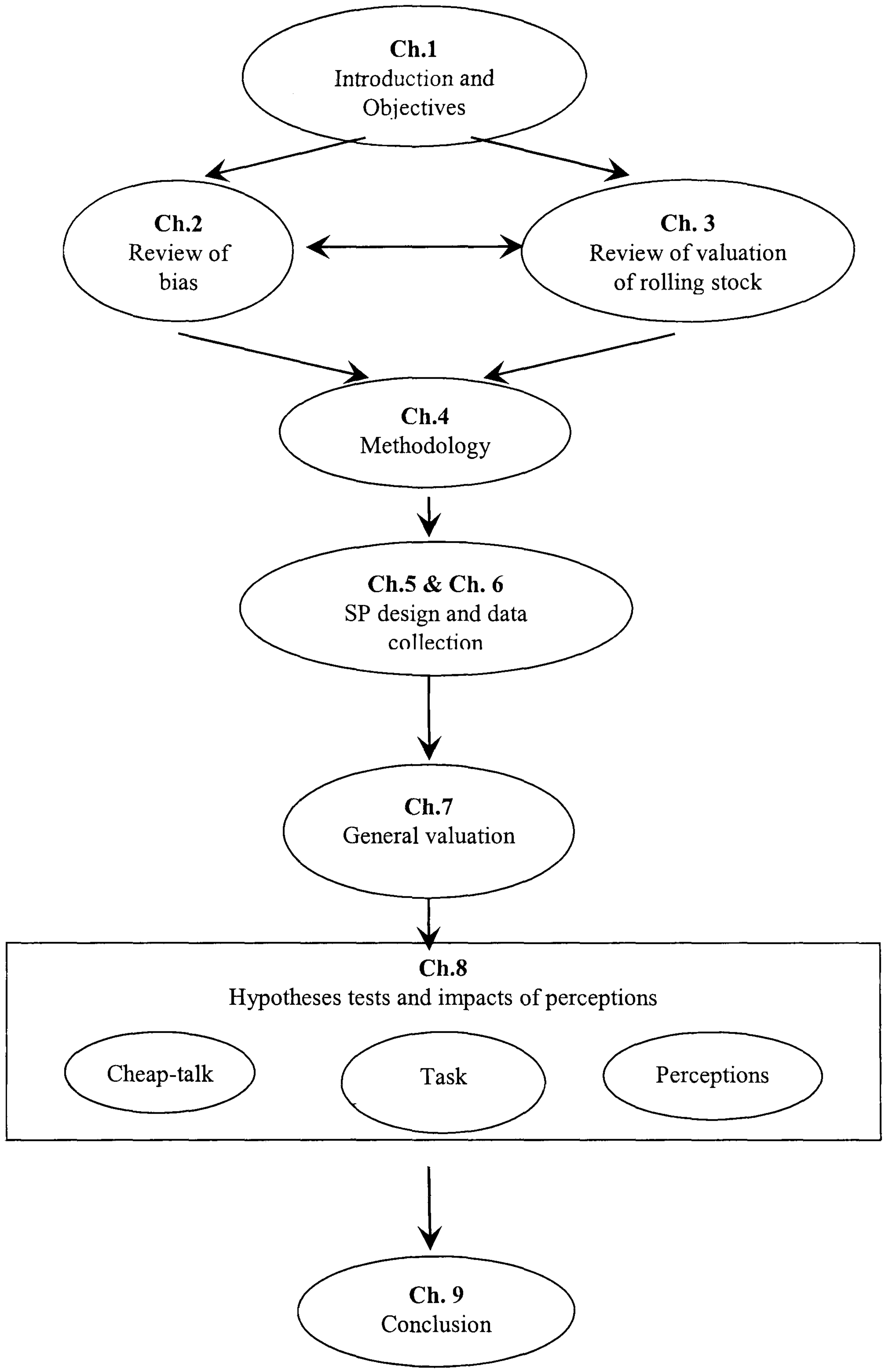


Figure 1.1 Outline of the thesis

Chapter 2

Review of Bias

2.1 Introduction

The objective of this chapter is to provide an overview of biases. The scope of this study will be defined and research hypotheses will be established based on the literature review. The review includes the general definition of bias in section 2.2 and the definition in the SP application in section 2.3. Section 2.4 reviews biases observed in the previous SP studies and develops a typology of biases, sufficient for the purpose of this thesis. Section 2.5 reviews the incentives to strategic bias survey answers in the transport field and suggests two possible methods to amend the bias in this study: Cheap Talk script and adding more attributes to mask the research aim. Section 2.6 introduces in detail previous applications of Cheap Talk script to amend the incentives to strategic bias. Section 2.7 presents a review of task complexity effects. Finally, section 2.8 summarises implications of the literature review for this study.

2.2 Statistical Definition of Bias

Bias is any systematic error that occurs in the estimates. Any factor or process that tends to deviate the results or conclusions of a test systematically away from the truth is called bias. Osterlind (1976, p.10) defines bias as a change in the accuracy of measurement.

> *"Bias is defined as a systematic error in the measurement process. It affects all measurements in the same way, changing measurement---sometimes increasing it and other times decreasing it. ...Bias, then is a technical term and denotes nothing more or less than the consistent distortion of a statistic."*

There are two types of bias: sample bias and estimation bias. Sample bias occurs when some members of the population are more likely to be chosen in the sample than others. For example, non-response bias can give a biased sample unless corrected for.

Estimation bias refers to an estimator that on average, for some reason, over or underestimates what is being estimated. Since there can be no "perfect" estimator that always gives the right answer, if the expected value of an estimator is equal to the parameter which it is supposed to estimate, the estimator is said to be unbiased; otherwise, it is said to be biased. According to this definition, the mean of any sample is an unbiased estimator of the population mean.

Suppose we are trying to estimate the parameter β using estimator $\hat{\beta}_n$ (that is, some function of the observed data), with distribution denoted as $f(\hat{\beta}_n)$ (i.e. $E(\hat{\beta}_n) = \int_{\beta_n=-\infty}^{\infty} \hat{\beta}_n f(\hat{\beta}_n) d\hat{\beta}_n$), if:

$$E(\hat{\beta}_n) = \beta$$ **Equation 2.1**

for all n, then this estimator is an unbiased estimator (Ben-Akiva and Lerman, 1985, p.13).

The estimation bias equals the difference between the expected value and the value of the quantity being estimated. The bias of $\hat{\beta}$ is defined to be

$$E(\hat{\beta}) - \beta = E(\hat{\beta} - \beta)$$ **Equation 2.2**

Figure 2.1 shows an example of the estimation bias. $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimates of β (potential true value) and their expected value are β_1 (where $\beta_1 = \beta$) and β_2, following the distribution of $f(\hat{\beta}_1)$ and $f(\hat{\beta}_2)$ respectively. From the above definition, a randomly chosen value from distribution of $\hat{\beta}_1$ is an unbiased estimator of β since $\beta_1 = \beta$; whilst a randomly chosen value from distribution of $\hat{\beta}_2$ is a biased estimator of β , that bias being ($\beta - \beta_2$).

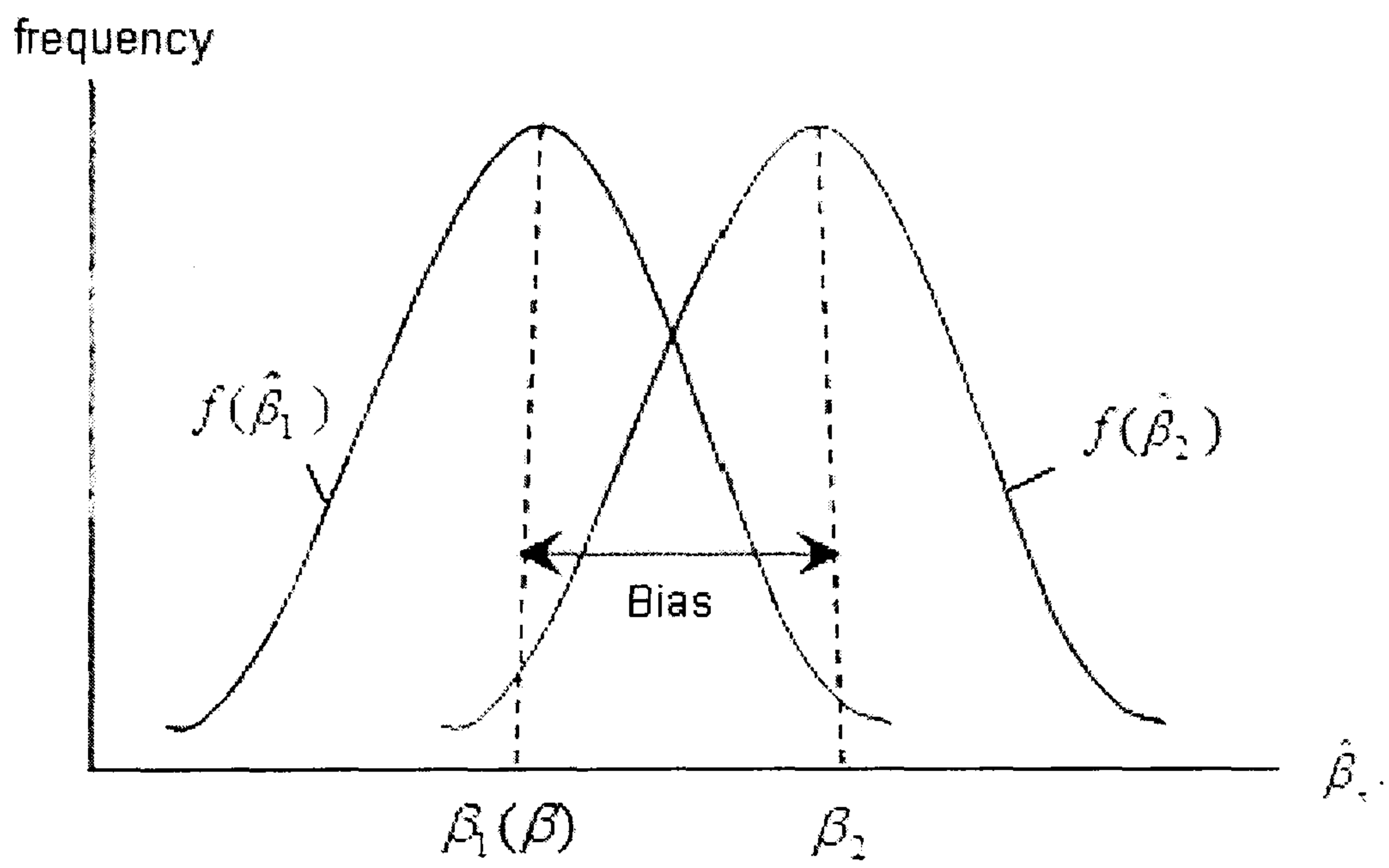


Figure 2.1 An example of estimation bias

Variance is used to measure the dispersion of a sample or population and generally denoted by $\text{var}(\hat{\beta}_n)$. Variances are indicative value of the variability of the β , which are actually observed. It is a measure of how good the sample is.

Both bias and variance refer to how far, on average, an observed value will be from the true value. Researchers obviously want to minimize both bias and variance. This section provides a general definition of bias in statistical meaning. In this study, biases in SP responses will be investigated, including sources, consequences and possible methods to amend the bias.

2.3 Errors in Stated Preference (SP) Method

2.3.1 Introduction of SP methods

Green and Srinivasen (1978, p.103) defined Stated Preference (SP) methods to "cover models and techniques that emphasize the transformation of subjective responses into estimated parameters". By using experimental design, researchers construct a series of hypothetical choices. Respondents are then asked to indicate their intention or preferences. By quantifying these underlying preferences, SP techniques can get the information about people's preferences, which may not easily be measured through observations of actual behaviour.

SP methods are based on some behavioural theories in which decision makers connect actions to consequences and then decompose consequences into attributes. Due to the discrete nature of respondents' behaviour, SP data analysis is usually based on random utility theory and the logit model, which is explained in details in McFadden (1973), Ben-Akiva and Lerman (1985), Louviere, Hensher and Swait (2000) and Train (2003), and the most recent and comprehensive details in Hensher, Rose and Greene (2005a). This method of analysis has been widely used in analysing and forecasting economic consumer behaviour in a wide variety of applications, including marketing research, travel demand, residential location choice, environmental economics and health economics.

The literature is rich with documented cases of the evolution of SP method (Wardman, 1987; Fowkes, 1991; Hensher, 1994; Fowkes, 1998; Adamowicz et al., 1998; Louviere et al., 2001; Hensher et al., 2005a). According to Wardman (1987), SP method can be traced back to studies in the area of mathematical psychology in the 1960's. Luce and Tukey (1964) introduced the concept of "Conjoint Measurement", in which alternatives can be viewed as the weighted combination of the various aspects or attributes.

The origin of SP methods can also be traced back to market research in the early 1970s, named as "Conjoint Analysis" and became widely used since 1978 (Kroes and Sheldon, 1988). According to Fowkes (1998), SP methods were first applied in the transport field in early 1980's (1982/3), for forecasting travel demand and behaviour where traditional travel demand models were inadequate, for example, due to poor quality or lack of data.

Some aspects of SP methods are common to the Revealed Preference (RP) methods. In the transport field, the pre-eminent method has traditionally been RP, where individuals are observed how they choose from the current options and give reported values for their chosen and rejected alternatives. The actual choice made reveals the importance of each attribute which characterises the alternative perceived by respondents. In contrast, SP methods provide individual hypothetical choices and the response supplied also indicates the importance attached to each attribute that characterises the alternative.

RP data has well known limitations in terms of understanding travel behaviour (Kroes and Sheldon, 1988; Pearmain and Swanson, 1991), mostly related to the cost and quality of data. The weaknesses of RP have led to the evolution of SP techniques. An important paper by Lerman and Louviere (1978) demonstrated the theoretical links between RP and SP. Compared with RP method, advantages of SP method are listed as below:

- As the researcher can precisely control the design by defining the choices offered to respondents, SP method ensure data of sufficient quality to construct good quality statistical models;
- Due to the control available to the researcher, the effects of correlation among variables can be avoided(Hensher and Louviere, 1983);
- SP method can deal with a variety of variables, such as some 'secondary' (latent) variables, like security, comfort and information. In reality, it is difficult to evaluate the impact of changes in these variables;
- Where an alternative is completely new, so that no RP data is available, SP method may represent the only practical basis for evaluation and forecasting (Louviere and Hensher 1983; Hensher, 1994);
- SP method is economical to apply, as respondents provide multiple observations in the interview;
- There is no measurement error in the independent variables.

Due to these advantages, SP techniques have become an attractive option in transport research.

2.3.2 The category of SP methods

Figure 2.2 shows Adamowicz et al. (1998)'s taxonomy of SP methods, based on the types of response data.

SP rating or scaling means that the respondents rate or score alternatives presented to them according to a numerical (e.g. on a scale 0-10) or semantic preference scale (response indicates strength and order of preference).

SP ranking means that respondents compare groups of alternatives against each other, so preference for alternatives can be ordered. Rank ordering can also be treated as a set of independent choices; therefore discrete choice models can be applied. However, SP ranking may not correspond to what respondents face in real life (Pearmain and Kroes, 1990) and has also been questioned in terms of reliability (Ortuzar and Garrido, 1991).

SP choice means that the respondents choose the best alternative from the set of possible ones. The process of estimation of individuals' preferences is based on random utility theory and uses discrete choice models.

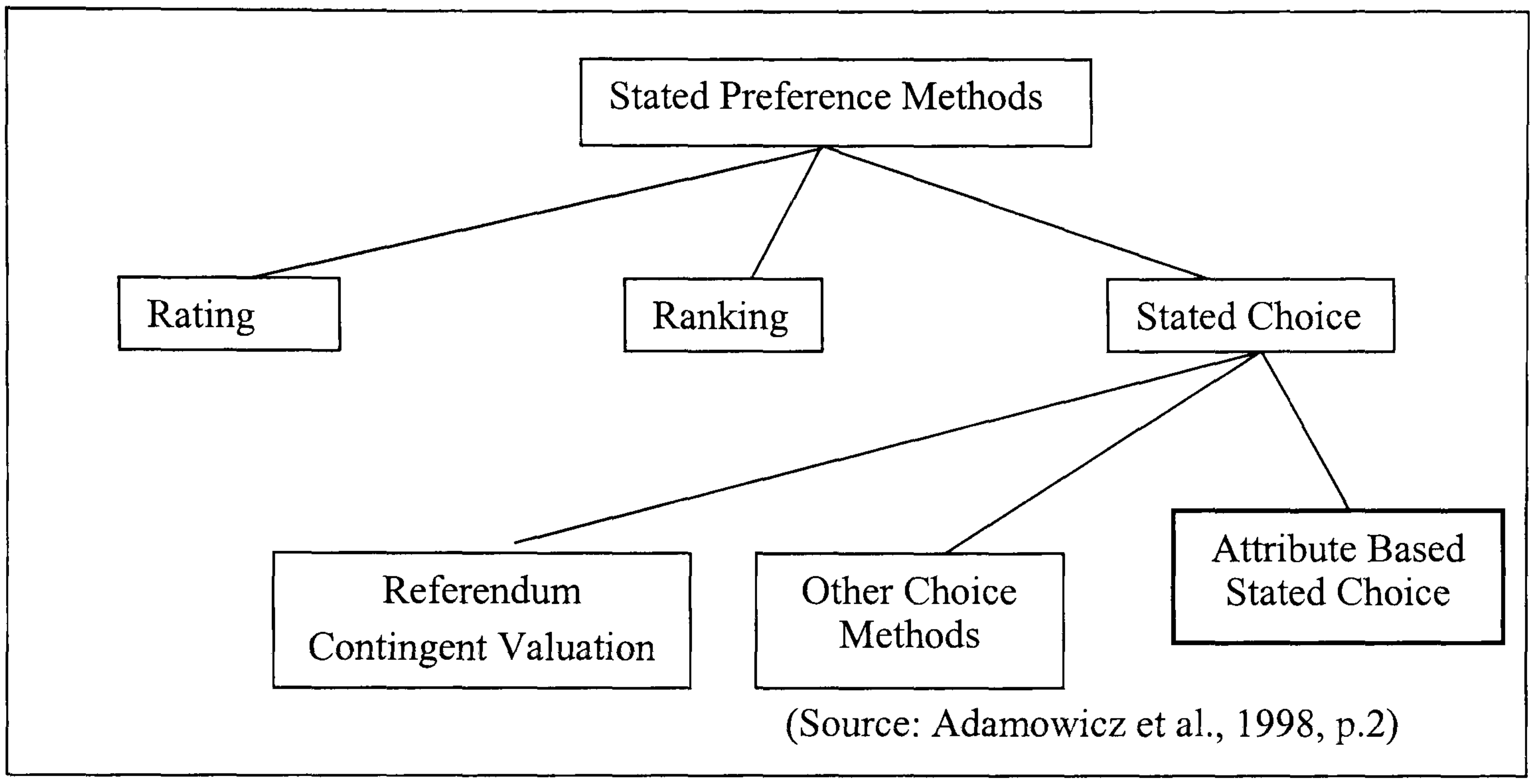


Figure 2.2 Taxonomy of SP methods

In this study, we focus on the response bias in the "Attribute Based Stated Choice" method, since this method is now the most popular form of SP method in transport and is growing in popularity in other areas such as marketing, geography, regional science and tourism. In the remainder of this paper when we refer to Stated Preference methods, we are referring to "Attribute Based Stated Choice" methods unless otherwise stated.

2.3.3 Errors in SP application

Literature has documented the errors that occurred in SP application (Ben-Akiva and Lerman, 1985; Bates, 1988). Figure 2.3 to illustrate the categorization of errors in SP methods (Bates and Terzis, 1997). There are two fundamental sources of errors found in the application: the data which may be unreliable in some way and the model which may be inappropriate to replicate the decision making. These two sources of error have further variants.

Data errors occur during the survey design and data collection. The measurement errors are the errors which happened in the independent variables (variables that are put in the model to "explain" the dependent variables). The response errors are the errors in the dependent variables (the quantities or choices which we are trying to model). There is another source of errors which is caused by the sampling, known as sample bias. It happens when selecting the sample, where some populations are more likely to be chosen in the sample than others.

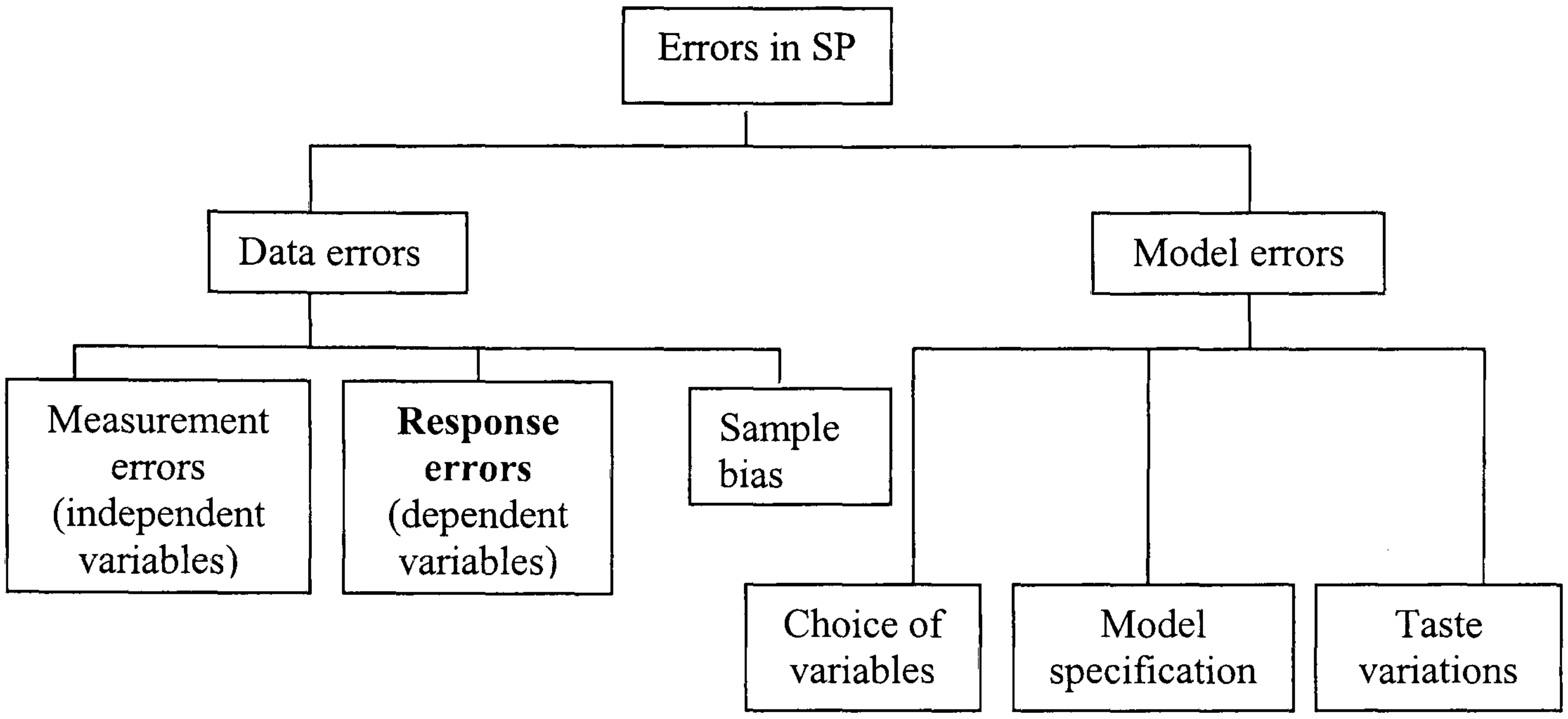


Figure 2.3 Errors in SP methods

With respect to the model errors, there are three types: the choice of variables in the model; the way in which the variables are combined (model specification) and the extent to which a given model is appropriate to different subsets of the population (taste variations).

In SP applications, researchers can control the experiments, in terms of attributes, alternatives and context. It is assumed the measurement errors are not the main concern of researchers, since all the attribute values are directly presented to respondents.

However, SP surveys provide individuals hypothetical choices; so response errors are a potentially serious source of errors in SP application. Bates (1988, p.64) stated that "*with SP, there would seem to be a further serious source of error, and that relates to the response variable itself.*" McFadden (1986, p.289) stated that "*Another issue is the stability of elicited preferences over the sequences of task performed by each subject. Factors such as learning, boredom, or anchoring to earlier tasks may distort the measurement of preferences, and cast doubt on the cognitive congruence of the time frames in which experimental versus market decisions are made*".

The present research aims to investigate the sources and possible consequence of response bias in the SP method. A desk study of possible response bias in SP applications will be presented in section 2.4.

2.3.4 Impacts of systematic errors on SP results

Explaining travel behaviour is to estimate quantitatively the relationship between dependent variables (utility/choice) and independent variables (attributes). The impact of errors (biases) on SP methods is demonstrated in following two aspects: estimation and forecast.

Estimation

In the SP methods, estimation bias refers to the biased estimate of attribute parameters in the utility function, as shown in Equations 2.3:

$$\text{If } \hat{\beta}_k \text{ is a biased estimate of } \beta_k\text{, then } E(\hat{\beta}_k) \neq \beta_k, \qquad \textbf{Equation 2.3}$$

where, $\hat{\beta}_k$ is the estimate value of β_k and β_k is the parameter of the kth attribute X_k. As the monetary value of attribute 'X' (VoX, more discussion in Section 4.4.4) is obtained by the ratio of marginal utilities (parameters) of attribute parameter estimates X ($\hat{\beta}_X$) and cost ($\hat{\beta}_{Cost}$), biased estimates of parameters will lead to biased monetary values.

This bias can happen to a single attribute. For example, individuals may bias their answers to prevent the increase of cost by giving 'cost' higher weight than actually is, thus leading to a lower WTP for the new product. This can also happen to a couple of attributes simultaneously.

Forecast

When the relationship between choice and attributes is estimated using the observed data, the results can be used to forecast the changes in the attributes or choices in real life. Since SP derived attribute parameter estimates are scaled according to SP error terms, which are unlikely to be equal to RP error terms, using these SP attribute parameters unadjusted for forecasting will give poor forecasts. If RP data is available, the SP attribute parameter estimates can be rescaled, but that was not the case in case study in this thesis.

Alternatively, monetary valuations can be used with known price elasticity to calculate attribute elasticity. Elasticity indicates sensitivity of demand to change in some variables, if all else remains constant. For example, in this thesis, to forecast the demand of the introduction of improved rolling stocks, PDFH (2005) suggested: firstly convert the stock improvement into an equivalent change in rail fare; and then the relevant fare elasticity is applied to calculate the expected demand increase. As the impact of the improved rolling stock is directly related to certain type of the stock, the monetary value of this stock is converted to a demand impact by using the fare elasticity.

Biased estimate of coefficients and valuations will lead the elasticity to be too high or too low, which indicates the model results tend to over or under predict actual changes in choice.

Therefore, biased estimate of the monetary valuation of improved rolling stock will lead to the biased demand forecast. This will be discussed in detail in chapter 3.

2.3.5 How to detect bias in SP experiments

SP surveys present respondents with some hypothetical scenarios. Normally, the output can be examined as below:

- Is it reasonable (sign and magnitude)? Can results get the right sign, for instance, cost parameter is always negative when evaluating the effect of introducing a new product.
- Is it robust? This refers to the reliability of the output.
- Is it consistent with the theory (such as economic theory)?
- By doing meta-analysis, which compares with other results (RP or different methods or different SP experiments), to test if the result agrees with them?
- Are forecasts consistent with past studies?

2.3.6 Summary and implications for this study

This section provided the concept of error in the SP experiment. Errors can be categorized into data errors and model errors in the SP application.

As researchers can control the SP experiment, in terms of the levels of attributes, attributes and alternatives, measurement errors are not the main concern. Due to the hypothetical nature of the SP survey, respondents are not committed to behave in accordance with their stated preferences; therefore, response errors are one of serious errors in the SP application. Response bias is that respondents, for some reasons, give the biased answers to the SP questions, thus leading to the biased estimate of the coefficients/valuation.

This study focuses on the response bias in the SP survey. In the next section, sources and possible consequences of bias in the SP application will be reviewed, and a typology of bias will be generated based on the review.

Next section presents a desk-study of bias that observed from previous SP studies in similar relevant work. The sources of biases are roughly categorized by two aspects: from experiment design and responses. The sources and possible consequences of bias are investigated and a typology of biases is produced.

2.4 Sources of Bias in SP Application

2.4.1 Bias from SP design

Bias can occur from the design and presentation method used in the SP practice. Rich literature in the SP application reports that SP responses are affected by the design. That is not to say that these biases only occur in SP data and a wider literature is examined, if only briefly.

Framing Effect Bias

Framing effect bias is so called because individuals may often respond differently to different descriptions of the same problem. For example 10 minute saving in travel time may be valued less on a 10 hour journey than on a 20 minutes journey. Ampt et al. (2000) found that weight attached to a rising of cost is larger than that attached to reduction in cost. Value of time is consequently higher for worsening (rising of cost) than for improvement (reduction in cost).

Cho (1998), in her PhD research, found that the coefficients were significantly more negative for the precisely known charges than for the imprecisely-known charges. She explained that this was consistent with the key features of Prospect Theory (Kahneman and Tversky, 1979). Prospect Theory describes how people make choices in situations where they have to decide between alternatives that involves risk. Prospect Theory states that, prior to making a choice, decision makers use heuristics to simplify the options available and set a reference point; then the options are assessed in relation to the reference point.

Packaging Effects

Packaging effects occur where respondents give the value of the package less than the sum of the values of its constituent parts, shown as Equation 2.4.

$$\frac{\text{Value of the package}}{\sum(\text{Values of its constituent parts})} < 1 \qquad \textbf{Equation 2.4}$$

Jones (1997) stated that: *"The 'packaging' problem arises when trying to value individual attributes of a journey that collectively contribute to one aspect of the journey experience, such as in-vehicle or the station environment. It commonly happens that the value derived from an SP experiment for an improvement in the level of each attribute in a cluster sums to an amount that is considerably different to the value which the same respondent ascribes to the package of improvements as a whole".*

Wardman and Whelan (2001) presented packaging effects observed from the studies on valuation of rolling stock. They concluded that the possible causes of the packaging effects include: interactions effects (the value of the package will be less than the sum of the values of its constituent parts); budget constraints, halo effects and the artificial nature of SP exercises. Table 2.1 shows the packaging effects observed from the valuation of rolling stock studies.

Table 2.1 Packaging effects observed from the valuation of rolling stock studies

Studies	Time	Packaging Effect Ratio
Steer Davis Gleave	1990	0.3
MVA	1992	0.62
MVA	1993	0.4-0.6
MVA	1993	0.74
Jones*	1997	0.4
Scandinavian study*	-	0.38-0.90

Note: Jones reports ratios for bus environment. The Scandinavian study estimated ratios between 0.38 and 0.90 according to circumstances
Source: Wardman and Whelan (2001, p.428)

Simplifying Bias

In SP experiments, respondents typically assess a number of alternatives and are asked to choose the most preferred alternative, including the choice not to choose any of the offered alternatives. Normally the alternative is defined by a set of attributes and each attribute is offered from a pre-specified set of levels and range of levels. This assessment is repeated a number of times up to the total number of choice sets that are being offered. Due to the limit of individuals' cognitive abilities, when the task is too complex for respondents, they may modify their decision strategy to simplify the task (Bradley and Daly, 1994). This has been called task complexity effect, which is supported by large empirical evidence (Swait and Adamowicz, 2001; DeShazo and Fermo, 2001; Caussade et al., 2005 and Hensher et al., 2007).

Unfamiliarity

If respondents have little experience of an attribute/alternative, they will value it differently from someone with experience, who is in better position to assess the importance. Benshoof (1970) study of motorists showed that the unfamiliar motorists did not accurately measure different route characteristics, which occurred when respondents had not experienced the route before. Wardman and Whelan (2001, p.423) conducted a review on stock valuations and found that if respondents are familiar with the rolling stock the survey presented; their valuation of new stock is lower than that from unfamiliar respondents. The valuation is 44% lower and the impact is significant. They concluded that unfamiliarity with the improved level of attributes would result in overestimation, which partly explained the inflated valuation of new stock.

Unrealistic Values Bias

Unrealistic values bias refers to the situation that respondents misinterpret or ignore an attribute when the set of value in a hypothetic scenario does not reflect the reality. A typical example is the lower estimation of the walk and wait time coefficient (Wardman, 2003). In SP studies, more attention is paid to the realism of cost and in-vehicle time (IVT), variation in walk and waiting time maybe presented unrealistically and would therefore be ignored. The consequence of this bias is that the coefficients of the ignored attributes will be more likely smaller than they would otherwise be.

This effect is likely to be reduced by appropriate instruction or guidance. When customizing the levels of all variables, the designer should pay attention to the combinations of the attribute levels and constraints of the experiment. Otherwise, respondents may ignore attributes, or interpret them in a different way.

2.4.2 Bias from SP response

The other source of errors is from unreliable data (section 2.3.3), which is 'wrong' answers from respondents. In the SP survey, respondents might not be committed to behave in accordance with their stated preferences; therefore, response errors are one of the serious concerns for SP researchers.

Habit Bias

Habit bias refers to the situation that respondents resist the challenge to their current behaviour in the SP survey (Wardman, 1986). For example, in SP studies of the acceptability of road user charging, there is a greater tendency to state the currently chosen alternative to be preferred and the coefficient estimates are distorted. Aarts and Dijksterhuis (2000) found that suppressing habitual response is difficult and often not successful under conditions of cognitive load, indicating that a transport model choice can become automatically associated with travel goals (e.g. have to go to universities).

Habit bias can be explained by the Cognitive Dissonance theory (Festinger, 1957), in which individuals avoided the mental disharmony. This theory describes the uncomfortable tension that may result from two conflicting thoughts at the same time, for instance, the information conflict with one's belief. So in the situation, there is the effort to ignore the information or reinforce one's belief.

Social Norms Bias

Social norm bias occurs when individuals incorrectly perceive (exaggerated and frequently overestimated) the attitudes and / or behaviours of peers and other community members to be different from their own. Social norms theory (peer effects) assumes that much of our behaviour is influenced by how other members of our social groups behave, and that our beliefs about what others do are often incorrect. This phenomenon has also been called "pluralistic ignorance" (Miller and McFarland, 1987) which lead individuals to act in ways that are inconsistent with their true beliefs and values. While in SP survey, some respondents may state their preferences based on the social norms rather then their true likes. Evidence shows that (Whelan, 2003) a car is strongly linked to feelings of independence and convenience. So it is difficult for the drivers to state to want to leave their car.

Status Quo Bias

Samuelson and Zeckhauser (1988) defined "Status quo bias" as where the respondent tends to choose the current state of affairs although it is no more attractive than other available alternatives. Status quo bias implies the resistance to change and a preference to stay with what people have. Some evidence shows that a status-quo bias sometimes does, sometimes does not exist, depending on prevailing conditions. It may occur if a decision maker conceives a loss in public, for instance, SP studies of the acceptability of road user charging (Schlag and Schade, 2000, p.317). The consequence of status quo bias is that the coefficients will be distorted.

Strategic Bias (Policy Response Bias)

If the aim of SP survey is to investigate a new policy, respondents' expressed choices or preferences may influence the way a policy maker introduces the new policy. Respondents will then have the incentive to bias responses strategically to obtain a more favourable outcome (Bonsall 1983, p.73; Wardman, 1986). For example, when introducing a new good, if there is any positive probability of wanting the new good at the stated price, the respondent should say "yes-would purchase." Their logic is that such response will encourage the company to produce the good, with respondents being able to decide later whether to purchase. Since increasing respondents' choice set in a desirable way increase utility, the optimal response is "yes".

Due to the hypothetical nature of SP survey, where respondents normally value some goods or policy which does not exist in the real market, they might perceive that their responses would affect the provision of good or policy. Therefore, they have various incentives to answer the question. Strategic bias is found in the environment studies (named as hypothetical bias), market research and transport when using SP as a method to achieve the WTP or marginal utility of attribute. A detailed review of strategic bias (incentive to bias) is presented in Section 2.5.

Affirmation Bias

Affirmation bias results from a tendency of respondents to agree with interviewer or analyst. It is a "well known hazard" (Bonsall, 1983, p.73) in attitudinal and SP research, especially in the case of personal interviews. Respondents use the questionnaire to express an opinion about the survey aim, thereby biasing the coefficients of variable. For example, if respondents perceived that the aim of research is to value the benefit from providing a new local rail station, it is likely that they will support rail compared with other modes.

Affirmation bias can occur in the situation that respondents are not sufficiently aware of the topic being surveyed. The fear of appearing uninformed may induce many respondents to conjure up opinions even when they had not given the particular issues any thought prior to the interview (Erikson, 1988). Respondents are prone to shape their answers to please either the interviewer or the sponsor, especially when they do not have a strong or well-considered view on the survey topic (Schuman and Presser, 1981).

The incentive to both policy response bias and affirmation bias is provided by the perceived aim of SP survey. The difference only lies in the fact that whether respondents can bring them the maximum benefit by the biased answer. In the former case, respondents strategically bias their answers for bringing them the maximum benefit. In the latter case, respondents distort their choices to agree with the interviewer and not necessarily bring them the maximum benefit.

2.4.3 Typology of biases in SP application

From the review, sources of bias in SP application can be roughly categorized to three main types: complexity, design misspecification and incentives to bias. We constructed Table 2.2 to show the sources of bias from these three categories.

Based on the Table 2.2, Figure 2.4 provides a typology for the bias in the SP application by following a decision making process. It is developed to illustrate the response process and possible outcome of the hypothetical survey. In the typology, three aspects are considered:

- The survey is a plausible description of the hypothetical scenarios;
- Respondents perceive the content of the survey as the researcher intended;
- Respondents stated their true preferences in the survey.

In practice, some biases or some wrong estimates are not just caused by a single reason, but result from combined reasons in various ways and degrees that produce different bias dimensions. For example, the lower estimate of value of time (VoT) might be caused by policy bias, or some lexicographic rule where respondents consider the cost as the most important attribute and ignore other attributes, therefore giving a higher weight to the cost.

Table 2.2 Typology of potential biases in SP application

1. Complexity
Bias in this case occurs when the task becomes complex respondents may make more errors or simplify their decision rule to make the task easier, thus giving an answer that differs from the true preference
A. **Lexicographic answers**: respondents evaluate the alternatives in terms of the most important attribute. B. **Inconsistency choices**: choices that violate the transitivity axiom of consumer theory
2. Design (Mis)specification
Bias in this category occurs when respondents do not respond to the correct valuation scenario (except in A). The other two situations are presumed that the intended scenario is correct and that the errors occur because respondents do not understand the scenario as the researcher intends it to be understood.
A. **Unrealistic values and variations biases**: respondents misinterpret or ignore an attribute when the set of values in a hypothetic scenario does not reflect reality. B. **Package effects:** respondents give the value of the package greater (less) than the sum of the values of its constituent parts. C. **Context misspecification bias**: respondents' perceived context of the SP experiments differs from the intended context: a. **Misunderstanding bias**: respondents may not fully understand SP survey and/or they may be fatigued from doing this exercise. Misunderstanding may cause a large amount of errors in the survey b. **Unconstrained bias**: respondents may disregard situational constraints. c. **Framing effects:** respondents value loss more highly than gains
3. Incentives to misrepresent response
Bias in this class occurs when respondents deliberately state their preferences differently than the true preferences.
A. **Strategic bias**: respondents state preferences that strategically differ from their true preferences (conditional on the perceived information) in an attempt to influence the provision of the good (policy) and/or the payment for the good to obtain a more favourable outcome. B. **Affirmation bias** (Compliance bias): respondents state preferences that differ from their true preferences in an attempt to comply with the presumed expectations of the interviewer.

2.4.4 Summary and implications for this research

This section presented a desk-study of sources and possible consequence of biases observed from previous SP studies. Based on the desk-study, the sources of bias can be categorized to three main types: design (mis)specification, task complexity and incentive to bias.

Amongst these, the issues of design/scenarios specification and task complexity have received a considerable amount of attention. On the other hand, and despite serious concerns in the early literature, the strategic biasing of responses tends to have been overlooked in recent times, particularly within the SP methodology. This thesis will contribute to this spirit.

In this study, it was decided to investigate the existence and consequence of incentives for respondents to strategically bias their answer. Possible methods suggested from the literature review will be examined and the possible impacts will be discussed. A detailed review of the incentive to strategic bias in the SP application is presented in Section 2.5, with illustration of examples in the transport and environment science field. Methods are suggested to amend incentive to bias. This defines the scope of this study which leads to the first research hypothesis.

As stated in chapter 1, we selected users' valuation of improved rolling stock as the SP experiment context, considering that biases found in the past relevant studies are suspected to be strategic bias. We will develop a series of SP experiment, introducing methods to amend the bias, to examine the research hypotheses.

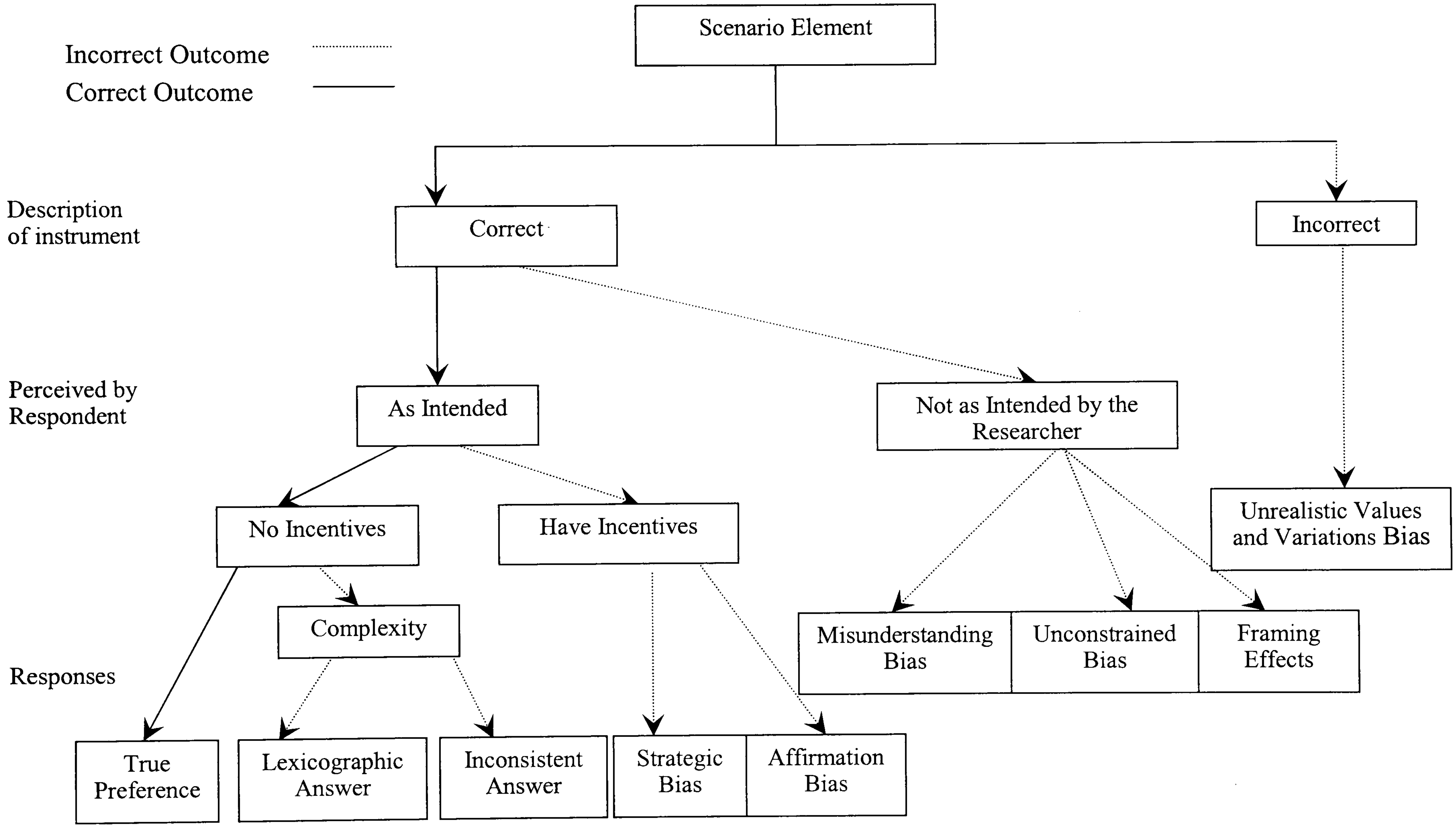


Figure 2.4 A typology of bias in the SP application

2.5 Review of Incentive to Strategic Bias

This section reviews incentives for respondents to strategically bias their answers in the SP application. It includes a brief introduction of incentive and the economic background of strategic behaviour. This section then summarises the concerns surrounding the extent to which the responses to hypothetical questions reliably reflect individuals' true preferences where there is an incentive to do so. The discussion is illustrated with examples from research in the transport and environment field. It suggests methods to amend the incentive to bias.

2.5.1 Definition of incentive and incentive compatibility

In the Cambridge dictionary (2007), incentive refers to "something which encourages a person to do something", "that which incites rouses or encourages a person". In economics, an incentive is any factor (financial or non-financial) that provides a motive for a particular course of action, or counts as a reason for preferring one choice over the alternatives. The study of incentive structure is central to the study of all economic activity (both in terms of individual decision making and co-operation and competition within a larger institutional structure). Economics literature documented incentive and its application in the mechanism design (Vickrey, 1961; Hurwicz, 1986).

Due to the hypothetical nature of SP experiment, respondents being surveyed might believe that their responses can influence actions taken by the government or business companies and so are encouraged by this perception to answer the question in the way to maximize their expected utility. There is the argument that if respondents behave strategically, their answer cannot provide useful information. This is obviously not true if optimal strategic behaviour coincides with truth-telling or if the influence of the strategic behaviour can be at least partially unravelled (Carson et al., 2000). Incentive compatibility is that in a choice situation, the optimal (and the dominant) strategy for the respondent is the truthful preference revelation. This concept was first introduced by Hurwicz (1972) in the mechanism design theory.

2.5.2 Economics background of incentive to strategically bias

It has long been recognised that some individuals will not reveal their true preferences when there is a benefit to be gained from not doing so. Samuelson (1954, p.388) stated that "now it is in the selfish interest of each person to give *false* signals, to pretend to have less interest in a given collective consumption activity than he really has"; whilst Bohm (1971) commented that "potential consumers of a proposed output of a public good have stated preferences which can only be expected to *over*estimate their true valuations. The simple reason is of course that the consequences as to their payments (e.g. a tax increase) have been left out of the process." These statements illustrate the classic free-rider problem and the reverse incentive to overstate values where payment is not expected.

Three divergent elements have emerged to challenge this orthodoxy: two are theoretical, the other is empirical.

Mechanism design theory

The first theoretical challenge assumes still that the free-rider problem is important but it seeks to develop 'incentive-compatible' mechanisms whereby individuals will be induced to reveal their true preferences. Mechanism design is a sub-field of Economics, which is the art of the designing rules of a game to achieve a specific outcome. A structure is set up in which each player has an incentive to behave as the designer intends, and seek to achieve the following basic outcomes: truthfulness, individual rationality, budget balance and social welfare. The application of the mechanism design theory includes the creation of market and auction, and also provision of the public good and the optimal taxation schemes by government.

Hurwicz (1972) introduced the concept of incentive compatibility and proved that "there cannot exist any informationally decentralized mechanism (or procedure) for resource allocation in private good economies that simultaneously yield Pareto-efficient allocations and provide sufficient incentives to consumers to honestly reveal their true preferences". Green and Laffont (1978) proved that the ideal preference revelation mechanism does not exist. Mechanism design theory shows that it is impossible to design an incentive–compatible responses format that allows for more than a binary response (including all multinomial and continuous response formats) if no restrictions are placed on individuals' preferences (Gibbard, 1973; Satterthwait, 1975). Even for binary choice format, it can be lack of incentive compatibility, such as survey questions for new private goods (Carson et al., 2000). Clarke (1971) and Groves (1973) showed an incentive – compatible mechanism with some assumptions on the restriction of individuals' preferences. However, the mechanisms developed were complex and a basic conflict inevitably remained between achieving a dominant equilibrium and achieving Pareto efficiency.

Ethical behaviour theory

The second theoretic challenge says that the free-rider problem may not exist, since more collectively conscious or ethical behaviour will guide individuals in their public good decision. Brubaker (1975) has criticised the conventional strategic bias literature for relying only upon individualist atomistic behaviour. Margolis (1981) has proposed notions of "behaving properly" or "sharing fairly" as separate arguments in individuals' utility functions. However, the actual extent of the operation of such collective or ethical preference in these analyses is left open and remains to be examined (Throsby and Withers, 1986).

Bohm's experiment

The third source of challenge is from empirical evidence, which says the problem of honest revelation of preference is not significant. Experimental economics have tested strategic effects in SP for public goods in a laboratory setting. Bohm (1972) compared the strategic effects of willingness to pay (WTP) for television programs between samples with different incentive schemes. Table 2.3 summarises his six experiments on the preference revelation methods. In his experiment, six different ways of asking respondents' WTP for the television programs were tested to see if hypothetical biases existed.

Table 2.3 Bohm CV contexts to testify the hypothetical biases and their effects

Approach	Experiment Context on WTP (the amount respondents have to pay)	Expected Responses
1	According to his maximum WTP as stated	Downward bias (Free-rider)
2	Pay in proportion to their stated maximum WTP	Downward bias
3	Uncertain, any one of several alternatives, the choice not yet being made	Uncertain*
4	A given amount, the same for all individuals	Bias in any direction
5	Will not have to pay or only pay negligible amount (Nothing)	Upward bias
6	Not ask the cost concerned with the volume of the public good increase	

Note*: In this situation, they will exaggerate their stated WTP to increase the possibility of the introduction of the new product; otherwise, for those who do not like to pay the WTP price, there is a possible tendency in the opposite direction.

For example, if the condition of providing the public good (television in this situation) depends on respondents' maximum WTP, then the demand of public good was biased downwards (free-riding); if the public good was provided for nothing, then the demand was biased upward. Although different preference revelation method had different direction of deviation in the experiment, his experimental results showed no significant difference (at the 5% level). In only one situation was the result significantly different from the first five approaches, which did not address the cost concerned with the volume of the public good increase (Approach 6).

Bohm's experiment provides a major advance to identify the extent to strategic behaviour in the presence of the public good. However, these empirical evidence is often "inconclusive", as stated by Throsby and Withers (1986) that the subject are "usually non-randomly chosen (e.g. students), only small groups are used in the laboratory setting. Therefore, Bohm's conclusion that the misspecification of the preference was less a problem than what believed by the economists is less convincing. He suggested that the test would seem to encourage further work in the field of experimental economics. From these early theoretical and empirical uncertainties, the existence of the free - rider problem is left as argument in the studies of demand provision of public/private good.

2.5.3 Incentive structure for preference of public good

Throsby and Withers (1986) provided a typology of incentive structure of the demand revelation for public good in the light of Bohm (1972)'s laboratory experiment. Table 2.4 reports the incentive towards strategic bias in the preference revelation for public good. They categorized the incentive structure along two dimensions: the payment liability and perceptions of the impact of responses on the good provision.

Table 2.4 Incentive towards strategic bias in the preference revelation for public good

Situation	Payment Liability	Perception on Provision of Good	Optimal Strategy (Incentive to Strategic Bias)
1	Have to pay	No impact	'strong' free-rider (underestimate)
2	Have to pay	Positive impact	'weak' free-rider (under/over estimate)*
3	No	No impact	truth telling
4	No	Positive impact	weak free-rider (over estimate)

Source: Throsby and Withers (1986), p. 308
* The relationship of stated to actual preference varies according to whether the individual values the public good at more or less than assigned cost. (Throsby and Withers, 1986, p310)

Payment liability represents whether or not individuals have the responsibility to pay for a unit of the public good, which is specified in the survey. Two possible cases can be presented: full liability (have to pay) and zero liability (do not have to pay). The perception of the impact of responses on the provision of good represents whether or not individuals believe that the supply of the public good may be influenced by their preferences, where there are two possibilities exist: positive impact where individuals believe their responses will have a positive effect on provision and no impact where individuals believe their responses will not affect provision. Throsby and Withers (1986) established a series of experiments following this incentive structure to evaluate art and proved the existence of strategic behaviour.

Bohm's experiment and the incentive schemes by Throsby and Withers (1986) suggested that the incentive properties in the revelation of preference depend on the payment and respondents' perceptions of the provision of good.

We have found that the early literature of examining the incentive compatibilities was mostly in the valuation studies of public good. In addition, Carson et al. (2000) contend that conventional CV techniques are not incentive compatible when considering the provision of a private good in a hypothetical context. They examined the incentive-compatibility by considering whether survey respondents will consider questions are consequential or not. They concluded that a survey is incentive compatible when: firstly, individuals perceive responses to the survey question as potentially influencing government or company action; secondly, individuals care about what the outcome of that action; and thirdly, information about the good, payment mechanism / vehicle and how the survey result be used in the future are provided to respondents in some certain way. For example, vague payment vehicles information runs the risk of overbidding and thus adding the possibility of introducing the new product.

2.5.4 Empirical evidence of incentive to strategic bias

In the transport context, the earliest methods used to obtain monetary valuations of attributes or to forecast likely behaviour were based around direct willingness to pay and stated intention questions. However, these methods were criticised in over predicting actual usage, for example to use a new transport service, often by a considerable amount, or yielding somewhat larger demand elasticities than other methods (Fowkes and Preston, 1991; Wardman and Shires, 2003). The reason is simply that respondents have an incentive to attempt to influence policy in their favour, since they are not committed to behave in accordance with their intentions and the policy is readily apparent. Similarly, differences were found between respondents' willingness to pay (WTP) for improvements and willingness to accept (WTA) a (similar sized) deterioration (Fowkes, 1995).

In the environmental literature, Contingent Valuation methods (CV) have been employed by economists to value changes for the goods not traded in the market-place, such as natural resources. However, the "basic issue is whether the necessarily hypothetical character of CV studies automatically renders their findings meaningless" (Mitchell and Carson, 1989, p171). A meta-analysis by List and Gallet (2001) found that respondents, on average, responding to the hypothetical situations overstated their preferences by a factor of about three. In response, the National Oceanic and Atmospheric Administration's (NOAA) blue-ribbon panel recommended that hypothetical bids be deflated using a "divide by 2" rule unless these bids can be calibrated using real market data (NOAA, 1996).

The concerns about bias in CV responses (Diamond and Hausman, 1994) are well documented and, as was the case in transport some years before, SP method was regarded as an advance (Wardman and Bristow, 2005, p.5). A paired comparison or multinomial choice format has been recommended as a means of reducing or eliminating the sensitivity of the estimate of the value of a particular good to the separate attribute in which it was valued. If all attribute variations are equally likely to occur, then the method is compatible with respondents revealing their true preferences. Adamowicz et al. (1999, p.467) stated that "Strategic behaviour should be minimal in SP tasks since the choices are made from descriptions of attributes and it will not be clear which choice will over- or under represent a valuation". List et al. (2006) addressed choice experiment may allow attenuation of the issue.

Recently, it has been found that choice experiments (CE) may also suffer from "the alleged problem with CV survey, namely, hypothetical bias" (Carlsson et al., 2005). Two distinct types of hypothetical bias can emerge: firstly, the decision to choose one alternative and secondly the intra-decision which is the marginal utilities of the attributes (List et al., 2006).

In the SP application in environmental science, Carlsson and Martinsson (2001) and Cameron et al. (2002) found significant difference in the marginal WTP in both a real and a hypothetical setting. Lust and Schroeder (2004, p.479) found that the predicted probability of purchasing one good (beef steaks) was generally higher (30%) in the hypothetical versus non-hypothetical setting; hypothetical total WTP for the good exceeded real WTP by 1.2 times (at the 5% level).

Empirical evidence in transport has found the existence of strategic bias. In public transport, Wardman and Whelan (2001) reviewed 45 SP values of new or improved trains from a large number of disparate studies. Not only did the values generally seem implausibly large, but they were found to be three times higher where the purpose of the study would have been readily perceived as valuing new trains. ATOC (2002), using the results from SP practice, advised that new trains will, on average, increase rail demand by 10%, this being around three times the demand which actually occurred in practice. This may have been because fares were regulated by government and hence there was an incentive for respondents to give high values to increase the chances that new trains are introduced, without financial consequences to themselves.

In the meta-analysis reported by Wardman (2001), there was a significant effect on the value of time if toll was the numeraire compared to other numeraires. When using a toll or road charge coefficient to calculate the value of time, it was found to be 19% lower than otherwise. Wardman and Bristow (2005, p.6) stated: *"Charging for the use of road space is a contentious issue and it is difficult to mask the purpose of studies dealing with it. SP studies covering tolling often detect a higher sensitivity to this than other cost variations whilst the sensitivity to tolls will be higher for their introduction to currently untolled roads than for variations on currently tolled roads or newly built tolled roads. These patterns are indeed evident in the literature."*

Furthermore, evidence showed that respondents overestimated the value of crowding, delaying, standing in crowded conditions and interchange during the journey. Respondents gave large standing penalties to 30 minutes standing in the survey, and "and even higher values" for the standing in crowded conditions (Wardman, 2003). The work covered 23 valuations from 8 studies. The mean value of standing time relative to seated time in the 20 instances where the purpose of the study would clearly have been seen as valuing overcrowding was 3.5. This fell to 2.7 across the three values from 2 studies where overcrowding was an element of a broader study looking at aspects of mode choice and interchange. However, in the research carried out on the London Underground on estimated of actual choice of train (LRT Operational Research 1988), standing time is valued at between 1.4 and 2.2 times seated time estimated the RP values of the penalty of having to change trains to be lower than the SP values, with the former averaging 5.4 minutes and the latter averaging 11.5 minutes. In a review of service quality values, Wardman (1998) found that the late arrival time was valued 7.4 times more highly than in-vehicle time. The excessive high value was suspected as respondents seriously biased their answer to reflect the inconvenience.

In summary, evidence from SP applications in environment and transport suggests the existence of strategic bias. CE can attenuate this bias; however, the application of this method has found that it still suffers from the existence of strategic bias.

2.5.5 Methods to reduce the bias

The existence of incentive to strategic bias has motivated a growing number of researchers to explore techniques to eliminate such bias, thereby providing methods from which unbiased estimate of WTP might be obtained from SP values. This section presents a review of these methods. It has been found that calibration techniques, counter-strategic warning message (Cheap Talk script) and masking research aim can eliminate strategic bias. These will be discussed in turn.

Calibration techniques

Bohm (1979) suggested an "interval" method whereby two samples be asked their WTP, with one sample being asked to pay an amount equal to the stated individual WTP, and the other to pay nothing. An incentive structure is established for under-and overstatement (see Table 2.3), respectively, of the value of the new good. If the WTP obtained from the two samples are the same, there is no misspecification of preference and a true response is obtained. If there is a difference, then the true response lies between the boundaries so established. The narrower the interval, the more precise the information. Throsby and Wither (1986) applied this method to evaluate art, and proved the existence of free-rider behaviour and strategic bias.

This "interval" method can be treated as an early use of calibration. Calibration techniques can be applied to obtain an unbiased estimate. Fox et al. (1999) applied within-sample techniques where responses to hypothetical and real valuation questions were combined. A calibration function is estimated which relates the difference in responses obtained in two experiments to respondent characteristics. Combining RP and SP responses is applied to adjust the values (Hensher et al., 1999; Louviere et al., 2000; Wardman and Whelan, 2001).

Calibration may prove to be a useful tool for ex post adjustments of SP values. However, its practicality appears to be limited by two aspects: firstly, the extent to which a calibration function derived for one good can be used to calibrate that of a different good (Fox et al., 1999); and secondly, the responses for the real situation (RP) can not always be obtained, considering the fact that SP surveys are designed to achieve responses to the potential change/good. Our SP experiment will be on the users' valuation of improved rolling stock, the RP data is unavailable. We will not use this method to amend the bias in this thesis.

Counter-strategic message

A second and very different approach to dealing with the strategic bias focuses on the design of the questionnaire. It might directly induce respondents to provide responses to hypothetical questions, as they respond to the actual situations.

Bohm (1984, p.140) stated that " to improve the performance of any method in which there are possible incentives for misrepresentation, measures should be taken a priori to reduce the inclination to give in to such incentives". He suggested two methods to counter-strategic. Firstly, respondents are told that their WTP statements are not anonymous but will be made public. Secondly, respondents are informed about: 1. the importance of being able to base collective decisions on consumer valuations instead of politicians' valuations or their interpretations of consumer valuations. 2. the implication of giving in to the misrepresentation incentives, namely that decision-making would be based on unrealistic data.

Recently, "Cheap-talk" script (Cummings and Taylor, 1999) can be found in the literature as the 'counter-strategic' message in the valuation studies. Cheap-talk is a term learned from game theory, and here refers to an explanation of previous bias with a warning message of the budget constraints introduced in the questionnaire. This script is provided prior to the hypothetical choices. Its application in the CV and recently CE studies found that a properly designed cheap-talk can effectively reduce or eliminate the bias caused by the hypothetical nature in survey.

The simplicity and cost effective feature of cheap-talk script makes it an attractive approach to reduce the hypothetical bias in the CV and SP studies. A review of cheap-talk application is presented in section 2.6. However, it is important to understand how it works in different context. In this study, we will introduce this method to detect the impacts of Cheap-talk script on amending the incentives to strategic bias in individuals' choice making.

Masking the research aim by introducing more attributes

Some other successful exploration of amending incentive to bias has been done by Wardman and Bristow (2003). They conducted a research on individuals' WTP for the improvement of aircraft movements (noise). Besides the contribution to this area, they also explored the incentives to bias by doing two SP exercises.

In the first SP exercise, more attributes considering the quality of life dimension were used to mask the purpose of the study, such as the safety (Crime) and education (School Pass Rate and Library). By doing that, it is difficult for respondents to perceive the main aim of these experiments, thus giving them no incentives to bias their answers. The second SP exercise is designed as a standard SP approach. There are two abstract alternatives (A and B), which offered trade-offs between money and aircraft movements. SP2 used the conventional SP design, which the aim of the experiments is straightforward to respondents. They found large differences in valuations of aircraft movement between these two SP exercise. The latter can be expected to attract more strategic bias and the higher values obtained by it are consistent with this. They concluded that the incentive for respondents to strategically bias their answers is that respondents being able to perceive the aim of research, thus providing the biased answers for some certain better outcome.

Their study suggested that where the objective of the exercise is obvious, especially where the issue is contentious, the strategic bias is likely to occur; more importantly, it provided an effective method to amend the incentive to strategic bias. However, introducing more attributes might add the risk of task complexity effect to the SP experiments. The literature of task complexity effects in SP application suggested that task load/SP design itself would impact on individuals' decision making. Increase of the task load would cause biased answers or less consistency in the choices.

In this study, this method will be introduced. More attributes will be introduced to some of the SP experiments. A discussion of the impacts of adding more attributes will be provided, both on the impact of amending the incentive to strategic bias and its possible impacts on the task complexity effects in individuals' choice making.

2.5.6 Summary and implication for this research

The above review presented the economic background of incentives to strategic bias in the hypothetical surveys, and illustrated with examples of strategic behaviour observed in the transport field. The review suggested the possible reasons for individuals to strategically bias in the hypothetical surveys are:

- Respondents perceive that their answer potentially affect the provision of the good by the government or company, thus having the incentive to bias their answer to increase the possibility of the introduction of the good, which believes to increase their utilities;

- The hypothetical nature of SP survey leaves the payment outside of the experiment; therefore, respondents do not have any financial consequence of their statement.

In the present study, the existence and possible consequence of the strategic bias in users' valuation of improved rolling stock will be examined. Methods to amend the bias will be tested. Among the three methods suggested in the review, the calibration method cannot be used due to the unavailability of the RP data. The second (Cheap Talk script) and third method (Adding more attributes to mask the SP survey aim) will be introduced into our SP experiments to detect their impacts on amending incentives to strategic bias in the context of users' valuation of rolling stock (a review will be presented in chapter 3).

Adding a CT script is a cost-effective and easily implemented method. A review of the origin and applications of CT scripts in the related studies is presented section 2.6. The implications for this study and research questions will be formalised and discussed based on the review.

Masking the SP survey aim by introducing more attributes also provides an effective method to amend the bias, however, introducing more attributes might add the task complexity to the SP experiments, and might change individuals' decision making process. Therefore, besides of the impact on the strategic bias, its impact on individuals' decision making process is examined in this study. Section 2.7 presents a review on the task complexity effects in the SP survey.

2.6 Review of Cheap Talk (CT)

2.6.1 Introduction of cheap-talk

To evaluate public good or private good which do not exist in the market, such as environmental or agricultural goods, Contingent Valuation (CV) experiments, and recently Choice Experiment (CE) are the main applied methodology. Both survey methods ask respondents to make hypothetical trade-off, a feature that enables researchers to obtain the marginal utilities (WTP) of attributes. The review in previous sections found both methods suffer from biases caused by the hypothetical nature of the survey. In the CV literature, this bias is called hypothetical bias. In our study, it is called strategic bias. This has motivated researchers to develop techniques that either eliminate or adjust for this bias. Some methods which have been suggested to amend the incentive to hypothetical biases are presented in Section 2.5.5. Among them, cheap-talk (hereinafter, CT) script seemed to be one of the most successful attempts and it was selected for further test in this present study.

Cummings and Taylor (1999) successfully employed "Cheap-talk" in the CV design to reduce the hypothetical bias and proved that their CT script is robust via different goods. This can be regarded as the first application of CT. The CT they used in the survey is a script describing the bias problem, and explicitly asking respondents to avoid overstating their true WTP. CT draws from lessons in experimental economics and psychology concerning the design of valuation institutions. Cummings and Taylor (1999, p.650) stated that: *"Cheap talk refers to the costless* *transmission of signals and information (i.e., cheap talk does not directly affect the payoffs of players in a game). Many game theorists typically use the term "cheap-talk" in referring to nonbinding communication of actions by two or more players in a game prior to their actual binding commitment, ..., the term "cheap-talk" is used in a parallel way, referring to nonbinding communication of actions by two or more players in an experiment prior to their hypothetical commitment."*

Following Cummings and Taylor, a series of CT designs were applied to the CV and CE studies. Using private goods, classroom experiments, or closely controlled field settings, the use of CT proved to be potentially successful (List, 2001; Bulte et al., 2005; Carlsson et al., 2005; List et al., 2006). While the hypothetical mean WTP without cheap-talk was significantly higher than WTP using actual economic commitments, the hypothetical WTP with CT script could not show a significant difference from the actual WTP. However, some of the results were not as good as expected. "The effectiveness of cheap-talk at attenuating hypothetical bias has been shown to be robust, but context dependent." (List et al., 2006, p.6).

This section reviews the previous use of cheap-talk in the CV and CE practice. It provides a brief introduction of previous application, and explains the rationale behind the CT. The factors that affect the effectiveness of CT in the previous use are discussed. The summary and implications for this study is provided.

2.6.2 Application of cheap-talk

NOAA (1994, p.23) proposed rules for the conduct of CV surveys for natural resource damage assessment, which require that "Prior to the value elicitation (in CV surveys), respondents shall be reminded of their budget constraints and their alternative expenditures. Respondents shall be reminded that their WTP for the environmental program in question would reduce their expenditures on other goods. This reminder should be more than perfunctory, but less than overwhelming."

Short Reminder

Prior to the CT method, as now understood, being introduced, short reminders were applied in the CV studies in response to NOAA protocol requiring reminding respondents of opportunity costs. Loomis et al. (1996) and Neil (1995) applied the short reminder statements about substitutes and budget constraints by explicit discussions of alternative goods and their costs prior to a valuation question for a specific good. They found that the short reminders were ineffective in removing hypothetical bias, although it could reduce the difference the hypothetical payments exceeded actual payments from 3:1 to 1.8:1. This is different from

Cumming and Taylor's CT script which contained not only the budget constraints/substitutes, but also an explicit discussion of the bias problem.

Cumming and Taylor's experiment with CT

Cummings and Taylor (1999) firstly introduced a CT script to their experiment. The script contained an explicit discussion of the hypothetical bias problem – what hypothetical bias is and why it might occur. Their results suggested that CT can effectively eliminate the hypothetical bias and proved robust among different (three different kinds of) goods. The efficiency of the CT script in eliminating hypothetical bias was tested in the 16 CV experiments of WTP for three different public goods (e.g. contribution to National Conservancy in Georgia). The respondents were undergraduate students in Georgia State University.

In the experiment, some individuals voted on the hypothetical referendum, whilst some voted on the real referendum where they need to actually pay at the end of the game if the referendum was passed. The CT method only differed from the hypothetical method only in that additional words were read to the respondents prior to their decision making. They found the existence of significant difference between the hypothetical and the real referendum, which indicated the existence of bias in the hypothetical referendum. The hypothetical YES responses were 17.3%, 16.7%, and 19.4% higher than the YES responses to the real referenda for three public goods respectively. Introduction of a CT effectively reduced/eliminated the hypothetic bias as the responses from hypothetical referendum with the CT script were indistinguishable to those from the real referendum.

This study can be seen as the first research on the application of Cheap Talk in seemingly eliminating the bias caused by the hypothetical nature of the valuation research. They suggested CT was a cost-effective and easily implemented method to amend the bias. Their results also suggested that their CT script did not "overcorrect" the bias, by which they meant that CT scripts did not introduce a bias and which just happened to offset the hypothetical bias found in the experiments. There are some limitations to this study: firstly, similarly as the experiment by Bohm, the application of the CT script in this experiment was conducted in a laboratory setting, respondents were students from university. The result of the experiment is less convincing as the sample is homogeneous. The second limitation lies in the fact that whether or not the CT script works in different contexts or for different kinds of good is left to be unknown.

Following on experiments on CT

Following on their experiment, there are a few applications of CT in the CV and CE studies, as summarised in Table 2.5. List (2001) expanded upon Cummings and Taylor's study by using their CT script in an auction for sports card (trading card with a sports-related subject) by personal interviews with sports card dealers and non-dealers (experienced and non-experienced respondents, respectively). It was discovered that the hypothetical bids were statistically higher than the actual bids, which indicated the existence of hypothetical bias in both dealer and non-dealer experiments. Results indicated that although cheap-talk mitigated hypothetical bias in the non-dealer experiments, it did not eliminate bias for dealers (experienced participants). List (2001, p.1498) provided an explanation that "the theory of value formation suggests that experienced bidders may not be easily swayed by the cheap-talk design as they have a well structured preference ordering for the good in question".

Poe et al. (2002) applied a short version of CT (truncated from Cummings and Taylor's script), and found that the short script failed on eliminating the hypothetical bias on valuing green power and tree planting in New York via CV survey. This finding is consistent with the results of Cummings et al. (1995) on the light CT script.

Murphy et al. (2003) tested the effectiveness of CT on the voluntary contribution to a Nature Conservancy in university. Respondents were asked a series of follow-up questions of their decision-making process and reaction to the CT script. About 56 percent of respondents reported in follow-up questions that they had reduced their payment in response to CT script. However, about 44 percent stated that they had already carefully considered their contribution decision and they were, therefore, not affected by the CT script. The results are consistent with List's arguments about why CT failed to eliminate hypothetical bias for respondents who had past experience with the good as their preferences were already well formed.

Aadland and Caplan (2003) found that the effectiveness of the CT varied by type of household. In particular, those households who might be expected to suffer the most from positive hypothetical bias also tend to lower their stated WTP the most in response to CT. They later expended the experiments (Aadland and Caplan, 2006) by crafting neutral cheap-talk statements rather than an explicit of biases with its direction and magnitude. This is motivated by the fact that researchers normally do not know *ex ante* whether hypothetical bias will exist or in which direction it will be. However, they found that shorter and neutral CT script appropriately tailored for phone interview worsen, but not eliminates/ reduces the hypothetical bias.

Brown et al. (2003) found that the long CT script was successful in a referendum, but only for higher payment amounts. More recently, Carlsson et al. (2005) and List et al. (2006) applied CT in the SP experiments to test the external validity. Both of them found the existence of significant positive hypothetical bias in the SP experiment on valuation of the WTP for private good (food) in the mail survey/actual market place. Short version CT is applied considering the limit of space in the questionnaire. They found that the CT can effectively lower the stated WTP in a hypothetical setting, and the values are statistically indistinguishable from actual responses.

Table 2.5 Previous applications of CT in CV and CE studies

Studies	Goods	Methods	CT Content	Results
Cummings et al. (1995)	Public Goods	CV-Referendum	1. Describes the hypothetical bias phenomena; 2. Discusses possible explanations for this phenomena-why different; 3 Requests that subjects vote in the upcoming hypothetical referendum as if it were a real referendum	1. The light CT have a worsening effect (increase the possibility respondent vote yes by 21%); 2. Heavy CT can effectively reduce the bias (reduce the possibility by 22%)
Neil (1995)			1. Remind respondents of budgetary substitute and cost	1. Found the existence of hypothetical bias; 2. Short reminder can not statistically eliminate the bias
Loomis (1996)	Private Goods		1. Remind respondents of budgetary substitute and cost; 2. Remind respondents to value the goods as if it were in the true valuation.	1. Found the existence of hypothetical bias; 2. Short reminder can reduce the hypothetical bias but not statistically
Cummings and Taylor (1999)	Natural Conservancy good	CV-Referendum	1. Describes the hypothetical bias phenomena; 2. Discusses possible explanations for this phenomena-why different; 3 Requests that subjects vote in the upcoming hypothetical referendum as if it were a real referendum	1. Existence of the positive hypothetical bias; 2. CT script effectively reduces the positive hypothetical bias; 3. The script is robust among different goods and experiments; 4. No "overcorrection" for subjects in response to the CT script
List (2001)	Private Goods	CV-Vickrey Auction	Followed Cummings and Taylor (1999) CT script	1. Existence of the positive hypothetical bias; 2. CT eliminated the positive bias in the non-experienced respondents, while did not eliminate the bias in the experienced group
Poe et al. (2002)	Public goods	CV-PPM (Provision point mechanism)	Short version of Cummings and Taylor (1999) CT script	Short version of cheap-talk can not eliminate the positive bias in the CVM
Murphy et al. (2003)	Natural Conservancy good	CV-VC(voluntary Contribution Method)	Followed Cummings and Taylor (1999) CT script; besides, they added following -up questions to probe how CT affects people making decisions	1. Existence of the positive hypothetical bias; 2. CT can eliminate the positive bias in the hypothetical valuation but there is a significant difference between the hypothetical valuation with the CT and the real referendum; 3. In the following up questions, 56% respondents stated that the CT reduce their WTP for the good, and the 44% stated that they already got an well-formed payment in mind, so the CT had no effect on them.

Table 2.5 Continued

Studies	Goods	Methods	CT Content	Results
Aadland and Caplan (2003)	Public Goods	CV-Referendum(Telephone survey and household interview)	Short version of Cummings and Taylor (1999) CT script	1. Effectiveness of CT script varies by different household; 2. This is the first case that short version CT script can effectively reduce the bias statistically.
Aadland and Caplan (2006)	Public Goods	CV-Referendum	Followed Cummings and Taylor (1999) CT script, except that they gave a neutral statement for the hypothetical bias	1. Existence of positive hypothetical bias; 2. Long CT design eliminated the bias; 3. No significant difference between hypothetical referendum with short CT and those without CT
Brown et al. (2003)	Public Goods	CV- Referendum	Followed Cummings and Taylor (1999) CT script,	CT can effectively eliminate the hypothetical bias but just when the payment is in the higher level.
Bulte et al. (2005)	Natural Conservancy good	CV- Referendum(television and online questionnaire)	Followed Cummings and Taylor (1999) CT script, but short version	1. Existence of positive hypothetical bias; 2. Stated values obtained using CT and consequential treatments are statistically indistinguishable, and significantly lower than values obtained using a hypothetical situation; 3. Short version CT can effectively reduce the hypothetical bias
			Consequentialism scripts (indicating the study results will be considered by policy makers)	
Carlsson et al. (2005)	Private Goods (food)	Stated Choice (mail survey)	Short version of Cummings and Taylor (1999) CT script	1. Existence of positive hypothetical bias in the WTP for food; 2. CT can effectively reduce the positive hypothetical bias in the mail survey
List et al. (2006)	Contribution to Public goods	Stated Choice (capital campaign	Short version of Cummings and Taylor (1999) CT script	1. Existence of positive hypothetical bias in the marginal valuation methods. 2. Responses across the real and hypothetical with CT treatments are not statistically different from each other. 3. The finding is robust across both inexperienced and experienced consumers. 4. Respondents in the hypothetical with CT treatment are more likely to make inconsistent decisions
	Sports Card	Stated Choice (actual marketplace)	Long version of Cummings and Taylor (1999) CT script	

From the review of previous applications of CT in CV and SC experiments, the effectiveness of CT are found to be sensitive to the script length and content, and some "produced undesirable results" (Aadland and Caplan, 2006, p.575). Researchers feel caution is warranted in using CT to correct for hypothetical bias until it can better understand how the length and content of CT statements influence the cognitive processes of survey respondents.

2.6.3 Rationale behind cheap-talk

Opaluch and Segerson (1989) noted that for unfamiliar goods, individuals may not know precisely their WTP, but can place it within a range, or ambivalence region. Hence, should the amount asked fall within this region, the person may become more uncertain about her response. When payments are real, the respondent may invest more cognitive effort to reduce the ambivalence region. This could lead to different responses in real and hypothetical settings. The rationale behind the CT script is to coax individuals to invest this cognitive effort even though payment is hypothetical.

Cummings and Taylor (1999) used the correction process in social judgement to explain the impact of CT on respondents' decision process. "The cheap talk script makes respondents aware of the potential influence of the context of a hypothetical referendum on their valuation of a good. In such cases, social psychologists find that subjects may "...effortfully subtract or partial out reactions toward the target" (Wegener and Petty, 1995, p.37). In other words, respondents may "effortfully" attempt to correct for the hypothetical nature of the referendum." (Cummings and Taylor, 1999, p.663)

Aadland and Caplan (2006, p.572) cited Fischhoff (2002)'s theory of how human cognition reacts to signals to explain the effectiveness of cheap-talk. "Artifacts (such as unexpected responses to cheap-talk) could emanate from the subtle ways that interviewers communicate their expectations...elicitation is a reactive measurement procedure...The process assumes that people sometimes need help, in order to understand what they believe and want. That help may include presenting balanced selection of opinions, lest clients miss a critical perspective just because it did not occur to them at the time."

2.6.4 Factors affecting the impact of cheap-talk

Content of CT

The effectiveness of CT script is found to be sensitive to the content and length. The review has found that CT scripts which are proved to be effective all contain: first of all, the budgetary constraints and substitutes; secondly, and the explicit explanation of hypothetical bias and introduction of its magnitude and direction observed from previous experiences.

Short CT exhibits mixed evidence of the effectiveness: in some experiments, the short CT can not eliminate the hypothetical bias, and in some cases can even exacerbate the problems (Cummings et al., 1995; Poe, 2002; Aadland and Caplan, 2003). Some recent CT applications in SC experiments found that short CT reduced the hypothetical bias in the response (Bulte et al., 2005; Carlsson et al., 2005; List et al., 2006).

Most previous research in CT was carried out by in-house or personal interview, only few experiments were carried out by telephone interview (Aadland and Caplan, 2003) or mail surveys (Carlsson et al., 2005; List et al., 2006).

Respondents

It has found that CT script is effective in some certain group of people. For example, List (2001) tested Cummings and Taylor (1999)'s CT script in the field with in-person interviews using a private good and found that it did not eliminate the hypothetical bias in all groups of respondents, especially ones who were experienced and already had a well-formed preference. Aadland and Caplan (2003) found similar results. Brown et al. (2003) and Murphy et al. (2003) found that Cummings and Taylor (1999)'s CT script eliminated hypothetical bias associated with higher payment levels, but was much less effective for lower payment levels.

2.6.5 Summary and implications for this research

This section reviews applications of CT script in the CV and CE studies: its origin and following application. The review found that the effectiveness of CT is sensitive to the script length and content. Researchers feel caution is warranted in using CT to correct for hypothetical bias until it can better understand how the length and content of CT statements influence the cognitive processes of survey respondents.

In this study, as suggested in section 2.5.6. a Cheap Talk script will be applied in the light of the script by Cumming and Taylor's, to identify if or not the strategic bias is existed in users' valuation of rolling stock and whether or not CT can be used to amend the bias by comparing responses to different experiments and the previous evidence (regarded as the real value). In addition, the effectiveness of CT among different population will be tested.

Different from previous experience, this study will also examine:

Firstly, further testing will be done to find out how respondents cope with this information in their choice making process using follow-up questions. For example, this study will investigate if adding CT script will add realism to SP survey by exploring respondents' perceptions of cost change in the SP choice for the introduction of the improved rolling stock.

Secondly, CT scripts are mostly applied in CV studies where only the impacts on the cost coefficient and WTP are investigated. There are more than one attribute in the CE. Recently, CT script has been applied in the CE. However, no research identified the impact of CT on the valuation of other attributes and/or if the impact exists, whether or not it is adding a different kind of bias to the experiment.

This study will therefore be unique and a major contribution to the literature.

2.7 Review of Task Complexity

2.7.1 Background

Section 2.5 reviews the existence and possible consequence of incentive to strategic bias in the SP surveys. Methods are suggested and discussed in section 2.5.5 to amend or adjust the bias. In this study, we will introduce a Cheap Talk script (see the review in Section 2.6) and adding more attributes to the SP experiment, which is motivated by Wardman and Bristol (2003)'s successful empirical evidence in valuing the aircraft noise (see Section 2.5.5). Their impacts on amending the incentives to strategic bias in the SP study will be discussed.

Traditionally, SP modellers have developed methods to explain respondents' decision behaviour, which tended to focus on the information supplied by the choices themselves, rather than the influence of the SP design can have. This is based on the assumption that when respondents make decisions, the quantity and structure of a choice will not affect their ability to choose the optimal choice. Recent studies in the SP application have found that the choice structure (instrument design) actually affects how respondents making decision (Bradley and Daly, 1994; Swait and Adamowicz, 2001; Deshazo and Fermo, 2002; Caussade et al., 2005). More specifically, the amount of information in the choice structure has an impact on the choice consistency and/or magnitude of valuation. This is knows as task complexity effects. Their results suggested that the assumptions typically made by SP-modellers could be inadequate in view of the limited ability of respondents to process information in the experiment. Robust models of respondents' behaviour should incorporate this task complexity effect (Swait and Adamowicz, 2001). The empirical evidence is supported by the behaviour theories and psychology theories, as we will briefly describe in section 2.7.2.

The present study will test if adding two more attributes can amend the incentive to strategic bias; however, adding more attributes might affect individuals' choice making in the way that it adds more task load/information to the experiment. Therefore, the impact of adding more attributes on the respondents' decision making will be tested.

This section reviews the task complexity effects, which includes a review of the related literature in the behavioural theory associated to this topic and the hypotheses to be tested. It also reviews the empirical examination of task complexity effects in the SP application. Finally it presents a review of the relationship between the task complexity effects and other types of bias in the SP application.

2.7.2 Behavioural theories on task complexity effects

Researchers suggested that decision makers are information processors with limited capabilities and resources, trying to make the best possible decisions within operational constraints.

Behavioural economists predict that an increase in choice set complexity will compromise choice consistency. Simon (1955) was the first to question the rationality of human behaviour. He suggested that consumers develop "satisficing" decision rules to avoid the full cognition cost of complexity by considering only a portion of the information available in the choice set. Heiner (1983) defined the C-D gap to be the gap between cognitive abilities of the decision maker and the difficulty of the decision. The C-D gap will happen when respondents find the task becoming complex. Heiner predicted that as this gap grows, consumers will find it "welfare-enhancing" to restrict the range of decision rules they consider. DePalma et al. (1994) identified individuals' imperfect abilities to choose within a traditional utility framework, and predicted that as choice complexity increases, the magnitude of sub-optional mistakes will increases, resulting in a lower consistency in responses.

In the psychology literature, Payne et al. (1992) defined a typology of decision strategies. The typology characterised decision strategies along three dimensions: basis of processing, the amount of information processing, and consistency of processing. Payne et al. stated that individuals construct strategies depending on the task demands and the information they are faced with. The full introduction of these six strategies can be found in Payne et al.

Some behaviour theories believe that the decisions may often be made by mixed strategies. For example, decisions may often be made in two stages: in the first stage, alternatives are screened by some non-compensatory process (e.g. elimination-by-aspects-EBA, Tversky, 1972), and in the second stage, remaining alternatives are evaluated in more detail, perhaps with a compensatory decision rule (Swait and Adamowicz, 2001).

Keller and Staelin (1987) suggested that the consistency of individuals' decisions is affected by the amount of information they must process, supporting the existence of an "information overload" effect. They proposed that complexity of choice experiments may hold a U-Shape relationship with decision effectiveness. That is, as the situations becomes more complex, individuals initially exert additional efforts and become more effective, until a point is reached at which their effectiveness begins to deteriorate.

In summary, psychologists and researchers in decision making behaviour support the view that decision makers are information processors with limited capabilities and resources, trying to make the best possible decision within their operational constrains. They suggested that there is a trade-off between cognitive effort and outcome accuracy. Based on the above theories, a few researchers have incorporated the task complexity within a framework in the SP application.

2.7.3 Empirical examination of task complexity effects in SP experiments

A review of examination of task complexity effects in SP practice is presented.

Measurement of task complexity in SP experiments

Researchers have explored the effects of increasing task load or complexity of choice sets on the consistency with which individuals make choices. These effects include:

- Number of attribute per alternative (DeShazo and Fermo, 2002; Arentze et al., 2003; Caussade et al., 2005; Hensher, 2006a);
- Number of alternatives (Malhotra, 1982; DeShazo and Fermo, 2002; Arentze et al., 2003; Caussade et al., 2005; Hensher, 2006a);
- Number of levels characterizing an attribute (Mazzotta and Opaluch, 1995; Caussade et al., 2005);
- and various measures of the correlational structure of information within alternatives, across alternatives and across a given attribute within the choice set such as the range of attributes (Dellaert et al., 1999; DeShazo and Fermo, 2002; Caussade et al., 2005; Hensher, 2006a).

In the present study, varying the number of attributes and information structure are applied in the SP experiments to test the research hypotheses (see chapter 1). A review of previous studies in these aspects is provided.

Impact of the number of attributes on responses

The number of attributes per alternative has shown a significant impact on SP responses. Previous studies investigate the impact of number of attributes in roughly two aspects. We can differentiate between these two main types of studies which explore how respondents deal with more complex experiments. However, they both reach different conclusions.

The first type of studies assumes that adding more complexity to the experiment, respondents attend to all information in the choice set but only increasingly make mistakes in processing that information (DePalma et al., 1994). The impact of number of attributes is detected by the variance of the random term in the utility function, such as through a Heteroskedastic Multinomial Logit model (HMNL, Swait and Adamowicz, 2001). Using a HMNL model, DeShazo and Fermo (2002, p.136) examined both complexity effects and consistency in choice making. The scale factor was parameterized as a function of the amount of information or correlation structure of the data. They found that when the number of attributes increase, the variance of random terms increased, which led to the conclusion that adding the number of attributes has an impact on the choice consistency.

Caussade et al. (2005) conducted a stated route choice experiment and investigated the influence of SC design complexity on individuals' ability to choose. Same as DeShazo and Fermo (2002)'s work, they examined the impact of the number of attributes by using the HMNL model. They found same results that adding the number of attributes had a clear detrimental effect on the ability to choose, contributing to a higher error variance.

Caussade et al. (2005)'s study differed from DeShazo and Fermo (2002)'s work in the way of varying the number of attributes in the SP experiment. DeShazo and Fermo (2002) varied the number of attributes by selecting only a few attributes from a larger set; whilst Caussade et al. (2005) varied the number by aggregating some attributes to be a less-attribute experiment. The former study might cause less-information (missing–data) problem that the absence of certain attributes in the less attribute condition might impact on the coefficient estimation. They solved the problem by incorporating dummy variables denoting the existence of missing variables in the utility function and later extended their study by using a rational-adaptive model to explain individuals attribute processing (DeShazo and Fermo, 2004). However, Caussade et al. (2005) study has not taken account of individuals' attribute processing strategy when facing different types of survey, they found that 'some respondents answered that they had aggregated or ignored some attributes in the attribute-rich condition design' (p.632). So respondents would behave differently in the attribute-rich/-less conditions.

The second type of study assumes that individuals recognise that their limited cognition has positive opportunity cost, thus steward them as efficiently as possible, which is know as a rationally-adaptive manner.

In the empirical demonstration, DeShazo and Fermo (2004) incorporated design features (measurements of task complexity) by adding the "propensity to attend" variables into the convention utility function to show that individuals are systematically adapting their pre-choice behaviour to the costs and benefits implied by various informational structures within the choice set. They stated (p.4): "Individuals will therefore allocate their attention across alternative–attribute information within a choice set in a rationally adaptive manner by seeking to minimize the cost and maximize the benefits of information evaluation".

Hensher et al. (2005b) explored individuals' information processing strategy (IPS) in the Stated Choice situations. They found that respondents have different information processing strategies of how specific attributes are processed, in terms of exclusion and inclusion. Their results suggested a sizeable difference in the mean value of travel time savings (VTTS) before and after accounting for the attribute processing strategy for each individual.

Later, Hensher (2006b) extended the research on individual's IPS by investigating the impact of rules such as attribute aggregation (i.e. travel time components) and reference dependency (such as contrasts of attribute levels in the SC design relative to recent experience) on preference profiles. They found that if the attributes levels deviate less from the reference (or experienced) level, it is more likely to produce lower mean value of travel time saving (VTTS) than those where the difference is greater. If an attribute has components that are potentially additive; the mean VTTS is higher when a respondent evaluates the components via an addition rule.

Hensher et al. (2007) provided a "stochastic" specification of attribute processing capable of accommodating the widespread consensus in the decision-making literature that decision-making is an active process which may require different decision making strategies in different contexts and at different stages of the decision process.

The second type of studies suggested that with increase of number of attributes, individuals might not attend to all the attributes, but use some decision strategies to minimise the cognitive cost, such as adoption of attribute processing strategies.

This section reviews two types of empirical evidence on how individuals deal with the complex SP (adding more attributes) experiment, which draws two different conclusions. In our study, as suggested in section 2.5.6, we will test the impact of adding more attributes on amending the incentive to strategic bias. However, whether or not this method will add the task load to SP experiments will be tested. We will examine the task complexity effects in the light of both studies. This study will try to find whether or not task complexity effects exist when add more attributes to SP survey and the possible impacts on respondents' choice making. The review in this section leads to our research hypothesis (H3B – see chapter1).

Impacts of the range of level for attributes

Another way of measuring the task complexity in the choice experiment is based on the range of levels for attributes. Mazzotta and Opaluch (1995) empirically tested the validity of Heiner's hypothesis concerning choice complexity, and strongly suggested the existence of a C-D gap. In their experiment, complexity was measured as the degree to which attribute levels differ across two alternatives. They incorporated the complexity of a choice task in the variance of a discrete choice model and show that the variation of levels for attributes in the choice set affects the parameter estimates.

Swait and Adamowicz (2001) studied this problem in more depth. They interpreted choice complexity by several measurements, and one of which is the distance between alternatives. They proposed the HMNL model (see section 4.3.5) to take account for the task complexity and consumer behaviour through a parameterisation of the scale factor as a function of the experimental features. By applying their model into the SP experiments, it supported that omitting the task complexity in the model will result in the biased estimate.

Fatigue effects and other exploration

In the empirical economics literature, Bradley and Daly (1994) were the first (in transport) to incorporate task complexity in a random utility model. They used scaling approach (see Section 4.3.3) to test for the fatigue effects in rank-order data, and concluded that scale effects existed. The increase in the task load of the choice set is correlated with increase in the variance of random component (i.e. error) of the logit models. As the ranking become lower and as the number of pair wise choices completed becomes greater, the amount of unexplained variance is shown to increase. Saelensminde (2001) used the scaling approach to investigate the differences in the amount of unexplained variance caused by inconsistencies in responses, showing that such a scaling effect existed, which also agreed with Bradley and Daly's finding.

2.7.4 The interaction effects between task complexity and other bias

Little research has been done to investigate the interaction effects of task complexity and other biases that could happen in the SP practice. From the previous evidence on the task complexity effects as shown above, the complexity of SP experiments influenced consistency of choice.

In related literature, the complexity of choice experiment has been related to the propensity to "avoid" choice by deferring or choosing the status quo. For example, Tversky and Shafir (1992) showed that when that choice environment is made complex (by adding alternatives or making the choice alternatives similar, but not identical), some individuals opt to delay choice, seek new alternatives, or even revert to a default (status quo) option. Similar results are found by Dhar (1997). Lust and Schroeder (2004, p468) proposed a hypothesis that the task complexity of the choice experiment "might serve to accentuate problems with hypothetical bias: subjects might behave inconsistently when they do not have to back up their choices with real commitments."

When the choice task gets more complex, individuals might take different strategy to cope with the task. In the present research, the interaction impacts of task complexity with the incentive to strategic bias will be explored.

2.7.5 Summary and implication to this study

In summary, related literature in economics and behaviour decision theory have convincingly illustrated how changes in task environment result in changes in decision-making. This provided us with a theoretical background for incorporating the impact of SP design on responses. In this study, two more attributes will be added to the SP experiment to test if it can amend individuals' incentives to strategic bias. However, the change of the choice experiment would add the risk of task complexity effects. Therefore, this study will investigate the existence and influences of the task complexity effects.

Based on the literature review, there are three ways for the forward testing: firstly, if or not respondents will make more mistakes in the experiment with two more attributes. This will be tested through the scaling approach by comparing the scale factors (inversely related to the variance of error term) of the simple and complex data sets. Secondly, if or not respondents will use different decision strategy in the experiment with two more attributes. DeShazo and Fermo (2004)'s method will be applied to detect the impact of design features on 'propensity to attend' the alternatives/attributes. Thirdly, some researchers (Hensher et al., 2007) suggested that the simplistic design may be 'complex' in a perceptual sense, since an individual expects more information which they know is relevant in making such a choice in real market setting. This study will explore individuals' perceptions of complexity by a following-up question. The relationship between the 'objective' and 'perceptual' complexity will be explored. The methodology of the tests will be introduced in chapter 4.

2.8 Conclusions and Implications for this Research

This chapter has presented general background and reviewed previous studies on bias in the SP application. Bias is defined in the context of SP methods. A typology of bias with explanation of the sources of bias is produced. Incentive to strategic bias and task complexity effects have been described and reviewed from theoretical and empirical evidence in this chapter, which led to the study aims, objectives and conceptual framework outlined in chapter 1.

SP methods have seen more and more applications in recent years, in transport, marketing and environmental science. However, bias is found in the SP practice, which is one of the main concerns of researchers, affecting the validity and reliability of SP results. The review of biases observed in the previous SP applications explored the sources of bias, which can be categorized as unrealistic design, incentive to strategic bias and task complexity effects. Amongst these, the issues of design/scenarios specification and task complexity have received a considerable amount of attention. On the other hand, and despite serious concerns in the early literature, the strategic biasing of responses tends to have been overlooked in recent times, particularly within the SP methodology. This study is motivated by the desire to investigate the incentives for respondents to bias their answer in the SP survey and methods to amend the bias.

Based on the literature review in this chapter, the reason suspected for the strategic bias is that respondents perceive that their answers might potentially affect the provision of the new good/service, thus having the incentive to shape their answers for some better outcome. It is also suspected that the hypothetical nature of SP survey leaves the payment outside of the experiment; therefore, respondents do not have any financial consequence of their statement. Large empirical evidence proves the existence of strategic bias in the SP application.

Cheap-talk and masking research aim by introducing more attributes in the SP experiment are suggested from the previous research that can be applied to eliminate the incentive to strategic bias. The effectiveness of these two methods in reducing/eliminating hypothetical bias warrants further testing as evidence is needed to find out how respondents cope with this information in their choice making process.

In this study, a Cheap-Talk script will be applied to the SP experiment to test:

- The existence of strategic behaviour, and whether or not adding a CT can amend the incentives to bias in the SP survey;
- The effectiveness of the CT among different population;
- The impact of the CT on the estimation of other attributes in the Choice Experiment, whether or not adding a CT will impose a different bias to SP responses;
- The impact of the CT on individuals' perceptions of the realism of SP survey by following up questions on perceptions of cost change.

This study also will test the impact of adding more attributes to the SP experiment. By adding more attributes, it may be hoped that respondents will exhibit less bias. This is partly be due to the extra effort required merely to complete the exercise with bias, but it is more likely to be due to respondents failing to see any single clear purpose to the experience. The following aspects will be tested:

- Whether or not adding more attributes will mask the research aim, therefore, amend the incentive to strategic bias in the SP experiment;

- Whether or not adding more attributes would add task complexity to the experiments?

- A following–up question on respondents' perceptions of complexity will be added to the experiment. The relationship of the 'objective' and 'perceptual' complexity will be explored.

The research hypotheses are proposed based on the literature review. The present research is focused on the methods to eliminate incentive to strategic bias and together with its impact on respondents' perceptions of task load and how they deal with additional information to the choice experiment. This study aims to identify the influence of different designs on the pattern of SP responses and to explore means of identifying and reducing strategic bias.

To test the research hypotheses, a series of SP experiments will be developed. The context of the experiment is selected to be users' valuation of the rolling stock. Chapter 3 presents a review of the previous studies in this context.

Chapter 3

Review of Users' Valuation of Rolling Stock

3.1 Introduction

The research hypotheses for the present study were outlined in chapter 1. The experiment context was selected as users' valuation of new rolling stock. The objective of this chapter is to present a review of previous studies of rolling stock users' valuation. Section 3.2 presents the general background for valuation of rolling stock studies. Section 3.3 reviews the previous cases, discussing the advantages and disadvantages of each method. Section 3.4 explores reasons for the variation of the valuations found. Section 3.5 ends the chapter with a summary and some implications for this study.

3.2 Background for Valuation of Rolling Stock

3.2.1 Background

Quality of service and passengers' priorities are important items, which many operators seek to quantify, on the grounds that concentration on the most important aspects may increase patronage and improve profitability. The Strategic Rail Authority (SRA) (2000, p.55) stated that "Provision of new, refurbished and improved trains is key to the Strategic Rail Authority's objective of securing a progressive improvement in the quality of services" and that "the investment in new trains since the start of franchising to over £2 billion", thus improving the passengers' satisfaction of the rail service.

The evaluation of investment in the improved rolling stock is an important issue. Prior to introducing new or refurbished rolling stock as part of their franchises, Train Operating Companies (TOCs) have often carried out studies (most of them were SP surveys) to investigate the public preference of the fleet and to test whether the improvement of the fleet and services would be enough to increase the fare and also the extent to which it would increase demand, thus to evaluate the cost and benefit of this investment.

The rolling stock value derived from SP results is used not only for welfare appraisal and pricing but also for rail demand forecasting. Using the fare elasticity, the value of rolling stock can be converted to a demand effect (ATOC, 2005).

3.2.2 Development history

A large number of studies have been conducted to investigate the valuation of rolling stock. According to Lawrence (1991), Research Projects (December 1968) produced one of the first rolling stock studies. They interviewed passengers, businessmen at work and people at home and asked them whether or not they would want to pay a supplement (dependent on the length of their journey) to get the new rolling stock or refreshment. 55% of non-daily users and 25% of commuters opted to pay the supplement. M.I.L. Research (1982) conducted similar interviews in London and the South East just before the introduction of new Class 455 trains. Passengers were asked whether they would prefer improved services at no extra cost (chosen by 61%) or the same service and 5% lower fares (chosen by 35%).

The results from early attitude studies identified travellers' preference regarding quality attributes. However, early studies were hampered by the limited description (only categorized by poor, fair, good, very good) of service quality (intangible attributes), and also the limit of methodologies. The elasticity produced "would almost be impossible to apply" (Lawrence, 1991, p.17).

SP methods were introduced to evaluate the new rolling stock by SDG (1983) and then widely applied by TOCs prior to the introduction of new stock. By providing respondents with hypothetical choices, they make trade-offs generally between cost, journey time and rolling stock types (overall or decomposed by service attributes such as layout, comfort or noise). Researchers then seek to find the valuation. Some studies provide a review of the valuation of rolling stock, such as MVA (1993) and Wardman and Whelan (2001). Reviews have found that "*SP provides the only reliable method of attaching valuations to quality improvement ... monetary values for time and frequency and other 'hard' variables derived from SP methods have been remarkably consistent among ...*"(MVA, 1993).

However, SP methods are criticised as causing bias due to their hypothetical nature, for example "*the valuation of soft variables, such as comfort and cleanliness have been rather less convincing, ...maybe over estimated*" (MVA, 1993).

Revealed Preference (RP) methods (Wardman and Whelan, 2001; Hague and Accent, 2002), demand impact analysis using real ticket sales data and market analysis (Accent, 2006) are observed to obtain the valuation of the improved rolling stock. A review of previous studies is presented in section 3.3.

3.2.3 The impact of improved rolling stock on rail passenger demand

Travel demand growth is not only influenced by factors such as income and population growth but also by the improvement in travel environment. Rolling stock quality tends to be a secondary factor compared to primary factors such as fares and journey time. Examples of other secondary variables are changes to crowding change, on-board facilities and cleanliness.

PDFH (2005) suggests the procedures to calculate the impact of rolling stock improvement on the passenger demand: firstly convert the improvement into an equivalent change in rail fare; and then the relevant fare elasticity is applied to calculate the expected demand increase.

Equation 3.1 (PDFH, 2005) shows the impact of fare on the passenger demand (elasticity) change:

$$I_F = \left(\frac{F_{new}}{F_{base}}\right)^{f_t} \qquad \textbf{Equation 3.1}$$

Here, I_F is the index for the change in volume due to changes in fare and fare related factors. F_{new} is the new average fare and F_{base} is the base average fare, and the ratio of the two is the uniform fare change. f_t is the overall fare elasticity which is the ratio of the incremental percentage change fare with respect to an incremental percentage change in another variable. Equation 3.2 shows how rolling stock benefits (RS) and other secondary variables (V_S) are allowed to amend the new fare:

$$I_F = \left[\frac{F_{new} - (RS + V_S) \times F_{base}}{F_{base}}\right]^{f_t} \qquad \textbf{Equation 3.2}$$

Here, *RS* is expressed as a proportion of the base fare. If the other secondary variables remain the same ($V_S = 0$), the impact of introduction of new rolling stock on the demand change can be estimated by the following equation.

$$I_F = \left(\frac{F_{new} - (RS \times F_{base})}{F_{base}}\right)^{f_t} \qquad \textbf{Equation 3.3}$$

Also, if $F_{new} = F_{base}$, then the impact on demand is $(1 - RS)^{f_t}$.

In this study, the impact of improved rolling stock on the demand forecast is analysed by using the PDFH (2005) manual. The demand impact of the new/improved rolling stock is related to certain type of rolling stock which can be reflected by the monetary valuation of this stock. The monetary valuation of rolling stock is converted to a demand effect by using the fare elasticity. The monetary valuation of the improved rolling stock is generated by the marginal utilities of the variables in the SP model. Therefore no scaling problems arise.

For example, using PDFH (2005) values, if other attributes remain same as before, the monetary value of rolling stock (RS) is 10% of the fare and fare elasticity is -0.6 for commuters (Non London area and journey distance < 20 miles). The change of RS will cause the demand change by $(1-0.1)^{-0.6}=1.06$. If the value of rolling stock improvement is 10% of the fare paid, it will cause the demand increase by 6%.

3.3 Previous Studies on the Rolling Stock

3.3.1 Stated Preference (SP) and Revealed Preference (RP) method analysis

A large number of studies (Kottenhoff and Lindh, 1996; Hensher 1998; Axhausen, 2002) examined the influence of improved transport infrastructure and supply on the route and mode choice using SP method. There are also large evidence (published/ unpublished) existed in UK (MVA, 1985; 1992; Oscar Faber TPA, 1994; Wardman and Whelan, 2001; Hague Consulting Group and Accent Marketing and Research 2002), investigating the impact of new rolling stock on passengers' demand and quantify users' valuation of improved rolling stocks.

The studies included in this review is selective and focus on primarily studies which are in the context of users' overall valuation of rolling stock and the impact on passengers' demand. Table 3.1 shows the previous accessible relevant studies, with the SP experiment context and valuation achieved from each study.

Using SP techniques, MVA (1985) explored the impact of replacing the train of slam door (old) with the sliding door type (new). They suggested the value of improving stock was significant, and equivalent to 8% of average fares. Some similar studies, in terms of SP survey context, were conducted as listed in Table 3.1 (Study 1 to 5), such as the studies by MVA (1990, 1992), Babtie (1993), Oscar Faber and TPA (1994). The difference among these studies is mostly in the rolling stock types and routes where the survey was conducted.

They applied conventional SP surveys in the study, which normally provided respondents the information such as rolling stock type, journey time, fare and some other service attributes, such as reliability (MVA, 1985) and headway (Babtie, 1993).

The new stock values varied by different types of trains provided in the survey. Approximately 10% (or more) of the average fare was obtained from the previous studies and the value varied by journey purpose and income. Business travellers were found to have a higher stock valuation, compared with commuters and leisure travellers. For example, MVA (1992) reported that Standard Business travellers would be willing to pay 13% of the average fare to improve the rolling stock, and the Leisure travellers only would be willing to pay 5% of the average fare. Higher income would be willing to pay more to improve rolling stock.

The value of rolling stock is normally presented by two ways in the model estimation and evaluation. The first method is to report the value of rolling stock by a specific constant term in the utility function, such as using by MVA (1985). The second method is to report the stock value by a proportion of the time unit value, so that the stock value is relative to the in-vehicle journey time (journey length), such as using by MVA (1992), Babtie (1993) and Wardman and Whelan (2001). The discussion of these two specifications is presented in section 4.4.1.

Table 3.1 Previous relevant SP studies of the valuation of new rolling stock

	Studies	Choice Experiments	Subject of Studies (Improvements)	Valuation
1	MVA (1985)	Journey time, fare, reliability of service, stock	New sliding door with old slam door stock	8% of fare; value of reliability to be very large
2	MVA (1990)	Journey time, fare and stock (old/new)	Mark III with Mark IV	12% (of fare) for 1st Class passengers; 13% for Std. Business; 5% for Std. Leisure
3	MVA (1992)	Journey time, fare and stock type/(service attributes)	Intercity evaluation Mark III with Mark IV	9% (of fare) for Std. Class; and 13.5% for First Class (FC) (varied by journey distance)
4	Babtie (1993)	Headway, journey time, cost and rolling stock	New electric trains with air conditioning and improved interiors and seating, compared with the existing Sprinters	1.22p per passenger minute, around 9% of the fare
5	Oscar Faber TPA (1994)	Journey time, fare and train type (basic (BR)/superior refurbishment(SR))	Northampton line stock refurbishment	1.0 p/min for SR (3% of fare) and 0.4p/min (7.5% of fare) for BS
6	Wardman and Whelan (1998, 2001)	RP Conventional SP (Time and cost) SP with service attributes	Express Sprinter vs. Sprinter	0.9% of fare
			Network vs Sprinter	0.7% of fare
			Express Sprinter vs. South East (SE) Slam Door	1.5% of fare
			SE Sling Door vs. SE Slam Door	0.6% of fare
			Network vs SE Slam Door	1.0% of fare
			Wessex Electric vs SE Slam Door	1.2% of fare
			Mark 2 vs SE Slam Door	1.4% of fare
			Mark 3 vs SE Sliding Door	1.5% of fare
			Mark 2 vs SE Sliding Door	0.7% of fare
			Mark 3 vs Networker	0.6% of fare
			Mark 2 vs Networker	0.6% of fare
			Mark 3 vs Mark 2	0.1% of fare
7	Hague Consulting and Accent Marketing and Research (2002)	Journey time, fare and stock type (characterised by stock attributes)	Class 170 Turbostar (Anglia)	16.0 Business(B), 13.2 Commuters (C), 18.8 Leisure (L)
			Class 170 Turbostar (ScotRail)	41.1 (B), 24.7(C), 31.6(L)
			Mark 2	18.0 (B), 7.2 (C), 3.6 (L)
			Class 158 Refurbished	20.2 (B), 13.2(C), 18.9(L)
			Class 321 sliding door	4.8 (B), 6.3(C), 9.3 (L)

Table 3.1 also lists some new evidence on exploring the valuation of new rolling stock, such as the study by Wardman and Whelan (2001). Besides the contribution of a very detailed review of previous rolling stock studies, they successfully applied RP methods into the research.

Wardman and Whelan (1998, 2001) provided a detailed review of the previous studies on the valuation of new rolling stock based on a set of previously unpublished SP and RP studies. After controlling the possible confounding effects due to individuals' heterogeneity, they explored the factors which contributed to the variation of the stock valuation.

The review covered 18 SP studies. They found that values of new rolling stock from SP studies were incredibly large in the situation that respondents can perceive the aim of studies (to be rolling stock studies). These high valuations were not supported by eight studies which were based on the analysis using ticket sales data. Among the eight studies, four found no significant change in demand after the introduction of new rolling stock, and the other four found that the demand increase was between 3% and 8% but with broad confidence intervals. The reason for this discrepancy was suspected to be strategic bias. In a related working paper, Wardman and Whelan (1998) commented that "*the main points of the previous research has tended to obtain what we believe to be too high monetary valuations of rolling stock improvement, yet the effects of such improvements expected on the basis of valuation studies have rarely been detected in the analysis of rail demand.*"

A regression model was developed to explain the variation of 45 rolling stock values (expressed as a proportion of the average fare paid). They found that:

- If the purpose of studies (stock valuation) could be perceived by respondents (denoting by a dummy variable in the regression model), the valuation obtained was three time higher. The impact was significant;

- Familiarity to the experiment context had a significant impact on values of improved rolling stock. Unfamiliarity among rail users would lead to inflated values, and the impact was significant at 1% level;

Based on the review of previous studies, they conducted a research to investigate values of a large number of different types of existing rolling stock. A novel feature of the study was the successful development of RP models based on actual choices between different rolling stock types to complement corresponding SP models. In the RP study, respondents who really had a chance to choose between train types were asked about the choices they really made. Cross-sectional RP approach was applied which involved the comparison of different stock types at a given point in time and undertaken at the level of individuals rather than rail ticket sales. By doing this, it could avoid the serious correlation problems and could get values of different stock type from individuals. Two SP studies were conducted: one conventional SP investigated the overall valuation of certain rolling stock (labelled) which was based on the same choice context as the RP approach; and another SP survey investigated service attributes.

Using RP data, disaggregate choice models were successfully developed to estimate the value of rolling stock. They found that the values from the RP study were slightly higher than that from SP, which can be partly explained by the non-response bias observed in the RP survey. They believed that the values obtained from RP were more reliable as the results from SP suffered from the problem of imprecisely estimate due to the relatively small sample size. Table 3.1 gives the values obtained from the joint RP-SP model. The rolling stock values vary with income and journey purpose, with commuters having the highest values, followed by leisure travellers and business travellers, and the values increase with income.

Compared with previous SP studies in the same context, values from this study are on the lower side but more in line with the demand impact studies. This is the first study where RP models were successfully developed to explain travellers' actual choices of rolling stock.

Similarly, Hague and Accent (2002) found new rolling stock values from the conventional SP results were found to be very high. Therefore, an RP study was conducted on London-Ipswich route as a supplement to the SP study. The RP survey collected information about respondents observed train, ticket choice and their journey purpose for their journey. Respondents were also asked to rate each of the alternative rolling stock types as part of the questionnaire.

However, the RP results were found to be disappointing that the values of rolling stock were implausible even when the coefficients were right sign and significant. The poor RP results can be explained by: firstly the small sample size, and secondly missing variables problem, for example, reliability and crowding which respondents stated were more important than train type and which may well correlate with train type; and thirdly, it was suspected that in the RP exercise, travellers were not aware that they faced trade-offs between different types of trains and other travel attributes.

3.3.2 Demand impact analysis

Rail ticket sales data has been regarded to provide a reasonably accurate account of rail demand. It is assumed that the econometric models based on the real data can provide plausible elasticity estimate. Therefore, a large number of studies have attempted to estimate the demand impact of the introduction of new rolling stock by an econometric analysis of rail tickets data, relating to the effects of fare and service quality, inter-modal competition and socio-economic factors (Glaister, 1983; Fowkes et al., 1985; Wardman, 1997; Koppelman and Sethi, 2005).

Earlier studies are seriously hampered by relatively small sample sizes, by strong correlations between changes in rolling stock and in other aspects of service quality such as frequency and travel time and by unreliable ticket sale data. PDFH (2005) listed the explanation for the failure of econometric analysis of ticket sales data to detect a rolling stock effect:

- Firstly, the relatively small magnitude of the rolling stock effect to be measured, in some cases, and the coincidence of the other effects.
- Secondly, fares are often increased in real terms when services are improved, and this may offset any rolling stock effect on volume, converting it into revenue effect;
- Thirdly, some services have been unreliable when first introduced, and any longer term effect, once the service has settled down, is then even more difficult to detect;
- And finally, similarly, if the introduction of new service leads to higher levels of crowding, this may offset the rolling stock benefits, and so suppress demand.

There is only one study on the introduction of Inter City 225 (TCI – OR, 1993) which successfully separated the effect of journey time and the effect of improved rolling stock using analysis of ticket sales data (cited from PDFH, 2005). The study concluded that the new rolling stock had generated an extra 4.4% (±2.2%) of revenue. Although the confidence intervals of the value are large, this is the only study which produced the independent estimates of the impact of timetable and the impact of new rolling stock.

3.3.3 Market and event studies

ACCENT (2006) conducted a research on the valuation of new rolling stock in various routes, including a market research and event study. The market research contained a self-completion questionnaire which asked respondents to state why they were making the journey and to say whether the new rolling stock has had an impact on their use of the train. The market research suggested a change in rolling stock may lead to a lift in total demand, between 0.7% and 2.0%.

In order to estimate the impact of new or refurbished rolling stock on demand change, they introduced a popular method from the finance literature, namely 'event studies'. It has been used to assess the impact on share prices of events, such as the announcement of takeovers, share buy–backs and changes to dividend policy.

By modelling the demand change before and after the introduction of new rolling stocks, event study interpreted the volume change due to the introduction of new rolling stock. The event study analysis suggested that a change in rolling stock may lead to a small uplift in total demand, between 0.7% and 11%. Dynamic demand models were applied to allow the presence of lagged responses as some consumers may not respond immediately to a change in an explanatory variable.

They suggested results from the event analysis should be combined with other market research findings and previous results. Firstly, the robustness of results from event analysis was undermined by data availability and model specification, for instance, some of the 'before' data (before the change of rolling stock) was not available for some routes in the analysis. Secondly, heteroskedasticity was found in some demand models which meant the variance of the error from the demand model was not constant and might be conditional on explanatory variables. Here, the error represented the difference between demand model results and the actual demand which was not explained by the model. The results might be biased upwards or downwards.

3.3.4 Summary

The measurement of new rolling stock valuation has relied mainly on SP techniques rather than econometric analysis of ticket sales data and some other methods. However, the values from SP studies were found to be higher than those from other evidence, especially when the issue of the SP survey was introduction of improved rolling stock. The possible reason for the inflated values is the existence of incentive to strategic bias in SP surveys.

RP method is a disaggregate analysis of individuals' actual choices, which is suggested as a promising method and has not been fully explored. However, it is not always the situation that respondents faced with the actual choices of different types of train; especially sometimes the value of new rolling stock is the interest. In addition, this method is suffered from non-response bias observed from previous evidence.

3.4 Factors Influencing the Valuation Variation

3.4.1 Individuals' socio-economic features

This study will examine the existence of strategic bias in the context of users' valuation of rolling stock. Based on the review in section 3.3, the stock values from SP studies are found to be much higher than those from other evidence. This section will discuss several factors influencing the variation of stock valuation.

Respondents' different characteristics partly explain the variation of stock values. Review of previous cases found that income and journey purpose affect individuals' choice making.

It has been found income has a strong effect on respondents' choice making and valuation of attributes (Fowkes, 1986; Wardman, 2001; Gunn, 2001; Wardman, 2004). The value of time studies found that with income increase, people are less sensitive to the cost in the choices.

Monetary value is obtained by the ratio of parameter estimates of a target attribute and cost. Hence, higher income group would have a higher value of new rolling stock.

Previous studies found respondents' journey purpose contributes to the variation of values of improved rolling stock. Wardman and Whelan (2001, p.423) conducted a meta-analysis and found that "*the absolute money values for standard class business were on average 186% larger than for standard class leisure and commuters had absolute values on average 33% larger than leisure travellers did ... and first class business values were on average 70% higher than for standard class business*", shown as follows:

$$VoS_{Business} > VoS_{Commuter} > VoS_{Leisure}$$

In summary, individuals' socio-economic features should be considered in the interpretation of valuation variations.

3.4.2 Familiarity effect

Travellers' familiarity of the rolling stock has shown a significant impact on the valuation variation. Kottenhoff and Lindh (1996, p.240) suggested that "*SP method is sensitive to the presentation; especially of a new product that the customer has not experienced (a difference of 49% is observed)*". The objective of their study was to explore the value and effects of introducing high standard train in Blekinge, Sweden. Two SP surveys were conducted at the time before (1991) and after (1992) the introduction of high speed rail. They found that the SP data from the 1992 revealed a much higher value for the new train (+ 124%). Besides an asymmetrical effect (Dargay, 1993) contributed to 75% of the difference, they detected a difference of 49% units which were due to the presentation of SP alternatives. Before the introduction of rolling stock, only text and colour pictorial information were provided in the SP survey, whilst in the survey after the introduction of rolling stock, respondents had the experience of new rolling stock. Therefore, they suggested respondents with the experience of new rolling stock would give a higher value to the new stock.

Wardman and Whelan (2001) found familiarity of the stock type contributed to the variation of the stock valuation negatively. From the regression model (section3.3.1), they found travellers who were familiar with the stock type in the experiment gave lower value to the stock. Familiarity reduced the weight by 44% (t=3.44). In their model, familiarity was measured by the fact that respondents would have been familiar with both rolling stock types. They concluded that "*unfamiliarity leads to artificially high values*" (p.424).

3.4.3 Experiment design effect

Many studies found that SP experiment context contributes to the variation of values. MVA (1993) did not detect difference in rolling stock values between the mode choice and abstract choice SP experiments. However, some researchers noted that the abstract choice contexts may lead to higher values than other choice contexts (Jones, 1997; Wardman and Whelan, 2001). Jones (1997) has stated that "*Past experience has shown that attribute values are often lower in between-mode than within mode exercises*". Wardman (1998) found that value of time values obtained from mode choice SP experiments were 16% lower (on average) than abstract choice experiment. Wardman and Whelan (2001) provided an interpretation of the inflated values observed from within mode SP experiments to be the existence of strategic bias. They found that most of the rolling stock studies used an abstract choice context, normally, a within-mode comparison of different train alternatives with only trade-offs among cost, journey time and stock types. This simplicity made the aim of SP studies very transparent, thus being easily perceived by respondents. Respondents might have incentives to strategically bias their answers for better outcome (introduction of new train) in the SP survey where there is no financial consequence of their statements.

Another design impact refers to the situation that some factors associated with improved rolling stock are better/worse than the current option. For example, if a new stock is often associated with reduced reliability and more crowding, whereupon the confounding effect would reduce the estimated values and the stock values would be relatively low.

Although SP techniques have the advantage that they can control for extraneous influences, compared with RP studies, there remains the concern that valuation of improved rolling stock could be influenced by some factors associated with improved rolling stock, such as reliability and crowding. MVA (1985) conducted a SP study and found that stock values were relatively low, at 5% and 8% of the average fare. This lower value can be partly explained by a worse reliability associated with the new rolling stock which individuals cared in their choice making. And service attributes in choice context made the aim of experiment less transparent; therefore, there was less chance that respondents could strategically bias their answer.

3.4.4 Existence of strategic bias

Wardman and Whelan (2001) addressed that the inflated stock values from the SP studies can be explained by the existence of strategic bias. They provided the following evidence:

Firstly, their meta-analysis (section 3.3.1) found if respondents could perceive the SP study aim, the value of new rolling stock was three times higher than those from who cannot perceive the aim. This was derived from a regression model based on 45 values of rolling stock. A dummy variable denoting whether the purpose of study can be easily perceived was incorporated into the model. The effect was significant where a 202% (t = 5.53) effect was obtained.

The second evidence is obtained by comparing values from SP studies with that from other evidence for the same type of rolling stock. They found there was no conflict between these two methods, although SP studies often yield large values of new/improved rolling stock. Table 3.2 shows values of new rolling stock obtained from several SP studies. The studies on effect of Mark IV relative to Mark III stock were compared.

Table 3.2 Evidence of effects of new rolling stock on rail demand (Mark III–Mark IV)

	Context	Research Type	Monetary Value (% of fare)	Demand effect
1	London inter-urban	SP study (stock valuation)	1st Class: 11% Std Class: 14%	11% 14%
2	London inter-urban	SP study (stock valuation)	1st Class: 12% Std Business: 13% Std Leisure: 5%	11% 9% 6%
3	London inter-urban	SP study (stock valuation)	1st Class: 13.5% Std Business: 12% Std Leisure: 9%	13.5% 8% 12.5%
4	London and non-London inter-urban	SP study (pricing)	1.5%*	1.5%
5	London inter-urban	Ticket sales data		To London 4% (±2%) From London 3% (±2%)

Cited from Wardman and Whelan (2001) Table 1 and 4.
*Rolling stock coefficients were not significant at the usual 5% level

In Table 3.2, it would be reasonable to take 10% of the fare as representative of the estimated value of this stock change from SP studies (except for study 4). For comparison, we calculated the demand impact using PDFH method (section 3.2.3). The fare elasticities were given as -0.6, -1.0 and -1.25 for commuters, business and leisure travelers respectively, as suggested by PDFH (2005). The introduction of new stock would, therefore, be equivalent to averagely 10% of demand increase. Only one exception was found for a lower SP value obtained (Study 4) where the aim of research was perceived by individuals to be pricing. On the other side, study 5, derived from ticket sales data, found that the demand impact by the replacement of rolling stock was 3-4%. PHFH (2005, Table B5.2) suggested the fare value for "*Replacement of Mark III (HST) trains with Mark IV (225) trains is 3.3%*".

The inconsistency between the higher SP values and values from ticket sales data was suspected to be the existence of strategic bias. The incentive to strategic bias is that respondents can easily perceive the aim of SP study to be rolling stock valuation. As respondents do not have the financial consequence in the SP survey, they would bias their answers for increasing the possibility of the introduction of new rolling stock.

Recall the review of strategic bias in section 2.5, it illustrates the existence of strategic bias in SP studies in environmental science, food science, marketing and transport research. The existence of strategic bias is found to be subjected to the experiment subject and different types of good. However, it is suspected that the strategic bias is mainly caused by the hypothetical nature of SP studies, where individuals have the incentive to strategically bias their answers for better outcome. The reason is mainly that the payment process is out of the SP experiment and individuals perceive that their response would have an impact on the provision of the good.

3.4.5 Summary

This section presents a discussion of the factors which contribute to the variation of new / improved rolling stock valuation. Individuals' socio-economic features, respondents' familiarity with the rolling stock in the survey (alternatives), SP experiment design and existence of strategic bias are observed from previous studies to have impacts on the value of rolling stock.

3.5 Conclusions and Implications for this Research

The main objective of this study is to examine the existence and possible consequence of the strategic bias in the context of users' valuation of rolling stock. This chapter provides a review of previous studies in this context. The review suggested that SP techniques were commonly used to investigate the quality improvements and valuation of rolling stock. Other methods such as RP methods, demand analysis and methods from marking research were observed to be used in the stock valuation studies. However, limitations of these methods were found from previous studies, such as: limit of the data availability, serious correlation among different attributes and larger intervals in the valuation estimation.

In the context of the valuation of rolling stock, the review found that monetary values of time and frequency from SP methods have been consistent among studies. However, the valuation of rolling stock was rather less convincing. The values from SP studies were much higher than those from other evidence when the issue was to introduce a new rolling stock or refurbishment, after controlling the impact of individuals' socio-economic features. The inflated values are explained as the existence of strategic bias by which respondents overestimate the value of rolling stock to increase the chance of the introduction of new rolling stock.

Based on the review in chapters 2 and 3, we decided to examine the existence and consequence of incentives to strategic bias in the context of users' overall valuation of improved rolling stock and methods to amend the bias. A series of within mode SP experiments will be developed to test the research hypotheses outlined in chapter 1.

The existence of strategic bias will be examined by comparing the values obtained from our study with some empirical evidence, for instance the PDFH (2005) recommended values (which assumed to be the true value). According to the definition of bias which is an estimator is on average over or underestimates what is being estimated, we will detect if the bias exists in the valuation. The review in chapter 3 suggested some factors that might contribute to the variation of rolling stock valuation, such as individuals' socio-economic features. We will control the impacts of those factors to avoid the possible confounding effects.

In our study, we will examine two methods to amend the incentives to strategic bias in SP experiment (see the discussion in section 2.8): cheap talk and adding more attributes to mask the research aim. Cheap-talk scripts (section 2.6) provide a means of overcoming the incentive to strategic bias. The simplicity and economic feature of this method makes it an attractive approach to reduce the strategic bias in the CV and SP studies. We will introduce a CT script in some of the SP experiments. By comparing the responses with and without CT script, we will detect the impact of CT on valuation of improved rolling stock and investigate if it is effective in reducing the bias (if exists). The development of the CT script will be presented in chapter5.

The literature review in chapter 2 suggested that masking the research aim by introducing more attributes as a method to eliminate the strategic bias. The present study explores the impact of adding more attributes on respondents' preferences of rolling stocks. In addition, we will examine if adding more attributes will add more task load to respondents, thus leading to the task complexity effects (see discussion in section 2.8).

The literature review in these two chapters suggested that respondents' perceptions of SP survey have impacts on their choice making, such as the perception of the cost change (payment) (see section 2.5), difficulty of choice making (section 2.7) and familiarity of stock types (section 3.4.2). We will introduce a series of follow-up questions to investigate impacts of respondents' perceptions on their choice making.

The development of SP experiment will be presented in detail in chapter 5. The development of Cheap Talk script and complex design will be discussed in chapter 5.

Chapter 4

Methodology of the Research

4.1 Introduction

Chapter 1 presented the objectives and hypotheses, based on the literature review discussed in chapters 2 and 3. To achieve the objectives and test the research hypotheses, the methodology is presented in this chapter as follows:

- Constructing the framework of the study on the basis of the literature review (chapters 2 and 3)
- Specifying the method and analysis (sections 4.2 - 4.4)
- Designing the survey form and SP experiment (chapters 5 and 6). This is an iterative process, in which the questionnaire and SP experiment were tested and developed through a set of pilot surveys until they were satisfactory for use in the main survey.
- Conducting data collection (chapter 6)
- Analysing the data and test the research hypotheses (chapters 7-8)

The objective of this chapter is to introduce the stated preference (SP) techniques and data analysis method. Section 4.2 presents the SP method: the concept, design process and then the simulation testing. Section 4.3 discusses the analytical issues involved in SP data analysis, the random utility theory and logit models. Finally, section 4.4 presents the development of the utility function containing the valuation of rolling stock.

4.2 Stated Preference Method Design

Stated Preference (SP) methods are a well known and widely used preference elicitation technique in transport studies. They are based on the observed responses of individuals who face a set of hypothetical scenarios, set up by researchers. Each scenario represents a package of different attributes. SP methods have been used to evaluate the effects of relevant attributes of a system on individuals' responses and provide forecasts of changes in demand and travel behaviour. Detail guidelines for SP experiment design can be found in Green and Srinivasan (1978, 1990), Bradley (1988), Fowkes and Wardman (1988), Pearmain and Kroes (1990), Hensher (1994), Fowkes (1998), Louviere et al. (2000) and Hensher et al. (2005). The design process of an SP experiment can be summarised in four steps, which will be introduced in turn.

4.2.1 Characterization of decision problem

The characteristics of the hypothetical scenarios are represented by attributes that influence preferences. Through focus groups/pilot surveys, literature reviews of previous studies and interviews with experts, a series of attributes can be selected to characterize the decision problem. To select the attributes, the researcher needs to, firstly, understand what the respondents need to have to make a decision; secondly, define the dimensions of the product to be evaluated; and then search for information on alternatives and attributes and finally make a choice set. The attributes selected need to be understood by individuals and be familiar to the respondents in their real life. For example, in mode choice studies SP exercises usually include in-vehicle time, out of vehicle time and the cost and quality of transport modes as attributes for each mode. The attributes in this study include the journey characteristics and service quality attributes of the train journey (see chapter 5).

In addition, the sources of individual heterogeneity (e.g. income, education, attitude towards target issues) need to be identified as they could lead to important behavioural differences.

4.2.2 Specification of the number and magnitude of attribute levels

As the number of the specified attributes and their levels increase, the number of the combined scenarios (such as from a fractional factorial design) also increase. Additionally, respondents have limited cognitive abilities (section 2.7), which must be taken into account when designing the choice experiments. Pearmain and Kroes (1990) suggested that in an exercise attributes should be limited at six or seven per alternative, and less if it includes unfamiliar variables. DeShazo and Fermo (2002) suggested having less than six attributes per alternative; otherwise, respondents may ignore the attributes or use certain decision heuristics. We will examine the impact of the number of attributes on SP responses.

The levels for attributes should be chosen to represent the relevant range of variation in the present or future market of interest. The variation of attributes values across scenarios need to be large enough for respondents to trade-off (Fowkes and Wardman, 1988), otherwise they may be ignored. The level of attributes can be tested by the simulation of responses (section 4.2.5), which allows the designer to improve the values of attributes before collecting data and pilot survey, which helps to understand if individuals could understand the survey and how they cope with the survey (e.g. format, questioning, presentation, survey conducting and response rate).

4.2.3 Experimental design development

Once attributes and associated levels have been determined, analysts typically use some form of orthogonal design to generate different combinations of attribute levels, named "scenarios" (or profiles) (Louviere, 1988). A "scenario" is a single attribute level combination in a complete factorial combination of attribute levels. A "design" is a sample of profiles which has a particular set of statistical properties that determines the utility specification(s) that can be estimated. Louviere et al. (2000) defined experimental design as a way of manipulating attributes and their levels to permit rigorous testing of certain hypotheses of interest. One of the crucial objectives of the experimental design is to create the choice set in such a way that the number of choice sets is minimized while being able to infer utilities for all possible scenarios – which implies keeping the choice task simple to respondents and at the same time being able to extract all the necessary information from choices (i.e. securing a high degree of design efficiency).

Factorial design has very attractive statistical properties from the standpoint of estimating the parameters of models to test hypotheses (Louviere et al., 2000). A factorial design is simply the factorial enumeration of all possible combinations of attributes. For example, if there are 3 attributes with 2 levels for each, then the factorial will be $2^3 = 8$, implying 8 possible combinations of attribute levels.

A complete factorial (full factorial) design is a design in which each level of each attribute is combined with every level of all other attributes. In other words, it contains all possible combinations of attribute levels. In this situation, the number of possible choice sets increase exponentially when the number of attributes and levels increases.

Fractional factorial design involves a selection or a subset (a fraction) of the original full factorial design, in which the properties of the full factorial design are maintained in the best way possible. Rather than random selection, statisticians have developed sampling methods that lead to practical and manageable designs with specific statistical properties. The advantage of fractional factorial design is that the number of scenarios can be dramatically reduced from the full factorial design, while it still ensures that the main effects of attributes are independent from the significant interaction effects, so that the main effects can be estimated efficiently. The loss of information can sometimes be significant, as fractional factorial designs limit the ability to take interaction effects into account. Louviere et al. (2000) argued that the exclusion of interaction effects does not necessarily lead to biased results.

Orthogonality is satisfied when the difference in levels of each attribute varies independently over choice sets, meaning that the levels of the attributes are independent of each other. However, pure orthogonality are only available for a very small number of very specific problems, the primary purpose is to optimize the design as best on can, by minimising multicollinearity (Kuhfeld et al., 1994).

Some research on non-orthogonality has been conducted. Among them, Fowkes (1993) stated that to achieve a precision with which target valuations are estimated, by allowing some 'tolerance' of the correlation levels among attributes (such as cost and time), it can help to get a smaller variance of the estimation of value of time. In addition, a D-optimal design technique (Hensher et al., 2005) minimised the asymptotic standard errors for each of the estimated parameters. The D-error of a design is generated by taking the determinant of the asymptotic variance-covariance matrix and applying a scale factor which accounts for the number of parameters to be estimated (including the constants). This design method has only been used widely subsequent to this study; and it requires specialized software which was unavailable to student at that time.

A fractional factorial design has been applied in the SP survey design, which will be presented in chapter 5. Additionally, prior to the data collection, simulation tests using boundary rays maps (Fowkes, 1985, 1998) and synthetic data sets (Fowkes and Wardman, 1991) were applied to test the SP design (section 4.2.5). These methods are well tested and understood, and can be relied upon to give satisfactory designs.

4.2.4 Questionnaire development

After the attributes and levels for each attribute are determined, and the set of hypothetical scenarios are developed, a questionnaire is developed to gather respondents' preferences for alternatives in the hypothetical scenarios. The questionnaire also asks for respondents' socio-demographic, attitudinal and perception information. The additional data is useful in the analysis of the SP data and in the explanation of the observed behaviour.

The hypothetical scenarios can be presented in the form of ranking, rating or choices. The review in section 2.3.2 has found that choice-based tasks are the most realistic and simplest for the individual to understand. They are also the simplest in terms of data analysis and for prediction. As a result, they are the most often used technique in SP studies. Choice-based tasks usually require respondents to choose between two options. The number of choice scenarios needs to be defined carefully to avoid the learning and fatigue effects (section 2.7). Previous studies found that the number of scenarios should be less than 16 for each individual (Swait and Adamowicz, 1997; Arentze et al., 2003).

4.2.5 Simulation test

Prior to the data collection, SP experimental designs may be tested using a simulation (Fowkes and Wardman, 1991). This section reports the procedures used in two types of SP design tests: boundary ray maps analysis and simulation tests using synthetic data.

Boundary Ray maps (Bin analysis)

Boundary ray map analysis is used as simulation test tool to test if SP design is robust enough to cover the target attribute valuations accurately (Fowkes, 1991). Boundary values show the relative valuations of attributes at which alternatives are equally valued.

Equation 4.1 and Equation 4.2 explain how to generate the boundary ray from the SP design. For example, there are three attributes included to explain respondents' preferences of the rolling stock: time, cost and headway in the utility function (will be introduced in section 4.3.2). There are two alternatives (i) in the experiment, namely trains S and P. We derived the relationship of Value of Time (VoT) and Value of Headway (VoH) by applying the following equation (see section 4.4.4). β_{ik} is the coefficient of the corresponding attribute 'k' in the utility function.

$$
\begin{aligned}
&\mathrm{U_i} = \beta_{i1}\mathrm{Time_i} + \beta_{i2}\mathrm{Cost_i} + \beta_{i3}\mathrm{Headway_i} \\
&\frac{\mathrm{U_i}}{\beta_{i2}} = \frac{\beta_{i1}}{\beta_{i2}}\mathrm{Time_i} + \mathrm{Cost_i} + \frac{\beta_{i3}}{\beta_{i2}}\mathrm{Headway_i} \\
&\frac{\mathrm{U_i}}{\beta_{i2}} = \mathrm{VoT} \times \mathrm{Time_i} + \mathrm{Cost_i} + \mathrm{VoH} \times \mathrm{Headway_i}
\end{aligned}
$$

Equation 4.1

Then the indifferent point is:

$$
\begin{aligned}
&\mathrm{VoT} \times \mathrm{Time_S} + \mathrm{Cost_S} + \mathrm{VoH} \times \mathrm{Head._S} = \mathrm{VoT} \times \mathrm{Time_P} + \mathrm{Cost_P} + \mathrm{VoH} \times \mathrm{Head._P} \\
&BVOT = \frac{Cost_S - Cost_P}{Time_P - Time_S} + VoH\frac{Headway_S - Headway_P}{Time_P - Time_S}
\end{aligned}
$$

Equation 4.2

For example if train S takes 20 minutes and costs 50 pence, and train P takes 30 minutes and costs 40 pence, and all the other (headway) equal, then the Boundary Value of Time (BVOT) is:

$$BVOT = \frac{Cost_S - Cost_P}{Time_P - Time_S} = \frac{50-40}{30-20} = 1p/\min$$

This indicates that individuals with value of time larger than 1 p/min will choose the train S and the remainder will choose train P.

Figure 4.1 shows an example of a boundary ray map. Each ray represents the boundary value of each hypothetical choice. Boundary rays help us to obtain useful information to be modelled, and avoid asking similar questions and to allow for variations in valuations.

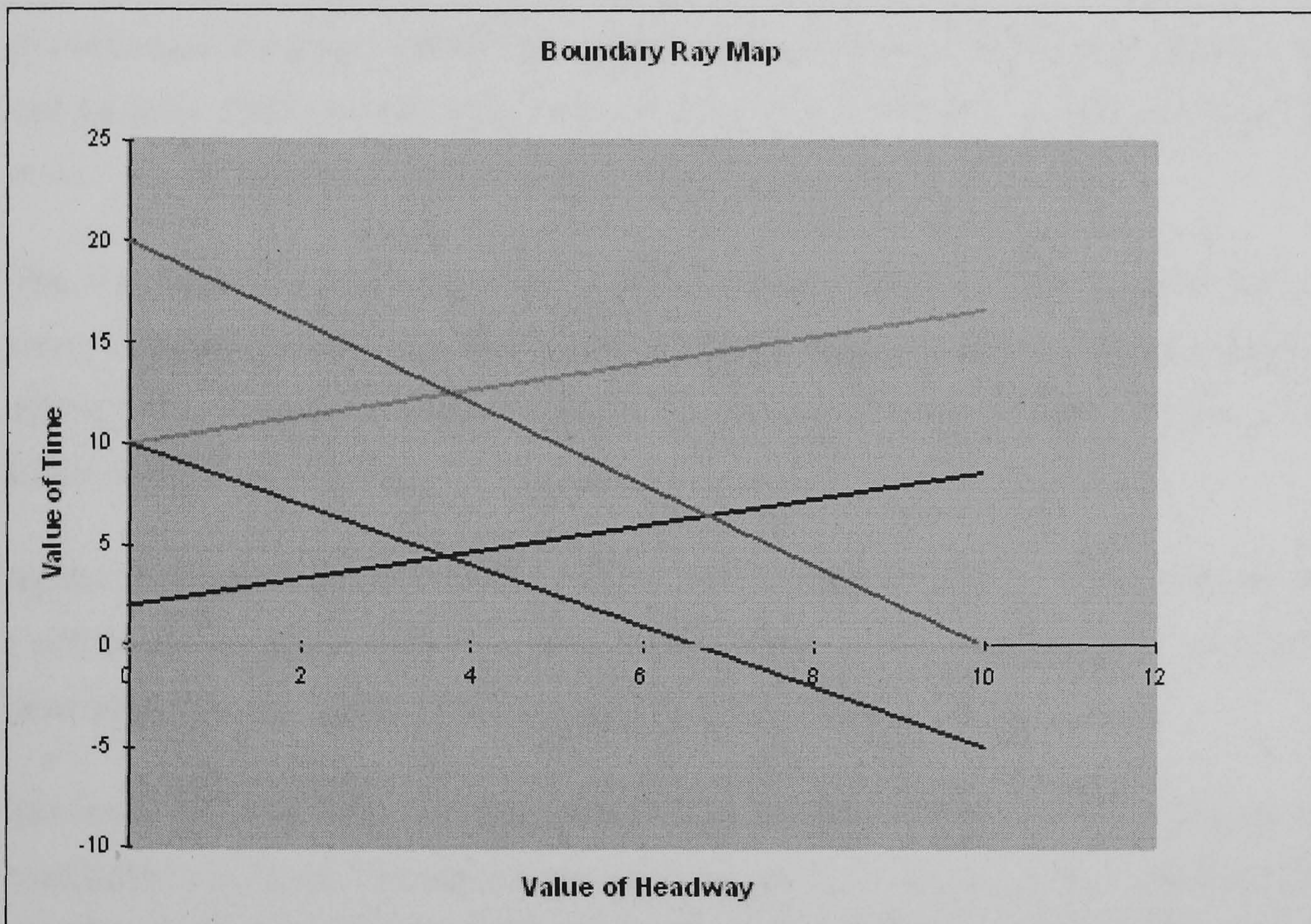


Figure 4.1 An example of boundary ray map

For each ray, the intercept and slope are decided by:

$$Intercept = \frac{Cost_S - Cost_P}{Time_P - Time_S} \qquad Slope = \frac{Headway_S - Headway_P}{Time_P - Time_S}$$

By adjusting the boundary rays generated by different SP scenarios, the interaction of boundary rays would be beneficial (Holden, 1992). A good spread of boundary values is expected to cover the range where the true value is expected to lie. Fowkes and Wardman (1988) stated that "the process of choosing the precise trade-offs to be presented in the SP experiment is both important and non-trivial, particularly given the assumed presence of inter-personal taste variation. The prime objective is to offer choices that will permit model parameters to be determined accurately."

Fowkes (1985) suggested "in order to obtain an accurate estimate of the respondent's relative valuation, we must present sufficient boundary values to make the inter-boundary value distance acceptably small. It will usually be thought desirable to have boundary values closer together where we are expecting to find actual values. This will not 'force' these values to be returned by the estimation, but will imply a lesser accuracy for values more sparsely covered."

Simulation by synthetic data

Prior to the real data collection, the simulation test by synthetic data is conducted to ensure that the SP design is capable of generating the accurate estimates of a series of relative values (Fowkes and Wardman, 1991). This method is based on the discrete choice theory (Ben – Akiva and Lerman, 1985), in which the random utility is a function of measurable utility and random error, ' ε ', the details of utility function will be introduced in section 4.3.1.

The simulation process consists of generating synthetic (artificial) responses to an SP design using accepted (known) values and then estimating the choices in the normal manner to see how efficient the designs are at extracting the specified parameters. Figure 4.2 presents a flow chart of the simulation process.

Synthetic responses are produced based on the assumption of utility maximisation where option i will be chosen if the utility of option i is higher than option j. Individuals' artificial utilities are generated from the measurable (known) utility and a random error.

The measurable utilities are functions of the attribute values in the SP design and a set of reasonable coefficients (accepted values) from previous studies of the variables. Normally the coefficient of cost is set to one; hence coefficients of other variables are known as the relative values of the variable. Accepted values in the simulation tests vary in a certain range to reflect different monetary values of time/services attributes due to the taste variation of respondents.

The error term is assumed to be a given distribution (see section 4.3.2) to represent the random effects for each alternative. Random errors and hence the simulated choices can be generated by many programs. This study uses Excel and FORTRAN. The assumptions made, regarding distribution, need not to be the same at the simulation stage as at the estimation stage, where the statistically 'best' distribution can be identified by suitable analysis.

This process is repeated many times to achieve a number of responses. These simulation responses were analysed using the ALOGIT program. The relative values estimated from the simulation responses are compared to the input values. The comparison of results indicates whether or not the design would be capable of generating the expected values. The simulation runs also test if the standard errors (t-ratios) are acceptable, and that they cannot be improved substantially by changing the design.

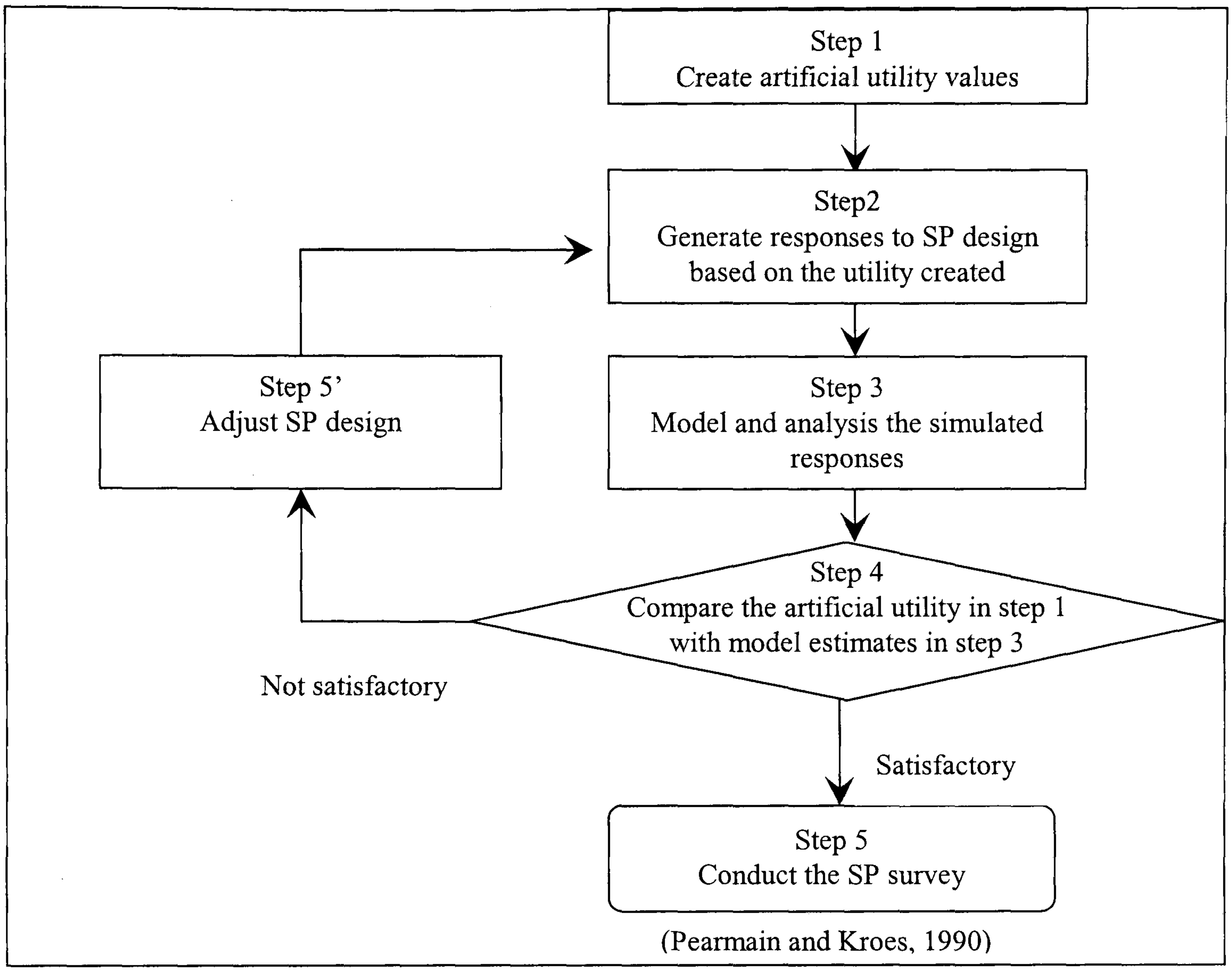


Figure 4.2 Flow chart for the process of the simulation test on the SP design

In this study, prior to the pilot survey and main data collection, each SP design was tested using the synthetic data. The bias has been simulated by specifying differences of monetary values of time and rolling stock in the SP design, and simulating to check that such differences can be detected as significant differences in estimated model coefficients. Simulation results showed that SP designs were statistically efficient for the data collection and intended analysis. The simulation test for the second pilot survey, which was later used in the main survey, will be explained in section 5.5.2 in detail as an example.

4.2.6 Pilot survey

Simulation only tests the statistical properties of SP design, but not the realism of the design. Sometimes the simulation test cannot guarantee the design will be problem free, particularly where there is a lack of previous information about magnitudes and ratios of coefficients in the study (Tudela, 2000). The simulation also cannot test whether individuals find the exercise realistic; therefore, a pilot survey is used to test the design and to find how individuals respond to the survey in terms of format, questioning, presentation, survey conducting and response rate.

4.3 Analytical Issues

4.3.1 Random utility theory

The SP method is based on the behavioural assumptions that decision makers connect actions to consequences and then decompose consequences into attributes. An individual's choice is assumed to depend on 'utility', representing the satisfaction or benefits received by the person from each alternative. The basic assumption is that respondents make their decision to maximise the utility (rational). Due to the discrete nature of decision making behaviour, SP data analysis is often based on the random utility theory and uses discrete choice models to measure the preference of individuals (McFadden, 1973).

Because of the deficiencies in potential observations, researchers cannot explain the choice behaviours precisely (McFadden, 1973; Manski, 1977; Ben-Akiva and Lerman, 1985). Therefore, the total utility (U_i) of an alternative i is composed of two parts: observable components (V_i) and unobservable components (ε_i), shown as following:

$$U_i = V_i + \varepsilon_i$$

Equation 4.3

The measurable component V_i is the representative or deterministic utility. A utility function is usually assumed to be linear in parameters, but need not be linear in the attribute. Equation 4.4 indicates that the attributes of the utility function are in a linear manner, where β_{ik} are the parameters/utility weights of the k attributes, X_{ik} for alternative i. It is also possible to specify non-linear transformations of attributes and their parameters to improve models.

$$V_i = \sum_k \beta_{ik} X_{ik}$$

Equation 4.4

The unobservable components ε_i are assumed to incorporate the inconsistency between the result and the actual values. Manski (1973) identified four distinct sources of randomness, as quoted in Ben-Akiva and Lerman (1985): unobserved attributes, unobserved taste variations, measurement errors, and model specification error. The taste variation can be addressed by segmentation analysis and advanced logit models. However, other errors are still assumed to be in the single additive element ε, with assumed known distribution. In order to overcome apparent behavioural inconsistencies within and between decision-makers, probabilistic choice theory was developed to allow a probabilistic process to account for unobserved variations.

4.3.2 Conventional logit models

The utility function is used to evaluate the ordinal utility scale by using attribute values. It is based on the assumption that the consumer is rational (his/ her decisions are consistent and transitive). An alternative i is chosen rather than alternative j from a choice set C_n (n alternatives available), if and only if:

$$U_i > U_j \qquad for\ all\ j \neq i \in C_n \qquad \textbf{Equation 4.5}$$

From equations 4.3 and 4.4, alternative i is chosen if:

$$V_i + \varepsilon_i > V_j + \varepsilon_j \qquad \textbf{Equation 4.6}$$

As discussed before, due to the uncertainty and complexity of human behaviour, the random terms ε_i are unknown. They may vary across alternatives and individuals. Therefore, a distribution for them is assumed and only a choice probability of occurrence can be obtained, as shown in equations 4.7 and 4.8:

$$P_i = P(V_i + \varepsilon_i > V_j + \varepsilon_j) \qquad for\ all\ j \neq i \in C_n \qquad \textbf{Equation 4.7}$$

$$P_i = P_i(\varepsilon_i - \varepsilon_j \geq V_j - V_i) = P_i(\eta \geq V_j - V_i) \qquad for\ all\ j \neq i \in C_n \qquad \textbf{Equation 4.8}$$

Where, $\eta = \varepsilon_i - \varepsilon_j$. The definition of the distribution of errors has important implications for the properties of the resulting choice model.

If the random error terms (ε) are independently and identically distributed (IID) with a Gumbel distribution, the choice model is called a Multinomial Logit Model (MNL). The origins of MNL can be traced back to the work of Luce (1959). McFadden (1973) derived it from random utility theories. In MNL, the probability P_i of choosing alternative i from the choice set C_n, given measured utilities V_j $(j \in C_n)$, is given by:

$$P_i = \frac{\exp(V_i)}{\sum_{j \in C_n} \exp(V_j)} \qquad \textbf{Equation 4.9}$$

The process of estimating utility parameters (β_{ik}) of MNL model is usually based on maximum likelihood estimation. This estimator is based on the idea that the values of parameters are most likely to occur for the observed sample. The parameter (β_{ik}) can be interpreted as an estimate of the weight of attributes in the utility function of alternative i. They can be allowed to vary across groups of respondents; for example, varying across their socio-economic characteristics.

The advantages of the MNL model are that it is easy to estimate, compared with other models (e.g. probit), and coefficients are easy to interpret, therefore, it is the most commonly used model in the SP techniques. The main drawback (limit) of this model is the basic assumption about the independence from irrelevant alternatives (IIA property) which are:

- The error components (ε_i) are identically and independently distributed IID) across alternatives; and
- The error components are identically and independently distributed (IID) across cases.

MNL assumes that the choice options are independent. For any two alternatives, the ratio of their choice probabilities (estimate of parameters) is unaffected by other alternatives in the same experiment choice set. Therefore MNL fails to take account of correlation between alternatives. The need to overcome the IIA property of the MNL has been the primary motivation for the development of the numerous generalized logit discrete choice model structures.

Binary logit model is the simplified MNL which is characterised as models explaining a binary (0/1) dependent discrete variables. The probability of alternative i being chosen can be expressed as Equation 4.10:

$$P_i = \frac{\exp(V_i)}{\exp(V_i) + \exp(V_j)}$$ **Equation 4.10**

4.3.3 Scale parameter

As explained in section 4.3.1, an analytical method used for explaining choice behaviour is discrete choice analysis based on random utility theory (RUM). In logit models, estimated coefficients are 'scaled' according to the variance of the unexplained error. In general, utility can be expressed as:

$$U_i^* = V_i + \varepsilon_i^*$$ **Equation 4.11**

Similar to Equation 4.3, the random utility (U_i^*) for alternative i, is broken down to a measurable part (V_i) and a random part (ε_i^*). In RUM, the choice models are derived by making assumptions about the distribution of the random effects. It assumes ε_i^* has variance ($\sigma^2 = \lambda^2 \times (\pi^2/6)$), and the variance can be expressed as:

$$\lambda = \frac{\pi}{\sqrt{6}\sigma}$$ **Equation 4.12**

As the scale of utility is irrelevant to behaviour, utility can be divided by λ without changing behaviour. Utility becomes:

$$U_i = V_i / \lambda + \varepsilon_i$$ **Equation 4.13**

where $\varepsilon_i = \varepsilon_i^* / \lambda$. Now the unobserved portion has variance $\pi^2 / 6$ (Gumbel distribution). If V_i is linear in parameters with coefficient β_i^*, the choice probabilities become:

$$P_i = \frac{\exp(V_i / \lambda)}{\sum_j \exp(V_j / \lambda)} = \frac{\exp((\beta_i^* / \lambda) \cdot X_i)}{\sum_i \exp((\beta_j^* / \lambda) \cdot X_j)}$$ **Equation 4.14**

Each of the coefficients are scaled by λ. The parameter λ is called the scale parameter, because it scales the coefficients to reflect the variance of the unobserved portion of utility. Only the ratio β_i^* / λ can be estimated which means β_i^* and λ are not separately identified (Ben-Akiva and Lerman, 1985). Usually, the model is expressed in its scaled form, with $\beta_i = \beta_i^* / \lambda$, which gives the standard logit expression.

$$P_i = \frac{\exp(\beta_i \cdot X_i)}{\sum_j \exp(\beta_j \cdot X_j)}$$ **Equation 4.15**

Figure 4.3, from Adamowicz et al. (1998), shows the influence of the scale parameter of a data set on the choice probabilities.

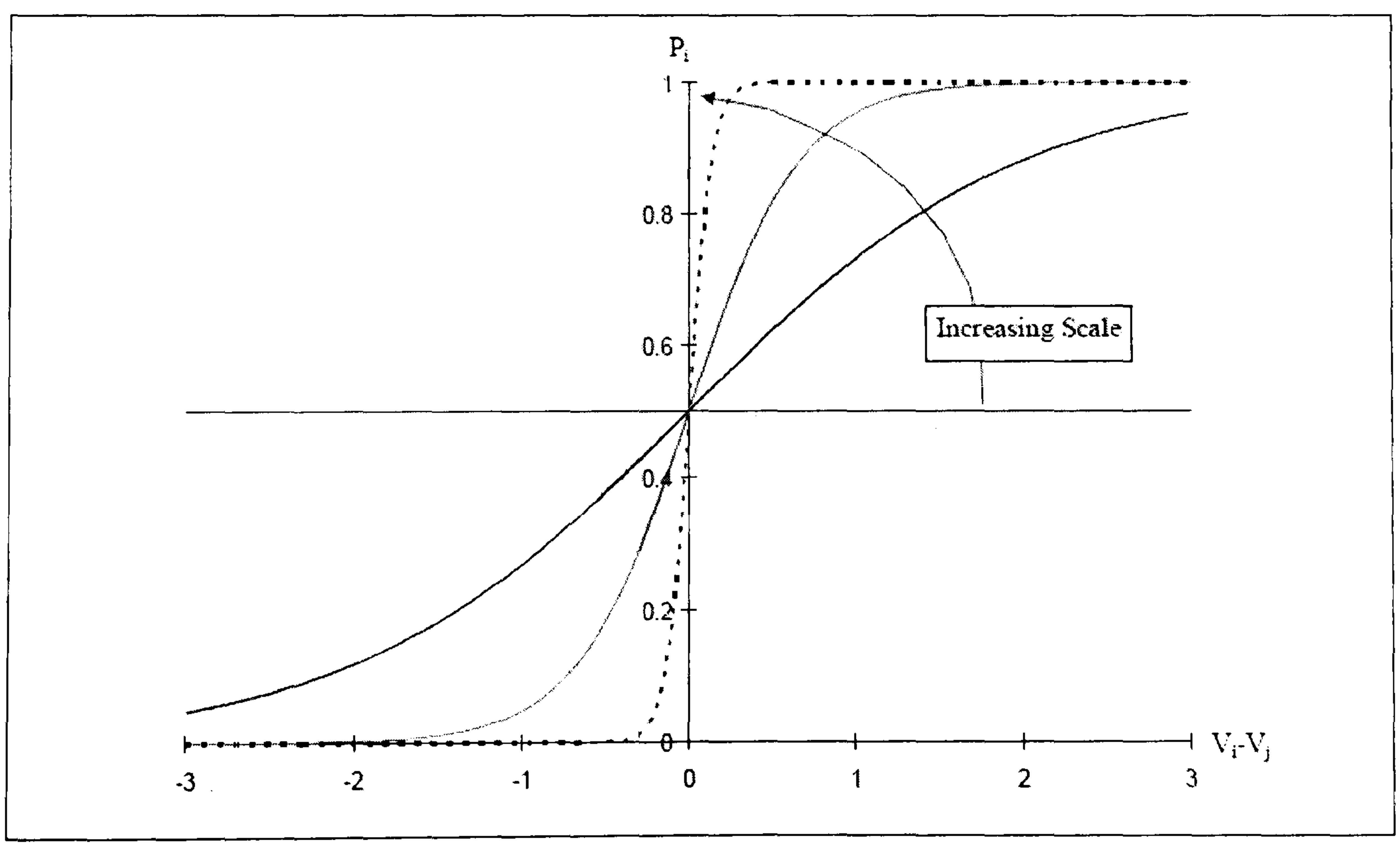


Figure 4.3 The effect of scale parameter on choice probabilities

When the scale is zero, the choice probability are equal (in the binary choices shown in the figure, both probabilities equal 0.5); as the scale grows, the choice model predicts more and more like a step function, perfectly discriminating between the two alternatives in the graph.

The scale parameter has two features. Firstly, from Equation 4.12, the scale parameter in the MNL model is inversely related to the variance of the error term. This means that the higher the scale, the smaller the variance, which in turn implies that models with better fit have larger scales. In the present study, this feature will be used in the hypothesis test in chapter 8.

Secondly, the scale parameter cancels out in the estimation of the relative valuation of attributes. In the linear additive utility function, the relative values are estimated as the ratio of the parameters (representing the marginal rate of substitution), for instance, the monetary value is obtained by the ratio of the attribute parameter and the cost parameter (see section 4.4.4).

$$\frac{\beta_i^*}{\beta_j^*} = \frac{\beta_i / \lambda}{\beta_j / \lambda} = \frac{\beta_i}{\beta_j}$$ **Equation 4.16**

The role of scale parameter has spawned several streams in recent empirical research. Among them, a) 'data fusion' in travel demand modelling and combining different source of data (Morikawa, 1989; Swait and Louviere, 1993; Hensher and Bradley, 1993; Hensher et al., 1999; Ortuzar and Willumsen, 2002); b) comparison and testing differences in various sources of preference data (Bradley and Daly, 1994; Swait and Adamowicz, 2001).

4.3.4 Combining different sources of data

Why do we need to consider scale factors in pooling the data?

As explained in section 4.3.3, the scale parameter scales the coefficients to reflect the variance of the unobserved portion of utility. The larger the variance of the error terms, the smaller the scale. In some situations, the variance of the error terms can be different for different segments or for different regions. For example, different data collection processes may influence choice variability differentially, and if this is not recognised, may confound the real behavioural role of observed and unobserved influences on choice.

For example (Train, 2003, p.45), for two data sources A and B, the original model is:

$U_i = \alpha T_i + \beta C_i + \varepsilon_i^A$ in City A (here i is alternative)

$U_i = \alpha T_i + \beta C_i + \varepsilon_i^B$ in City B where $\varepsilon_j^A \neq \varepsilon_j^B$

Label the ratio of variance as

$$\theta = Var(\varepsilon_i^B) / Var(\varepsilon_i^A)$$ **Equation 4.17**

If the utility for respondents in City B is divided by $\sqrt{\theta}$, doing so allows us to get the same variances for the two cities:

$$Var(\varepsilon_i^B / \sqrt{\theta}) = (1/\theta) Var(\varepsilon_i^B) = [Var(\varepsilon_i^A) / Var(\varepsilon_i^B)] Var(\varepsilon_i^B) = Var(\varepsilon_i^A)$$

where θ is scale factor. Not considering θ in this situation leads to the over/under estimation of coefficients. Another example is that more complex designs might be expected to yield a smaller scale due to a larger amount of random errors. Failure to account for the difference in scale could lead to its effect appearing in the coefficient estimation. In the present study, the scale factor impact is considered in the data analysis in chapters 7 and 8.

Combining different sources of data

Morikawa (1989) noted that the basic scale identification problem applied to a single preference data sources, but the ratios of scale parameters in two or more sources of data could be identified, which led to sequential (Swait and Louviere, 1993) and simultaneous estimation (Morikawa, 1989; Hensher and Bradley, 1993; Bhat, 1995) methods for combining different sources of preference data.

In this study, Simultaneous Estimation (SE) is applied to estimate the scale factors. A brief introduction of the process (for more details see Ortuzar and Willumsen, 2002, p.218) is given. All data are analysed simultaneously in a single model. Each choice is treated separately by specifying a dummy nest.

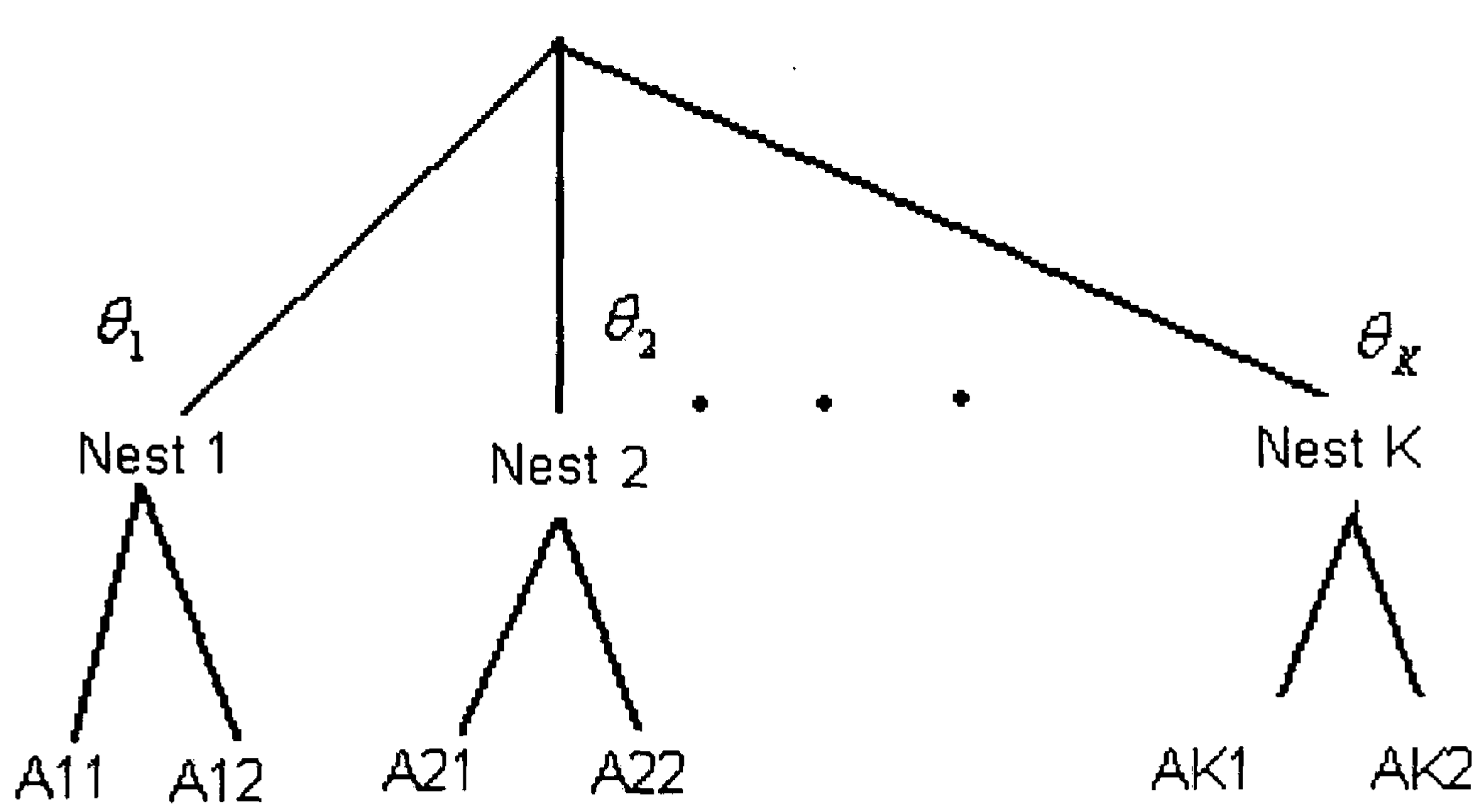


Figure 4.4 Tree structure to combine different sources of data

Figure 4.4 demonstrates the tree structure used to combine different sources of data. Suppose there are K preference data sources to be combined (called K nests). There are two alternatives in each choice set, such as A11/A12 in the nest1, AK1/AK2 in the nest K. The estimation problem involves imposing an equality restriction on the coefficients of attributes of the K preference data sources, and the estimation of K-1 additional scale factors (θ_1, $\theta_2,\ldots,$ θ_K). One scale factor is fixed, say $\theta_1 = 1$ (normalized to Nest 1), and the K-1 other parameters are inverse variance ratios relative to the reference (fixed) data source. The corresponding unrestricted model frees attribute coefficients and scale factors for the K data sources by estimation ($\theta_K \beta_K$), k=1,...K.

A Hierarchical Logit model (McFadden, 1981) has normally been estimated. In an upper nest in a hierarchical logit, we represent the utilities (*expected maximum utility* -EMU) of alternatives in the lower nest as:

$$EMU = \ln[\exp(U_1) + \exp(U_2) + \ldots + \exp(U_K)] \qquad \textbf{Equation 4.18}$$

In lower nests, the utility of the alternative has been scaled to the upper nest by the scale factor θ.

$$\log sum = \theta_k \ln[\exp(V_i)] = \theta_k V_i \qquad \textbf{Equation 4.19}$$

This index $\theta_k V_i$ is commonly referred to as the *inclusive value (IV),* or alternatively, the *logsum* or *expected maximum utility*, whereas θ_k is the coefficient of the inclusive value. The parameter estimated in the $EMU(\theta)$ is a scale factor. Unlike the conventional hierarchical logit model, the scale factor θ estimated is not constrained to be less than 1. An example of combining different data set is presented in section 7.3.2.

4.3.5 Heteroskedastic Multinomial Logit framework

Swait and Adamowicz (2001) formulated a Heteroskedastic Multinomial Logit (hereafter, HMNL) model which allows the variance of the random components of the utility to vary across individual/observations. HMNL has also been called Parameterised heteroskedastic MNL (PHMNL) model (Hensher et al., 1999, p.209) and Covariance heterogeneity (CovHet) fixed effects HEV model (Louviere et al., 2000, p.195).

The choice probabilities for the HMNL model are given as follows:

$$P_i = \frac{\exp(\lambda V_i)}{\sum_{j \in C_n} \exp(\lambda V_j)} \qquad \textbf{Equation 4.20}$$

Where λ_s is the scale parameter for data source, s. The heteroskedastic logit model is based on parameterization of the scale of the multinomial logit model as follow:

$$\lambda = Exp(\tau + \alpha Z_s) \qquad \textbf{Equation 4.21}$$

Where Z is a vector of individual and trip related characteristics; $\tau; \alpha$ are parameters to be estimated; and the exponential function to ensure λ is non-negative. If $\tau; \alpha$ in Equation 4.21 turn out to be zero, then λ equals one and the MNL model is obtained shown as Equation 4.10.

In the HMNL model, the scale parameter (across all the alternatives) is not a constant term, but "a function of alternative-specific variables: attributes associated with an alternative and each sampled individual can be introduced as sources of scale decomposition, adding useful behavioural information of sample heterogeneity" (Louviere et al., 2000, p.195). HMNL model is different from the HEV (Bhat, 1995) model. The latter allows different scale parameter for different alternatives in a choice set, whereas the former allows different scale parameter across choice situations, but same across all the alternatives in the choice sets.

Swait and Adamowicz (2001) applied this model to analyse how SP task complexity and RP choice environment (e.g. market structure) influence levels of variability in preference data. They proposed ways of measuring the task complexity and choice environment, which were constant across alternatives, hence, scale parameters in their model varied across individuals and SP replications, instead of alternatives. Later, DeShazo and Fermo (2001) proposed an exponential functional form for the scale parameter to prevent λ from being negative:

$$\lambda_K(C_K) = \exp[\sum_{l=1}^{m} \gamma_l \cdot C_l] \qquad \textbf{Equation 4.22}$$

The HMNL model allows the analyst to incorporate the complexity and cognitive burden of the experiment through an appropriate parameterisation of the scale parameter. In addition, socio-economic variables (Caussade et al., 2005) can be introduced into function of scale parameter to obtain a fully heteroskedastic model, i.e. due to different respondents' social-economic variables, the nth choice situation within a SP questionnaire for two individuals will have a different scale parameter.

The HMNL model has been applied to detect the impact of task complexity of SP design (Swait and Adamowicz, 2001; DeShazo and Fermo, 2002; Caussade et al., 2005), choice environment (Koppelman and Sethi, 2005), and respondents' characteristics (Caussade et al., 2005) on the choice making. The potential difference in error variances is incorporated in the HMNL model as a parameterization of scale parameter of the utility function.

In the present study, HMNL model is used to explore the impact of SP design, respondents' socio-economic features and their perceptions on the variability of the SP responses. The analysis will be presented in chapter 8.

4.3.6 Taste variation among individuals

Logit models can represent systematic taste variation, but only with respect to observed factors (commonly socio-economic information and trip characteristics), using segmentation analysis. However, they cannot handle unobserved or purely randomly variables (Train, 2003).

Segmentation analyses can be applied to find the taste variation among groups of respondents, which are normally conducted by estimating separate models for each group (for example, used in Preston and Wardman, 1991 and Accent and HCG, 1994) and by using incremental factors (MVA et al., 1987; Louviere, 1988).

The first method would reduce the significance of coefficients (MVA et al., 1987) as only a small number of observations are available for each segment. The incremental factors allow different marginal utilities across segments of the sample. They can be specified as:

$$\sum_{y=1}^{n-1} \gamma_y d_{ky} X_{ik} \qquad \textbf{Equation 4.23}$$

Where: γ_y is an incremental factor for the k^{th} attribute (X_k) and d_{ky} is a dummy variable denoting whether an observation is in y^{th} group of n groups in a category. If so, d_{ky} is equal to one, otherwise zero. One of the groups in the category is arbitrarily chosen as the base. The incremental effects for other groups are relative to this base, so only n-1 dummy variables are defined. The utility function of alternative i is:

$$V_i = \sum_k \beta_{ik} X_{ik} + \sum_{y=1}^{n-1} \gamma_y d_{ky} X_{ik} \qquad \textbf{Equation 4.24}$$

In Equation 4.24, the coefficient of X_{ik} for the base group is β_{ik} , and the coefficient of X_{ik} for y^{th} group in the category is $\beta_{ik} + \gamma_y$. This approach can indicate the sign and size of any effect from the segmentation variable, provided it is statistically significant, and can be tested on the fit of different models, for example, between the basic model and the segmentation model or among the different segmentation models (Wardman et al., 1998).

4.3.7 Model estimation and comparison

<u>Model Estimation</u>

As estimated parameters have associated standard errors, a coefficient is considered to be significantly different from zero at the 5% confidence interval when its corresponding t-ratio (the ratio of the mean value to its standard error) has an absolute value greater than 1.96. Values of t-ratio as low as 1.6 are sometimes accepted, representing the 90% confidence interval, if the sign is correct and magnitude (e.g. implied values) seems plausible.

The overall model goodness-of-fit is indicated by likelihood-ratio index, $\rho^2(C)$, which is similar to R^2 for a linear regression model. $\rho^2(C)$ values between 0.2 and 0.4 indicate an extremely good fit (Louviere et al., 2000). These results of the estimation process: values of parameters, their t-ratios and likelihood indexes can be estimated using available computer software programs, for example, ALOGIT (Hague Consulting Group, 2003), BIOGEME (Transport and Mobility Laboratory, 2005) and GAUSS (Aptech Systems, 1997).

Model Comparison – Log likelihood ratio test

In the cause of model development, we need to compare our model with a more constrained version which assumes that some of the coefficients in the former model are zero. A Log Likelihood ratio test (Train, 2003, p.75) is applied to compare models. The test statistic is twice the difference between the log likelihood values of the models at convergence. The likelihood ratio [$-2(LL(\hat{\beta}^H) - LL(\hat{\beta}))$] is distributed χ^2 with degrees of freedom equal to the number of restrictions (parameters in the basic model which are constrained to zero) implied by the null hypothesis, where $\hat{\beta}^H$ is the (constrained) maximum value of the likelihood function under the null hypothesis H, and $\hat{\beta}$ is the unconstrained maximum of the likelihood function.

4.3.8 Repeated measurement effect

One of the features of SP experiments is that multiple responses are obtained from each respondent. The data analysis assumes that these responses are independent. This may lead to the "repeated measurement" problem, which is defined as the problem of different variances (heteroscedasticity) and correlation of repeated observations from each individual.

It is believed that this problem will result in the upward biased values of the t-ratios (under – estimating the standard errors), but does not affect consistently the estimated parameter values of logit models. The simplest solution for correcting the t-ratio involves dividing the t-ratios by a correction factor. A suggested factor is the square root of the number of observations per individual (Bradley and Daly, 1993). However, researchers argued the factor is too high. Some other factors mentioned are third root or the fourth root of the number of repeated questions per individual (Bates and Terzis, 1997). However, this method has been criticised because this is an extreme correlation treatment when all responses per individual are perfectly correlated (Abdel-Aty et al., 1995).

Moreover, Cirillo et al. (2000) applied re-sampling techniques, Bootstrap and Jackknife, to solve the problem. The Bootstrap method creates a completely new sample each time for each Bootstrap sample by drawing randomly with replacement within the sample. The Jackknife method uses the same dataset as the original data set but deletes small parts of the data in each Jackknife sample. The results of applying the Jackknife method confirmed that the estimated coefficient values were unbiased, but the variance estimates were varied.

These two methods were examined by Ortuzar et al. (1997) who found that neither was reliable. They found that most coefficient values were smaller than those from traditional approach, but in the case of specific constant, the t-ratios were larger in minimum distance method than those in the traditional approach.

Ouwersloot and Reitveld (1996) applied the minimum distance method (sub-samples), taking only one observation from each respondent for each sub-sample, in order to avoid correlation between the responses. Models were analysed for each sub-sample, and then a final model for the whole sample was developed later using minimum distance estimator. They found that in their data the repeated measurement effect was modest and statistically insignificant.

In summary, it is believed that effect of repeated measurements does not significantly affect values of coefficients in the logit model, with only a small effect of reducing the significance of variables in the model. In the present study, the repeated measurement problem has been handled using the Jackknife method provided in the ALOGIT program.

4.4 Model Specification for Valuation of Rolling stock

4.4.1 Two types of model specification for estimation of rolling stock

Wardman and Whelan (2001) conducted a review of previous studies on the valuation of rolling stock improvements. They reported two types of model specification for obtaining the valuation of rolling stock. The valuations of rolling stock are often reported as a proportion of the fare paid and sometimes as an impact on the value of time (VoT).

The first type of model specification (I) reports the valuation of the improved rolling stock as a constant term in the utility function, as shown in Equation 4.25. ASC_i represents the alternative specific constant term, which refers to the preference for the improved rolling stock. In the utility model, the ASC term also captures the average effect on the utility of all factors that are not included in the model (Train, 2003). The X_{ki} represents variables in the SP design such as time, cost and service attributes. Hence, the monetary value of improved rolling stock compared with the current rolling stock is $ASC_i / \beta_{\text{cost}}$.

$$U_i = ASC_i + \sum(\beta_{ik} \cdot X_{ik}) \qquad \textbf{Equation 4.25}$$

This model specification was commonly used in the early studies on the valuation of rolling stock. The advantage of this model specification is that it is easy to interpret and transfer the results. The limitations of this specification are that the valuation of rolling stock is not allowed to vary with journey time (Wardman and Whelan, 2001, p.417).

The second method (II) assumes that the value of improved rolling stock is related to the in-vehicle time. We might expect that the stock valuation increases with the duration of the journey. This requires the specification of interaction terms which are the product of the dummy variables denoting the rolling stock d_i and the train journey time T_i. γ is the parameter used to measure this interaction effect. A utility function with such a feature is:

$$U_i = \gamma d_i Time_i + \sum(\beta_{ik} \cdot X_{ik}) \qquad \textbf{Equation 4.26}$$

The monetary value of improved rolling stock relative to the current rolling stock would be $\gamma Time_i \,/\, \beta_{\text{cost}}$, which varies with the length of journey. This formulation is entirely equivalent to the value of train time varying according to the type of rolling stock. The first study to explore the second model specification was MVA (1991) and it was subsequently examined by Babtie (1993), MVA (1992) and Oscar Faber TPA(1994). This model specification can be extended to cover both a constant and time dependant component (MVA, 1993).

4.4.2 Income effects

Income has been found to have a strong effect on respondents' decision making and valuation of attributes. In value of time studies, it has been found that as the income increases, people become less sensitive to cost (Wardman, 2001). As the monetary value is obtained by the marginal utilities of the target attribute and cost coefficient, differences in income leads to the variation in the monetary values (Fowkes, 1986; MVA, 1987; Hague Consulting Group and Accent, 1999; Wardman, 2001; Gunn, 2001; Wardman, 2004).

In this study, the income effect is tested together with the journey purpose effect (section 4.4.3) to avoid the confounding effects prior to the subsequent analysis of SP design impacts on the valuation of the improved rolling stock. The income effects are examined by two methods: segmentations and income elasticity.

Segmentation

Equation 4.27 is the segmentation method which has been applied to obtain the effect of income on the estimation of coefficients. Dummy variables for different income band 'd_{incm}' are incorporated into the utility function to obtain income incremental effects on the cost coefficient, where γ_{incm} are the corresponding coefficient for income group 'm'.

$$U_i = ASC + \sum(\beta_{ik} \cdot X_{ik}) + \sum(\gamma_{incm} d_{incm} \text{cost}_i)$$

Equation 4.27

Income Elasticity

'$\kappa \dfrac{\text{cost}}{Y^{\varpi}}$' is added into the utility function to investigate the income elasticity of value of time. The term is derived from:

$$\frac{\partial VoT}{\partial Y}\frac{Y}{VoT} = \varpi$$ **Equation 4.28**

Wardman (2001) conducted a meta-analysis and found that the value of time income elasticity was positive and is approximately 0.5. Equation 4.29 investigates the value of time (VoT) income elasticity. Y represents the household income.

$$U_i = ASC + \sum(\beta_{ik} \cdot X_{ik}) + \kappa \frac{\text{cost}_i}{Y^{\varpi}}$$

Equation 4.29

4.4.3 Journey purpose effects

Previous studies have found respondents' journey purpose have significant effects on their travel decision making, thus leading to the variation of valuations (section 3.4.1). For example, a meta-analysis (Wardman, 2001) found that business travellers have higher value than other types of travellers. Compared with leisure travellers, commuters have a 16% higher value of in-vehicle time (IVT), and a 26% higher value of headway.

As journey purpose contributes to the variation of the valuation in the SP experiment, the impact of journey purpose on the estimation is explored prior to the research hypotheses testing, to avoid confounding effects.

4.4.4 Monetary values from the model estimation

The monetary values can be derived by the ratio of the marginality utility of the target variable and cost. For instance, the value of time is the ratio of time and cost coefficients. Equation 4.30 shows how to obtain the monetary value of the k^{th} attribute X_k, β_{x_k} represents the coefficient for the target variable and β_{cost} is the cost coefficient.

$$VoX_k = \frac{\beta_{X_k}}{\beta_{cost}}$$ **Equation 4.30**

To generate the interval of values, two methods can be used, which will be introduced in turn.

Fowkes's method (1998)

Fowkes (1998) derived a formula for the variance of the VoT, which can be obtained by the variance of X term and cost together with their covariance, shown as below:

$$Var(VoX) = \frac{\beta_X^2}{\beta_{cost}^2}\left[\frac{Var(\hat{\beta}_X)}{\beta_X^2} + \frac{Var(\hat{\beta}_{cost})}{\beta_{cost}^2} - \frac{2Cov(\hat{\beta}_{cost}, \hat{\beta}_X)}{\beta_{cost}\beta_X}\right]$$ **Equation 4.31**

This method is very simple and easy to account for the incremental effects of individuals' socio-economic features. For instance in model specification I (Equation 4.25), where the stock value is estimated as a constant, the monetary value of the improved rolling stock can be obtained from the ratio of the ASC term and the cost coefficient (incorporating the income incremental effects). Equation 4.32 shows how to obtain the monetary value of the attributes for the different income bands, and the standard deviation.

$$VoX_{km} = \frac{\beta_{X_k}}{\beta_{cost} + \gamma_{incm}}; \qquad \text{here, let } a = \beta_{cost},\ b = \gamma_{incm},\ c = \beta_{X_k}$$

$$Var(a+b) = \mathrm{var}(a) + \mathrm{var}(b) + 2\,\mathrm{cov}(a,b)$$

$$\mathrm{cov}(a+b, c) = \mathrm{cov}(a,c) + \mathrm{cov}(b,c)$$

$$Var(VoX) = Var\left(\frac{c}{a+b}\right) = \frac{c^2}{(a+b)^2}\left[\frac{Var(c)}{c^2} + \frac{Var(a+b)}{(a+b)^2} - \frac{2Cov(a+b,c)}{(a+b)c}\right]$$

Equation 4.32

Here, γ_{incm} measures the incremental effect of income group m on the cost. The variance of the monetary value is derived from Equation 4.30 and the statistical knowledge. The variance (for a single attribute) and covariance value in the formula are generated by ALOGIT program.

Armstrong et al.' method (2005)

The maximum likelihood estimation method yields coefficients that are asymptotically distributed multivariate normal (Ben-Akiva and Lerman, 1985). Consequently, the point estimate of the monetary value (such as VoT) is a random variable by an unknown PDF (the probability distribution for the ratio between two normally distributed variables is unknown a priori) (Armstrong et al., 2001, p.144). For example:

$VoT = \frac{\beta_t}{\beta_c}$, Where VoT is the point estimate of the value of time, which is the ratio between the parameters of time β_t (β_{time}) and cost β_c (β_{cost}). Garrido and Ortuzar (1993) derived the upper and lower bounds for the interval as follow:

$$V_{S,I} = (\frac{\beta_t}{\beta_c}\frac{t_c}{t_t})\frac{(t_t t_c - \rho t^2}{(t_c^2 - t^2)} \pm (\frac{\beta_t}{\beta_c}\frac{t_c}{t_t})\frac{\sqrt{(\rho t^2 - t_t t_c)^2 - (t_t^2 - t^2)(t_c^2 - t^2)}}{(t_c^2 - t^2)}$$ **Equation 4.33**

Where t_t and t_c correspond to the t-statistics for β_t and β_c, respectively; t is the critical value of t given the degree of confidence required and sample size and ρ is the coefficient of correlation between these two parameter estimates. Armstrong et al. (2001) found that this method represents the case studies more accurately; however, this method is tedious and needs considerable computing efforts.

This study will compares the values obtained from the model estimation, using these two methods in section 7.4.5.

4.4.5 Measure the effect of design factor

The previous section has presented the method for SP data analysis, based on random utility theory. The utility represents the preference of an individual for a chosen option, according to its attributes. The choice is based on an assumption that the individual maximises his/her own utility. From the literature review in chapter 2, respondents might have different incentives to over/under estimate the valuation of new product for certain outcome. In this research, to test the research hypotheses (chapter 1), a series of SP experiments are designed (chapters 5 and 6). Among them, cheap talk and more attributes to amend incentive (to strategic bias) were added into the experiments to explore their impacts on respondents' decision making.

To measure the impact of the design factors, three aspects in the utility model are investigated: the alternative specific constants, scale parameter and qualitative analysis of the relationship between two nominal variables.

Alternative specific constant

As explained in the section 4.3.1., the observed part of utility can be linear in parameters with a constant, as shown in Equation 4.4.

$$V_i = \sum_k \beta_{ik} X_{ik} + ASC_i$$ **Equation 4.34**

The alternative specific constant (ASC) is the constant term in discrete choice models. In the situation that the alternative is labelled (i.e. in this experiment, the alternatives are labelled to be two types of trains), the ASC explains the preference difference between the two alternatives. The ASC for an alternative captures the average effect on utility of all factors that are not included in the model (Train, 2003).

In addition, if a constant is included in the model, it symbolizes the tendency, for example, to choose the specific alternative. For example, DeShazo and Fermo (2004) includes the ASC term to explain the "propensity to attend" behaviour of respondents. They tested the task complexity effect of the standard deviation among the attributes, by incorporating an ASC term '$\delta_c d_{STALjk}$' into the utility model, which characterizes the cognitive cost associated with evaluating one alternative. The δ_c measures individuals' propensity to attend to that alternative, when the cognitive cost of evaluation increases.

In this research, cheap-talk and complexity are included in some of the SP experiments, together with some follow-up questions on individuals' perceptions. To identify impacts of the above factors on individuals' "propensity to attend" one alternative (i.e. one type of rolling stock), ASC together with alternative specific dummy terms are included to utility function, as shown in Equation 4.35:

$$V_i = \sum_k \beta_{ik} X_{ik} + ASC_i + \delta_{Design} d_{Design} \qquad \textbf{Equation 4.35}$$

Here, d_{Design} are the dummy variables denoting the design factors (cheap-talk script and complexity) and δ_{Design} are the coefficients of the design effects.

Scale parameter

Another method of measuring the effects of design factors is by testing the scale parameter in the discrete choice models. As explained in sections 4.3.3 and 4.3.4, the scale parameter in the utility function explains the variance of the unobserved factors. The scale parameter in the MNL model is inversely related to the variance of the error term. This indicates that the higher the scale, the smaller the variance, which in turn implies that high-fit models have larger scales.

Heteroskedastic Multinomial Logit model (section 4.3.5) parameterizes the scale parameter to explain the heterogeneity of individuals' decision making. In the research, the effects of design factors are tested by their impacts on the scale parameter.

$$\lambda = Exp(\tau + \alpha d_{Design}) \qquad \textbf{Equation 4.36}$$

Scale parameter λ is not a constant term, but explained by τ and α . α measure the effects of design factors (denoted by dummy variable d_{Design}). The exponential form is used to avoid the scale parameter to be negative. If α is negative, it means adding the design factor leads to the smaller scale parameter, which implies more variance in the unobserved factor, and vice versa.

Relationship between two nominal variables

Cramer's V coefficient (Bryman and Cramer, 2005) is a statistic measuring the strength of association or dependency between two (nominal) categorical variables in a contingency table.

For the sake of clarification, the notation in this section is defined differently from that in the context of utility function. Table 4.1 lists one example given the above information. Suppose X and Y are two categorical variables that are to be analyzed in an experimental or observational data with the following information: X has M distinct categories or classes, labelled $X_1,...,X_M$; Y has N distinct categories labelled $Y_1,...,Y_N$.

Table 4.1 Example of the contingency ($M \times N$) table

X/Y	Y_1	Y_2	...	Y_N	$\sum$
X_1	n_{11}	n_{12}		n_{1N}	$\sum n_{1.}$
X_2	n_{21}	n_{22}		n_{2N}	$\sum n_{2.}$
...			n_{ij}		$\sum n_{i.}$
X_M	n_{M1}	n_{M2}		n_{MN}	$\sum n_{M.}$
$\sum$	$\sum n_{.1}$	$\sum n_{.2}$	$\sum n_{.j}$	$\sum n_{.N}$	$\sum n_{ij}$

N pairs of observations (x_k, y_k) are taken, where x_i belongs to one of the M categories in X and y_i belongs to one of the N categories in Y. A contingency ($M \times N$) table can be formed where cell (i, j) contains the count n_{ij} of occurrences of Category X_i in X and Category Y_j in Y where $n = \sum n_{ij}$.

Suppose that the null hypothesis is that X and Y are independent random variables. Based on the contingency table and null hypothesis, the chi-square statistic (χ^2) can be computed by

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ **Equation 4.37**

Where O is the observation in the table, and E is the expected frequency in any cell and obtained by the equation:

$$E_{ij} = \frac{\sum n_{i.} \times \sum n_{.j}}{\sum n_{ij}}$$ **Equation 4.38**

Then, Cramer's V is defined to be:

$$V = V(X,Y) = \sqrt{\frac{\chi^2}{n \min(M-1, N-1)}}$$ **Equation 4.39**

Cramer's V coefficient varies between 0 and 1 to indicate the strength of relationship between two nominal variables that have more than two categories (Bryman and Cramer, 2005). The closer V is to 0, the smaller the association between the categorical variables X and Y. On the other hand, V being close to 1 is an indication of a strong association between X and Y. If $X = Y$, then $V = V(X,Y) = 1$.

4.5 Summary of the Methodology

This chapter has presented the related methodological issues used in the research. Firstly, the SP method was presented, covering the processes of SP design and simulation tests.

The technique of discrete choice analysis was also presented. The analysis technique is based on random utility theory, in which individuals are assumed to maximise their utilities by choosing the option with the highest utility to them. In this study, the analysis method used is based on logit family of models. A binary logit model is used in analysing the data. The scale parameter of the utility function represents the variance of unobserved factors. In the present study, responses from different SP designs are combined prior to the research hypotheses tests, by allowing for a different scale factor for each group. Segmentation models, based on incremental factors, are used to investigate the taste variation among different groups of individuals. The Heteroskedastic Multinomial Logit (HMNL) model was introduced, which will be used to determine the impacts of design factors on the variability of SP responses by parameterization of the scale parameter.

Finally, this chapter has reviewed the utility functions for the valuation of the improved rolling stock from previous studies. The impacts of socio-economic factors are incorporated into the utility function, to avoid potential confounding effects. Methods to estimate the monetary values of attributes are presented, with the calculation of the standard error.

Chapter 5

SP Experiment Design

5.1 Introduction

To test the research hypotheses, a suite of SP experiments were designed. This chapter presents the design of SP experiment and development of the questionnaire for identifying the influence of different designs on the pattern of SP responses. The development of the SP design followed the process discussed in chapter 4, and was refined and improved through two pilot surveys which are explained in detail in chapter 6.

Section 5.2 reports the SP experiment outline; a brief description of each experiment is discussed. Section 5.3 presents the characteristics of the survey form used in the data collection. Section 5.4 presents the design of SP experiment, involving the key elements: alternatives, attributes and their levels. Section 5.5 demonstrates the simulation tests and improvement of SP design. Section 5.6 explains the development of the cheap-talk script for testing the research hypothesis. Section 5.7 describes the presentation and measurement of task complexity (adding two more attributes to mask the research aim) in the experiment. Section 5.8 discusses the post – questionnaire questions for probing respondents' perceptions and choice making. Section 5.9 ends the chapter with a summary of the SP design for main data collection.

5.2 SP Experiment Outline

To test the research hypotheses, cheap talk script and task complexity (adding two more attributes to mask the research aim) were incorporated into SP experiments. The experiment context was selected to be users' valuation of rolling stock. An experiment outline combining these two factors is listed in Table 5.1.

Table 5.1 SP experiment outline

SP Design	Adding Cheap Talk (CT)	Adding Two More Attributes (to masking the aim)
SP1	No	Three attributes
SP2	Yes	Three attributes
SP3	No	Five attributes
SP4	Yes	Five attributes

The 'orthogonal plan' design allowed for the interaction between two factors to be estimated. There were four SP experiment designs: SP1 and SP2 were simple design, while SP3 and SP4 were complex design. A cheap-talk script was added to SP2 and SP4.

5.3 SP Survey Form

A set of paper-based questionnaires was designed for data collection. Examples of the SP questionnaires are shown in Appendix A and B. The questionnaire contained four main parts:

Part 1 was designed for gathering the basic information of the journey, including time (journey time), cost (fare), journey purpose, starting station and final station. The questions focused on the actual travel choice which was relevant to the study context.

Part 2 was the main SP survey part. The design was a conventional SP choice exercise, including nine hypothetical scenarios. Respondents were asked to state their preferences of two types of rolling stock: Super Sprinters versus Pacers. For better understanding of the difference between the improved train and current train, word descriptions accompanied by pictorial information were provided before SP choices. Part 2 was different among the four SP designs. As illustrated in the research outline, there were 3 attributes in the simple design: time, cost and headway (presented by service frequency); whilst two more attributes punctuality and crowding were added into the complex design.

Part 3 included some questions about personal characteristics, such as: gender, age and annual household income.

The fourth part was applied to explore respondents' decision making process and their understanding/ perceptions of the experiments. Furthermore, at the end of the questionnaire an open space was provided for respondents to add any comments which they had.

5.4 SP Experiment Design

5.4.1 Selection and presentation of alternatives

In the present research, the SP experiment was conducted in Greater Manchester, UK. Most of the trains running on the selected commuters' routes in Greater Manchester are Pacers (Class 142, 144). Improved trains, Sprinters (Class 150) and Super Sprinters (Class 153 and 156), also run on these routes, so that most travellers will have had experience of both Pacers and the improved trains.

Two types of train were chosen as the alternative in the SP experiments. Initially, Pacers (Class 142/144) and Sprinters (Class 150) were chosen as the two alternatives. After the pilot survey, it was found that Super Sprinters started to appear on the commuter route. We replaced Sprinters with Super Sprinters as the improved trains.

The 'Pacers' was originated with a project by British Rail (BR) to create a train, with low running cost, for use on rural and suburban services. At that time, BR was under increasing financial pressure from the government including proposals to cut more rail lines. BR set a challenge to several companies to design a cheap, lightweight train similar to rail buses. Since then, over 200 Pacers trains have been built, with many of them continuing to be in service 20 years later.

In Greater Manchester area, Class 142 is a common sight in service on the non-electrified heavy rail routes. It was built between 1985 and 1987. It is updated from the Class 141, and has a capacity of 106 passengers per two car set. Currently, franchisee Northern Rail Ltd. operates them on the route.

Although, Pacers are economical, there are limitations of using bus parts for railway use. The greatest complaint is the ride smoothness. Pacers use a basic four-wheel two-axle configuration, often resulting in a rough ride, especially over points and around tight curves, which has given rise to the nickname 'Nodding donkeys' due to the up and down motion on uneven track. Other performance problems include poor acceleration and poor reliability for some units. Train Operating Companies (TOCs) are researching ways of trying to replace the Pacer, although little progress has yet been made.

The Sprinter is a family of diesel multiple unit trains in use on the UK railway system. They were introduced at the same time the existing Pacers were built, but construction of Sprinters lasted until the early 1990s. Sprinters can be seen operating in almost every part of the UK, from rural branch lines to commuter expresses into London. They have conventional coaches, offering a much better ride than Pacers, and consequently are somewhat more expensive to run.

Class 150 Sprinters were built from 1985 to 1987, and feature high density 'five across' seating, with two door per coach for quick loading. Unlike the class 150s, the 156s (Super Sprinters) have a single leaf sliding door at either end of each coach - this feature reflected the anticipated longer journeys (with fewer stops) that the Class 156 was supposed to operate. Class 156 vehicles, built from 1987 to 1989, are longer than Class 150 vehicles, but have 4 across seating giving similar seating capacity.

For better understanding of the two alternatives, pictorial information was provided as shown in Figure 5.1. The inner and outer layout of the train were provided in the picture.

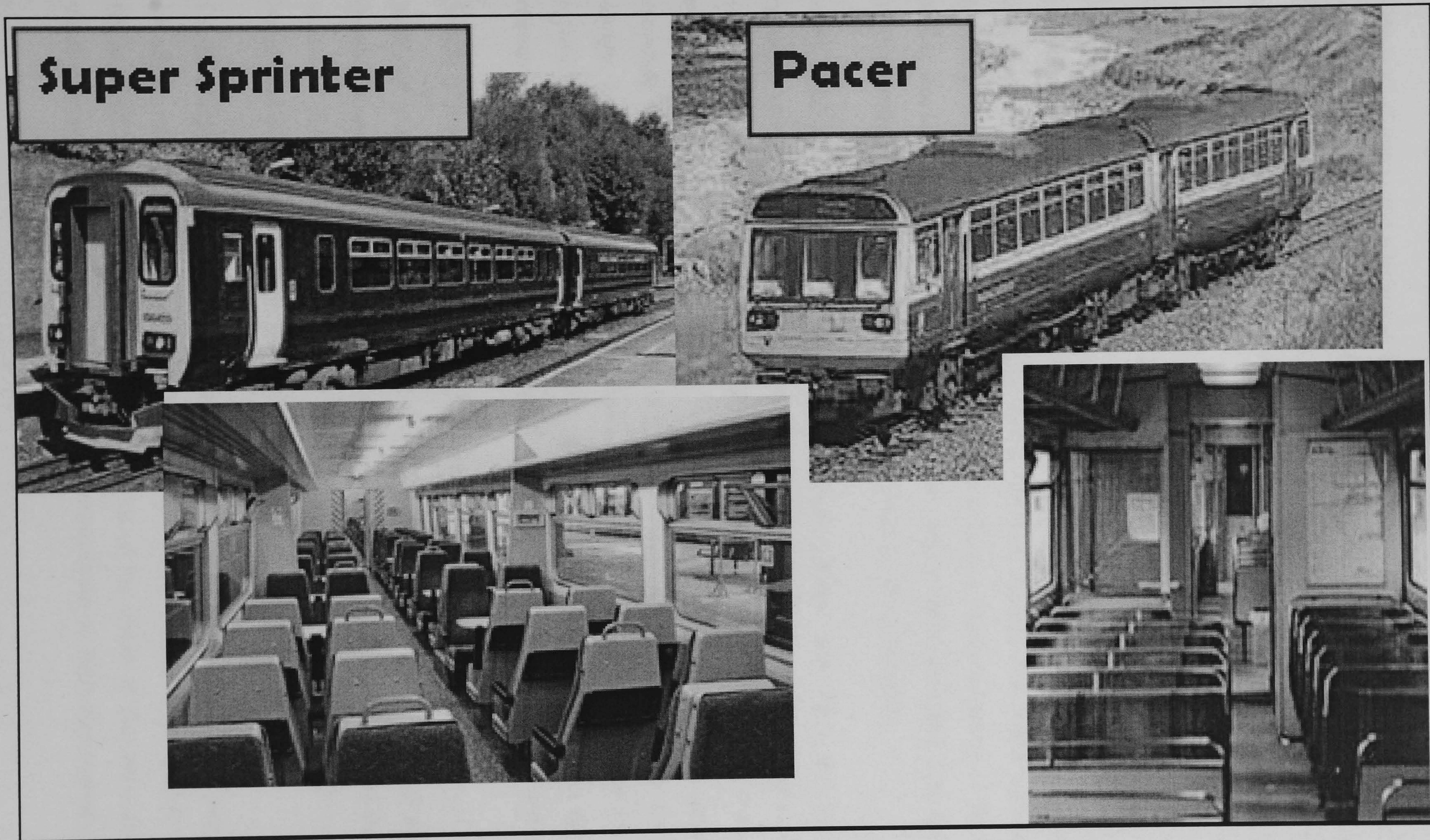


Figure 5.1 Stock types – Super Sprinters versus Pacers

Some word description was added as a supplement to the pictorial information. The development of the word description was through the pre-pilot and pilot survey, and also satisfied by the Train Operation Company. The process is described in chapter 6. In the main survey, the following word description was added in the questionnaire:

"One way of improving a train service is introducing newer trains. We would like to know how you react to such a change by presenting you with sets of fictional options. Imagine you have the choice between two types of train: Super Sprinters (New) and Pacers (Old). Super Sprinters are air-conditioned and Pacers have the alternative of opening the window. The trains' pictures are presented below."

5.4.2 Attributes in SP experiment

Based on review of previous studies on the valuation of rolling stock (see Chapter 3), a set of SP experiment was used to investigate effects of different designs on responses. Cost and journey time were obvious attributes to include. Some attributes such as cleanliness, ventilation, noise, crowding and smoothness of ride were suggested to have a strong influence on the valuation of rolling stock. Some attributes relevant to the rail service were also taken into account in the previous experience such as service frequency, reliability and punctuality.

After much thought, three common attributes were chosen: cost (fare), in-vehicle time and headway (presented by frequency). Two more service attributes were included in the complex design to investigate the impact of adding some attributes on the valuation, to see if it could mask the aim of research (valuation of rolling stock). From the review, punctuality and crowding were found to be very important for travellers, thus being added into the complex SP design. In addition, adding two attributes increases the risk of task complexity effect. The review in Section 2.7 suggested that with increase of the number of attributes, the variance of random error terms increases, which leads to the less accurate estimation or biased estimates.

5.4.3 Presentation and customization of attributes

Attribute levels should be realistic and acceptable to respondents. Levels of SP attributes were developed through two pilot surveys, which are presented in chapter 6. Variations of attribute values across scenarios were kept large enough for respondents to trade off; otherwise they may be ignored (Fowkes and Wardman, 1988).

Levels of attributes were set based on the actual journey characteristics. In the present study, the SP survey location was selected to be Greater Manchester area as currently both rolling stocks (Pacers and Super Sprinters) are running on most of the routes in this area. Journey time is usually between 20 and 50 minutes and cost is around £4-£6 for the return journey.

Table 5.2 shows the set of attribute levels for the base group. It was found to be satisfactory in the second pilot survey, and then was applied in the main survey. The base group was also used for developing the other three sets of attribute levels in the main survey (see section 5.4.5).

Table 5.2 Attributes and levels for base group SP design

Attributes	**Levels** 1	2	3
Basic attributes (SP1- 4)			
Journey time	25 30	20 25	15 20
Cost	20	50	100
Headway	-15	-5	10
Additional attributes (SP 3/4)			
Punctuality	always on time	1 out of 5 times delay for 10 minutes	2 out of 5 times delay for 10 minutes
Crowding	Enough seats	1 out of 5 times stand for the whole journey	2 out of 5 times stand for the whole journey

The second pilot survey was conducted in Oldham Mumps. An introduction of the second pilot survey is presented in section 6.3. The location was selected to be Oldham Mumps, as rolling stocks of interest were in operation; and the estimated sample size was sufficient enough to see some statistically significant effects of different SP design.

The journey time is between 16 and 25 minutes (depends on the service) from Oldham Mumps to Manchester. The service frequency is four times per hour. For the current service (Pacers), the levels of attributes were selected based on the real journey characteristics. For the improved service (Super Sprinters), the in-vehicle journey time was assumed to be slightly quicker or equal to the current service, and the single fare was higher or equal to the current service for the realistic reason. The levels for the service attributes (frequency, punctuality and crowding) for improved service were selected to be either better or worse than the current service.

In the SP design, the attribute "journey time" for both alternatives was kept absolute as shown in Table 5.2. The review in chapter 3 found that the valuation of new rolling stock is closely related to the journey time spent on the type of train. For cost and headway, differences were taken between the two alternatives. 'Journey time' is presented in minutes. 'Cost' is specified in pence. In the SP survey form, cost is valued by pounds for easy understanding. Service frequency is quantified by the time period between two sequential trains (headway), such as 'every 15 minutes'.

The presentation and measurement of crowding and punctuality were difficult. In previous relevant research (Wardman, 2004), punctuality and crowding were given as combination of probability of occurrence and length of time. Punctuality was presented as an amount of time delay with a given frequency, for example, '1 out of 5 times delay for 10 minutes'. Crowding was presented by a given frequency and standing time, such as '1 out of 5 times standing for the whole journey'. Table 5.3 shows an example of the choice in the SP survey. In all cases, Option A referred to Super Sprinters and Option B to Pacers.

Table 5.3 Example of the choice in the SP experiment

CHOICE 7	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	20 minutes	30 minutes
Single Fare	£2.00	£1.80
Frequency	Every 20 minutes	Every 10 minutes
Punctuality	2 out of 5 times delay for 10 minutes	Always on time
Crowding	Enough seats	2 out of 5 times stand for whole journey
Preference	☐	☐

5.4.4 Combination of different levels and attributes

The levels and attributes were combined by a fractional factorial design. Table 5.4 shows levels and attributes in the initial SP design for the base group.

Table 5.4 Levels of attributes for the base group SP design

	Improved Rolling Stock	**Current Rolling Stock**	
	Values	Values	Difference
Time (minute)			
Level 0	25	30	-
Level 1	20	25	-
Level 2	15	20	-
Cost (pence)			
Level 0	200	180	20
Level 1	250	200	50
Level 2	300	200	100
Headway (minute)			
Level 0	15	30	-15
Level 1	15	20	-5
Level 2	20	10	10

In SP experimental design, an orthogonal method is usually applied, which means that there is zero correlation between the different levels of explanatory variables. The attributes presented to respondents are varied independently from one another (Louviere et al., 2000). Fractional factorial SP experimental designs involve showing respondents only a subset of the full set of options. This method is useful when a full factorial design has too many scenarios due to the fact that complete factorial design includes all the combinations. The number of combinations is normally decided by the number of attributes and levels in SP design (Permain and Kroes, 1990;

Louviere et al., 2000). The fractional factorial design can reduce the number of scenarios dramatically whilst covering most of the main effects, independently from the significant interaction effects. The catalogue of experimental plans provided by Kocur et al. (1982) was applied in the present research.

A fractional factorial design was applied into each of the SP design. There were 4 attributes for each alternative with 3 levels each; therefore, 9 scenarios were generated in the simple SP experiment. 18 scenarios were generated in the complex SP design. Considering the task load, these 18 options were split to two questionnaires by choosing 9 for each. Table 5.5 is an example of fractional factorial design by applying the levels of attribute shown in Table 5.4.

Table 5.5 Combination of attributes and levels of base group SP design (band B)

Combination of SP attributes and levels (A: improved train; B: current train)						
Scenarios	Time A	Time B	Cost A	Cost B	Headway A	Headway B
	minute		pence		minute	
1	25	30	200	180	15	30
2	25	25	250	200	20	10
3	25	20	300	200	15	20
4	20	30	250	200	15	20
5	20	25	300	200	15	30
6	20	20	200	180	20	10
7	15	30	300	200	20	10
8	15	25	200	180	15	20
9	15	20	250	200	15	30

5.4.5 Development of SP design for other bands

18 rail stations were first selected for main survey data collection. The criteria to select the location were: firstly, the routes should have the old rolling stock in current operation, so that it is likely that rolling stock can be improved; secondly, according to the annual report from GMPTE, the boarding number of the station should be more than 100 in the morning peak hours. The main survey was eventually conducted in 14 rail stations as by the time, enough responses were achieved. The main survey process is discussed in detail in chapter 6.

A map of the Greater Manchester rail network is shown in Figure 5.2. Each station has its own journey characteristics in terms of in-vehicle time and service frequency. As the level of attributes in the SP experiment should be realistic and acceptable to respondents, 18 different SP designs should be generated in the experiment. To simplify the task, the 18 locations were divided into 4 different bands which share similar journey characteristics, named from A to D. Table 5.6 shows the category of four bands and their journey characteristics.

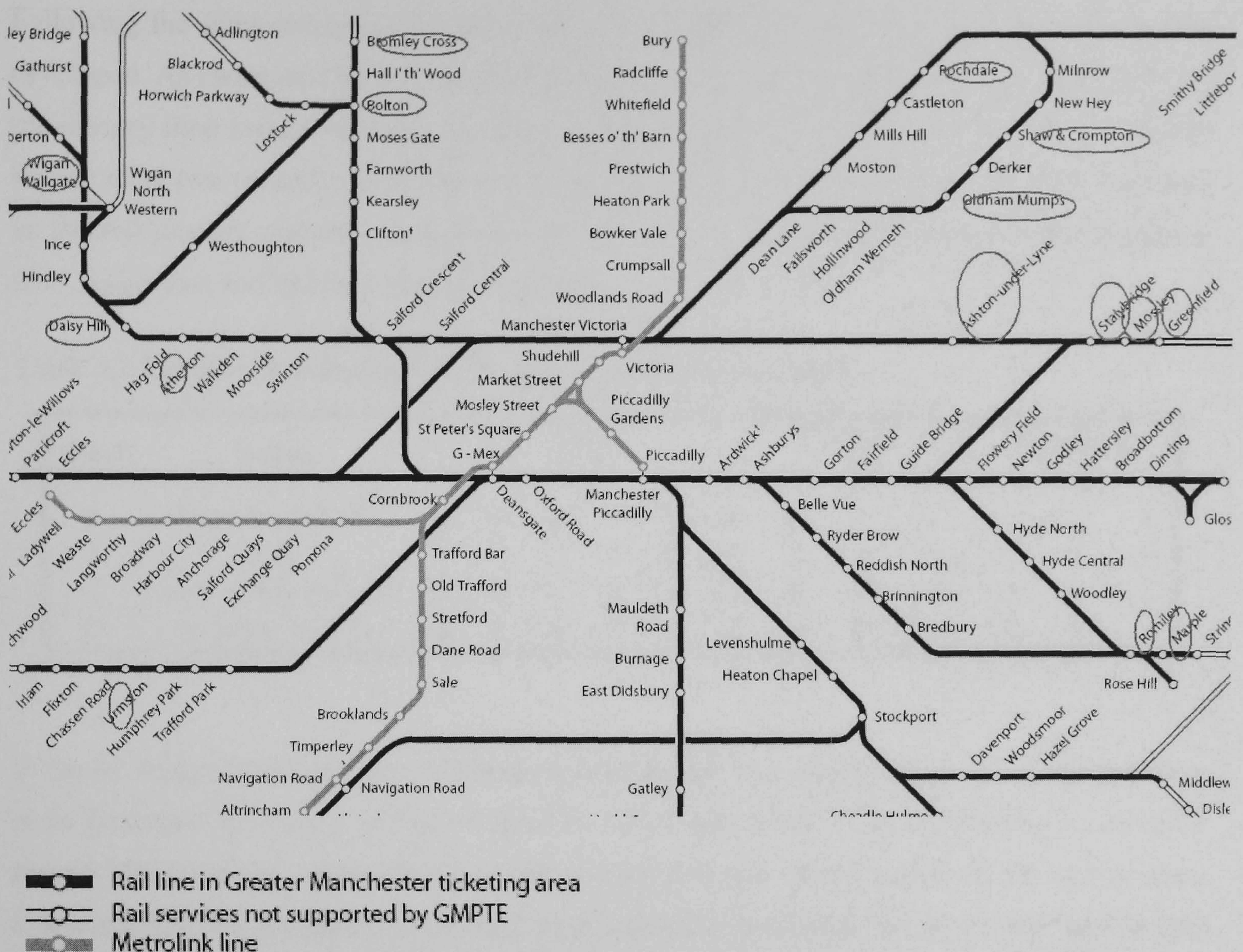


Figure 5.2 Railway map in Greater Manchester

Table 5.6 Information of stations in the main survey

Band	Journey Time	Frequency	Locations	Boarding Number
A	9-15	5/h	Ashton, Mills Hill	300
B (Base)	16-25	3/h-6/h	Rochdale, Stalybridge, Mossley, Oldham Mumps, Urmston	1600
C	26-35	2/h-6/h	Marple, Romiley, Atherton, Irlam, Greenfield, Shaw, Bromley Cross, Daisy Hill	1700
D	35-45	4/h	Hindley, Wigan Wallgate	600

Note: 'Boarding Number' is the number of travellers boarding the train in the morning peak hour; the figure is obtained from GMPTE annual report 2003.

In Table 5.6, band A has the shortest journey time, and band D has the longest journey time. The category of SP design bands also reflects the journey distance, where stations in band A have the shortest journey distance to Manchester city.

The design for band B (base group) was applied to the second pilot survey in Oldham Mumps. The estimation of the coefficients gave all correct signs with reasonable values. The design can cover the target monetary values, and also can be understood by the respondents. The SP design for the base group was shown to be suitable for the main survey.

Following the same principle, the SP design of the other three bands in the main survey was developed. As mentioned before, the levels for the journey time was kept separate to account for the journey time impact on stock valuation. Difference between the two alternatives was taken for the other two variables (cost and service frequency). The levels for attributes were set based on the real journey characteristics, as shown in Table 5.6. Table 5.7 shows the levels of journey time, single fare and headway for the four bands.

Table 5.7 Levels of attributes for different experiment locations

Band	Design	Time						Cost			Headway		
		L 1		L 2		L 3		L 1	L 2	L 3	L 1	L 2	L 3
A	Scaled to B (1/2)	8	10	10	15	15	20	10	25	50	-10	0	5
B	Base	15	20	20	25	25	30	20	50	100	-15	-5	10
C	Same difference as B	20	25	25	30	30	35	20	50	100	-15	-5	10
D	Scaled to B (5/3)	30	35	35	40	40	50	50	80	150	-20	-10	15

In the SP design for other bands, differences between the two alternatives for cost and headway were generated by scaling to that of Band B, which was found to be satisfactory in the pilot survey. The absolute values for the level of each attribute varied based on the real journey characteristics. For example, the journey time in band A is around half to that of band B (see Table 5.6); therefore, the difference of cost and headway between the two alternatives was specified as half to that of base band. The differences were rounded for better understanding by respondents. Similarly, in the development of SP design for band C, same difference between the two alternatives was used. In band D, the difference between the two alternatives was scaled to band B by (5/3).

In the complex design for each band, levels for punctuality and crowding were kept same as the base group, as shown in Table 5.2. Punctuality and crowding were presented as the combination of a given frequency and length of journey time.

The development of SP design for bands A, C and D is presented as follows. Some of the levels were adjusted to obtain a good spread of boundary ray map (see section 4.2.5). The design for bands A, B, C and D were tested through simulation tests, which are presented in section 5.5 and Appendix C. The boundary ray map showed that the SP design can capture the "target value". Simulation tests by synthetic data sets demonstrated that the coefficient estimation was significantly precise.

Development of SP design for band A

In Table 5.6, the journey distance between stations in band A and Manchester city are normally the shortest. The journey time is around 9-15 minutes and the service frequency is about 5 times per hour. Base on the actual journey characteristics, levels for attributes in band A are selected, as shown in Table 5.8.

Table 5.8 Levels of attributes for band A

	Improved Rolling Stock	Current Rolling Stock	
	Values	Values	Difference
Time (minute)			
Level 0	8	10	-
Level 1	10	15	-
Level 2	15	20	-
Cost (pence)			
Level 0	160	150	10
Level 1	225	200	25
Level 2	250	200	50
Headway (minute)			
Level 0	20	30	-10
Level 1	20	20	0
Level 2	15	10	-5

Compared to the actual journey characteristics, the journey time of stations in band A (9-15 minutes) is around half to that of band B (16-25 minutes). In the SP design, the level of journey time for the improved and current trains was kept separate. Differences of cost and headway between the two alternatives were specified as half to that of the base band (band B) and the differences were rounded for better understanding by respondents. For example, for the attribute headway, as the difference of 7.5 minutes is difficult to be perceived by respondents, the difference of 10 minutes was used instead of 7.5 minutes. The combination of attributes and levels for band A is reported in Table 5.9.

Table 5.9 Combination of attributes and levels of band A

Combination of SP attributes and levels (A: improved train; B: current train)						
Scenarios	**Time A**	**Time B**	**Cost A**	**Cost B**	**Headway A**	**Headway B**
	minute		pence		minute	
1	8	10	160	150	20	30
2	8	15	225	200	15	10
3	8	20	250	200	20	20
4	10	10	225	200	20	20
5	10	15	250	200	20	30
6	10	20	160	150	15	10
7	15	10	**200**	**250**	15	10
8	15	15	160	150	20	20
9	15	20	225	200	20	30

Development of SP design for band C

For the stations in band C, the in-vehicle time is normally 26 to 35 minutes; and the frequency of service is about 2 to 6 times per hour. The journey characteristics of stations in band C are similar to that of band B; therefore, differences of cost and headway between the two alternatives in band C were kept same as band B. Absolute values of attribute levels varied by the actual journey characteristics in this band. Table 5.10 and Table 5.11 present the levels of attributes for band C and the combination of attributes and levels.

Table 5.10 Levels of attributes for band C

	Improved Rolling Stock	**Current Rolling Stock**	
	Values	Values	Difference
Time (minute)			
Level 0	20	25	-
Level 1	25	30	-
Level 2	30	35	-
Cost (pence)			
Level 0	200	180	20
Level 1	250	200	50
Level 2	300	200	100
Headway (minute)			
Level 0	15	30	-15
Level 1	15	20	-5
Level 2	20	10	10

Table 5.11 Combination of attributes and levels of band C

Combination of SP attributes and levels (A: improved train; B: current train)						
Scenarios	**Time A**	**Time B**	**Cost A**	**Cost B**	**Headway A**	**Headway B**
	minute		pence		minute	
1	20	25	200	180	15	30
2	20	30	250	200	20	10
3	20	35	300	200	15	20
4	25	25	**200**	200	15	20
5	25	30	300	200	15	30
6	25	35	200	180	20	10
7	30	25	**200**	**250**	20	10
8	30	30	200	180	15	20
9	30	35	250	200	15	30

<u>Development of SP design for band D</u>

The actual journey time for stations in band D is 35-45 minutes with 4 times per hour service frequency. The journey time in band D is about (5/3) times of that of band B. In the development of SP design, differences of cost and headway between the two alternatives in band D was generated by scaling to that of band B by a factor of 5/3. Rounded number was taken for value of levels for better understanding by respondents.

Table 5.12 Levels of attributes for band D

	Improved Rolling Stock	Current Rolling Stock	
	Values	Values	Difference
Time (minute)			
Level 0	30	35	-
Level 1	35	40	-
Level 2	40	50	-
Cost (pence)			
Level 0	250	200	50
Level 1	330	250	80
Level 2	450	300	150
Headway (minute)			
Level 0	10	30	-20
Level 1	10	20	-10
Level 2	30	15	15

Table 5.13 Combination of attributes and levels of band D

Combination of SP attributes and levels (A: improved train; B: current train)						
Scenarios	**Time A**	**Time B**	**Cost A**	**Cost B**	**Headway A**	**Headway B**
	minute		Pence		minute	
1	30	35	250	200	10	30
2	30	40	330	250	30	15
3	30	50	450	300	10	20
4	35	35	330	250	10	20
5	35	40	450	300	10	30
6	35	50	250	200	30	15
7	40	35	**300**	**450**	30	15
8	40	40	250	200	10	20
9	40	50	330	250	10	30

5.5 Simulation and Improvement of SP Design

SP experimental designs may be tested by a simulation test (Fowkes and Wardman, 1988). This would allow the design to be improved in specifying magnitude of attribute levels and combining the levels, before the data collection. This section reports the simulation of SP design, through boundary ray maps and synthetic data analysis.

5.5.1 Boundary ray map (Bin analysis)

Boundary ray maps were constructed, and simulations run, as part of tests to see if the design can cover the target attribute valuations sufficiently accurate (Fowkes, 1991). By adjusting the boundary rays generated by different SP scenario, the interaction of boundary rays would be beneficial (Holden, 1992). Section 4.2.5 explained how to construct the boundary ray map.

Given the value of headway, a boundary ray for each hypothetical scenario can be obtained. The boundary ray map for band B is presented as an example. The initial SP design of band B is reported in Table 5.5 (Section 5.4.4). The boundary ray map generated by initial SP design is shown in Figure 5.3. Each ray represents a scenario. The area achieved by the rays is expected to cover the accepted values (true value).

In Figure 5.3, some of the boundary rays are not satisfying, negative value of time (VoT) was obtained from the initial design such as Scenario 1 and 3. The SP design was then adjusted (Fowkes, 1998). Table 5.14 reports the improved SP design for band B. Figure 5.4 demonstrates the boundary ray map for the adjusted design. Comparing Figure 5.3 and 5.4, the interaction area of the improved SP design is better in the way that it covers the target values better than the initial design. The design was shown to be suitable for the main survey.

Table 5.14 Improved design of SP experiment (band B)

Combination of SP attributes and levels (A: new train; B: old train)						
Scenarios	**Time A**	**Time B**	**Cost A**	**Cost B**	**Headway A**	**Headway B**
	minute		pence		minute	
1	25	30	**300**	180	15	30
2	25	25	**200**	**250**	20	10
3	25	20	**220**	200	15	20
4	20	30	250	200	15	20
5	20	25	300	200	15	30
6	20	20	200	180	20	10
7	15	30	300	200	20	10
8	15	25	200	180	15	20
9	15	20	250	200	15	30

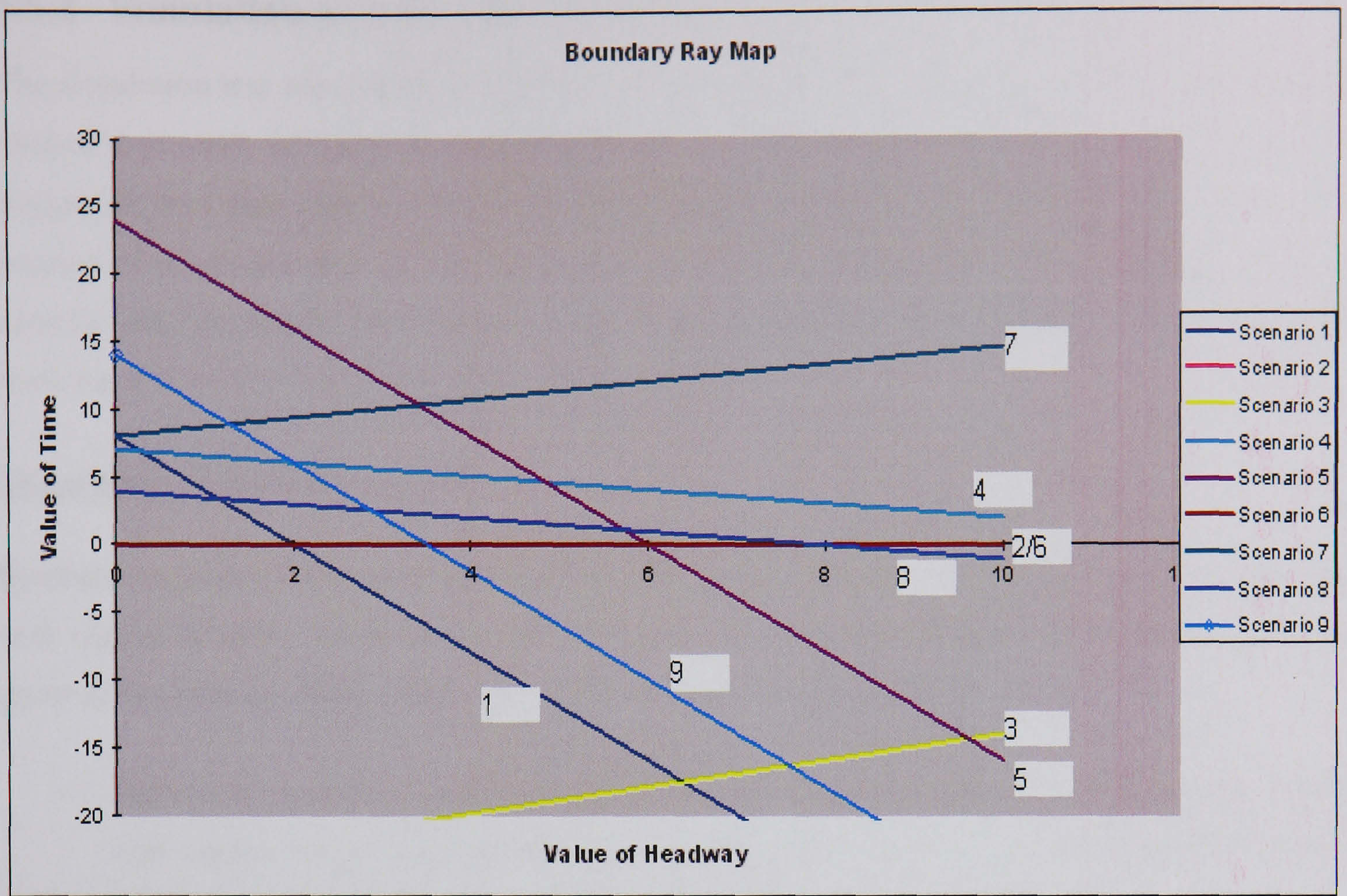


Figure 5.3 Boundary ray map for the initial SP design (band B)

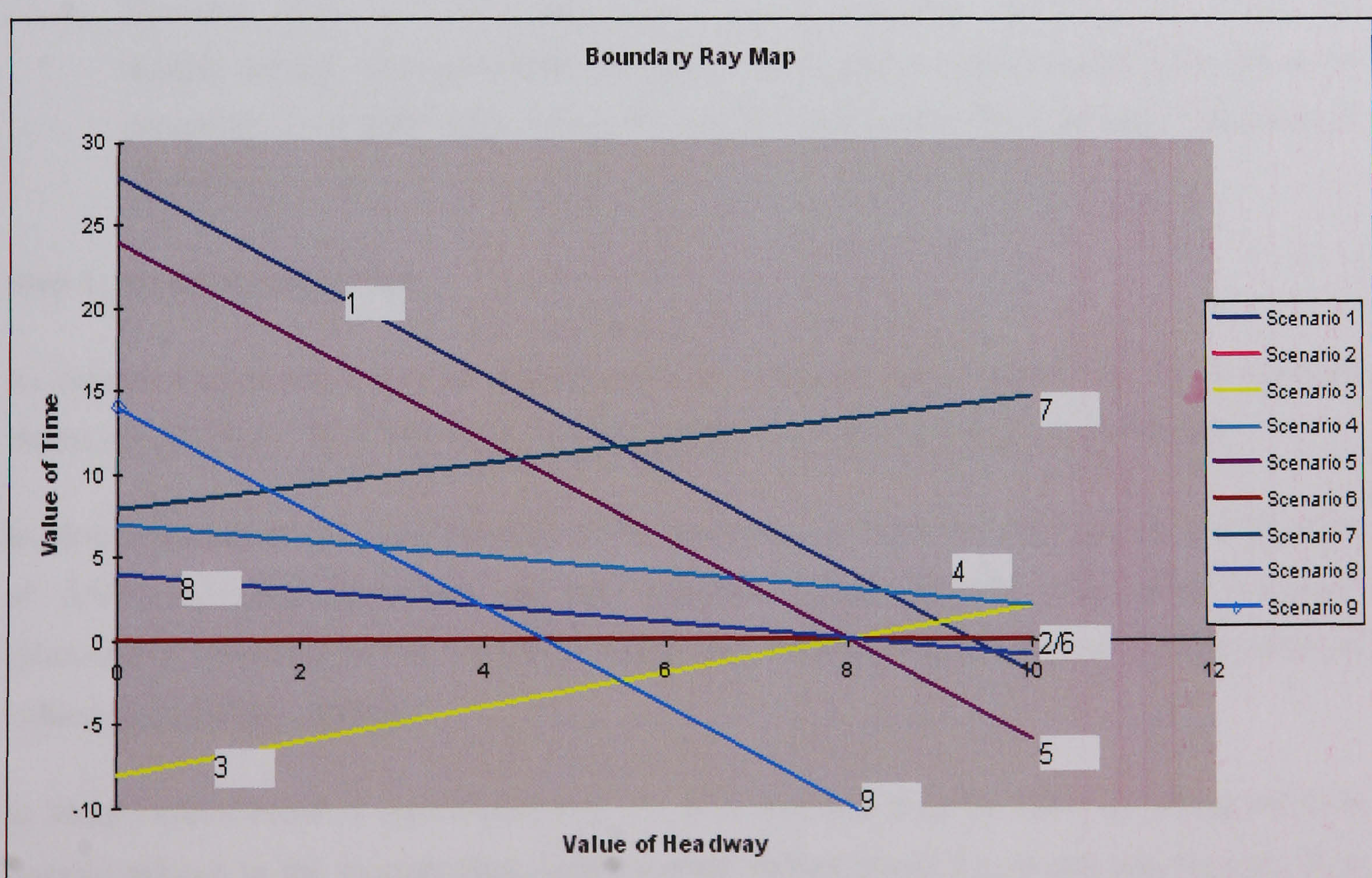


Figure 5.4 Boundary ray map of the improved SP Design (band B)

5.5.2 Simulation tests by synthetic data

The simulation test was conducted for four bands by using the synthetic data. It was to test the statistical property of the SP design, and also to examine if the standard errors and t-ratio were acceptable and that they could not be improved substantially by changing the design. The process of simulation tests by synthetic data was discussed in section 4.2.5. Simulation tests for band B (both the simple and complex design) are shown here. The simulation tests for the other three bands are shown in Appendix C.

Generation of synthetic data sets

Synthetic responses were produced based on the utility maximisation assumption in that option i with respect to option j will be chosen if the utility of option i is higher. In the present study, the generation of the synthetic data sets contains the following three parts:

- Following the well established method of Fowkes and Wardman (1988), the logit model was chosen to simulate and estimate the responses, which was a function of attribute levels of the design, attribute valuations (coefficients) and an error term.
- Each coefficient was based on the results from actual observations (i.e. consistent with past studies).
- The error terms were estimated from $\varepsilon = Ln[-Ln(random \quad number)]/\lambda$; where the random number was generated between 0 and 1, and λ was determined as the scale parameter $\lambda = \pi/(\sqrt{6} \cdot SD)$, where SD was the assumed level of standard deviation of the error ε_i.

Step 1: Model Specification

As explained in section 4.4.1, the valuation of new rolling stock can be analysed by two types of model specification. Thus, two model specifications were applied in the simulation test.

In Model specification 1 (see Equation 4.30), the utility of improved rolling stock is a function of ASC (indicating preference for the improved trains), journey time, cost, headway, (punctuality, crowding in the complex design) and error term. The improved rolling stock is valued by pence per journey.

In Model specification 2 (see Equation 4.31), it is assumed that the value of rolling stock is directly related to the journey time in the certain rolling stock (i.e. pence per minute). It is expected that people in a better service would have lower value of time, as they would be willing to pay more for one minute saving in a 'less comfortable' service.

Step 2: Selection of the input values in the generation of the synthetic data

The input value for each coefficient is selected based on the review from previous studies. The cost coefficient was set to '-1', and then the other coefficients were equal to the monetary values observed from the previous studies.

From the review of urban rail travel, the average value of time (VoT) for Commuting, Leisure and Business are 7.2, 6.3, 19.2 pence/minute respectively (Wardman, 2004). In the meta-analysis of service attributes, Wardman (2004) found that the frequency is valued at 0.8 minute of in-vehicle time with a narrow confidence interval of $\pm 0.8\%$. From the convention in the railway industry in Great Britain, the time value of headway in the range 0.4-0.7 (ATOC, 1997) has been applied in using. The review by Wardman and Whelan (2001) of the rolling stock studies found that in the previous SP studies, on average, the improved rolling stock has a value equivalent to 21% of fare for business travellers and 12% for leisure travellers.

In the present study, punctuality was presented as the combination of given frequency and the length of time (late time in relation to the timetable). In the data analysis, the 'punctuality' was quantified by the expected value (Wardman, 2004) which is obtained by the delay time multiplied by the given frequency. For example, when the level for punctuality is '1 out of 5 times delay for 10 minutes', the expected value for the 'punctuality' is $10 \times (1/5)$ which is 2 minutes. Wardman (2001, p.112) conducted a meta-analysis on 14 cases where the expected value was used, and obtained 26.79 (p/min) for the average value of punctuality.

Crowding was quantified by the expected value in the study. Wardman (2003) conducted a meta-analysis on the valuation of standing time, which covered 23 valuations from 8 studies. It was found that the mean value of standing time relative to seated time in the 20 instances was 2.7 – 3.5, depending on different types of SP survey. PDFH (2005, Table B5.1) recommended that the value of standing time for commuter is around 10 pence per minute.

The selection of input values is based on the results from the above mentioned relevant studies. From the review, individuals' heterogeneity, (for instance, journey purpose and income) may have an impact on the variation of the valuation. In simulation tests, the input value varies for each test to cover certain areas to take account of the taste variation among respondents.

Step 3: The error term in the utility function

In this study, the data has been analysed by MNL model and HMNL model (see chapters 7 and 8). The error term is assumed to be Gumbel distributed. The standard deviation was selected in the estimation of the random error to obtain the goodness-of-fit index ($\rho^2(C)$) close to 0.1-0.2, which is normally assumed to be good in the choice models (Louviere et al., 2000, p.55).

Creation of the synthetic data sets was conducted by the program source code written in FORTRAN. An example of the generation of the data sets for the band B complex design is attached in Appendix C. This includes the program file (appendix C-2) to create the responses, and the examples of created responses (appendix C-3).

The simulation tests for each SP design were based on 100 respondents initially, which generated 900 preference observations. The number of respondents was decided to obtain a sufficient sample size for statistical significant estimation. These simulation responses were analysed and then estimated coefficients were compared with the input values.

Simulation results

The synthetic data were analysed by ALOGIT program (HCG, 2000). The estimated results were compared with those of the input values. Table 5.15 and 5.16 demonstrate an example of the simulation tests process for band B. The simulations tests were conducted for both simple and complex design, by applying model specification I and II.

Different sets of valuations were specified for each set of design: (VoT, VoH, VoS, VoP, VoC); for example (8, 5, 20, 20, 10) and (10, 8, 20, 20, 10). Two runs for each design were conducted. The mean of relative values yielded from the two runs was calculated and compared with the input values which were selected from previous studies. The coefficient estimation was correct sign with a significant t – ratio ($t>1.96$). The estimated values from the synthetic data were not statistically different from input values.

The simulation results showed that SP designs could capture a reasonable range of relative valuations and were proved to be statistically satisfactory for data collection and the intended analysis.

The bias has been simulated by specifying differences of monetary values of time and rolling stock in the SP design, and simulating to check that such differences can be detected as significant differences in estimated model coefficients. From the simulation test, the SP design used in the data collection can well detect the difference.

MNL was used during the design process as it was satisfactory for the task. The assumption made, regarding distribution, need not to be the same at the simulation stage as at the estimation stage, where the statistically 'best' distribution can be identified by suitable analysis.

Table 5.15 Simulation results from SP design for base group (band B) – Model Specification 1

Input Values			Estimated Coefficients								Estimated Values				
VoT	VoH	VoS	Time	(t)	Cost	(t)	Headway	(t)	ASC	(t)	VoT	VoH	VoS	ρ^2(c)	SD
8	5	20	-0.1634	(-9.2)	-0.0205	(-8.8)	-0.0993	(-10.2)	0.5306	(5.0)	7.979	4.846	25.908	0.1330	60
8	5	20	-0.1774	(-9.9)	-0.0210	(-8.8)	-0.1003	(-10.2)	0.4906	(4.6)	8.468	4.788	23.418	0.1423	60
											8.223	**4.817**	**24.663**		
5	3	20	-0.1317	(-8.2)	-0.0251	(-11.4)	-0.0693	(-7.8)	0.5436	(5.1)	5.251	2.764	21.675	0.1227	50
5	3	20	-0.1373	(-8.5)	-0.0255	(-11.7)	-0.0759	(-8.4)	0.3733	(3.5)	5.376	2.972	14.616	0.1304	50
											5.314	**2.868**	**18.145**		
6	4	10	-0.162	(-9.6)	-0.0269	(-12.0)	-0.1048	(-10.8)	0.1969	(1.9)	6.034	3.903	7.333	0.1363	60
6	4	10	-0.1355	(-8.4)	-0.0230	(-10.6)	-0.0865	(-9.4)	0.3437	(3.3)	5.902	3.768	14.970	0.1197	60
											5.968	**3.836**	**11.151**		

Note: The value of the improved rolling stock transferred to be positive

Input Values					Estimated Coefficients						Estimated Values						
VoT	VoH	VoS	VoP	VoC	Time	Cost	Headway	ASC	Punctuality	Crowding	VoT	VoH	VoS	VoP	VoC	ρ^2(c)	SD
6	4	20	20	10	-0.0808 (-7.6)	-0.0120 (-7.9)	-0.0629 (-9.5)	0.2046 (2.5)	-0.2779 (-12.1)	-0.1238 (-12.4)	6.72	5.23	17.02	23.12	10.30	0.1662	100
6	4	20	20	10	-0.0915 (-8.6)	-0.0129 (-8.6)	-0.0518 (-8.0)	0.1673 (2.0)	-0.2469 (-10.9)	-0.1282 (-13.1)	7.11	4.03	13.00	19.18	9.96	0.1679	100
											6.92	**4.63**	**15.01**	**21.15**	**10.13**		
8	5	20	20	10	-0.1504 (-12.4)	-0.0201 (-11.7)	-0.1010 (-12.3)	0.4215 (4.6)	-0.3465 (-12.5)	-0.1617 (-13.1)	7.48	5.02	20.97	17.24	8.04	0.1533	120
8	5	20	20	10	-0.1104 (-10.1)	-0.0145 (-9.2)	-0.0569 (-8.4)	0.3233 (3.8)	-0.3392 (-14.1)	-0.1134 (-11.0)	7.62	3.93	22.31	23.41	7.83	0.1854	120
											7.55	**4.48**	**21.64**	**20.32**	**7.94**		
10	8	20	20	10	-0.1139 (-9.7)	-0.0099 (-6.2)	-0.0893 (-12.1)	0.2140 (2.5)	-0.2199 (-9.4)	-0.1116 (-10.7)	11.56	9.06	21.71	22.31	11.32	0.1847	120
10	8	20	20	10	-0.0960 (-8.6)	-0.0109 (-7.1)	-0.0820 (-11.9)	0.3175 (3.8)	-0.2266 (-10.1)	-0.0926 (-9.5)	8.80	7.51	29.10	20.77	8.49	0.1531	120
											10.18	**8.29**	**25.41**	**21.54**	**9.90**		

Table 5.16 Simulation results from SP design for base group (band B) – Model Specification 2

Input Values			Estimated Coefficients								Estimated Values				
VoTA	VoH	VoTB	TimeA	(t)	Cost	(t)	Headway	(t)	TimeB	(t)	VoTA	VoH	VoTB	ρ^2 (c)	SD
7	5	9	-0.1628	(-8.3)	-0.0228	(-8.4)	-0.1072	(-10.1)	-0.2085	(-10.2)	7.128	4.694	9.129	0.1482	60
7	5	9	-0.1363	(-7.2)	-0.0202	(-7.9)	-0.1130	(-11.0)	-0.1761	(-9.2)	6.738	5.586	8.705	0.1502	60
											6.933	**5.140**	**8.917**		
5	3	7	-0.0962	(-5.5)	-0.0246	(-9.8)	-0.0761	(-8.2)	-0.1498	(-8.4)	3.913	3.096	6.094	0.1025	50
5	3	7	-0.1456	(-8.0)	-0.0271	(-10.5)	-0.0814	(-8.6)	-0.1924	(-10.4)	5.365	2.998	7.089	0.1328	50
											4.639	**3.047**	**6.592**		
12	7	15	-0.1483	(-7.5)	-0.0118	(-4.4)	-0.0916	(-8.3)	-0.1871	(-9.3)	12.621	7.797	15.923	0.1549	100
12	7	15	-0.1407	(-7.1)	-0.0142	(-5.3)	-0.0971	(-9.0)	-0.1831	(-9.1)	9.901	6.830	12.885	0.1397	100
											11.261	**7.314**	**14.404**		
8	5	8	-0.1903	(-10.3)	-0.0239	(-10.2)	-0.1235	(-12.2)	-0.1938	(-10.8)	7.956	5.163	8.102	0.1862	60
8	5	8	-0.1412	(-8.4)	-0.0188	(-8.6)	-0.0981	(-10.5)	-0.1421	(-8.9)	7.523	5.226	7.571	0.1252	60
											7.739	**5.195**	**7.836**		

Input Values					Estimated Coefficients						Estimated Values						
VoTA	VoH	VoTB	VoP	VoC	TimeA	Cost	Headway	TimeB	Punctuality	Crowding	VoTA	VoH	VoTB	VoP	VoC	ρ^2 (c)	SD
7	5	9	20	10	-0.0667 (-5.9)	-0.0094 (-5.9)	-0.0541 (-7.8)	-0.0912 (-8.6)	-0.2489 (-10.8)	-0.1228 (-11.6)	7.11	5.77	9.72	26.55	13.10	0.1509	100
7	5	9	20	10	-0.0958 (-8.2)	-0.0144 (-8.9)	-0.0684 (-9.6)	-0.1235 (-11.2)	-0.2605 (-11.2)	-0.1068 (-10.2)	6.66	4.75	8.59	18.12	7.43	0.1598	100
											6.89	**5.26**	**9.16**	**22.33**	**10.26**		
5	3	7	30	10	-0.0644 (-5.8)	-0.0119 (-7.3)	-0.0319 (-4.8)	-0.0876 (-8.5)	-0.3539 (-14.8)	-0.1208 (-11.6)	5.43	2.69	7.39	29.84	10.19	0.1792	100
5	3	7	30	10	-0.0600 (-5.3)	-0.0117 (-6.9)	-0.0349 (-5.0)	-0.0849 (-8.1)	-0.4031 (-15.7)	-0.1381 (-12.3)	5.13	2.98	7.25	34.42	11.79	0.2080	100
											5.28	**2.84**	**7.32**	**32.13**	**10.99**		
12	7	15	20	10	-0.1266 (-9.5)	-0.0104 (-5.8)	-0.0745 (-9.0)	-0.1595 (-12.3)	-0.2052 (-8.3)	-0.1052 (-9.0)	12.16	7.16	15.32	19.71	10.11	0.1840	120
12	7	15	20	10	-0.1380 (-10.0)	-0.0120 (-6.5)	-0.0799 (-9.2)	-0.1734 (-12.8)	-0.2186 (-8.5)	-0.1215 (-9.8)	11.48	6.64	14.43	18.19	10.11	0.2063	120
											11.82	**6.90**	**14.87**	**18.95**	**10.11**		

5.5.3 Pilot survey

Sometimes the simulation test cannot guarantee the design will have no problems, particularly where there is a lack of previous information about magnitudes and ratios of coefficients in the study (Tudela, 2000). In addition, the simulation cannot test whether individuals find the exercise realistic, in this sense, a pilot survey is suggested to test the design, and also find how individuals respond to the survey as a whole (e.g. format, questioning, presentation, survey conducting and response rate). Two pilot surveys were conducted for the above purpose, which are explained in detail in chapter 6.

5.6 Development of Cheap Talk (CT)

5.6.1 Introduction of Cheap Talk

A cheap-talk script was added into some of the SP experiments. Cheap-talk draws lessons from experimental economics and psychology concerning the design of valuation institutions. Cummings and Taylor (1999) successful employed "cheap-talk" (hereafter, CT) in the Contingent Valuation (CV) study to reduce hypothetical bias and found that the CT script was robust via different goods. This was the first successfully application of CT script in reducing the hypothetical bias. The CT script they applied in the survey was a script explicitly describing the hypothetical bias problem and asking respondents to avoid overstating their true willingness to pay (WTP).

The review of previous studies in section 2.6 found that an effective CT script contains an explicit explanation of the previous hypothetical bias (its magnitude and direction) and budgetary constraints and substitutes (Cummings and Taylor, 1999; List, 2000). The review also found that attention should be paid to the wording of CT script to avoid over-correcting the bias or introducing new bias.

The effectiveness of CT script is sensitive to the length and content. Short version CT shows mixed evidence of effectiveness: in some experiments, the short version CT cannot eliminate the hypothetical bias, but exacerbate the problems (Cummings et al., 1995; Poe, 2002; Aadland and Caplan, 2003). Some recent experience in Stated Choice experiments found that short version cheap talk can reduce the hypothetical bias in the response (Carlsson et al., 2005; List et al., 2006). Most of the previous studies were conducted by in-house or personal interview, only few by telephone interview (Aadland and Caplan, 2003) and mail surveys (Carlsson et al., 2004; List et al., 2006).

In the present study, considering the time and budget constraints, paper-based mail-back questionnaire was chosen for data collection. A long version CT script is not applicable in the limit of space. Therefore, a short version of CT script is developed for research hypotheses testing.

5.6.2 Presentation of previous bias in Cheap Talk script

In an effective CT script, warning message is an essential element. The present experimental context is the valuation of the improved rolling stock.

Review (see Chapter 3) of previous studies on the valuation of rolling stock has found that monetary values of rolling stock from SP studies are much higher than that from the other evidence (such as: revealed preference data or demand analysis using ticket sales data) when the issue of the survey is to introduce a new rolling stock or refurbishment. This can be explained as the strategic bias by which respondents overestimate new stock values to increase the chance that new stock can be introduced (Wardman and Whelan 2001).

The CT script in this study explains the overestimation of improved stock values observed from the review of previous studies.

5.6.3 Development of the Cheap Talk script

Through pilot surveys, the CT script had been revised and refined (see chapter 6), so that the wording could be understood by respondents and did not cause clear bias. The CT script which was applied in the main survey is presented in the following:

> "Previous surveys have sometimes found that people say they would be happy to pay extra for improved trains but when the fare is raised and the improved trains are provided, people say they would prefer the cheaper fare with the old trains.
>
> Bearing this in mind, as you read through the following choices, please imagine you will actually have to pay the fare stated."

5.7 Presentation of Complex Design

5.7.1 Masking the aim of research

This study tests the incentives for individuals to strategic bias in the SP survey. Besides the introduction of a CT script in the SP experiment; the impact of introducing more attributes to mask the aim of survey is being explored.

In the SP experiment, respondents perceive that their responses have impacts on the provision of the new good, thus having an incentive to bias their answers for certain better outcome. By adding complexity to the SP task, it may be hoped that respondents will exhibit less bias. This may partly be due to the extra effort required merely to complete the exercise without bias, but it is more likely to be due to respondents failing to see any single clear purpose to the exercise.

This method is motivated by Wardman and Bristow (2003)'s successful empirical evidence to amend individuals' incentives to strategic bias. The details of the study are presented in section 2.5.5. In their experiment, large differences were found in valuations of aircraft movement between an SP exercise which aimed to conceal the purpose of the survey by placing aircraft movements alongside a wide range of quality of life variables and a more conventional SP exercise which offered trade-offs between local taxes and aircraft noise where the purpose of the study would have been obvious.

5.7.2 Task complexity effects

The review of decision behaviour (in section 2.7) has found that respondents are subjected to certain cognitive capability in making the choice. Many studies have detected impacts of task complexity on SP responses, which were discussed in the literature review (section 2.7). It was found that the number of attributes per alternative has a significant impact on respondents' choice making in the SP experiments. In the empirical evidence, the number of attributes per alternative was normally varied between 3 and 6. With the increase of the number of attributes, the level of error variance increases which indicates the lower precision of estimation (DeShazo and Fermo, 2002; Caussade et al., 2005). Some research has detected the variation of valuation by adding the number of attributes to the SP experiment (Hensher et al., 2007).

In this study, the context of SP experiment is the valuation of improved rolling stock. In-vehicle time, fare and service frequency (headway) were suggested from the review to have strong influence on respondents' choice making. These three attributes[1] were chosen in the experiment.

From the review of previous relevant studies, service attributes such as 'Punctuality' and 'Crowding' were suggested to have strong impact on the travellers' decision making. These two attributes were added to some of the SP experiments to test both impacts: 1. if adding of these two attributes can distract respondents' attention from valuation of rolling stock (mask the research aim); and 2. if adding two more attributes could cause task complexity effect as suggested by previous research. The impact of adding two more attributes on both the estimation precision and valuation variation are examined.

[1] From the SP design point of view, we said that there were four attributes (see section 5.4.4). This is because that journey time for both alternatives was kept separate to capture the impact of journey time on the valuation of improved rolling stock. However, in the SP exercise, respondents need to consider/trade-off three attributes in their choice making which are journey time, cost and headway.

The presentation of 'Punctuality' and 'Crowding' and levels for these two attributes are presented in section 5.4.3. The simulation tests of complex design by using synthetic data sets are presented in section 5.5.2.

5.8 Presentation of Perception Question

Respondents make their decision based on their perceptions of the SP experiment (Powe et al., 2005). To probe respondents' choice making process and to gain a better understanding of the influence of perceptions on the SP responses, three follow-up questions were added into the survey. Respondents' opinions of task complexity, familiarity of the alternatives and the perception of potential price increase by the introduction of the new trains were explored.

5.8.1 Difficulty of making choice

The SP experiment tests whether adding some attributes to the survey can make the experiment less transparent. In the suite of SP experiments, some questionnaires contained five attributes as explained in the previous section, whilst some of the questionnaires included three attributes. One question exploring the relationship between the perceived task complexity and the number of attributes in the choice task was provided as follows:

"Did you feel it difficult in making the choices?"

Four options were provided to respondents as 'Yes, very', 'Yes, quite', 'Yes, a little' and 'No'.

5.8.2 Familiarity of experiment subject

The familiarity of the experiment subject is expected to have an important effect on respondents' decision making. In Benshoof (1970) study of motorists, evidence showed that the motorists did not accurately measure different route characteristics. This occurred when the respondents had not experienced the route. Wardman and Whelan (2001) have found that the variation of stock valuation can be explained by individuals' familiarity with the stock types (see section 3.4.2). A meta-analysis was conducted on previous values. It was found that if respondents were familiar with the rolling stocks that the survey presented, they gave them lower stock valuation. Wardman and Whelan concluded that unfamiliarity with improved levels of attributes would result the overestimation of the coefficients.

In the present research, the familiarity of stock types was measured to see if respondents can distinguish Super Sprinters from Pacers with the help of word description and pictorial information. The influence of familiarity of stock types was explored using the follow question:

'Did you feel confident that you could distinguish Super Sprinters from Pacers with the help of information provided?'

Four options provided to the respondents were 'Not at all', 'Not sure', 'Fairly' and 'Very' in the initial pilot survey.

From the pilot survey, it was found that to put the option 'Not sure' between 'Not at all' and 'Fairly' is confusing as the latter two terms represents the ability of distinguishing the difference between rolling stocks increase. 'Not sure' is different from those two. In the main survey, 'Not sure' was left at the last of the options.

5.8.3 Perception on the potential price increase

The review of incentive to strategic bias in section 2.5 found that the payment has a strong impact on individuals' choice making (Bohm, 1971; Throsby and Withers, 1986; Carson et al., 2000). More specifically, incentive compatible can be achieved when respondent feels that the payment is compulsory (assumed other criteria are satisfied). In this situation, respondents would give their true WTP towards the valuation of new product, rather than giving the upward bias. However, if respondents believe the payment is voluntary, they have a strong incentive to strategically bias their WTP to increase the possibility of the provision of the new product.

If respondents perceived that the price is the variable most likely to vary in the new policy, reduction of the price or cost is preferred rather than increase. This can be explained by "public preferring the cheapest option, although this could stem from a low value of time" (Wardman, 1987). Wardman (2001) conducted a meta-analysis on values of time from SP studies and found that there is a significant effect on the value of time if toll is the numeraire compared to other numeraires. When using a toll or road charge coefficient to calculate the value of time, it is found to be 19% lower than otherwise. This can be regarded as the strategic bias.

From the above theoretical and empirical evidence, respondents' perceptions of potential price increase due to the introduction of improved rolling stock are expected to have some impact on their decision making. Question probing their perceptions of price increase is given as follows:

> 'How likely do you think it is that fares would increase if newer trains would be introduced?'

Four options from 'Not at all' to 'Very' were provided to respondents.

This question also helps to test the impact of the cheap-talk on respondents' perceptions of cost change. As explained before, CT was introduced in some experiments to test if it can amend respondents' incentive to overestimate the values of the improved rolling stock (Section 5.6).

5.9 Conclusions on the SP Experiment Design

This chapter has presented the SP survey design, following the design process discussed in chapter 4. The questionnaire of the survey included four parts. The first part was designed to gather basic journey information of respondents. The second part was a traditional SP experiment design. Fractional factorial design was applied to the basic design. Word and pictorial information of the alternatives were provided before the SP choices. The third part was to gather social economic information of respondents. The fourth part contained three follow-up questions for probing respondents' decision making process and their perceptions of the survey.

To test research hypotheses, a suite of SP experiments were developed. More specifically, cheap-talk and task complexity (to mask research aim) were introduced into the SP experiments. Cheap-talk is a warning message of previous bias and reminding of the budget constraints. Two more attributes were added into the SP experiments to identify the incentives for individuals to strategic bias their answer, and to examine the impact of task complexity on SP responses.

Prior to the main data collection, the design and survey was tested by simulation and pilot surveys. The questionnaires have been revised and refined through two pilot surveys. The final pilot survey was found to be easily understood by respondents and capable to test research hypotheses. This process is discussed in detail in chapter 6. The main survey data collection and data description are presented in chapter 6.

Chapter 6

Data Collection and Description

6.1 Introduction

The main data collection was carried out in Greater Manchester from October to December 2005 using paper-based self – completion questionnaires. Prior to the main data collection, pilot surveys were conducted in April and August 2005 respectively. In this chapter, the process of the data collection is described.

Paper-based questionnaires were designed and developed which were tested by simulation tests and two pilot surveys. The first pilot survey was undertaken in the Institute for Transport Studies (University of Leeds) and outside of the University to investigate the presentation and measurement of SP design, and also the factors affecting respondents' choice making. The second pilot survey was carried out in August 2005 to test the presentation, levels and attributes of SP design. In addition, two methods, namely cheap-talk and task complexity (adding more attributes to mask the survey aim), were examined. The final version of the SP questionnaire was used for the main survey conducted in Greater Manchester.

Section 6.2 demonstrates the process of a first pilot survey. After the first pilot survey, the SP design and questionnaire were improved. The second pilot survey was conducted in Oldham Mumps in August 2005. Section 6.3 describes the second pilot survey and implications for the future main survey. Section 6.4 presents the main survey process, followed by the description of the sample characteristics in section 6.5. Section 6.6 ends the chapter with the conclusion of data collection and description.

6.2 First Pilot Survey

The first pilot survey was conducted in April 2005. It was a simple pre-stage of the second pilot survey. The main purposes of the first pilot study were as follows:

- to test the presentation of SP attributes and levels for simple SP design;
- to test if respondents can understand the questionnaire;
- whether or not they could perceive the difference of rolling stocks by the information provided in the survey;

- to test if respondents' choice making were affected by the cheap-talk (CT) script.

The survey was carried out among the ITS PhD group and some friends outside of University who commute to work by train or underground. In ITS, it was conducted through an in-hall survey and followed by focus group discussions on the survey design. Twelve PhD students who commute by train frequently or within the last month participated in the in-hall survey. In addition, mail-back questionnaires were distributed among some friends who commute to work by train or underground. Eight people mailed back the completed survey form.

6.2.1 Survey form

The survey form was based on the principles outlined in Chapter 5. The questionnaire consisted of four parts. The first part collected information of actual journey, while the second part was the SP questions. The third part was designed for obtaining the socio-economic information of respondents. The last part contained some follow-up questions on the survey and respondents' perceptions of the experiment.

A cheap-talk script was included in the second part of the questionnaire the in-hall survey to test its impact on respondents' choice making,. Considering the small sample size, the CT script was added in the middle of the nine hypothetical scenarios (between the fourth and the fifth choice), rather than putting CT in some questionnaires and compare responses with /without CT script. The first four choices were called part 1 and the others were called part 2. The nine scenarios were arranged randomly in the survey. A set of 9 questionnaires were used in the first pilot survey to make sure that all the scenarios are presented equally in part 1 and part 2. Each questionnaire was given a number to avoid confusion.

For the outside University first pilot survey, the questionnaire was distributed to some friends who commute by train or underground. It was difficulty to control the process of SP experiment; therefore, the CT script was added at the beginning of SP experiment. Suggestions and refinement of the script were asked in the follow-up questions.

6.2.2 Presentation of rolling stock information

The presentation of rolling stock information was tested in the first pilot survey. Descriptive and pictorial presentation of the two different types of rolling stock, Pacers and Sprinters were provided to respondents in separate boxes. The difference between rolling stocks was described in terms of seat style, noise level and smoothness of ride.

Noise level and smoothness of ride are difficult to describe. Although there are some metric scale to define noise with decibel scale, and smoothness of ride with vibration, these are not easily interpreted by respondents. In previous research, the terms were generally presented in the following way:

- Use a 1-10 scaling between two alternatives (very poor and very good) (Wardman and Whelan, 1998).

- Presentation and levels of specific service attributes. For example, the smoothness of ride, some research used "comfortable, uncomfortable", "nothing noticed, just noticeable, noticeable, slightly uncomfortable, and rather comfortable" (British Rail Standard); some research used "can drink water/write, difficult to drink water/write".

In the first pilot survey, 'very noisy/noisy' were used for Pacers and Sprinters respectively, and to describe the 'smoothness of ride', 'difficult to drink/write' and 'can drink/write' were given to the Pacers and Sprinters. This description was found to be subjective opinion, and was adjusted in the second pilot survey, as shown in section 6.3.3. Picture information of the rolling stocks, both outer and inner layout, was presented in the survey.

Except for the difference of rolling stocks described in both word and pictorial format about the service quality, the journey time, single fare and frequency of service are different for each SP scenario in the simple design. For better understanding of the journey characteristics, they were defined before the SP experiment as follows: '**Journey Time: the amount of time spent travelling on train; Fare: this is how much you pay for a one-way single journey; Frequency of the service: this is the frequency with which the train service operates.**'

In the SP experiment, the format for each scenario was tested:

Format 1:

Single Fare	Type of Train	Journey Time	Service Frequency	Choice
£2.50	Sprinter	20 mins	1 services / hour	O
£2.00	Pacer	25 mins	2 services/ hour	O

Format 2:

Option	Option A	Option B
Stock	Sprinter	Pacer
Journey Time	20 minutes	25 minutes
Single Fare	£3.00	£2.00
Frequency	Every 15 minutes	Every 30 minutes
Choice	O	O

6.2.3 SP design

Table 6.1 shows the attributes and levels of the SP design in the first pilot survey. The figures in 'bold' are adjusted figures to help obtain a better boundary ray map. The simulation tests proved that it can cover the target variables and can detect the difference between different rolling stocks as discussed in section 5.5.2.

Table 6.1 Design of SP experiment for the first pilot survey

Combination of SP attributes and levels (A: Improved train; B: current train)						
Scenario	**Time A**	**Time B**	**Cost A**	**Cost B**	**Head. A**	**Head. B**
	minute		pence		minute	
1	25	30	**300**	180	15	30
2	25	25	**200**	**250**	20	10
3	25	20	**220**	200	15	20
4	20	30	250	**180**	15	20
5	20	25	300	200	15	30
6	20	20	200	180	20	10
7	15	30	300	200	20	10
8	15	25	**300**	180	15	20
9	15	20	250	200	15	30

6.2.4 Presentation of Cheap Talk

A cheap-talk script was added into the questionnaire, as shown in the quotation. This script followed the CT script developed by Cumming and Taylor (1999), which was proved robust in three different kinds of good in CV studies (see section 2.6). It includes two parts: reminder of hypothetical bias and constraints (budgets and time constraints).

> "Before stating your preference in the following situations, pleasing keep in mind that in the previous surveys, we have found that some people who say they are going to take the new train service sometimes behave differently from what they said when the new rolling stock became available. This is what we call as Hypothetical Bias. The reason for this is that when the rolling stock is introduced, you would have to actually pay more. So please consider that the increased money you spend on the travel expenses will be unavailable for your other expenses. We would like you to take this account and read through the survey imagining that you will actually be paying for the new services."

Some follow-up questions were added to test the impact of cheap-talk on decision making, and also test if it can be understood by respondents.

6.2.5 Follow up questions about the survey

In the first pilot survey, some follow up questions were provided to find respondents' perceptions of the SP survey. In the in-hall study in university, this part was carried out by a focus group discussion after respondents completed the survey form; while questions were provided in the last part of the questionnaire for the respondents outside of university.

Questions asking about the format of survey were included to determine if respondents could understand the experiment, such as the presentation and scaling of the variables in the survey. Respondents were asked to state their preferred format for the SP scenarios. Respondents were also asked to indicate if they could perceive the difference between the rolling stocks from the information provided to them.

Questions were also asked to know if there were any important variables missing relating to travellers' choice making in the SP design. For example, one question was: 'When you plan a journey, what factors do you consider? (Please rank the following factors: 1 being the most important and so on).

Journey Time ☐ Fare ☐ Comfort of the train ☐ Operating Company ☐
Frequency of the service ☐ Reliable arrival time at destination ☐
Convenience of departure Time of Train ☐ Others (Please specify) ☐ ________________

Other questions on respondents' perceptions of survey were presented in this part, such as the difficulty to make a choice in the SP experiment and their perceptions of the cheap talk.

6.2.6 First pilot survey data collection and description

The first pilot survey was conducted in April 2005 in and out of university. Twelve PhD students and eight commuters outside of university participated in the survey. Among them, 65% of the respondents were male and most of them were between 20 to 40 years old. The income distribution showed that 60% of the respondents had their annual income under £20k, and 40%, between £21 to £35k. The income distribution is distorted because most of the respondents were research students. The average income is expected to be higher than this in the main survey where the respondents are normally working-class people.

Respondents were asked if they could perceive the difference between the trains by the text description and pictorial information. Respondents who favour the pictorial description have given comments such as:

> "Word description doesn't give any information on comfort; picture gives more information, but I think you also need a short description, i.e. as now!".

Respondents who are against the pictorial information have the comments such as:

"Pictorial information is confusing and easy to bias" and "can be unclear."

Most of the respondents think that both the pictorial and wording information are necessary:

"I think both of them should help to make a more conscious choice."

Respondents were asked to indicate the importance of factors during the journey decision making. Table 6.2 shows the average score which represents the importance of each factor in respondents' decision making on their journey by train. Among them, 'Fare', 'Journey time' and 'Reliable arrival time (Punctuality)' were the most important factors being considered by the respondents during their decision making, followed by the service frequency. In the present research, the above factors were involved in the SP experiments. All important factors are included, which avoid the omitting variable bias (Louviere, 1998; Louviere et al., 2000).

Table 6.2 Factors impact on respondents' decision making regarding to the rail journey

Factor	Ave. Score (Importance)
Reliable arrival time (Punctuality)	3
Fare	1
Comfort of the train	6
Journey Time	2
Frequency of the service	4
Operation Company	7
Convenience of departure time of the train	5

Respondents were asked if they could understand the presentation of attributes and the levels for each attributes. They all felt that the survey form was easily understood. Most of respondents preferred the second format (section 6.2.2) for presentation of SP choices. On average, it took them about 15 minutes to complete the questionnaire.

Respondents were asked their perception of the CT script. Some respondents felt that CT did not affect their choice making. Some of their comments are listed as below:

'I just look at the table.'; 'It is good additional information, but it doesn't affect my decision.', 'Didn't bother to read/understand'. Some of the respondents are familiar with the cheap- talk concept, 'I already took into account the information mentioned in the cheap- talk', 'I know that before and I was taking it into account'.

There were also comments on the length of CT such as 'the cheap-talk can be put in more than one paragraph, considering the fact that I just read half the sentence of this paragraph.'

Considering respondents characteristics, they might have not an incentive to bias their answer. No effect of CT was detected. From the first pilot survey, the wording of CT was tested and some of the suggestions were taken to refine the script.

The data from the 12 questionnaires from the survey inside university was analysed in the ALOGIT program. Responses and comments from the survey outside university were only taken for suggestions and refinement of the questionnaire for the subsequent survey, rather than putting into the model analysis.

The model results show poor estimation, therefore are not shown in the thesis. The poor estimation can be explained for the following reasons:

- The sample is small for calculating any values with significance.
- The data used are collected from PhD students studying in transport research. Some of the respondents understand biases in SP and the concept of CT, which could cause them to answer differently.

6.2.7 Implications for subsequent survey

The first pilot survey has revealed several problems in the survey design and development of the questionnaire.

Firstly, some respondents felt that cheap-talk was too long; hence they ignored most of the information provided to them. The survey format was paper-based mail-back questionnaire. Most of the respondents in the real survey are commuters who are very busy during the travel in the morning peak hour. Therefore, they might ignore the messy information when completing the questionnaire in a hurry. For the next survey, the cheap-talk script was modified to be more concise and was put in a box for emphasis purpose.

Secondly, as respondents might come from different educational backgrounds, using simple English was suggested in the survey, by getting rid of technique terms. Suggestions on the wording of the survey were incorporated into the next questionnaire.

Thirdly, it was suggested that the survey format should be neat, that it should be limited to a consistent style to a maximum of two styles. For example, different formats were used to display the information such as using box, picture, bold and italic in the SP (second) part of the questionnaire. In the next survey, the format should be kept more consistent and concise.

Finally, the word description of the train information was suggested to be subjective. It was adjusted in the next pilot survey.

6.3 Second Pilot Survey

6.3.1 Background information

After the first pilot survey, the survey design was refined. The second pilot survey was carried out on 8^{th} of Aug. 2005 in Oldham Mumps. The aim of this survey was:

- To test the design of SP and whether or not the presentation can be understood
- To test the response on different designs and adjust the wording.
- To estimate the survey response rate and task load for the main survey.

A total of 170 questionnaires were distributed in the railway station. This section presents the practice and analysis of the second pilot survey and some suggestions for the main survey. This section gives the basic introduction of the second pilot survey; Section 6.3.2 discusses the development of the SP survey; section 6.3.3 reports the field work; followed by the data description in section 6.3.4; section 6.3.5 part gives the data analysis and some of the research hypothesis testes. Finally, suggestions for the main survey are discussed.

At the outset of the second pilot survey, two criteria were specified for the selection of stations/routes to be used in the pilot survey:

- The routes should have the old rolling stock currently in operation, for which to be improved.
- The sample size should be large enough to show some statistically significant effects of different design, and but too large (which is better to be used in the main survey).

The service line between Oldham Mumps and Manchester satisfied these criteria and was selected. In the peak hour, there were approximately 150 passengers boarding the train to Manchester. Currently, the journey time is around 19-20 minutes to Manchester and the frequency of the service is four services per hour during the day time.

Northern Rail Ltd. took the franchise since 2004. The rolling stocks running on this route are Pacers and Sprinters. From an on-site visit, the number of these two stocks was observed to be about equal. Some of the Pacers (Class 142/144) have been refurbished. The most obvious change is that the bus style seats have been replaced by more comfortable seats.

In the second pilot survey, self completion questionnaires were handed out to customers on the platform. As the train is expected to be crowded in the morning peak hour, respondents would prefer to finish the questionnaires at their free time and were therefore given the FREEPOST envelope with each questionnaire.

6.3.2 Improved SP design and Cheap Talk

The SP experiment design has been adjusted based on results of the initial pilot survey and further simulation tests. First, the levels of attributes were adjusted. For example, the level of the cost attribute in scenarios 4 and 8 were changed back to the original orthogonal design (see Table 5.5 for the initial design) to avoid the correlation problem found in the simulation. Table 6.3 reports the SP design for the second pilot survey.

Table 6.3 SP Design of the second pilot survey

Combination of SP attributes and levels (A: improved train ; B: current train)						
Scenarios	Time A	Time B	Cost A	Cost B	Headway A	Headway B
	Minute		pence		minute	
1	25	30	**300**	180	15	30
2	25	25	**200**	**250**	20	10
3	25	20	**220**	200	15	20
4	20	30	250	200	15	20
5	20	25	300	200	15	30
6	20	20	200	180	20	10
7	15	30	300	200	20	10
8	15	25	200	180	15	20
9	15	20	250	200	15	30

The cheap talk script has been modified to have a concise format as shown in the following:

> "Previous surveys have sometimes found that people say they would be happy to pay extra for improved trains but when the fare is raised and the improved trains are provided, people say they would prefer the cheaper fare with the old trains.
>
> Bearing this in mind, as you read through the following choices, please imagine you will actually have to pay the fare stated. "

The script has been split into two paragraphs. The first paragraph explains the bias found in the previous studies. The second paragraph explains the constraints and warning message for the SP experiment. The script was put in a box for emphasis purpose.

6.3.3 Refinement and adjustment of the SP questionnaire

The second pilot survey needs to be authorised by GMPTE and Northern Rail Ltd (NR). This SP study investigates the potential change to the infrastructure. The company's concerns and suggestions have been considered in the SP questionnaire. Adjustments to the questionnaire have been agreed with the TOC's requirement, so as not to cause unrealistic expectation of their customers, but still be suitable for our research aim.

Firstly, regarding the information for different rolling stock, NR addressed that 'the way of describing both types of train is very subjective' and 'is in danger of influencing the option of respondents even before started answering the questions'. For example, terms such as 'noisy' and 'very noisy' are very subjective in describing the noise level of the rolling stock. The TOC informed us that some of the Pacers have replaced the bus style seats with new seats. The company also advised removing the word description to avoid a negative view of the Pacers from respondents when answering the questions.

Secondly, some of the words in the choice questions were not accurate, such as 'Reliability' should be replaced by 'Punctuality' in this SP context. Reliability has a broader meaning.

Thirdly, the company stated that "the formulation of the question suggests that new trains will be introduced to the route". This would give passengers' expectation and suggests that in the near future they may benefit from the new trains on the route. From the company's point of view, this kind of expectation is dangerous. They suggested to replace the phrase "if the new trains were" (in Question 13) by "if new trains would be" to avoid any confusion that a new train will be introduced. Another example is the word "Choice" has been be replaced by "Preference", which emphasised on the hypothetical situation.

Fourthly, they insisted to add the following sentence in the introduction part of the survey. "This research is not commissioned by Northern Rail Ltd and is conducted for scientific / academic purposes only and Northern Rail Ltd will not be using the results from this survey for any purpose. Northern Rail Ltd has authorised the survey, but has no involvement with the survey content and distances itself from any suggestions made in the survey." It is concerned that this sentence would give respondents a strong incentive that answering this questionnaire will have no influence with their future journey. This incentive might cause respondents to ignore the questionnaire if they think their input in the survey is 'useless'. After negotiation, this sentence was kept in the footnote at the last page as the company requested.

Except for the above questions, TOC had some other questions regarding to the design and social economic questions. For example, they pointed out that they never ask customers about their household income, as income is a personal question. The purpose of the income question is for segmenting the sample in the future. After negotiating with the company, the question can stay in the questionnaire, providing the option 'don't want to say' to respondents.

After taking into account the concerns from the train company, the Train Operation Company (TOC) gave permission to carry out the survey using the approved version of the questionnaire.

6.3.4 Second pilot survey data collection and description

After the permission was issued, the section pilot survey was conducted in Oldham Mumps on 8th August 2005. In the survey, there were four types of questionnaires being sent out. Table 6.4 presents the SP questionnaires using in the survey. The suite of questionnaires was applied to the main survey, as explained in chapter 5.

Table 6.4 SP questionnaires in the second pilot survey

SP Questionnaire	Description
S1	Simple design with 3 attributes
S2	Simple design + Cheap talk script
S3 (S3A/S3B)	Complex design with 5 attributes
S4 (S4A/S4B)	Complex design + Cheap talk script

The SP questionnaire S1 was a simple design with three variables which were journey time, single fare and headway (presented by frequency in the survey form). S2 was a simple design plus a cheap-talk script before the SP choices. S3 (S3A/S3B) was a complex design with the same three attributes as S1/S2 and two additional attributes (punctuality and crowding). By using fractional factorial design, 18 options were generated. The whole choice set was split into S3A and S3B, with 9 options each to reduce the task load. S4 (S4A/S4B) was same as S3, which were complex design, but with a cheap-talk script.

In each questionnaire, the nine options were arranged randomly to avoid similar options being placed close to each other (order effect). Then the questionnaires were folded and put into a freepost envelope. The order of distributing the questionnaires were arranged as: S1, S3A, S2, S4A, S1, S3B, S2, S4B and S1 etc…to make sure that each type of the SP questionnaires has been sent out equally.

Survey field work

The pilot survey started at 7:15am at the railway station in Oldham Mumps. When sending out the questionnaires, a brief oral description of the survey was given. Most of the people took the questionnaire. In the morning period, approximately 5 people refused to take the questionnaire.

By 9:00am, approximately 70 questionnaires were handed out, which were much less than the expected number of people boarding the train. The GMPTE annual report states there are more passengers (approximately 150) boarding the train during the morning peak hours. The potential reasons might be the survey was undertaken during the summer holiday, and lots of people might take a holiday. Secondly, the 7:15am service to Manchester was missed during the survey which accounted for 20 to 30 people boarding the train.

On that day, most people were observed to arrive on the platform 5 or 10 minutes earlier than the train departure time, but not arrived at the last minute before trains' departure.

After 9:00am, it can be split to two time periods. The first period start from 9:00am to 11:00am, which a lot of people still boarded the train. During that time period, 70 questionnaires were given out. From 11:00am to 2:00pm, there were fewer people boarding the train and only 30 questionnaires were given out. During this time, some of them agreed to answer the questionnaires on the platform.

Because there were fewer boarders during the period (11:00am to 2:00pm), questionnaires were handed out on the other platform where passengers got off the train from Manchester. It was very difficult to stop those people and ask them to complete the questionnaires or even have the questionnaire brought back and completed at their convenient time.

Data description

170 questionnaires were sent out in the second pilot survey. 60 questionnaires were mailed back, generating a response rate of 34%.

The data was screened[2] by getting rid of any of the questionnaires that were incomplete and the ones which gave very illogical answers. After cleaning the data, 55 of the questionnaire were found satisfactory. This gave a 32% response rate. Due to the small sample size, the four different SP experiments are combined and called 'ALL' (see Section 6.3.5). Respondents' journey characteristics and their personal characteristics are summarised as below:

- **Sample characteristics**: 53% of the respondents were commuters. As the second pilot survey was conducted in summer holiday, it is expected to have a higher percentage of commuters in the main survey. The percentage of female and male in the sample were 53% and 47%, respectively. Most of respondents were in the age group of 18-25 and 26-35, the total percentage was about 60%. Most of respondents were in the income categories of £10K - £20K (35%) and £21K -£35K (24%).

- **Ticket information**: the type of tickets was split as follows: 27% used Standard day return ticket, 15% used monthly Train Card, and 11% used weekly Train Card. Very few of them used annual card. It was not difficult for them to estimate/compare their daily cost with the single journey fare provided in the SP choices.

---

[2] In the data filtering process, we have an argument whether or not to keep the respondents who chose just one alternative for all the nine options, for example, always chose Sprinter. After the discussion and careful consideration, we kept them in the analysis, as they also show the true preference of respondents.

- **Survey and questionnaires design**: 71% of the respondents felt no difficulty to make the choice. Most of the respondents (more than 90%) felt fairly or quite easy to distinguish the difference between the stock types. More than 90% of the respondents perceived that the fare would likely be increase if new trains were introduced.

6.3.5 Models and results from the second pilot survey

Data enrichment and consistency test before pooling the data

The small sample size made the separate analysis lack statistical significance and variance, which is shown in Table 6.7. SP responses from 4 different SP questionnaires were pooled together to make a bigger sample size. Coefficients of the logit model are scaled relative to the errors in the SP response (see section 4.3.3), which generates the scale factor problem when combining different data sets. The process of combining data sets is presented in section 4.3.4.

A hierarchical logit model was built allowing scale factors for different data sets. Each alternative in each group was set as one root. There were 8 roots altogether. The scale factor for S1 was set as unity (reference) and all other groups were scaled to S1. Table 6.5 reports the scale factors for different groups in the second pilot survey.

Table 6.5 Scale factors for different group

SP Design	Scale Factors		
S1 (Reference)	1	t(0)	t(1)
S2	0.8022	2.1	0.5
S3	0.6812	2.2	1.0
S4	0.8069	2.2	0.5

't(0)' refers to the t-test of the coefficient of scale factor against 0, and 't(1)' is the t-test against unity (reference). The scale factors for S2, S3 and S4 were all not significantly different from unity, which indicated that the levels of error variance were the same across different SP designs in the second pilot survey. Therefore, four groups of SP responses can be combined directly.

Model Specification

The choice observations were analysed by model specification 1 (Equation 4.30), that the value of the new rolling stock is assumed to be a constant value. The value of improved rolling stock is measured by incorporating a constant term (ASC) into the utility function. Table 6.6 defines attributes and monetary values in the data analysis.

Table 6.6 Definition of the attributes and monetary values

Terms	Description
Time	In vehicle journey time
Cost	Single fare per journey
Head.	The headway of the service
Punc.	Punctuality of the service
Crow.	Crowding of the service
ASC	The preference for improved rolling stock
VoT	Value of Time, respondents' willingness to pay (WTP) for one minute saving in the train journey
VoH	Value of Headway, respondents' WTP for one minute saving of the service headway
VoS	Value of Rolling Stock, respondents' WTP for the stock improvement per single journey
VoP	Value of Punctuality, respondents' WTP for improving the service punctuality (reduction of the delay in the journey)
VoC	Value of Crowding, respondents' WTP for reducing the crowding (standing during the journey).

In the model analysis, time is in units of minute and cost is specified in pence. Headway is quantified by the time period between two trains, for instance, the level of 'every 15 minutes' is given 15 minutes in the utility function. Punctuality is presented as an amount of time delay with a given frequency, for instance, '1 out of 5 times delay for 10 minutes'. Punctuality is quantified by the expected value which is calculated as the delay time multiplied by the given frequency. For example, if the level of the punctuality is '1 out of 5 times delay for 10 minutes', the expected value is $10 \times (1/5)$ which is 2 minutes. Crowding is present as standing time with a given frequency. Here, the standing time is the in-vehicle journey time specified in the choice Crowding is quantified by the expected value in the data analysis. Table 6.7shows the coefficient estimation from the separate and joint analysis by using model specification 1.

Table 6.7 Analysis of the second pilot survey (t-ratio)

	S1		S2		S3		S4		ALL	
Obs.	99		135		117		144		495	
ASC	0.470	(1.3)	0.141	(0.4)	0.178	(0.6)	0.108	(0.3)	0.223	(1.6)
Time	-0.096	(-1.8)	-0.162	(-3.5)	0.007	(0.2)	-0.060	(-1.4)	-0.073	(-4.0)
Cost	-0.020	(-2.6)	-0.025	(-3.0)	-0.007	(-0.9)	-0.017	(-2.8)	-0.017	(-6.1)
Head.	-0.061	(-2.1)	-0.074	(-2.5)	-0.028	(-0.8)	-0.067	(-2.6)	-0.057	(-5.8)
Punc.					-0.335	(-2.9)	-0.371	(-3.6)	-0.402	(-5.7)
Crow.					-0.054	(-1.0)	-0.026	(-0.7)	-0.036	(-1.2)
ρ^2(C)	0.0980		0.1536		0.1715		0.1860		0.1325	

Initially, the analysis was applied to each group (S1, S2, S3 and S4) and to the combined data sets defined as "ALL". The data was analysed in ALOGIT program. 'Jack-knife' was applied to avoid the repeated measurement problem (see section 4.3.8). The model's goodness of fit (ρ^2(C)) was about 0.1 to 0.2, which is good in the logit models. Coefficients obtained from the data analysis were of right sign with reasonable values. For example, the sign of time coefficient was negative which represents the disutility of journey time during the journey. Some of the coefficients were not significant by the t statistic test.

Monetary Values

Monetary value of each attribute can be obtained by the marginal utilities of target attributes and cost (see section 4.4.4). The value is interpreted as individuals' willingness to pay (WTP) to improve the attribute by one unit. Table 6.8 reports the monetary values derived from the model estimation from the combined dataset (Table 6.7).

Table 6.8 Monetary values derived from model estimation

Characteristic	Unit	Respondents WTP (Pence)
In vehicle Journey Time	One min. saving	4.2 (p/min.)
Service Frequency	One min. saving	3.3 (p/min.)
Rolling Stock	Pacers to Sprinter	13 (p /single journey)
Punctuality	Pence/minute	23.5 (p /min to avoid one min. delay)
Crowding	Pence/minute	2.1 (p /min to avoid standing on the train)

The value of time (VoT) is 4.2p/min. The value for headway is 3.3p/min, which is approximately 0.78 times of the VoT. These values are in line with previous research of VoT (Wardman, 2004), although the VoT from the second pilot survey was slightly lower. No obvious bias was found. Due to the small sample size, segmentation was not undertaken. The income effects was not taken into account, but will be analysed in the main survey. Passengers are willing to pay around 13 pence per journey to improve the rolling stock from Pacers to Sprinters. The valuation varies among different SP designs.

Impact of Cheap Talk

Dummy variable (d_{CT}) denoting the impact of cheap-talk script was added into the ASC term in the utility function. Table 6.9 shows the coefficient estimation of the CT impact in the combined groups S1/S2 (simple design) and S3/S4 (complex design) and 'All'.

Table 6.9 Coefficients produced from SP

	Time	Cost	Head.	ASC	Punctuality	Crowding	d_{CT}	Obs.	Rho-s
S1&S2	-0.120 (-3.4)	-0.022 (-4.6)	-0.064 (-3.3)	0.332 (1.2)			-0.051 (-0.2)	234	0.098
S3&S4	-0.047 (-1.6)	-0.015 (-3.5)	-0.057 (-3.4)	0.243 (0.8)	-0.769 (-6.2)	-0.191 (-1.7)	-0.105 (-0.4)	261	0.171
ALL	-0.079 (-3.5)	-0.018 (-5.7)	-0.060 (-4.7)	0.334 (1.6)	-0.815 (-6.8)	-0.196 (-1.7)	-0.162 (-0.7)	495	0.135

CT script showed a negative effect on the ASC term, which indicated that with the CT script, respondents gave a lower value to the improved rolling stock. The impact was not significant at the 5% level. However, the small sample size did not provide enough data to conclude this as it showed some trend of expected effects.

6.3.6 Lessons learned from the pilot survey

- **Survey and Questionnaire Design**: the SP design satisfied the purpose of the pilot survey. Respondents understood the SP experiment and questionnaire, and most of them perceived the difference between trains by the pictorial and text description. The sensible results from the estimation of second pilot survey indicated that the design and survey presentation is satisfactory for the main survey data collection.

- **Rolling Stock**: The main stock types running on this route are Pacers, Sprinters and Super Sprinters. After the pilot survey (8^{th} Aug. 2005), the pictorial information of stocks has been changed from Sprinters to Super Sprinters, considering the fact that most of commuters are familiar with Pacers and Super Sprinters.

- **Response Rate**: the response rate is approximately 34%, and 32% after screening the data, by getting rid of all the incomplete and illogical ones. The estimated response rate is around 25-30% for the main survey. This response rate is satisfactory. The higher response rate can be attributed to the short explanation when giving out the questionnaire or passengers are willing to help a student to finish the project. In the main survey, the short explanation on the survey is kept.

- **Survey Time**: it was observed that in the morning peak hours, 7:00am to 9:00am, and off-peak hours, 9:00am to 11:00am, there were lots of passengers boarding the train. But after that, there were very few people. It is suggested the period to carry out the main survey should start from 7:00am to 11:00am. At the bigger stations like Bolton or Wigan Wallgate, the survey could be extended the whole day to get a bigger sample.

- **Task Load**: in the pilot survey at Oldham Mumps station, a total of 170 questionnaires were sent out between 7:15am and 2:00pm. During the peak hours from 7:00pm to 9:00pm, there were around 20-30 passengers boarding the train for every service. Station similar to Oldham Mumps in terms of boarding number, one interviewer can handle the task. Assistant is needed at the larger stations if the boarding number is greater than 200 during the morning peak hours.

- **Passengers**: in the pilot survey, passengers came earlier to wait for their train. This gave us enough time to distribute the questionnaires.

- **Travel to Survey Location**: it was exhausting to travel very early in the morning from Leeds to Manchester. Considering that a few passengers boarded the train around 7:00am from the on-site visit, it was recommended to stay in Manchester and travel to each location from Manchester in the main survey.

6.3.7 Summary from pilot survey

The second pilot survey was successfully conducted at Oldham Mumps railway station. The response rate was 35%. The main objective for this research is to determine effects of the SP design on the bias in responses. The pilot survey was the first stage of the SP experiment. The conclusions from the pilot study were:

Firstly, a difference in valuation of rolling stock was found. Sprinters were valued on average at 13 pence per one way journey.

Secondly, values of time and headway found were generally in line with (slightly lower than) PDFH recommended value (at 4.2 p/min for journey time, and 3.3 p/min for headway). Segmentation test was not undertaken as the sample size was not big enough to achieve all the significant values.

Thirdly, the cheap- talk script has shown some effects on the response, but sample size was too small to confidently conclude there is significant effect.

The SP design and survey presentation was satisfactory for the main survey data collection.

6.4 Main Survey

6.4.1 Objective of main survey

This research aims to identify the influence of different designs on the pattern of SP responses and to explore means of identifying and reducing bias in SP experiments. More specifically, the hypotheses tested in the experiment were:

- To test if respondent perceive the aim of the research, they would bias their responses strategically for certain better outcome;
- To test if the cheap talk script can amend the incentives for respondents to bias;
- To test whether or not adding more attributes would amend the incentive for respondents to bias their answer, and if task complexity of the SP design affects the valuations implicit in the SP responses.

6.4.2 Main survey field work

Selection of railway stations

At the outset of the survey, two criteria were specified for the selection of stations/routes to be used in the data collection, which was explained in section 5.4.5.

18 stations satisfied the criteria and were selected as the location for the main survey. The time for the main survey was divided into two periods: starting from end of October 2005 for three consecutive weeks and beginning of December for another week. The reason for the split was firstly, to allow some time to get the permission for Stalybridge station which is operated by the Transpennine Express (TPE) during the start of the main survey on Oct. 31st 2005. Secondly, considering the time and budget limit of this PhD project, the surveys carried out in the first three weeks were at bigger stations (in terms of the boarding number in the morning peak hours). If the response rate from the first three weeks was found satisfactory, the survey in the last week could be cancelled.

The half term holiday was avoided considering the fact that in the holiday, people's travel patterns are expected to be different, especially for commuters.

SP survey questionnaires

As explained before, the band reflects the journey distance of each rail station, named from A to D. For each band, there are four different types of SP questionnaires: simple and complex, with and without Cheap Talk. A brief introduction of the survey form can be found in Section 5.3. The questionnaire was coded by the following way (X represents the band name, A to D):

- XS1: was a simple design with three variables which were journey time, single fare and headway (presented by frequency in the survey form);
- XS2: was a simple design plus a cheap-talk script prior to the SP choices;
- XS3/4: was a complex design with the same three attributes as S1/S2 and two additional attributes (punctuality and crowding). By using fractional factorial design, 18 options were generated. The whole choice set was split into XS3 and XS4, with 9 options each to reduce the task load;
- XS5/6: was same as XS3/4, which was complex design, but with a cheap-talk script.

In each questionnaire, there were two alternatives which are Pacers and Super Sprinters, with nine hypothetical scenarios.

Distributing the SP questionnaire and controlling of the SP experiment

The main survey was carried out at 13 stations. Self-completion questionnaires with freepost envelope were handed out to the customers on the platform from 7:00am to 11:00am. The information of the main survey is presented in Table 6.10.

For example, the SP design of Band C was used in the survey conducted in station Romiley. Four different sets of questionnaires were distributed on the platform. The order of distributing the questionnaires was arranged as: CS1, CS3, CS2, CS5, CS1, CS4, CS2, CS6 and CS1 etc (repeating). This ensures that each type of SP questionnaires has been sent out equally. A short explanation of the survey was given to respondents when giving out of questionnaires. Our target population was the users of those trains, so we were happy to survey all those encountered waiting for those trains.

In the questionnaire, sufficient information was sought from respondents to enable confounding effect to be unscrambled during the data analysis. The impact of respondents' socio-economic features on their responses and perceptions of the SP experiments will be examined and discussed in chapters 7 and 8.

The task was conducted from Monday to Thursday each week. In the first week, around 688 questionnaires were sent out. In the second week, the survey was carried out in bigger stations such as Rochdale, Wigan Wallgate and Marple (the boarding numbers are all more than 350 in the morning peak hours). One assistant joined to help with the survey. 1080 questionnaires were sent out. In the third week, 620 questionnaires were sent out.

Response rate

In total, 2788 questionnaires were distributed during the main survey, and 1322 questionnaires were mailed back. The response rate was 47.7 %. The questionnaires which were incomplete or contained illogical answers to other questions, such as journey information/social economic information were removed from the sample. After the data consistency test and clarification, 1222 questionnaires were usable.

The response rate is high compared with the pilot survey and some other previous studies. The reasons can be concluded as: firstly, the short explanation might help to get the high response rate. The short contact with commuters helps them to understand what the survey is about. Secondly, commuters care about their daily travel. They might perceive that they could express their ideas about their journey through this questionnaire. Thirdly, the layout and format of the questionnaires might give respondents the impression that it is professional.

Table 6.10 Information of the main survey

Stations	Survey Date	Journey Time To Manchester	Questionnaire (Design Band)	Inbound Borders	Total Sent out	Total Received	Response Rate	Usable	Rate
Ashton	31/10 (Mon.)	9-14mins	A	210	240	85	**35%**	79	93%
Greenfield	01/11 (Tue.)	25-27mins	C	167	153	90	**59%**	82	91%
Atherton	02/11 (Wed.)	25-30 mins	C	231	165	91	**55%**	81	89%
Bromley Cross	03/11 (Thu.)	29-32 mins	C	173	130	67	**52%**	64	96%
Shaw	07/11 (Mon.)	28 mins	C	244	127	51	**40%**	45	88%
Rochdale	08/11 (Tue.)	16-22 mins	B	355	323	148	**46%**	134	91%
Wigan	09/11 (Wed.)	35-45 mins	D	484	320	137	**43%**	128	93%
Marple	10/11 (Thu.)	25-28 mins	C	353	340	202	**59%**	189	94%
Daisy Hill	14/11 (Mon.)	28-35 mins	C	164	153	70	**46%**	65	93%
Romiley	15/11 (Tue.)	25-27 mins	C	212	182	89	**49%**	84	94%
Urmston	16/11 (Wed.)	17 mins	B	157	125	61	**49%**	55	92%
Mossley	17/11 (Thu.)	21-23 mins	B	188	160	78	**49%**	70	90%
Stalybridge	07/12 (Wed)	15-20 mins	B	480	350	153	**44%**	146	95%
					2768	**1322**	**48%**	1222	93%

Note: 'Inbound Borders' is obtained from GMPTE annual report

6.5 Data Description

There were 1222 usable questionnaires from the main survey. Sample characteristics are given in the following tables.

6.5.1 Socio-economic characteristics

Table 6.11 Social characteristics of respondents

Respondents Characteristics	Number of Respondents 1222	%
Gender		
Female	684	56
Male	529	43
No answer	9	1
Household income per annum		
Less than £10k	184	15
£10k - £20k	415	34
£21k - £35k	345	28
£36k - £50k	95	8
Over £50 k	64	5
Do not want to say	115	9
No answer	4	0
Age		
Under 18	17	1
18 – 25	266	22
26 – 35	324	27
36 – 50	368	30
51 – 59	163	13
60 - Over	81	7
No answer	3	0

Table 6.11 presents the basic social characteristics of the sample in the main survey. 56% of the respondents are female, compared with 43% male.

Most of the respondents' annual income is in the category of £10-20k (35%) and £21- 35k (28%), 15% of the respondents whose annual income is less than £10k. 8% of respondents has the annual income from £36 -50k. 5% of respondents whose annual income lies in the category of over £50k. There are about 9 % of the respondents who did not want to specify their annual income in the survey.

30% of the respondents belong to the age category 36-50, and 27% in the age group 26-35, and 22% from 18-25.

We cannot find previous studies/data to compare the sample characteristics.

6.5.2 Journey characteristics

In the questionnaire, respondents were asked to indicate their daily journey, in terms of the ticket types, journey purpose and frequency. The journey characteristics information is reported in Table 6.12.

Table 6.12 Journey details of the respondents

Respondents Journey Characteristics	Number of Respondents (1222)	%
Ticket		
Standard Day Single	75	6
Standard Day Return	293	24
Cheap Day Return	98	8
Rail Ranger	16	1
Day Saver Ticket	12	1
Weekly Season Ticket	188	15
Monthly Season Ticket	308	25
Annual Season Ticket	119	10
County Card	38	3
Other Ticket	75	6
Is your ticket paid by others?		
Yes	110	9
No	1112	91
Journey Purpose		
Commuting to/from Work	860	70
Employer's Business	52	4
Personal Business	34	3
To/from School/University	148	12
Visiting friends/relatives	26	2
Sport/Entertainment	9	1
Shopping	75	6
Other Purpose	18	1
Journey Frequency		
5 or more times a week	722	59
2 to 4 times a week	247	20
Once a week	46	4
Once every two weeks	27	2
Once a month	58	5
Less Frequent/First Time	122	10

Most of the respondents use season ticket (Weekly, Monthly or Annual, County Card), still there are a few respondents who use Standard Day Return ticket. 91% of them paid for the ticket themselves. Most of the respondents were commuters (70%), as expected. Compared with the pilot survey, the percentage of commuters is higher which is same as expected. The journey information of respondents shows that most of them are regular train users.

Table 6.13 reports the relationship between respondents' journey purpose and ticket type. The percentage of the number of respondents within each category is presented in the bracket. The journey purpose in the present research has been categorized into the following five types:

commuters, travelling for employer business (EB), travelling for personal business (PB), to and from school/college (School), and leisure and others (contains shopping, visiting friends, sports and entertainment and others).

Table 6.13 Relationship between the journey purpose and ticket type

Ticket Type	Commuters		EB		PB		School		Leisure		Total	
		(%)		(%)		(%)		(%)		(%)		%
Standard Day	42	(3.4)	8	(0.7)	2	(0.2)	15	(1.2)	8	(0.7)	75	(5.5)
Standard Day	154	(12.6)	30	(2.5)	8	(0.7)	61	(5.0)	40	(3.3)	293	(20.7)
Cheap Day	29	(2.4)	4	(0.3)	10	(0.8)	18	(1.5)	37	(3.0)	98	(5.0)
Rail Ranger	4	(0.3)			6	(0.5)	1	(0.1)	5	(0.4)	16	(0.9)
Day Saver	4	(0.3)	3	(0.2)	1	(0.1)	2	(0.2)	2	(0.2)	12	(0.8)
Weekly Season	175	(14.3)	2	(0.2)	1	(0.1)	9	(0.7)	1	(0.1)	188	(15.3)
Monthly Season	279	(22.8)	1	(0.1)	-		28	(2.3)	-		308	(25.2)
Annual Season	115	(9.4)	1	(0.1)	-		3	(0.2)	-		119	(9.7)
County Card	31	(2.5)	-		-		7	(0.6)	-		38	(3.1)
Other Ticket	27	(2.2)	3	(0.2)	6	(0.5)	4	(0.3)	35		75	(3.3)
Total	860	(70.4)	52	(4.3)	34	(2.8)	148	(12.1)	128	(2.9)	1222	1

Most of the commuters were using season ticket as shown in the table. Among them, monthly season ticket is the most common ticket type that commuters were using, followed by the weekly and annually season tickets. A few of the commuters were observed to use standard day return tickets in the study. People who travel to and from school/college were using season tickets quite often. For the other three groups (EB, PB and Leisure), day tickets were more often than season tickets.

Table 6.14 presents the relationship between the journey purpose and the reimbursement of ticket. Business travellers normally get their tickets reimbursed, compared with other groups.

Table 6.14 Relationship between the journey purpose and reimbursement of the ticket

Number of Respondents (1222)	Paid by him/herself		Ticket being reimbursed		Total
		%		%	
Commuters	820	95%	40	5%	860
Employer's Business	13	25%	39	75%	52
Personal Business	25	74%	9	26%	34
To and from School	138	93%	10	7%	148
Leisure travelling	116	91%	12	9%	128

Individuals' journey time has been asked in the questionnaire. A mean journey time of 29.50 minutes is obtained with the standard deviation as 0.68.

6.5.3 Perception of the survey and trains by respondents

Respondents make decision based on their perceptions of the SP experiment. Three follow-up questions were provided in the survey to explore individuals' decision making and impacts of perceptions on their responses. Among them, respondents' opinions of task complexity, familiarity of alternatives, and the perception of potential price increase were explored. Table 6.15 reports the perception of respondents in the study.

Table 6.15 Respondents' perceptions of the SP experiment

Perceptions	Number of Respondents 1222	%
Did you find it difficult in making the choices?		
Yes, very	35	3
Yes, quite	170	14
Yes, a little	382	31
No	623	51
No answer	12	1
Did you feel confident that you could distinguish Super Sprinters from Pacers with the help of the information provided?		
Not at all	25	2
Fairly	366	30
Very	780	64
Not sure	38	3
No answer	13	1
How likely do you think it is that fares would increase if new trains would be introduced?		
Not at all	17	1
Slightly	104	9
Moderately	244	20
Very	848	69
No Answer	9	1

Respondents' perceptions of the main survey and trains were found to be agreeing with that in the second pilot survey.

- 51% of the respondents felt no difficulty to make the choice. 31% of the respondents perceived a little bit difficulty in the choice making. Task complexity effect on respondents' decision making is presented in Chapter 8.
- With the help of the information presented in the survey, most respondents could distinguish the Super Sprinters from Pacers.
- As expected, 69% of the respondents perceived the price would increase if the newer trains were introduced.

The analysis of the relationship between individuals' perceptions and SP design is presented in chapter 8.

6.6 Summary of the Data Collection

This chapter presented the process of SP survey development. The SP experiment and questionnaires were designed and tested by a series of pilot surveys between March and August 2005. Pilot surveys showed that the designed questionnaire and SP experiment were suitable for use in the main data collection. The main survey was conducted between end of October and December 2005 in Greater Manchester by paper based self-completion questionnaires.

Throughout the pilot studies, many developments of the questionnaire and SP exercises were produced. The SP design was tested and improved from the simulation tests and pilot surveys, so that they can produce statistical values in the data analysis. The formats of questionnaire and SP exercises were changed until they were the least complicated for the respondents and were still capable to produce all the information needed in the analysis. The format of questionnaire was adjusted to be satisfied by the train operation company (TOC) to not causing customers' expectation due to the hypothetical characteristics of SP experiment, but still try to identify the incentives to the policy bias.

Results of the data analysis are presented and discussed in chapters 7 and 8. Chapter 7 demonstrates the users' overall valuation of the improved rolling stock. It is a general analysis from the SP responses. Chapter 8 explores impacts of cheap-talk script and adding more attributes to the SP experiment on the responses, together with the influence of individuals' perceptions on their choice making. Research hypotheses are tested in chapter 8.

Chapter 7

Valuation of Rolling Stock

7.1 Introduction

In chapters 5 and 6, the design of the questionnaire for the data collection was described. It was designed to gather the following information:

- Journey and socio-economic information for a better understanding of commuters' travel behaviour;

- SP data to find the effects of SP designs on the responses relating to rolling stock valuation;

- Post-questionnaire questions on respondents' perceptions of the SP survey for a better understanding of respondents' decision making process.

In addition, space was provided at the end of the survey form for respondents to add their comments.

The objective of this chapter is to present and discuss respondents' preferences and valuation of rolling stock from the main survey. A base model is estimated in this chapter using conventional logit methods. This model controls for several factors that might potentially confound the subsequent testing of research hypotheses on estimation bias. Chapter 8 will extend the analysis from this baseline position.

Section 7.2 explores the model specification, which accounts for the socio-economic factors. Section 7.3 demonstrates the procedure of pooling the data from different SP experiments and gives results from the pooled model. Model specification I has shown to be better than specification II from log-likelihood ratio tests, so is chosen as the preferred model. Section 7.4 presents the results from the preferred model and derives monetary values from the model parameters. Section 7.5 examines the model by comparing values obtained from this study and previous evidence. Section 7.6 ends the chapter with conclusions about respondents' preferences and valuations of the rolling stock.

7.2 Model Specification

7.2.1 Introduction of model specification

Respondents were asked to state their preferences between Super Sprinter and Pacer in the SP survey. The difference between these two types of trains was explained both verbally and by pictorial information. Nine hypothetical scenarios were presented in the survey with combinations of different attributes and levels. Respondents were asked to make binary choices between stock types during the survey. The binary choices were each described by 3 (simple design) or 5 (complex design) attributes. A series of logit models have been applied to analyse respondents' valuations of the improved trains (Super Sprinter) and other attributes.

This section illustrates two types of model specification (section 4.4.1) for analysing the valuation of rolling stock. The review (chapter 3) of stock valuation studies found that income and journey purpose contribute to the variation of valuations. To avoid confounding effects in further hypotheses testing, these two factors are taken into account in the base model analysis.

7.2.2 Two types of model for estimation of rolling stock

Two types of specification (see section 4.4.1) are applied to the present study to interpret respondents' preferences for rolling stocks.

The first model specification assumes the stock value as an absolute value per trip which is represented by the alternative specific constant term (ASC). This specification was commonly used in the early stock valuation studies. The advantage of this specification is that it is easy to interpret and transfer results. The limit of this specification is that the valuation of rolling stock is not allowed to vary with journey time (Wardman and Whelan, 2001, p.417).

Reporting the value as a percentage of the fare makes the valuation more transferable. Individuals' preferences of the improved rolling stock is affected by the journey characteristics such as time and cost, and service attributes such as headway, punctuality and crowding. The utility function is shown in Equation 4.25 (section 4.4.1). The monetary value of the improved rolling stock (Super Sprinter) (VoS) can be obtained by Equation 4.30 (section 4.4.4).

The second method assumes that the value of improved rolling stock is influenced by the journey duration. It might be expected that the stock valuation increases with the duration of the journey. The utility function with such a feature is shown in Equation 4.26 (section 4.4.1)

The monetary value of improved rolling stock relative to the current rolling stock would vary with the length of journey, which can be obtained by the difference between valuations of the in-vehicle time for two different rolling stocks multiplied by the journey length (journey time t), shown that:

$$\boxed{VoS_S = (VoT_S - VoT_P) \times t}$$

Equation 7.1

where $VoTs$ is the value of in-vehicle time in the 'Super Sprinter' train and VoT_P is the value of in-vehicle time in the 'Pacer' train.

7.2.3 Impacts of socio-economic factors

The review of rolling stock studies found that income and journey purpose have strong impacts on the variation of stock valuation. The impacts of income and journey purpose are investigated at the outset of analysis to avoid potential confounding effects. The two factors are incorporated into the model by segmentation analysis (see section 4.3.6).

7.3 Combine Different Source of Data

7.3.1 Different data sets in the main survey

As explained in the previous chapters (section 6.4.2), the main survey was conducted in thirteen different locations in Greater Manchester. The journey made from each station has its own characteristics in terms of journey time, frequency and others. To make the SP experiment more realistic, the 13 locations were divided to four bands in terms of journey distance, named A, B, C and D. The stations in each group have similar journey characteristics. For each group, there are four different questionnaires which are called S1, S2, S3 and S4:

S1: Simple design (journey time, fare, frequency, train types)

S2: Simple design with an additional paragraph added to the questionnaire, warning people of the strategic bias, known as cheap-talk script.

S3: Complex design, with the same three attributes as in S1 and added two more attributes: punctuality and crowding. By using fractional factorial design, 18 options were generated. We split the whole group to two with 9 options each to reduce the task complexity.

S4: Complex design with cheap-talk script.

There were 1222 usable questionnaires included in the data analysis. The following table presents the number of respondents in each group:

Table 7.1 Number of respondents in each group

Group No.	**AS1**	**AS2**	**AS3**	**AS4**	**CS1**	**CS2**	**CS3**	**CS4**	**Total 1222**
	19	22	20	17	170	148	158	164	
	BS1	**BS2**	**BS3**	**BS4**	**DS1**	**DS2**	**DS3**	**DS4**	
	89	119	94	96	30	28	28	20	

As mentioned in chapter 6, band A and D were conducted at a single station each. We combined group A with group B to be Short Journey distance group (S), and groups C and D to be Long Journey distance group (L), for example, AS1 and BS1 to be SS1, AS2 and BS2 to be SS2...CS1 and DS1 to be LS1, CS2 and DS2 to be LS2 etc.

Prior to the testing of our research hypotheses seeing the effects of SP design on responses, different data sets need to be combined and compared. Before pooling the data, the scale factor problem should be considered (Morikawa, 1989; Swait and Louviere, 1993; Hensher et al., 1999) to allow for the differences in the variance among datasets (section 4.3.4); otherwise, the estimation of the coefficients from different source of data would be biased. In some situations, the variance of the error terms can be different for different segments. If this is not recognised, it may confound the real behavioural role of observed and unobserved influences on choice.

As demonstrated in section 4.3.3, there exists an inverse relationship between scale factor and the variance error term. When the scale factor is larger for one source of data, the error variance is smaller for this source of data. For example, a more complex design is expected to yield a smaller scale due to a large amount of residual variation. Failing to account for the difference in scale could lead to its effect appearing in the coefficient estimates.

7.3.2 Pooling data

Simultaneous estimation (Ortuzar and Willumsen, 2002) has been applied to determine the scale factors. The process of pooling data was discussed in detail in section 4.3.4.

There are 8 different SP designs with combination of short and long journey distances (S/L), simple and complex design (S1, S2/ S3, S4) and the design with/without cheap-talk script (S2, S4/S1, S3). There are two alternatives in each choice (1/2).

A Hierarchical Logit model has been estimated which pools the binary choice experiments and, by specifying dummy nests as demonstrated in Figure 7.1, allows the inclusive value parameter to serve as a rescaling parameter.

The data set has been manipulated into a hierarchical format, which allows only one nest to be available at a time, while other nests are not available during the procedure. In this example, the parameters are all scaled to be in the units of the first SP exercise by scaling factors θSS2 to θLS4 (i.e. assumes that $\theta SS1 = 1$). Table 7.2 lists the scale factor for each data set when combining the data.

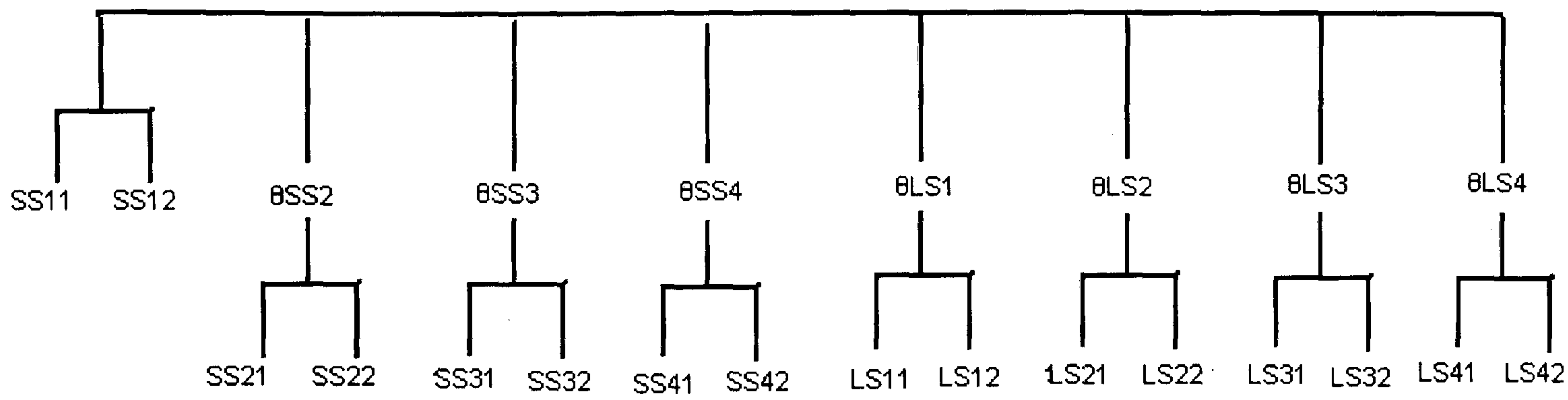


Figure 7.1 Artificial tree structure to obtain scale factors

Table 7.2 Labelling the scale factors

SP Experiment	Journey Distance	Design	Cheap Talk	Scale Factor
SS1	Short	Simple	N	1
SS2			Y	θSS2
SS3		Complex	N	θSS3
SS4			Y	θSS4
LS1	Long	Simple	N	θLS1
LS2			Y	θLS2
LS3		Complex	N	θLS3
LS4			Y	θLS4

Different data sets can only be combined without rescaling if the scale factors for those data sets are not significantly different from each other. For example, SS1 and SS2 can only be combined without rescaling if θSS2 is not significantly different from unity which is the scale factor of SS1. If the scale factors are not significantly different from each other, the level of variance for them is also not significantly different. In this situation, the datasets can be pooled directly without causing any other confounding effects caused by variance.

7.3.3 Model specification I

In this section, pooling data for the model specification I is presented.

The utility of rolling stock choice is a function of journey characteristics and service attributes. Time is in units of minutes, and cost is specified in pence. Headway is presented by frequency in the SP experiment, which is quantified by the time period between two subsequent trains. For example, if the level for headway is 'every 15 minutes', 15 is being used in the utility function. Punctuality is presented as an amount of time delay with a given frequency, for example, '1 out of 5 times delay for 10 minutes'. The 'punctuality' is quantified by the expected value which is obtained by the delay time multiplied by the given frequency. For example, when the level for punctuality is '1 out of 5 times delay for 10 minutes', the expected value is $10 \times (1/5)$ which is 2 minutes delay. Crowding is quantified by the expected value, which is generated by the given frequency and standing time. Here, the standing time is equal to the in-vehicle time which is specified in the choice. For example, if the in-vehicle time is 20 minutes, when the level for crowding is '1 out of 5 times standing for the whole journey', the expected value is $20 \times (1/5)$ which is 4 minutes standing.

The standard logit model and hierarchical logit model are estimated in ALOGIT (HCG, 2003). The results reported here have been Jack-Knifed (Cirillo et al., 2000) to overcome the repeated measurement problems (section 4.3.8). In SP experiments, there are multiple observations per individual. The repeated measurement problems would happen, where there is correlation between the responses within individuals. It is believed that this problem results in upwardly biased values of the t-ratios (underestimating the standard errors), but does not affect consistently the estimated parameter values of logit models.

Table 7.3 shows results from our model analysis, allowing different scale factors for the 8 data sets. In the table, the t–ratio is against zero or unity, as appropriate. t (0) shows the t-statistics against zero. t(1) shows the t-ratio against unity, to test if the scale factor is significantly different from that of the reference group (SS1) which is set at 1.

Model 7-1 is the standard logit model pooling all the data, without account of different levels of variance (scale factors). Model 7-2 has 7 scale factors to allow different error variances for each data set. In Model 7-2, SS1 is chosen as the reference nest, and the scale factors are obtained in ALOGIT. Among the 7 data sets, the scale factors for SS2, LS1 and LS2 are not significantly different from 1 which indicates that their variances are not significantly different from the reference nest. However, scale factors for SS3, SS4, LS3 and LS4 are significantly different from unity (scale factor for the reference data set SS1) according to t-statistic tests. The t-tests against unity for those data sets are all larger than 1.96, which indicates that the null hypothesis of same scale factor across data sets can be rejected at the 5% level. From the standard errors, the scale factor for SS3 is not significantly different from that of SS4, and scale factor for LS3 is not significantly different from that of LS4. Therefore, SS3 & SS4 and LS3 & LS4 are combined separately in the subsequent analysis.

A likelihood ratio (LR) test has been applied to establish whether the model has been significantly improved (i.e. whether the improvement is significant with 7 extra scale parameters). The test statistic is twice the difference between the log likelihood values of the models at convergence. The calculated statistic (52.8) is significantly larger than the χ^2 critical value at the 5% (14.07) level with 7 degrees of freedom. Therefore, the null hypothesis that constraints on the coefficients are zero can be rejected at the 5% level. Model estimation is significantly improved by allowing different error variance for different data sets.

Table 7.3 Results from model specification Ⅰ

Estimation of Coefficients	Model 7-1 Coef. (t-ratio)		Model 7- 2 Coef. (t-ratio)			Model 7- 3 Coef. (t-ratio)		
Rolling Stock								
RS	0.3491	(8.46)	0.3664	(8.05)		0.3909	(9.67)	
Time	-0.0705	(-12.48)	-0.0915	(-7.62)		-0.0978	(-11.60)	
Cost	-0.0138	(-18.95)	-0.0171	(-8.60)		-0.0182	(-15.93)	
Headway	-0.0555	(-17.60)	-0.0678	(-8.40)		-0.0724	(-15.49)	
Punctuality	-0.3289	(-21.08)	-0.4807	(-8.26)		-0.5092	(-10.25)	
Crowding	-0.1017	(-22.02)	-0.1461	(-7.05)		-0.1538	(-8.75)	
Scale Factors				**t(0)**	**t(1)**		**t(0)**	**t(1)**
SS1 (Reference)			1					
θSS2			1.0747	(7.57)	(0.53)			
θSS3			0.7213	(6.05)	(-2.34)			
θSS4			0.7591	(8.14)	(-2.58)			
θLS1			1.0368	(7.06)	(0.25)			
θLS2			1.1226	(5.82)	(0.64)			
θLS3			0.6263	(6.64)	(-3.96)			
θLS4			0.5608	(7.09)	(-5.55)			
θSLS12(Reference)						1		
θSS34						0.6998	(8.40)	(-3.60)
θLS4						0.5617	(9.15)	(-7.14)
ρ^2 (C)	0.1102		0.1138			0.1136		
LL (C)	-6621.8		-6595.4			-6596.8		
LL test statistic			52.80 (vs. M7-1)			2.76 (vs. M7-2)		
Degree of Freedom			7			5		
χ^2 Critical Value (5%)			14.07			11.07		

The relative values are compared between M7-1 and M7-2. The monetary value of the improved rolling stock (VoS) and in-vehicle time (VoT) are derived from the model estimation, by using Equations 4.30 and 4.31. Table 7.4 presents the comparison of the values.

Table 7.4 Comparison of monetary values from M7-1 and M7-2

	M7-1			M7-2			Diff.
	value	s.e.	t-ratio	value	s.e.	t-ratio	t-ratio
VoS	25.3	2.96	8.55	21.4	1.95	10.97	1.56
VoT	5.1	0.34	15.12	5.4	0.26	20.90	-0.81

The monetary values of the improved rolling stock and in-vehicle time demonstrate differences from the two models. This indicates that the heterogeneity of different sources of data should be taken into account when pooling the data; otherwise it would distort the valuation estimation.

Different data sets have been combined into three groups, which share the similar levels of variance (scale factors are not significantly different). Model 7-3 is the constrained model which the data sets have been combined into three groups: Simple for both short and long journey (SLS12), Complex for short journey (SS34) and Complex (LS34) for long journey.

The scale factors of SS34 and LS34 are smaller than 1 (the scale factor for the reference group), at 0.70 and 0.56 respectively. This confirms what is expected: the complex design (with two more attributes) will cause more variance in the error term, thus yielding a smaller scale factor. The scale factors for the two complex groups are similar, which is encouraging, because they are same type of SP exercises, so are expected to have similar scales. From the likelihood ratio test, Model 7-3 is chosen as the preferred model.

Within Model 7-3, the estimation of coefficients is reasonable, with expected signs. The disutility of travel will fall as the train improves; therefore, the coefficient for the improved rolling stock is positive.

7.3.4 Model specification II

This section explores the second model specification, which assumes that the value of the improved rolling stock depends on the length of the in-vehicle time. In the SP design, the time coefficient is treated differently for each alternative (Chapter 5 and 6), assuming that respondents have different values of in-vehicle time (VoT) in different types of trains. Therefore, the time coefficients are kept separate for each alternative in the utility function, as shown in Equation 7.1.

Table 7.5 shows the results of applying the second model specification. The time coefficient for Pacers (current trains) is slightly higher than that of Super Sprinters (improved trains). This indicates that commuters prefer the improved train, as they are willing to pay more for the time savings in the current train.

The value of the improved rolling stock can be obtained by the difference of VoT in each pair, given certain in-vehicle time. For example, the VoT can be obtained by the ratio of the time and cost coefficients in Model 7-5, as shown below:

$$VoT_{SS} = \frac{\beta_{time}}{\beta_{cost}} = \frac{0.0903}{0.0182} = 4.96\,pence/\min \qquad VoT_P = 5.76\,pence/\min$$

The standard errors for those two valuations are 0.34 and 0.29, which yield the t-ratio for these two values are 15.0 and 19.9 respectively. For a 25-minute journey, the value to improve the train from Pacer to Super Sprinter is 20.05 pence for single journey, if we assume all the other factors are the same.

$VoS_{SS} = (VoT_{SS} - VoT_P) \times 25 = 20.05$ (pence/single trip)

Table 7.5 Results from model specification II

Estimation of Coefficients	Model 7- 4 Coef. (t-ratio)			Model 7- 5 Coef. (t-ratio)		
Time (Super Sprinters)	-0.0884	(-7.00)		-0.0903	(-9.84)	
Time (Pacers)	-0.1030	(-7.90)		-0.1049	(-12.00)	
Cost	-0.0179	(-8.74)		-0.0182	(-15.36)	
Headway	-0.0708	(-8.37)		-0.0722	(-14.76)	
Punctuality	-0.4966	(-8.30)		-0.5031	(-10.08)	
Crowding	-0.1516	(-7.08)		-0.1526	(-8.61)	
Scale Factors						
SS1 (Reference)		t(0)	t(1)		t(0)	t(1)
θSS2	1.0693	(8.01)	(0.52)			
θSS3	0.6998	(6.30)	(-2.70)			
θSS4	0.7396	(8.06)	(-2.84)			
θLS1	0.9261	(7.33)	(-0.58)			
θLS2	1.0290	(5.47)	(0.15)			
θLS3	0.6083	(6.75)	(-4.35)			
θLS4	0.5363	(7.24)	(-6.26)			
θSLS12(Reference)				1		
θSS34				0.7133	(8.18)	(-3.29)
θLS34				0.5676	(8.88)	(-6.76)
ρ^2 (C)		0.1135			0.1132	
LL (C)		-6597.8			-6599.7	

Comparing results from these two model specifications, the log likelihood is slightly better in M 7-3 (-6596.8) than M 7-5 (-6599.7). Both models have the same number of coefficients. It can be concluded that M7-3 is better than M7-5 in explaining the travellers' behaviour. M 7-3 is chosen as the preferred model and the base model for the further analysis.

7.4 Individual's General Preference of the Rolling Stock

The review of previous studies on the valuation of rolling stock found that income and journey purpose contribute to the variation of the stock valuation (chapter 3). Prior to the exploration of design impacts on SP responses, those factors are incorporated into the model analysis in order to avoid the potential confounding effects. The methods used to analyse income and journey purpose effects are discussed in detail in section 4.4.

Section 7.4.1 reports the income effect on the variation of the valuation, without taking account of the journey purpose impact. Section 7.4.2 presents the examination of journey purpose effect on the SP responses. A preferred model is obtained in section 7.4.3, and the valuations derived from the preferred model are reported in section 7.4.4 and 7.4.5.

7.4.1 Income effect

The income effect is analysed by two methods: segmentation and income elasticity.

Income incremental effects on cost

Income is expected to have a strong influence on the individual's sensitivity to cost; with richer people being less sensitive to the price increase. The income incremental effects are analysed by the segmentation method, which is discussed in section 4.3.6.

Table 7.6 shows the impact of income on the estimation of coefficients. Model 7-6 presents the income incremental effect on the cost coefficient. In the survey, respondents were asked to state their household annual income. Five categories were provided starting from less than £10k to over £50k. The option of 'don't want to say' was provided to the respondents who did not want to provide the income information.

In Model 7-6, the base income group combined income bands 'less than £10k' and 'don't want to say'. In a model in which these two categories are kept separate, the coefficient for the option 'don't want to say' is not significant (t=0.31) at the 5% level, which indicates that the coefficient for this category is not significantly different from that of the reference group ('less than £10k'). In addition, in the same model, the LR test ($\rho^2(c)$ is 0.1187, likelihood is -6559.0) cannot rejected the null hypothesises that the coefficient for the population who do not want to specify their income information is equal to zero. Therefore, these two income categories are combined to be the reference income category.

In Model 7-6, significant incremental effects of income on the cost coefficient are found. The coefficients are positive which show that compared with the reference income group (<£10k), people with higher income are less sensitive to the cost change. With the increase in the income, the cost coefficients decrease. The marginal utility of money will fall and monetary values will increase as income increases. This leads to higher value of time for the high income group. The income effects on the monetary value are shown in section 7.5. The model with income effects (M7-6) is significantly improved, compared to the base model (M7-3). The LR test is 75.4, which is larger than the χ^2 critical value (9.49) at the 5% level with 4 degrees of freedom.

Income Elasticity

Model 7-7 explores the income elasticity of value of time. Instead of the segmentation method, a term ' $\kappa \frac{\text{cost}}{Y^{\omega}}$ ' has been incorporated into the utility equation (section 4.4.2). Then, $\frac{\partial VoT}{\partial Y}\frac{Y}{VoT} = \omega$, where Y represents the household income and ω refers to income elasticity.

In the survey, six income categories were specified, including one representing cases where information on household income has not been supplied. From the previous research (Wardman and Bristow, 2004, p.14), the income effects were analyzed by specifying an actual income level based on the mid-point of the income category. This allows the examination of whether income per household provides a better account of households' willingness to pay.

In M 7-7, cost is expressed in pence. The income is the average income for each category. For 'less than £10k', £5000 is chosen as the medium income. And for 'over £50k', £60,000 is chosen as the medium income. For people who do not want to provide the income information, firstly, the average income of the whole income category '£25,000' is chosen. This is following the method using in the study by Wardman and Bristow (2004). The best goodness of fit was 0.1161. Recall the analysis in M 7-6, the income incremental effect for income band 'don't want to say' is not significantly different from the reference income band; and is therefore combined with the base income category 'less than £10k'. In M 7-7, '£5,000' is given to the combined income category, and the best goodness of fit (0.1172) is obtained. M 7-7 is a significant improvement over the model without consideration of income effects (M 7-3); from the LR test.

Table 7.6 Impact of income on the estimation of coefficients

	Model 7-3 Coef.	(t-ratio)	Model 7-6 Coef.	(t-ratio)	Model 7-7 Coef.	(t-ratio)
Income Incremental Effects						
Base (<£10k)			0			
Inc2(cost) (£10-£20k)			0.0018	(1.50)		
Inc3(cost) (£21-£35k)			0.0051	(3.59)		
Inc4(cost) (£36-£50k)			0.0086	(5.45)		
Inc5(cost) (£over 50k)			0.0125	(5.59)		
Income Elasticity						
Cost/income$^{0.18}$					-0.1043	(-14.78)
Estimation of Coefficients						
Rolling Stock	0.3909	(9.67)	0.3899	(9.76)	0.3922	(9.58)
Time	-0.0978	(-11.60)	-0.0982	(-10.99)	-0.0989	(-11.02)
Cost	-0.0182	(-15.93)	-0.0216	(-12.95)		
Headway	-0.0724	(-15.49)	-0.0729	(-14.36)	-0.0731	(-14.66)
Punctuality	-0.5092	(-10.25)	-0.5153	(-10.57)	-0.5127	(-10.48)
Crowding	-0.1538	(-8.75)	-0.1557	(-8.68)	-0.1554	(-8.81)
Scale Factors						
θSLS12(Reference)	1		1		1	
θSS34	0.6998	(8.40)	0.7038	(8.68)	0.6999	(8.64)
θLS34	0.5617	(9.15)	0.5473	(9.52)	0.5550	(9.56)
ρ^2 (C)	0.1136		0.1187		0.1171	
LL (C)	-6596.8		-6559.1		-6570.0	
LL test statistic			75.4 (vs. M7-3)		53.6 (vs. M 7-3)	
Degree of Freedom			4		1	
χ^2 Critical Value (5%)			9.49		3.84	

Table 7.7 shows the searching process to find the best fit of the model which considers the income elasticity ω. Different values were given to test the model. It has been found that when the income elasticity is 0.18, the goodness of fit for the model $\rho^2(C)$ is the best. The searching process finds a range of values for ω which leads to the best $\rho^2(C)$. For example, when it is set as 0.2, the goodness of fit for the model is 0.1172. However, the log likelihood of the model achieves the best value (-6570.0), when the value of ω is set as 0.18.

Table 7.7 Searching process for the income elasticity

ω	**0.30**		**0.20**		**0.18**		**0.15**	
Coefficients		t-ratio		t-ratio		t-ratio		t-ratio
$\cos t / inc^{\omega}$	-0.2978	(-13.69)	-0.1250	(-14.60)	-0.1043	(-14.78)	-0.0793	(-15.02)
RS	0.3740	(9.18)	0.3900	(9.52)	0.3922	(9.58)	0.3948	(9.64)
Time	-0.0900	(-10.34)	-0.0979	(-10.93)	-0.0989	(-11.02)	-0.1002	(-11.16)
Headway	-0.0686	(-13.89)	-0.0725	(-14.53)	-0.0731	(-14.66)	-0.0737	(-14.82)
Punctuality	-0.4914	(-10.35)	-0.5101	(-10.47)	-0.5127	(-10.48)	-0.5156	(-10.46)
Crowding	-0.1492	(-8.72)	-0.1547	(-8.81)	-0.1554	(-8.81)	-0.1562	(-8.81)
Scale Factors								
θSLS12(Ref.)	1		1		1		1	
θSS34	0.7253	(8.58)	0.7028	(8.63)	0.6999	(8.64)	0.6956	(8.61)
θLS34	0.5716	(9.58)	0.5565	(9.58)	0.5550	(9.56)	0.5525	(9.50)
ρ^2**(C)**	0.1159		0.1172		0.1172		0.1171	
LL(C)	-6579.3		-6570.2		-6570.0		-6570.9	

The segmentation analysis of income elasticity by the journey purpose was conducted. The parameter estimation for each journey purpose is significant, and model has been significantly improved (ρ^2(C) = 0.1222). However, the search process found that the best fit was achieved when ω was equal to 0.18 for each journey purpose group.

Wardman (2001) conducted a meta-analysis of VoT, which covered 20 cross-sectional studies. The household income elasticity for in vehicle time (IVT) was found to be 0.6. Mackie et al. (2003) found that in the 'preferred constant elasticity model specification of the AHCG data set, the income elasticities were +0.36 for commuting and +0.16 for other, both with respect to household income'. The cross-sectional income elasticity tends to be somewhat below unity (Gunn, 2001). In the environmental studies conducted by Wardman and Bristow (2004), an income elasticity of 0.7 was obtained.

In the present study, the value of income elasticity 0.18 is below unity; which agrees with the previous evidence. It is similar to the values obtained from the Tyne crossing data sets (Wardman, 2001), where an income elasticity 0.20 was obtained for commuters and 0.30 for the leisure group. However, the income elasticity from this study is lower than the values obtained from the previous meta-analysis; which are normally around 0.5.

The reasons suspected for this low income elasticity are: firstly, the sample may well include some low income users who have high value of time such as students. Secondly, in practice, household income is used as a segmentation variable (Wardman, 2001). In the present study, household income rather than individual income has been requested in the survey form. Previous studies have obtained higher income elasticities by using the individual income in the model specification (Fowkes, 1986).

7.4.2 Journey purpose effect

In addition to the income effects, the journey purpose has shown a significant impact on explaining the variation of valuations. As discussed in section 4.4.3, the impact of journey purpose on the estimation of coefficient has been incorporated into the model. In this study, the journey purpose has been categorized into 5 types which are 'Commuters', 'Employer's Business (EB)', 'Personal Business (PB)', 'to and from School/College (School)' and 'Leisure'. 'Leisure' group contains 'Visiting friends/relatives', 'Sport/Entertainment' and 'Shopping'. 'Commuters' group is selected as the reference group in the segmentation model.

People with different journey purposes are expected to have different sensitivity to time, headway, punctuality and crowding. The segmentation analysis of journey purpose is presented in the following preferred model.

7.4.3 Preferred model

Table 7.8 shows the estimation results for the segmentation models. M 7-9 is the full MNL model which incorporates the impact of journey purpose and income on the estimation of the coefficients. From the LR test, M 7-9 is significantly improved than M 7-3. M 7-10 is the constrained model in which the non-significant variables are removed, and some of the variables are combined if the coefficients from separate analysis are not significantly different from each other, which are presented in italics in the table.

Comparing M 7-10 with M 7-3, the LR test is 223.6. The χ^2 critical value is 21.03 with 12 degrees of freedom at the 5% level. M 7-10 is significantly improved from M 7-3, which is chosen as the preferred model for further analysis.

Findings from the preferred model

From M7-10, income shows a significant incremental effect on the estimation of the cost coefficient, which is same as expected. Richer people are less sensitive to cost changes. Journey purpose demonstrates significant influence on the estimation of time, headway, punctuality and crowding coefficients.

Table 7.8 Segmentation model on impacts of income and journey purpose

	Model 7- 3 (Reference MNL)		Model 7- 9 (Full Model)		Model 7- 10 (Preferred)	
ASC Segmentation		(t-ratio)		(t-ratio)		(t-ratio)
Base (Commuters)	0.3909	(9.67)	0.3949	(7.25)	0.4048	(7.39)
+ Leisure			-0.2796	(-2.22)	-0.3273	(-3.98)
+ Employer's Business			-0.3352	(-1.46)	*-0.3273*	*(-3.98)*
+ Personal Business			0.1245	(0.54)		
+ To/from School			-0.3093	(-2.33)	*-0.3273*	*(-3.98)*
+ Reimburse			0.2803	(1.72)	0.2806	(2.20)
+ Male			0.1273	(1.98)	0.1268	(1.97)
Cost						
Cost (Base)	-0.0182	(-15.93)	-0.0205	(-13.40)	-0.0206	(-13.76)
+ Cost - Inc3 (£21-35k)			0.0031	(1.95)	0.0031	(1.93)
+ Cost - Inc4 (£36-50k)			0.0064	(3.46)	0.0063	(3.34)
+ Cost - Inc5 (over 50k)			0.0099	(4.34)	0.0100	(4.41)
Time						
Time (Commuters)	-0.0978	(-11.60)	-0.1012	(-10.46)	-0.1007	(-10.50)
+ Leisure			0.0543	(4.75)	0.0504	(3.27)
+ Employer's Business			-0.0222	(-1.16)	-0.0250	(-1.89)
+ Personal Business			-0.0234	(-1.05)	*-0.0250*	*(-1.89)*
+ To/from School			-0.0235	(-1.34)	*-0.0250*	*(-1.89)*
Headway						
Headway (Commuters)	-0.0724	(-15.49)	-0.0770	(-13.88)	-0.0771	(-14.75)
+ Leisure			0.0316	(3.41)	0.0309	(3.11)
+ Employer's Business			-0.0013	(-0.09)		
+ Personal Business			-0.0085	(-0.44)		
+ To/from School			0.0032	(0.31)		
Punctuality						
Punctuality (Commuters)	-0.5092	(-10.25)	-0.5815	(-10.14)	-0.5793	(-10.39)
+ Leisure			0.2642	(3.58)	0.2617	(3.52)
+ Employer's Business			0.1450	(0.94)	0.1351	(2.33)
+ Personal Business			0.1655	(1.37)	*0.1351*	*(2.33)*
+ To/from School			0.1382	(2.02)	*0.1351*	*(2.33)*
Crowding						
Crowding (Commuters)	-0.1538	(-8.75)	-0.1549	(-7.66)	-0.1526	(-8.50)
+ Leisure			-0.0475	(-1.38)	-0.0494	(-1.48)
+ Employer's Business			-0.0097	(-0.20)		
+ Personal Business			0.0595	(1.39)		
+ To/from School			0.0089	(0.37)		
Scale Factors						
θLS12	1		1		1	
θS34	0.6998	(8.40)	0.6849	(8.73)	0.6872	(8.89)
θL34	0.5617	(9.15)	0.5486	(9.06)	0.5492	(9.16)
ρ^2 **(C)**	0.1136		0.1289		0.1286	
LL (C)	-6596.8		-6482.9		-6485.0	
LL test statistic			227.8		- 4.23	
Degrees of Freedom			25 (vs. M 7-3)		13 (vs. M 7-9)	
χ^2 Critical Value (5%)			37.65		22.36	

The alternative specific term (ASC) is incorporated into the model to represent individuals' preferences for the improved rolling stock. Segmentation on the ASC term has been conducted by the journey purpose and individuals' characteristics, such as the gender and reimbursement of the ticket. It is expected that respondents whose tickets are being reimbursed will have a higher value of rolling stock because they are less sensitive to the fare change. From the model estimation, a positive effect (0.2806, t = 2.20) is found for the group whose tickets have been reimbursed. Males demonstrate a higher preference for improved rolling stock than females.

Travellers with different journey purposes show different level of preference for the improved rolling stock. Among them, commuters and PB travellers give the highest coefficient for the improved rolling stock; which led to higher monetary values. The possible reason is that commuters are assumed to be more frequent travellers, so that they are more sensitive to the improvement of the train condition. Travellers for leisure, EB and School have a relatively lower value for improved rolling stock. The lower value for the business group (EB and PB) can be explained as their tickets are normally reimbursed (refers to Table 6.14), and M 7-10 takes account of the incremental effects of reimbursement into the model.

The segmentation on the time, headway, punctuality and crowding coefficients by journey purpose has been done to identify the taste variation among groups of people. People who travel for business and school/college have the highest value of time coefficient. Compared with other groups, leisure travellers have a lower value of headway. Commuters give punctuality the highest value. Leisure travellers are the most sensitive to the crowding in their journey. The monetary values and time units' value for variables are calculated in section 7.4.4.

7.4.4 Values from preferred model

This section presents the monetary values derived from the preferred model (M 7-10). The impacts of journey purpose and income on the variation of the valuation are discussed.

The monetary value of the variables are obtained by the ratio of the marginal utilities of target coefficient and cost coefficient, using Equation 4.30 (section 4.4.4). For example, the value of in-vehicle time (VoT) is obtained by the ratio of marginal utilities of time and cost. The incremental effect of the income and journey purpose is taken into account when we calculate the monetary value.

The values of variables can be estimated in units of in-vehicle time. This is obtained from the ratio of marginal utilities of variables (such as headway) and time, shown in Equation 7.2:

$$VoX_{k\cdot IVT} = \frac{\beta_{X_k}}{\beta_{\text{time}}} = \frac{VoX_k}{VoT} \qquad \textbf{Equation 7.2}$$

The time units' value presents the marginal utility of time effects on the values of other variables. The time units' value does not need to adjust for GDP and price index; therefore it may be compared with previous evidence (see more details in section 7.5.1). For example, the VoH is found to be 0.78 times of value of IVT for rail users (Wardman, 2004) if the journey distance is 2 miles.

Average value of the improved rolling stock (VoS)

The average VoS is achieved from the weighted mean of individuals'/samples' VoS. Each individual/sample VoS is the marginal utilities of the constant term and cost coefficients, which may include the incremental effects of the design factors and respondents socio-economic information, as shown in Equation 7.3.

$$VoS = \frac{\sum (VoS_m \cdot n_m)}{\sum n_m}$$ **Equation 7.3**

In Equation 7.3, subscript m represents the category of respondents from which the value is derived. n is the sample size for this category.

Value of in-vehicle time (VoT)

Table 7.9 shows the VoT derived from the preferred model (M7-10), with the standard error in the brackets. Income contributes to the variation of VoT as expected in that the value of time increases with the increase of income. Richer people have a higher VoT compared with people from lower income groups.

VoT varies with journey purpose as expected. Among them, business travellers and people who travel to and from school have the highest VoT; commuters' VoT is higher than that of leisure travellers. School travellers show a high VoT. This partly explains the low income elasticity of the study (section 7.4.1). It would be expected that students have lower incomes. Students have a higher VoT, thus yielding low income elasticity. The VoTs are compared with the previous evidence in section 7.5.1.

Table 7.9 Monetary value of rail travel time for different journey purpose (p/min)

Journey Purpose	Commuters			EB/PB/School			Leisure		
		(s.e.)	(t)		(s.e.)	(t)		(s.e.)	(t)
Income (<£20k)	4.89	(0.36)	(13.6)	6.10	(0.56)	(10.9)	2.44	(0.71)	(3.4)
Income (£21- 35k)	5.75	(0.69)	(8.3)	7.18	(1.02)	(7.0)	2.87	(0.88)	(3.3)
Income (£36-50k)	7.04	(1.02)	(6.9)	8.79	(1.43)	(6.1)	3.52	(1.11)	(3.2)
Income (Over 50k)	9.50	(1.80)	(5.3)	11.86	(2.42)	(4.9)	4.75	(1.60)	(3.0)
ave.	5.66			6.54			2.64		

Value of the improved rolling stock (VoS)

Table 7.10 shows the VoS according to income and journey purpose. The VoS varies with income in the expected manner. Journey purpose demonstrates a significant impact on the variation of the VoS. Commuters and personal business (PB) travellers have a higher VoS compared to the other group. Employer business (EB) travellers have a very low value. One possible reason is that their tickets are normally paid by others. This is confirmed by the data description (Table 6.14), which demonstrates that 75% of the EB travellers get reimbursement of their tickets.

The incremental effects of gender and reimbursement of ticket are found to be significant in the model estimation (refers to Table 7.8). It is found males normally have a higher preference for improved rolling stock. And travellers whose tickets are reimbursed have a higher willingness to pay for the improved stock. The impacts on the variation of the VoS are calculated and presented in Table 7.10.

Table 7.10 VoS for different journey purpose (p/single trip)

Journey Purpose	Income (<£20k)	Income (£21- 35k)	Income (£36-50k)	Inc 5 (Over 50k)
Commuters/PB	19.65	23.13	28.31	38.19
Leisure/EB/School	3.76	4.43	5.42	7.31
+ Reimburse	13.62	16.03	19.62	26.47
+ Male	6.16	7.25	8.87	11.96

In Table 7.11, the VoS is adjusted by the gender for all journey purpose categories. The VoSs for PB and EB groups are adjusted by incorporating the incremental effect of reimbursement, as it is found that lots of travelers in these two categories get reimbursement of their tickets.

Table 7.11 Joint VoS by gender and reimbursement (p/single trip)

Journey Purpose	Income (<£20k)	Income (£21- 35k)	Income (£36-50k)	Income (Over 50k)	Average (% of Fare paid)	
Commuters	21.83 (2.75)	27.00 (4.18)	34.68 (4.39)	48.37 (10.21)	26.39	8.8%
Employer Business	16.07	19.38	25.68	33.08	21.16	7.1%
Personal Business	25.23	31.01	39.59	54.99	26.87	9.0%
School/Leisure	5.85 (2.98)	7.36 (3.63)	10.96 (4.64)	13.22 (6.29)	6.25	2.1%

PB travellers have the highest value, which is on average 26.87 pence per single trip (ppst). Commuters are willing to pay 26.39 ppst. EB travellers on average would like to pay 21.26

pence to improve the train. Finally, people who travel to and from school and leisure travellers have the lowest WTP for the improved stock.

For this study, the average cost is £2.5-3.5 per single trip. If the average fare is set as £3, the equivalent value for the improved rolling stock for 'Personal Business' is 9.0% of the fare paid. The values are examined in section 7.5.2.

Value of other variables

Time units' values are calculated. By transferring the monetary values to the time units' value, it avoids adjusting for GDP growth and removes the income effect. Table 7.12 shows the value of different variables relative to the IVT, varying by different journey purpose category.

Table 7.12 Implied IVT values of other variables

	Rolling Stock	Headway	Punctuality	Crowding
Commuting to and from work	4.66	0.77	5.75	2.52
Employer's Business	3.47	0.61	3.53	2.21
Personal Business	4.11	0.61	3.53	2.21
To/from School	1.02	0.61	3.53	2.21
Leisure	2.36	0.92	6.31	5.02

The time units' value of improved rolling stock varies across different journey purposes. PB travellers have the highest value of improved rolling stock in time units. Commuters have the second highest value of the improved rolling stock at 4.66 times of the VoT.

Headway, presented through a frequency attribute in SP experiments, determines how long passengers will have to wait at a station, and how closely they can time their departure or arrival to their ideal requirements. The VoH varies with different journey purpose. It has been found that commuters have a higher VoH than business travellers, and leisure travellers have the highest VoH. The marginal utility of time effect on the headway (VoH_{IVT}) for commuters is 0.77, which agrees with the previous evidence (Wardman, 2004).

Two more service attributes (punctuality and crowding) have been added into the complex SP experiments for testing the research hypotheses. In the experiment, the punctuality is presented as an amount of time delay (late time in relation to timetable) with a given frequency, for example, '1 out of 5 times delay for 10 minutes'. The punctuality is modelled by the expected value (Section 7.4.3). From values obtained from Table 7.12, commuters have a higher time value of the punctuality compared with respondents who travel for business. The leisure group has the highest time value of punctuality relative to the VoT. The reason suspected is that the leisure group has a very low VoT, yielding a higher time unit value of punctuality.

The value of crowding is calculated in time units, as shown in Table 7.12. In the analysis, crowding is presented by a given frequency and the length of standing time, and quantified by the expected value. The expected value is equal to the standing time (specified by the IVT) multiplied by the given frequency. Therefore, the standing time value is the incremental value of standing time relative to seated time. For example, if the IVT value is 5p/min, and standing value is 2p/min, then the true standing value is 7p/min. Leisure travellers have the largest penalty of crowding (5.12 times of the value of IVT). Commuters have a lower value of crowding compared with leisure and EB travellers.

In summary, income incremental effects have been found for each value, which indicates that richer people have a higher value for time savings and higher penalty for service deterioration. Journey purpose shows significant effect on the value of the time and other service attributes. Normally, people who travel for business have a higher VoT and VoH compared with other groups. This agrees with the previous evidence (Wardman 2001, 2004). Values in time units have been calculated for comparison, as they do not need to adjust by GDP and is not affected by different income categories.

Further examination of the preferred model is presented in section 7.5, through the comparison of values obtained from this study with the previous evidence. A discussion of various discrepancies is also presented.

7.4.5 Interval of the values

To calculate the interval of the values, two methods are discussed in section 4.4.4. This section compares the intervals of the values using these two methods.

Fowkes (1998) derived a formula for the variance of the value (Equation 4.31). This formula is assumed that the parameter estimates follow a normal distribution. This method is very simple to apply and easy to account for the incremental effects of socio-economic features.

The second method (Armstrong et al., 2001, see Equation 4.33) takes into account of the fact that the maximum likelihood estimation method yields coefficients that are asymptotically distributed multivariate normal (Ben-Akiva and Lerman, 1985). Consequently, the point estimate of the monetary value (such as VoT) is a random variable by an unknown PDF (the probability distribution for the ratio between two normally distributed variables is unknown a priori). It was found that this method represents the case studies more accurately; however, it is tedious and needs considerable computing efforts (Armstrong et al., 2001, p.144).

Two methods are compared using an example of VoT (commuters) derived from the preferred MNL model (M 7-10). Table 7.13 shows the comparison of the intervals of VoT. The significance level is set as 5% for both cases. Figure 7.2 demonstrates the comparison.

Table 7.13 Comparison of methods to achieve the intervals of VoT

VoT (Commuters)	Point estimate of VoT		Interval of VoT			
			Fowkes' method (F)		Armstrong et al. method (A)	
	value	(s.e.)	Low	High	Low	High
Income (<£20k)	4.89	(0.36)	4.18	5.40	4.16	5.58
Income (£21- 35k)	5.75	(0.69)	4.40	7.10	4.56	7.34
Income (£36-50k)	7.04	(1.02)	5.04	9.04	5.38	9.65
Income (Over 50k)	9.50	(1.80)	5.97	13.03	6.8	14.85

From the table and figures demonstration, some findings can be achieved:

- Using Armstrong et al.'s method, the interval's mid-point tends to be equal to the point estimate of the VoT (mean value) in the lower income categories (£ 36-50k), and is smaller than that in the higher income categories. By using Fowkes' method, the interval's mid-point is equal to the mean value of VoT.

- The interval obtained by Armstrong's method is similar as that by Fowkes' method in the lower income category, but tends to be slightly wider in the higher income category.

- Smaller confidence interval is obtained in the lower and mid-level income category than highest income category by using two methods.

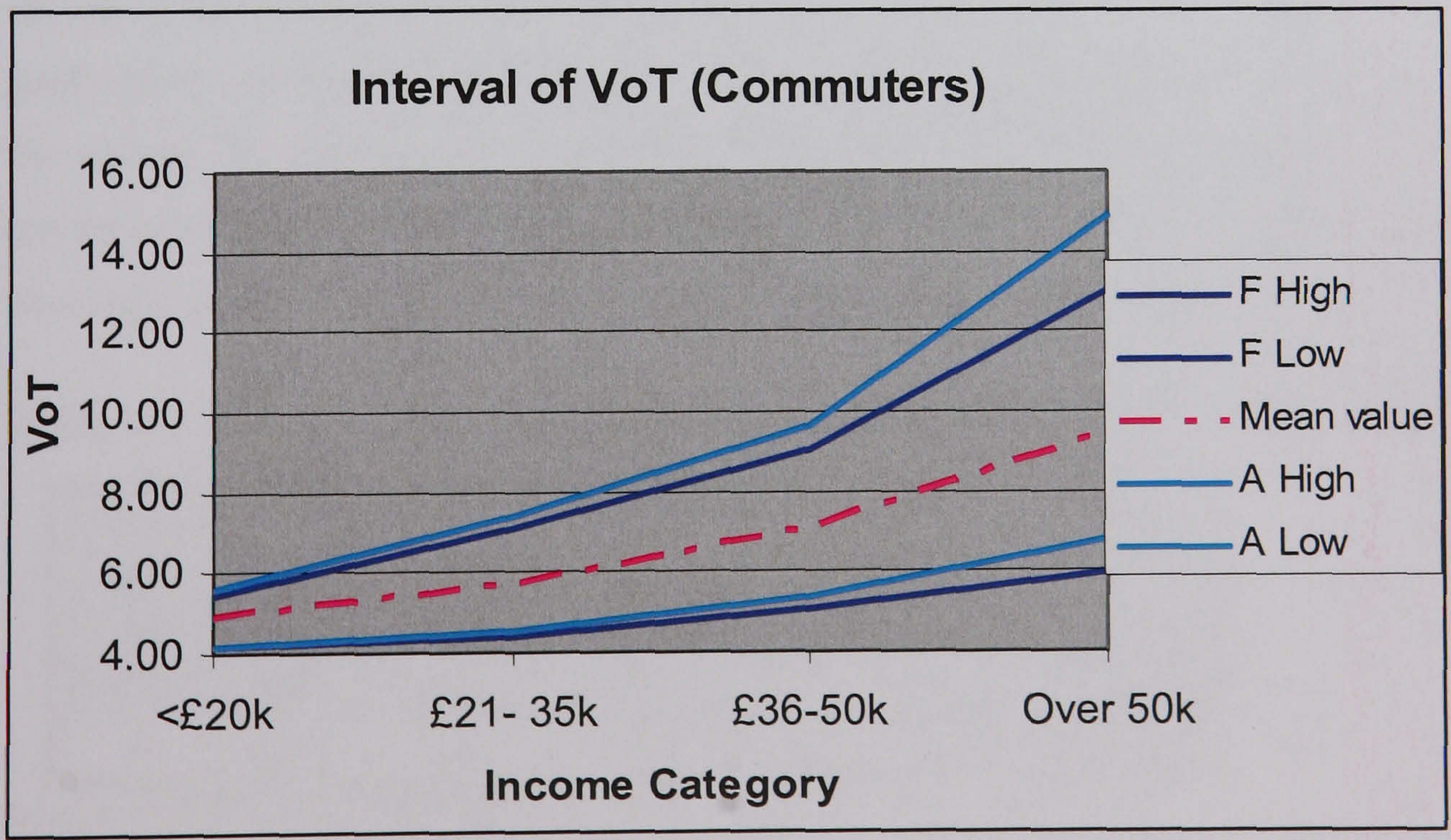


Figure 7.2 Comparison of methods to achieve the interval of VoT

These findings are consistent with Armstrong et al. (2001). Using Armstrong et al.' method, the VoS point estimate (mean value) is always smaller than the interval's mid-point, and only tends to become equal to it for a larger number of observations. In the current study, the lower income is the majority of the whole sample (refer to Table 6.12). They also found that smaller confidence interval is derived from more significant parameters. In the current study, (refer to Table 7.9), the standard deviation is getting larger for the higher income category.

Considering the large sample size in this study, we can conclude that Fowkes' method provides an easy and practical way to obtain good confidence intervals for the monetary value. Therefore, the comparison of the intervals in the following sections is based on this method.

7.5 Comparison with the Previous Evidence

7.5.1 The value of in-vehicle time

Numerous studies have been conducted on the VoT (see MVA et al., 1987; Mackie, et al., 2003; Wardman, 2004). The VoT is found to vary by the different income, journey purpose and journey distance categories. Wardman (2004, p.368) reviewed the previous studies on the VoT and obtained the average VoT for Commuting, Leisure and Business at 7.2, 6.3 19.2 pence/minute respectively. These values account for income variations across years by assuming a GDP elasticity of one. The VoT for commuters from this study is in line with the empirical values found from the meta-analysis; whilst the values for leisure and business group are slightly lower.

Table 7.14 presents the recommended value of rail travel time from PDFH (2005). The selected values in the table include the distance, journey purpose and 'South East' effect. The 'South East' effect refers to the finding that the IVT is higher for London and South East travellers. As the survey is carried out in Greater Manchester, the journey distance of respondents is commonly shorter than 25 miles. Therefore, the values of time for the category that journey distance equal or shorter than 20 miles are selected for comparison.

Table 7.14 Selected values of rail travel time (pence per minute at 2000 Q4 levels)

Distance	Business		Leisure		Commuting	
(miles)	1st Class	Std Class	South East	Non SE	South East	Non SE
10	40.4	19.0	8.1	6.9	8.9	7.7
25	47.7	22.5	9.5	8.2	10.6	9.1
50	70.2	33.1	14.0	12.1	15.5	13.3

Source: PDFH (2005) Table B 3.13

GDP growth factor needs to be taken into account. It has been found that in the previous studies GDP affects the VoT (MVA et al., 1987; Hague Consulting and Accent, 1999; Wardman, 2001, 2004). The elasticity of GDP is given as 0.723 in the meta-analysis by Wardman (2001, 2004). The VoT in Table 7.14 is updated by taking account of the growth of GDP and changes in prices. In the PDFH (June 2005, B3, p19), it is recommended that business values are adjusted directly but non-business values are increased using the GDP elasticity of 0.723. The following equation is applied to obtain the updated value as suggested in the PDFH book.

$$VoT\ (\text{in 2005 Q4 prices}) = VoT(\text{ in 2000 Q4 prices}) \times (\frac{\text{RPI index at 2005 Q4}}{\text{RPI index at 2000 Q4}}) \times (\frac{\text{GDP}_{2005}}{\text{GDP}_{2000}})^{0.723}$$

Equation 7.4

RPI Index at 2000 Q4: 172.1

RPI Index at 2005 Q4: 193.6

Real GDP per capita 2000: 17581

Real GDP per capita 2005: 19393

For example, the VoT for commuting (journey distance is 25 miles) is:

$$VoT\ (\text{in 2005 Q4 prices}) = 9.1\ \text{p/min} \times (\frac{193.6}{172.1}) \times (\frac{19393}{17581})^{0.723} = 10.99\ \text{p/min}$$

Table 7.15 Comparison with the updated recommended value (p/min at 2005 Q4 levels)

Distance	Business			Leisure		Commuting	
(miles)	1st Class	Std Class	This study	Non SE	This study	Non SE	This study
10	48.8	22.9	6.54	8.3	2.64	9.3	5.66
25	57.6	27.2		9.9		11.0	

The values from the present research are compared to the updated values, as shown in Table 7.15. The average VoT obtained from this research is lower than the recommended values by PDFH for the each journey purpose category. This difference may be due to Greater Manchester having lower income than average.

Table 7.16 Selected values of time in the context of rolling stock studies (p/m)

	Inc 1 (<£10k)	Inc 2 (£10-30k)	Inc 3 (£30-40k)	Inc 4 (>£40k)
Business	8.5	9.9	12.4	17.7
Commuting	4.9	5.7	7.2	10.3
Leisure	4.9	5.7	7.2	10.3

Sources: Wardman and Whelan (2001) Table 9

In the context of valuation of the improved rolling stock, Table 7.16 presents selected values from the empirical evidence by Wardman and Whelan (2001). The values in this study are in line with the values from the study by Wardman and Whelan (2001), although the value for the leisure group is slightly lower than their results.

Discussion on the lower value of IVT

The VoT from this study is obtained from the SP data, which is found to be slightly lower than previous empirical evidence. Wardman (2001, 2004) found that the VoT being estimated from SP data is lower than that from other sources of data (the effect is -0.13). This implies an average value which is 15% lower than values estimated on RP data.

It is argued that a strategic response bias is more likely to influence the cost term. Wardman(2001, p.120) stated that "If respondents regard cost to be more likely to vary than other attributes, it will attract more strategic bias and the resulting greater sensitivity to cost variations will imply lower monetary valuations." In the present research, cheap-talk message is added into some of the SP experiments (see chapter 5). The impacts will be discussed in chapter 8. It is expected that adding this message would affect respondents' sensitivities to the cost attribute. Higher cost sensitivity can also be explained by respondents' simplification strategy in the SP exercise (Wardman, 2001). Cost is expected to be less likely ignored than other attributes, its coefficient will be relatively larger and hence the monetary values smaller.

Further analysis will be conducted in chapter 8, to test if strategic bias and task complexity effects exist in the SP responses.

7.5.2 The value of the improved rolling stock

In this study, the value of the improved rolling stock (VoS-Super Sprinters versus Pacers) is obtained (refer to Table 7.11). The average valuation is around 8% of the fare, varying by journey purpose and income category.

No previous evidence can be found in exactly the same experimental context (replacing Pacers with Super Sprinters). Table 7.17 presents some selected values from the previous improved rolling stock studies which are relevant to the current experimental context.

The value from the present study is consistent with a few previous studies. For example, in the first study, the fare change is 8% for the new sliding door stock replacing older slam door stock. In study 2, for mode choice users, between coach and rail, it is valued 7% for replacing Pacers with Electric Sprinters. And in Suffolk Rail Study, the value of new electric trains with air conditioning and improved interiors and seating is equal to a 9% increase in fare. However, when compared with Study 5, the value obtained from this study is much higher than the values found by Wardman and Whelan (2001). Study 6 presents the recommended values of improved rolling stock recommended by PDFH (2005). Compared with the recommended values, the monetary value of this study is much higher.

Table 7.17 Selected value of improved rolling stock

	Studies	Stock Types	Money Value	Time Value
1	Passenger reaction to new suburban carriages	New sliding door stock replacing 20-30 year old slam door stock	8%	
2		Sprinters –Pacers (Car-rail)	0	
		Electric Sprinters –Pacers (Car-rail)	4%	
		Sprinters –Pacers (Coach-rail)	0	
		Electric Sprinters –Pacers (Coach-rail)	7%	
3		Existing diesel Sprinters v new air conditional electric stock		10%
4	Suffolk Rail Study	The value of new electric trains with air conditioning and improved interiors and seating, compared to the existing Sprinters,	9%	
5	Valuation of Rolling Stock by Wardman and Whelan(2001)	Express Sprinter v Sprinter	0.9%	1.9
		Net worker v Sprinter	0.7%	0.8
		Express Sprinter v SE Slam Door	1.5%	3.0
6	Values recommended by PDFH (2005)	None air-conditioned modern sliding door South East electric multiple units replacing non air-conditioned slam door electric multiple units.	2.5%	
		Air-conditioned sliding door diesel multiple units (158/159 Express Sprinters) replacing non air-conditioned sliding door diesel multiple units (150/156 Sprinters)	1%	
7	Present Study	Super Sprinters vs. Pacers in Greater Manchester	8%	

Study 1, 2 and 3 are selected from the review by Wardman and Whelan (2001) Table 1
Study 6 is selected from PDFH (2005) Table B5.2

The above comparison is based on the overall VoS. The VoS varies according to a few factors (See review in Chapter 3), such as journey length, income and journey purpose.

In this study the average journey time is around 25 - 30 minutes. Time units' values of rolling stock from previous studies (Wardman and Whelan, 2001) are selected for comparison, with account of the journey length (40 minutes) and income categories, as shown in Table 7.18.

In the present study, time units' values of the improved rolling stock for commuters (4.66) and leisure travellers (2.36) are generally in line with the values listed in Table 7.18 for the same category. The business travellers have a higher time unit's value of the stock than that found by Wardman and Whelan (2001).

Table 7.18 Comparison of the values of rolling stock in the time unit

	Wardman and Whelan(2001)				Present Study
	<£10k	£10-30k	£30-40k	>£40k	
Business	0.7	0.8	1	1.5	4.11
Commuting	2.3	2.7	3.4	4.8	4.66
Leisure	1.3	1.5	1.9	2.7	2.36

Sources: Wardman and Whelan (2001) Table 9

In summary, the monetary values of rolling stock for commuters and business travellers from this study are consistent with the previous relevant SP studies; however they are found to be larger than the recommended values from the PDFH, possibly due to strategic bias. A discussion of the inconsistency is presented in the following section.

Discussion

The VoS are presented in Table 7.11 for different income bands and journey purpose categories. On average 8% of the fare is obtained for the improved rolling stock (Super Sprinters) from the current stock (Pacers). This value is consistent with the previous relevant SP rolling stock studies; however, it is much higher as the PHFH (2005) recommended values.

Chapter 3 presents a review of factors influencing the variation of rolling stock valuation. From a demand analysis using ticket sales data before and after the introduction of the improved rolling stock, the VoS is about one third of the value obtained from the SP estimates. This is possibly due to strategic bias in the SP studies (Wardman and Whelan, 2001).

Wardman and Whelan (2001) conducted a meta-analysis on the valuation of the improved rolling stock, which covered 18 SP studies. They found that the VoS was much higher if respondents could easily perceive the aim of the study to be the stock valuation. They concluded that many SP surveys on the valuation of rolling stock were biased because the purpose of the study was clear (to test the market for new rolling stock) and this induced strategic responses. The incentive for respondents to strategically bias their answer was suspected to be that respondents perceive that their responses have an impact on the provision of improved rolling stock, thus they have an incentive to overestimate its valuation.

The PDFH stated "valuations of new or improved rolling stock in excess of 10% of the fare have been routinely obtained. In some cases, the values far exceed this amount. If rolling stock really was valued so highly, we would expect to detect its effect on demand from analysis of ticket sales data. After all, reliable fare elasticities are commonly estimated to ticket sales data containing variations in fare generally somewhat less than 10%."

In this study, the average value of the improved rolling stock is 8 % of the fare, after controlling for the impact of income and journey purpose. For commuter, 8.8% of fare is obtained for improved rolling stock. These values are smaller than 10% which is encouraging, and consistent with values obtained from the previous SP studies in the relevant contexts. However, they are higher than the recommended values by PDFH which are obtained from combined evidence (for example, ticket sales data analysis). For leisure and school travellers, the value is 2.1%. This agrees with the recommended values in the PDFH book.

Is the higher value caused by a strategic bias as suggested from the previous research? In the present study, two factors: cheap talk and adding more complexity to mask the aim of research are added to the SP experiments as explained before (Chapter 5). The impacts of two factors on the valuation of rolling stock will be explored and presented in the next chapter.

7.5.3 The value of service attributes

The Value of Headway (VoH)

The value of headway is presented in time units (Table 7.12), varied on the journey purpose effect. 0.77 is obtained for the commuters group. 0.61 is obtained for the business group (EB and PB) and school travellers. The leisure group has the largest time value of headway (0.92).

The meta-analysis of service attributes carried out by Wardman (2004) found that the VoH is varied on journey distance, purpose and mode. Table 7.19 presents the comparison of the IVT value of headway between the selected values from the meta-analysis and the present study.

Table 7.19 Comparison of the implied IVT values of headway

Miles	Previous Study*		Present Study	
	Business	Non-business	Business	Non-business
2	0.96	0.78	0.61	0.77
10	0.70	0.57		
50	0.51	0.41		

Source*: Wardman (2004) Table 12

It has been found that the magnitude of values in this study generally agree with the empirical evidence. In the meta-analysis by Wardman (2001), the headway is valued at 0.8 min of in-vehicle time with a narrow confidence interval of ± 0.8%. Conventionally, the railway industry in Great Britain has used values in the range of 0.4-0.7 (ATOC, 1997).

In summary, the value of headway in this study agrees with the previous empirical evidence, and also with the expectation that the VoH is less than the VoT and that the difference can be substantial (Wardman, 2004, p375).

The Value of Punctuality (VoP)

In the complex design SP experiments, two more service attributes have been added into the survey form for testing the research hypotheses. Punctuality is presented as an amount of time delay (late time in relation to the timetable) with a given frequency, for example, '1 out of 5 times delay for 10 minutes'. Punctuality is modelled by the expected value (Section 7.3.3), varied on different journey purposes.

Wardman (2001, p.112) conducted a meta-analysis and obtained 26.79 pence for the average VoP from 14 cases where the expected value was used. The overall VoP is 29.47 pence which is consistent with the meta-analysis result.

Due to the different ways used to present the delay/delay distribution, the value/penalty of the delay is different across studies. The recommended VoP in the PDFH is 2.5 to 3.0 times of the IVT unit. In the meta-analysis by Wardman (2001), the overall time valuation of late time is valued as 7.40 (S.D. =3.86).

In this study, the value of punctuality for commuters is 5.75 times of the value of IVT. Business travellers (EB and PB) and people who travel to and from school value the penalty of delay as 3.53 times of in-vehicle time saving. Leisure travellers have the largest value of punctuality at 6.31 times of IVT value. The values are higher than the recommended value presented above.

The reasons for the larger time value of the punctuality are suspected as: firstly, most of the respondents are morning commuters, and their journey distance is generally short (average 30 minutes journey). Commuters are expected to have a very high penalty for a delayed arrival. The second possible reason is due to the serious punctuality issue observed during the on-site work. During the main survey, due to the engineering work conducted in Huddersfield, a delay occurred on the relevant routes. This is also confirmed by the comments from the respondents. It is suspected that the VoP in this study is slightly biased upwards. Respondents would like to pay more for an improvement in punctuality. The higher time unit value of the punctuality for leisure group is suspected to be due to their much lower VoT (section 7.4.4), which yield a very high time units' value of punctuality.

The Value of Crowding (VoC)

In the SP experiment, the crowding is presented as the standing time in the train. As explained in section 7.4.4, the VoC is obtained by the incremental value of standing time plus the in-vehicle time. Table 7.20 shows the VoCs, in both monetary and IVT values with account of different journey purposes.

Table 7.20 Value of crowding (standing) from present study

	Monetary Value (p/min)	Time Value (minutes of IVT per min.)
Commuters	15.04	2.52
PB/EB/School	13.36	2.21
Leisure	14.16	5.02

Wardman (2003) showed that respondents overestimated the VoC. The work covered 23 valuations from 8 studies. The mean value of standing time relative to seated time in the 20 instances where the purpose of the study would clearly have been seen as valuing overcrowding was 3.5. This fell to 2.7 across the three values from 2 studies where overcrowding was an element of a broader study looking at aspects of mode choice and interchange. However, from the RP analysis (LRT, Operational Research, 1988), the standing time was valued at between 1.4 and 2.2 times seated time.

In market research by MVA (1991) on rolling stock studies carried out around London, it was found that the value of crowding was 1.2 times journey time for commuters, 2.7 times journey time for business and 3.2 times journey time for leisure. In this study, the IVT value of crowding is 2.52 for commuters and 5.02 for leisure travellers. This is higher than the values found by the MVA studies.

Table 7.21 Recommended crowding penalty for passengers outside of London (p/min)

Load Factor (Seating)		Leisure	Business	Commuting
100%	Stand	22.0	48.0	6.5
120%	Stand	26.4	50.5	7.5
140%	Stand	30.8	53.0	8.5
160%	Stand	-	-	9.5

Source: PDFH (2005) Table B5.1

Table 7.21 shows the recommended crowding penalty for passengers outside of London in 2000 quarter four price and income index from PDFH (2005). The value of crowding here is presented by considering the loading factors and journey purposes. After adjustment by inflation, the commuters' crowding penalty in this study is consistent with the recommended value in the PDFH. However, for Leisure and Business travellers, the penalties of crowding are less than the recommended value.

PDFH (C4, p.25) stated that: "The empirical evidence is not always entirely plausible, and it is by no means consistent across different studies. There is the possibility that the results from these SP studies have been influenced by response bias". A common feature of previous crowding analysis is that the over crowding is valued very highly. This partly reflects that respondents gave high penalty to the standing in the vehicle.

In this study, a possible reason for a lower VoC for leisure travellers is that most of the leisure travellers in the sample travelled after 9:30 in the morning (off peak hour). During off peak time, the train has enough space. Hence, there is no concern of crowding for leisure travellers. Another reason suspected is that the aim of survey is to investigate the VoS. In this situation, respondents have no incentives to bias their answer towards overcrowding in the survey (Wardman, 2001).

7.6 Summary

This chapter presented the initial analysis of respondents' preference of different rolling stocks. Respondents generally preferred the Super Sprinter to the Pacers.

In order to test the research hypothesis, different SP experiments were developed. To pool the data from different SP designs, scale factors were taken into account to allow different error variance across datasets. A hierarchical MNL model was developed allowing different levels of error variance in the data sets. The pooled model was significantly improved compared with the model without account of the heteroskedasticity of different data sets. Smaller scale factors were obtained for the complex design, which implied larger error variance in the responses from complex SP experiments.

Two model specifications were tested in the analysis. It was found the first model specification where the rolling stock is estimated as a constant term was better than the other one. Therefore, the first model specification was chosen as the preferred one.

In the initial analysis, income and journey purpose were incorporated into the model to identify individuals taste variation, thus avoiding the confounding effects. From the model estimation, income showed significant incremental effects on the cost coefficient estimation. Different journey purposes were found to have a significant impact on the coefficient estimation of time, headway, punctuality and crowding.

Most of the monetary values of different attributes in the survey were found to be reasonable and in line with the previous empirical evidence. The value of improved rolling stock was found to be slightly higher than the recommended values from PDFH (2005).

In summary, a base model was set initially by using the conventional logit model. The base model controlled several factors which might potentially confound the subsequent analysis of estimation bias, such as the income and journey purpose impacts. From the comparison with the previous studies, it can be concluded that the initial model is a reasonable model to form the basis for more detailed analysis on the research hypotheses.

Chapter 8 extends the base model, and investigates the impacts of SP design on the estimation bias. Research hypotheses are tested to examine the existence and possible consequence of the bias. Some measures to amend the bias such as cheap-talk and adding more attributes to mask the research aim are examined to identify their impacts on the SP responses. The impacts of individuals' perceptions on the variability of SP responses are discussed.

Chapter 8

The Impacts of SP Design on Responses

8.1 Introduction

Chapter 7 established a base model for users' valuation of the improved rolling stock. The base model controlled several factors (i.e. income and journey purpose) which cause the variation of valuations to avoid their potential confounding effects. This chapter explores the effects of design factors (cheap-talk and complex design) on SP responses from the base model.

Section 8.2 presents qualitative analyses of the design impact on respondents' perceptions. Some first impressions are obtained. Choice models are developed to explore the effects of cheap talk and complex design in section 8.3. A standard MNL model is applied to find the impact of design factors on the valuation. A Heteroskedastic Multinomial Logit (HMNL) model is developed to find the impact of design factors on the precision of model estimation. The heterogeneity of individuals' capabilities of making choice is incorporated into the utility function by the parameterization of the scale parameter. Sections 8.4 and 8.5 summarise the impact of cheap-talk and complex design on SP responses. Section 8.6 explores the influence of individuals' perceptions on their responses, both in the variation of the valuation and the consistency of choice making. Section 8.7 presents the interpretation and discussion of the result. Section 8.8 ends the chapter with a conclusion of effects of SP design on responses.

Figure 8.1 presents the outline of model development within this chapter. The qualitative analysis on the impacts of design factors on SP responses is conducted in Stage 1 (Section 8.2). Stage 2 examines the impacts of cheap-talk and complex design on SP responses, in terms of variation of the valuation and precision of model estimation (sections 8.3-8.5). Stage 3 explores the influence of respondents' perceptions on SP responses (section 8.6). Conclusions will be arrived at based on the findings from the model estimation.

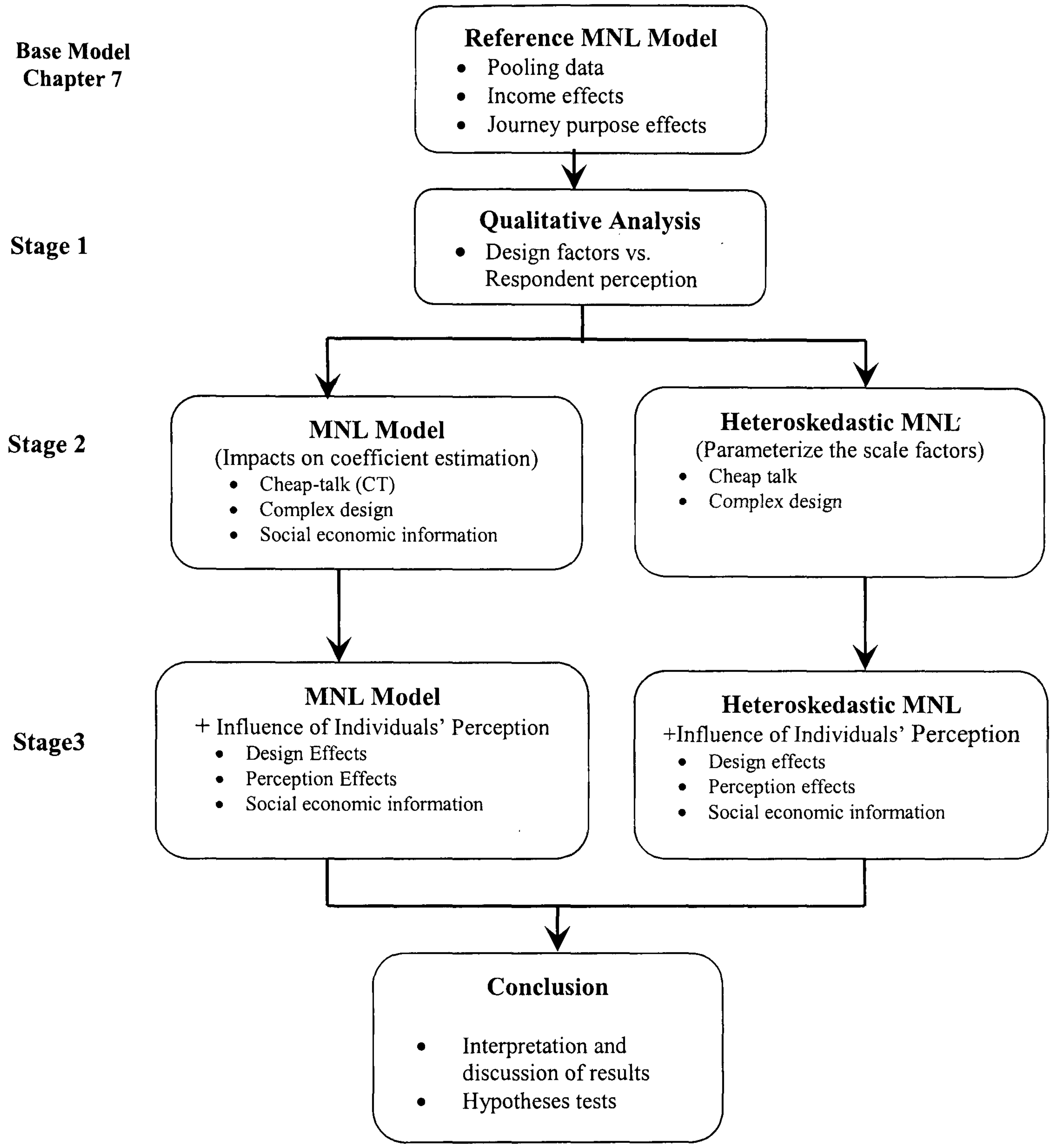


Figure 8.1 Model development outline of Chapter 8

8.2 Qualitative Analysis

8.2.1 Introduction of design factors in the SP experiment

Chapter 1 presented the research hypotheses for this research. A series of SP experiments were developed (chapters 5 and 6) to test these hypotheses. Among them, a cheap-talk (CT) script and adding more attributes (namely Complex Design CD) to amend incentives to strategic bias were added into the SP experiments to detect their impacts on individuals' choice making processes. The development of two design factors (CT and CD) was presented in chapters 5 and 6. Prior to the further model analysis, some qualitative analyses are conducted.

8.2.2 Qualitative analysis of cheap talk

In some of the SP experiments, a CT script was provided to respondents prior to SP choices to see if CT can amend the incentive to strategic bias. The CT script in this study is a message that discusses the bias found in previous studies and reminds respondents of the budget constraints in the valuation. The development of CT script is demonstrated in section 5.6.

A qualitative analysis is conducted to see if adding a CT script would change individuals' perceptions of the survey, for example, their perceptions of the cost change for the introduction of the improved stock. One follow-up question was provided in the SP questionnaire to probe respondents' perceptions of the potential price increase:

- "How likely do you think it is that fares would increase if new trains would be introduced?"

Four options from "Not at all" to "Very" are presented to respondents.

Table 8.1 and Figure 8.1 show the relationship between the adding of CT script and respondents' perceptions of the potential price increase. Among the 1222 respondents in the study, 607 respondents completed a survey without a CT warning message; whilst 615 of them had the CT script in the questionnaire. From the table and figure demonstration, adding CT does not show a significant impact on respondents' perceptions in each category.

Table 8.1 The influence of cheap-talk on respondents' perceptions of price increase

Perceptions of Price Increase	Not at all		Slightly		Moderately		Very likely		No Answer		Total
	(1)	(1/6)	(2)	(2/6)	(3)	(3/6)	(4)	(4/6)	(5)	(5/6)	(6=1+2+3+4+5)
No Cheap Talk	11	2%	47	8%	113	19%	432	71%	4	1%	607
With Cheap Talk	6	1%	57	9%	131	21%	416	68%	5	1%	615

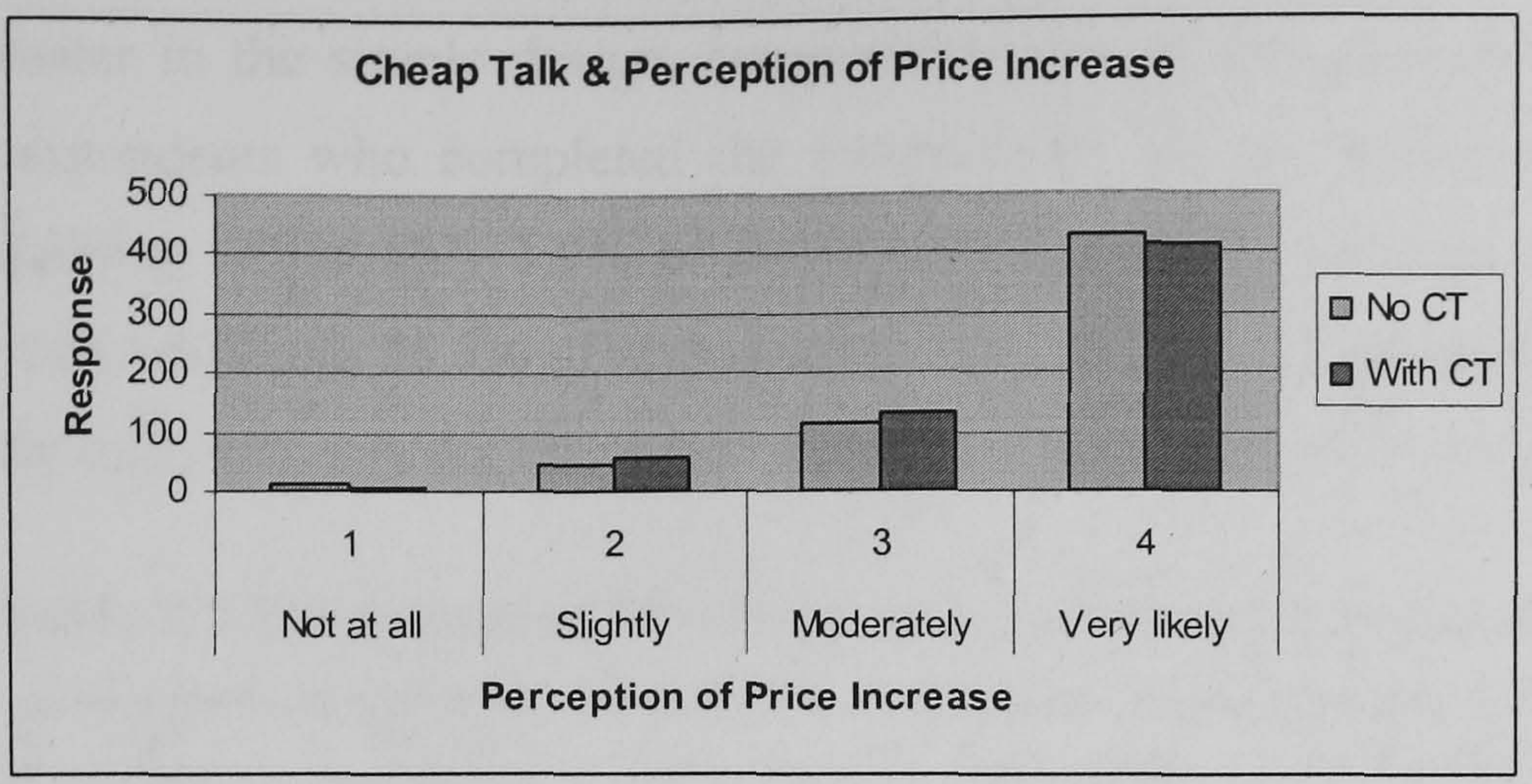


Figure 8.2 The influence of cheap-talk on respondents' perceptions of price increase

Moreover, the strength of the relationship between adding CT and the perceived chances of price increase is calculated by the Cramer's V coefficient (Bryman and Cramer, 2005). The process of computing Cramer's V coefficient is explained in section 4.4.5. Cramer's V coefficient varies between 0 and 1 to indicate the strength of relationship between two nominal variables that have more than two categories. The closer V is to 0, the smaller the association between the two categorical variables.

The coefficient is 0.05, which demonstrates that the relationship overall is very low between the adding of CT script and perceived chances of price increase. From the above analysis, no clear evidence implies that adding CT has significantly affected respondents' perceptions of the potential price increase. Section 8.6.2 presents more discussion from the model analysis.

8.2.3 Qualitative analysis of complex design on SP responses

In some of the SP experiments, two more attributes are added into the SP design, to detect its impact on amending incentives to strategic bias in the choice making (see section 5.7).

Adding two more attributes to the SP experiments might add to the task load of respondents, thus leading to the task complexity effect in their choice making. The review of task complexity effects (section 2.7) indicates that the number of attributes affects the precision of coefficient estimation (i.e. the variance of the error term in utility function increases) and causes differences in the valuations between the complex and simple designs. A question probing respondents' perceptions of difficulty in choice making is provided in the questionnaire.

- "Did you feel it difficult in making the choice?"

Four options are provided to respondents: 'Yes, very', 'Yes, quite', 'Yes, a little' and 'No'.

Table 8.2 presents the relationship between the number of attributes and individuals' perceived difficulty in the choice making. It is clearly shown that respondents felt the choice making was easier in the simple design, compared to that in the complex design. For example, 3.9% of respondents who completed the complex SP choices selected 'very' difficult in the choice making; whilst only 1.9% of the respondents who completed the simple SP choices selected 'very' difficult. 56.5% of respondents in the simple experiment felt 'no difficulty' in completing the task, and only 46.4% of the respondents in the complex experiment felt not difficulty.

Table 8.2 The impact of SP design on respondents' perceptions of difficulty (%)

SP Design	No. of Attributes	Yes, Very		Yes, quite		Yes, a little		No Difficulty		Total
		(1)	(1)/(5)	(2)	(2)/(5)	(3)	(3)/(5)	(4)	(4)/(5)	(5)=(1+2+3+4)
Simple	3	12	1.9%	75	12.1%	183	29.6%	349	56.4%	619
Complex	5	23	3.9%	95	16.1%	199	33.7%	274	46.4%	591

The Cramer's V coefficient is computed for the relationship between the complexity of the SP design and respondents' perceived difficulty to make the choice. 0.11 was obtained which illustrates that the relationship is low but exists. The χ^2 produced in the computing process is 14.87 with 3 degrees of freedom (χ^2 critical value is 7.81 at the 5% level). This indicates the complexity of SP design can affect respondents' perceptions of difficulty in the choice making.

Based on the findings from qualitative analyses, in the next sections, models are developed to explore the influence of design factors on users' valuation of improved rolling stock.

8.3 Models Exploring the Impact of Design Factors on SP Responses

8.3.1 The effect of cheap talk

The second research hypothesis (H2, see section 1.2.3) states that "adding the cheap talk script can amend individuals' incentive to strategic bias". Bias refers to the overestimation of the improved rolling stock valuation in the SP experiment. This hypothesis is tested from three aspects:

<u>Cheap talk & ASC</u>

The ASC term is included into the utility function to capture the difference of alternatives (Train, 2003), and the tendency to choose the specific alternative (DeShazo and Fermo, 2004). In this study, the ASC term refers to the preference for the improved rolling stock – Super Sprinters. $\delta_{CT} d_{CT}$ is included in the utility function as part of the constant term to characterize the impact of the CT script associated with values of the improved rolling stock. d_{CT} is the dummy variable denoting the existence of the CT script. δ_{CT} measures respondents' tendency to attend to the alternative (the improved rolling stock in this experiment) with the CT script.

Cheap talk & other attributes

This is to examine if adding CT affects the taste parameter of attributes in the utility function, thus affecting the valuation of the improved rolling stock. For example, the CT script reminds respondents of the budget constraints in the SP experiment. It is expected that CT increases respondents' sensitivity to the cost coefficient. This test is conducted by including $\gamma_{CT.X_{ik}} d_{CT} X_{ik}$ in the model. $\gamma_{CT.X_{ik}}$ investigates the incremental effect of CT on the estimation of attribute X_{ik} (alternative i, and the k^{th} attribute) in the utility function.

Cheap talk & estimation precision

This is to examine if adding CT script affects individuals' decision strategy or causing more/fewer errors in SP responses. The test is conducted by a Heteroskedastic Multinomial Logit (HMNL) model. The model specification and estimation are presented in section 8.3.4

8.3.2 The effect of adding two more attributes to the SP experiment

Two more attributes are added into some of the SP experiments to see if adding more attributes can mask the research aim and then amend the incentive for respondents to strategically bias their answer. By adding more attributes, it may be hoped that respondents will exhibit less bias. This is partly due to the extra effort required merely to complete the exercise with bias, but it is more likely to be due to respondents failing to see any single clear purpose to the experiment.

The potential drawback of this method is that more attributes might add to the task complexity for respondents, which is found to affect the estimation precision and/or magnitude of valuations. The research hypothesis (H3B) states that "*An increase in the number of attributes will always increase the variance of error terms.*"

We call the SP designs with two more attributes 'complex' in the subsequent analysis. Similarly, the impact of complex design will be examined through two aspects: on the estimation of coefficients (through the incremental effects on the attributes) and the choice consistency (through the scale factor by using the HMNL model).

In this study, respondents' actual perceived difficulty was explored with follow-up questions, which will be demonstrated in section 8.6.3. The initial qualitative analysis (section 8.2.3) found that the number of SP attributes significantly affects respondents' perceived difficulty of the SP experiment.

8.3.3 MNL model estimation

Model specification

A utility function is initially set for each response, as in Equation 8.1. The first part of the utility function is the same as the base model in the previous chapter, accounting for the influences of respondents' characteristics, denoting by dummy variable d_{ky} (whether an observation is in y^{th} group of n groups in a category). The incremental effect of factor d_{ky} on the estimation of attribute X_{ik} is measured by γ_y. And the impact of d_{ky} on the estimation of ASC is measured by δ_y Design factors are included in the model by dummy variables, relative to a base scenario. The dummy variables denoting the factors are defined in Table 8.3.

Table 8.3 Definition of dummy variables of SP design factors

Design Variables (d_{Design})	
d_{CT}	is 1 if the survey contains the cheap talk
d_{CD}	is 1 when the design is complex (with two more attributes)
d_{JD}	is 1 when the journey distance is long

$\delta_{Design} d_{Design}$ is included into the utility function as part of the constant term to characterize design factors' impact associated with evaluating one alternative. δ_{Design} measures respondents' tendencies to attend to the improved rolling stock when the design is in a specific category d_{Design}. The design factors include the cheap talk (CT), complex design (CD) and journey distance (JD), which are represented by the dummy variables defined in Table 8.3.

$$\boxed{\begin{aligned} U_i = ASC + \sum \beta_{ik} X_{ik} + \sum (\gamma_y d_{ky} X_{ik}) + \sum (\gamma_{Design} d_{Design} X_{ik}) \\ + \sum (\delta_y d_y) + \sum (\delta_{Design} d_{Design}) \end{aligned}}$$

Equation 8.1

Model estimation

The estimation is based on 10885 preference observations from 1222 individuals. Initially, a full MNL model is established using Equation 8.1. Dummy variables representing the design factors are included in the utility function. The likelihood ratio (LR) test is applied to detect model improvement. 41.0 is obtained which is significantly larger than the χ^2 critical value at 5% (28.9) with 18 degrees of freedom. The model account for the design factors is statistically improved compared with M 7-10, which indicates that including the design factors in the model estimation can better explain respondents' choice making.

The non-significant variables are then removed from the full MNL model. After several attempts, M 8-2 is selected as a preferred model, shown in Table 8.4. The interpretation of the coefficient estimates for attributes is provided below.

Findings from the MNL model

- The coefficient for the dummy variable denoting the presence of the CT script in the SP experiment shows a negative sign in the estimation of the constant term (ASC, refers to the preference for the improved rolling stock); however the impact is not statistically significant (t=-0.34) at the 5% level (M 8-2a), thus being removed from the model. This implies that CT does not show a significant impact on the estimation of ASC term.

- CT shows significant incremental effects (t>1.96) on coefficients of time, cost and crowding attributes: increases the value of time and cost coefficients and decreases that of the crowding coefficient.

- Adding more attributes to the SP design (complex design) does not show a significant impact on the estimation of the ASC term. Some interaction effect (positive) of long journey distance and complex design on the estimation of the ASC is found from the model estimation, which leads to a higher valuation of the improved rolling stock (VoS) in the experiment in this category.

- Adding more attributes to the SP experiments shows a significant incremental effect on the cost coefficient. This is demonstrated by a negative value (-0.0016) of the incremental effect, although the impact is not significant at the 5% level.

- Similar to M7-10, the scale factors for the complex design (θS34 / θL34) are smaller than that of (θLS12) the simple design. Recall that the scale factor inversely relates to the variance of the error term in the utility function; this indicates that complexity contributes a larger variance in the error term.

- The full model analysis found that the interaction effect of CT and adding more attributes on the estimation of other attributes (represented by the incremental effects) is not significant, thus removing them from the subsequent model analysis.

In summary, the model taking account of design factors is significantly improved (at the 5% level) compared to the reference MNL model. It can better explain individuals' decision behaviours. The detailed interpretation of findings with discussion is presented in the sections 8.4 and 8.5.

Table 8.4 Initial standard logit model allowing the difference among different datasets

Estimation Coefficients (t-ratio)	Model 8-2a (MNL)		Model 8-2b (MNL)	
Time				
Time (Commuters)	-0.0875	(-7.53)	-0.0880	(-10.95)
+ Leisure	0.0511	(3.23)	0.0512	(3.25)
+ EB/PB/School	-0.0258	(-1.94)	-0.0258	(-2.12)
Cost				
Cost (Base)	-0.0187	(-9.55)	-0.0186	(-12.36)
+ Cost - Inc3 (£21-35k)	0.0031	(1.94)	0.0031	(2.20)
+ Cost - Inc4 (£36-50k)	0.0064	(3.32)	0.0064	(2.41)
+ Cost - Inc5 (over 50k)	0.0101	(4.37)	0.0099	(2.48)
Headway				
Headway (Commuters/EB/PB/School)	-0.0774	(-15.27)	-0.0772	(-14.44)
+ Leisure	0.0308	(3.01)	0.0308	(3.02)
Punctuality				
Punctuality (Commuters)	-0.5966	(-9.23)	-0.5952	(-10.37)
+ Leisure	0.2707	(3.56)	0.2721	(4.63)
+ EB/PB/School	0.1390	(2.27)	0.1396	(2.91)
Crowding				
Crowding (Commuters/EB/PB/School)	-0.1817	(-7.32)	-0.1814	(-8.61)
+ Leisure	-0.0551	(-1.56)	-0.0564	(-1.49)
ASC Segmentation				
Commuters/PB	0.4033	(5.81)	0.3915	(8.73)
+ Leisure/EB/School	-0.3314	(-4.00)	-0.3314	(-3.72)
+ Reimburse	0.2914	(2.22)	0.2856	(2.88)
+ Male	0.1209	(1.79)	0.1203	(1.40)
Design Variables on ASC				
+ CT (Adding Cheap Talk)	-0.0247	(-0.34)		
+ JD*CD (Long Distance * Complexity)	0.1813	(1.38)	0.1771	(1.29)
On the Other Attributes				
+ CT*Time	-0.0270	(-3.55)	-0.0255	(-2.65)
+ CT*Cost	-0.0034	(-2.23)	-0.0034	(-2.02)
+ CT*Crowding	0.0487	(3.26)	0.0487	(2.86)
+ CD*Cost	-0.0015	(-0.99)	-0.0016	(-1.01)
Scale Factors				
θLS12	1		1	
θS34	0.6667	(8.71)	0.6606	(10.70)
θL34	0.5358	(8.46)	0.5357	(11.21)
ρ^2 **(C)**	0.1303		0.1303	
LL (C)	-6472.2		-6472.8	
LL test statistics	15.45		24.47	
Degree of Freedom	12 (vs. full MNL)		5 (vs. M 7-10)	
χ^2 Critical Value (5%)	21.0		11.07	

8.3.4 Heteroskedastic multinomial logit model estimation

The Heteroskedastic Multinomial Logit (HMNL) model (Swait and Adamowicz, 2001) allows the analyst to incorporate the complexity and cognitive burden of the SP experiment through an appropriate parameterisation of the scale parameter (see section 4.3.5).

HMNL model is selected to analyse impacts of SP design features and individuals' heteroskedacity on choice making. This model is found to be commonly used to test the task complexity effect in the literature (see section 2.7). As the scale parameter is inversely proportional to the variance, we would expect that the complex decision choice process would lead to higher error rates and thus a lower scale parameter. Simply put, when the task get more complex, respondents would make more errors in their choice making, which leads to a large error variance in the utility function. Researchers can detect the level of error variance by seeing the magnitude of the scale factor.

The strength of HMNL model is that the scale parameter (across all alternatives) is not a constant term, but is decomposed to be a function of alternative-specific variables and/or individuals' decision behavioural information. The scale factor is different across observations. This model is also used to identify the influence of respondents' perception (section 8.6) on the error variance, thus investigating the influence on choice making consistency.

Model specification

The probability function is shown in Equation 8.2. There are two alternatives in the utility function: improved trains (denoted by S-Super Sprinters) and the current trains (P-Pacers). The model specification is the same as the MNL model (M8-2), differs in the scale factor. As stated above, the scale factor in the MNL model is a constant term for each dataset. In the HMNL model, the scale parameter is not a constant term across the dataset, but is a function of the following factors:

- SP design – presence of Cheap Talk, the number of attributes and the range of attribute levels across two alternatives;
- Individuals' socio-economic information

$$P_S = \frac{\exp(\lambda_s V_S)}{\exp(\lambda_s V_S) + \exp(\lambda_s V_P)}, \quad \lambda = \exp(\alpha_{Design} d_{Design} + \alpha_y d_y)$$

Equation 8.2

Here, λ is the scale parameter, which is a function of the above factors, measured by parameter α. d_y is the dummy variable denoting the category of individuals' socio-economic features, and d_{Design} is the dummy variable denoting SP design features, shown in Table 8.3.

Different levels of attributes (bands) are set in the SP design to make the experiment more realistic. The band represents the journey distance length, where A refers to the shortest and D is the longest. The band in SP design also partly captures the range of levels for attributes (cost and headway) between two alternatives. The time coefficient is kept absolute for the two alternatives in the SP design; whilst difference are taken for cost and headway coefficients for two alternatives (see sections 5.4.4 and 5.4.5). Band A has the smallest difference of attributes, where the comparison between the two alternatives is the smallest, band B and C has the same difference of attribute range, and band D has the largest difference in levels of attributes.

The exponential function is applied to ensure λ larger than zero (DeShazo and Fermo, 2002). If all the parameters α turn out to be zero, then λ equals one and the MNL model is obtained. Table 8.5 presents the interpretation of the estimated parameter from the HMNL model analysis on choice consistency, assuming all the other parameters in the λ function are equal to zero.

Table 8.5 The interpretation of the parameter from the HMNL model results

Parameter of the factor	Scale parameter λ_s	Impact on the precision of model estimation
$\alpha > 0$	$\lambda_s > 1$	Fewer errors in the model analysis
$\alpha < 0$	$\lambda_s < 1$	More errors in the model analysis
$\alpha = 0$	$\lambda_s = 1$	Does not change the level of errors

Model estimation

To estimate the HMNL model, several software packages were tried. Alogit is a friendly and powerful tool. In the present study, all the MNL models were estimated by this package. The Jack-knife function in the software can be used to solve the repeated measurement problems. However, the current version does not enable the estimation of the parameterization of the scale factor. BIOGEME (Bierlaire et al., 2004) was initially used to analyse the HMNL model. However, it requires very complex model specification to parameterize the scale factor. Finally, we adapted a code in GAUSS (Aptech Systems, 1997) from Caussade et al. (2005), using the MAXLIK routine to maximize the values of the log likelihood function. Test models were estimated using GAUSS and BIOGEME. The same specification led to the consistent results from different software packages. Therefore, in this study, all the HMNL models were estimated by the GAUSS package. It took a long time (>24 hours) to reach the convergence.

Table 8.6 shows results from the HMNL model (M 8-3a and M 8-3b). In M8-3, rather than allowing difference of the datasets by constant scale factors for each group (as with the previous standard logit model), the heterogeneity of the respondents different ability of handling the choice task is incorporated into the utility function by parameterization of the scale parameter.

A LR test was conducted. The log-likelihood of M 8-2b (-6472.8) is smaller than that of M 8-3a (-6450.5). These two models are not directly nested as the scale factor in the MNL model is constant for each dataset (see section 7.3.2); however, the scale factor in the HMNL model is a function of design features and individuals' socio-economic characteristics, which is different across each choice set. For example, in the MNL (M8-2b) model the scale factor (θL34) for the dataset from the simple design of long journey distance is 0.5357. In the HMNL model, the scale factor for this data set is decomposed as design features: complex design (two more attributes) and long journey distance (reflected by the design band).

During the data analysis, another HMNL model was estimated with constant scale factor for each dataset, and then the scale factor within each dataset was decomposed to capture the design features and individuals' different capabilities of choice making. This model is more like a Heteroskedastic Nested Logit model which has the same model specification as M8-3a, but has two more scale factors for different datasets (SS34, LS34). The log-likelihood of this model is -6443.44, which indicates that the model significantly improves, compared to M8-2a and M8-3a. However, the t-ratios for these two scale factors are not significantly different from zero, therefore being removed from the preferred model, and this model is not presented.

An approximate likelihood ratio test has been conducted. The likelihood ratio $[-2(LL(\hat{\beta}^{H}) - LL(\hat{\beta}))]$ is 44.6, which is larger than the χ^2 critical value at 5% (9.5) with 4 degrees of freedom. M 8-3a (HMNL) is better, compared to M 8-2, as explaining individuals' behaviours. This indicates that allowing for heteroskedacity across observations improves the model fit and interpretation. M 8-3b is a constrained HMNL model, which the non-significant variables ($t<1.60$) are removed and similar coefficients are combined. From the likelihood ratio tests, M 8-3b is selected as a preferred HMNL model.

In the process of estimation of HMNL models, the repeated measurement problem cannot be overcome in the GAUSS program. In the review of repeated measurement (see section 4.3.8), it is believed that the repeated measurements effect does not significantly affect values of coefficients in the logit model, and only has a small effect in reducing the significance of variables.

Table 8.6 Estimation results from the MNL and HMNL models

Estimation Coefficients	Model 8-3a (HMNL)		Model 8-3b (HMNL)	
Time				
Time (Commuters)	-0.1505	(-7.51)	-0.1458	(-7.82)
+ Leisure	0.0851	(3.92)	0.0825	(3.89)
+ EB/PB/School	-0.0529	(-3.08)	-0.0502	(-3.00)
Cost				
Cost (Base)	-0.0343	(-9.26)	-0.0324	(-9.87)
+ Cost - Inc3 (£21-35k)	0.0068	(4.19)	0.0064	(4.08)
+ Cost - Inc4 (£36-50k)	0.0102	(3.95)	0.0103	(4.13)
+ Cost - Inc5 (over 50k)	0.0158	(4.58)	0.0156	(4.68)
Headway				
Headway (Commuters/EB/PB/School)	-0.1371	(-9.91)	-0.1301	(-10.48)
+ Leisure	0.0518	(3.71)	0.0486	(3.55)
Punctuality				
Punctuality (Commuters)	-1.0004	(-8.62)	-0.9440	(-9.10)
+ Leisure	0.5022	(4.38)	0.4624	(4.32)
+ EB/PB/School	0.2403	(2.69)	0.2222	(2.70)
Crowding				
Crowding (Commuters/EB/PB/School)	-0.3197	(-8.10)	-0.2953	(-8.32)
+ Leisure	-0.0862	(-2.08)	-0.0879	(-2.22)
ASC Segmentation				
Commuters/PB	0.7620	(7.33)	0.7460	(9.09)
+ Leisure/EB/School	-0.5820	(-4.52)	-0.5686	(-4.52)
+ Reimburse	0.4698	(2.95)	0.4238	(2.79)
+ Male	0.0597	(1.32)		
Design Variables on ASC				
+ CT (Adding Cheap Talk)	-0.1226	(-1.09)		
+ JD*CD (Long Distance * Complexity)	0.2307	(1.47)		
- On the Other Attributes				
+ CT*Time	-0.0449	(-2.46)	-0.0380	(-2.29)
+ CT*Cost	-0.0052	(-2.32)	-0.0055	(-2.75)
+ CT*Crowding	0.0897	(3.12)	0.0779	(2.98)
+ CD*Cost	-0.0010	(-0.50)		
Parameterization of Logsum				
Adding two more attributes	-0.5964	(-7.72)	-0.5320	(-7.28)
Distance Band B	-0.6015	(-6.21)	-0.4895	(-5.59)
Distance Band C	-0.4551	(-4.99)	*-0.4895*	*(-5.59)*
Distance Band D	-0.9175	(-7.25)	-0.8799	(-7.00)
Income Group 4 & 5	0.1180	(1.55)	0.1409	(1.86)
Adding of CT	0.0367	(0.65)		
ρ^2 **(C)**	0.1333		0.1325	
LL (C)	-6450.5		-6455.9	
LL test statistics			10.89	
Degree of Freedom			6(vs. M 8-3a)	
χ^2 Critical Value (5%)			12.59	

Findings from HMNL model

As stated above, the HMNL model (M 8-3) has the same model specification as the MNL model (M8-2). All parameters in M 8-3 have the expected and the same sign as that from the M8-2 estimation, which confirms the findings from M8-2 (section 8.3.3). These will not be repeated here. The valuations obtained from M8-3 and M8-2 will be compared in section 8.3.5.

In contrast to the MNL model, the HMNL model can also capture the impact of design features and individuals' heterogeneity by decomposing the scale parameter as a function of factors of interest. In this study, the impacts of the presence of the cheap-talk, adding two more attributes, the range of the attributes and individuals' income are tested on the error variance. Before interpreting estimated coefficients, recall that there exists an inverse relationship between scale parameter and the variance of the random component: when a variable affects the scale parameter negatively, it affects the model precision positively.

Adding two more attributes to the SP experiments has a significant negative effect on the scale parameter. The estimated coefficient corresponding to the number of attributes is -0.5320 with a t-statistic of (-7.28). Adding two more attributes has a clear detrimental effect on respondents' ability to choose, contributing to a higher error variance. This concurs with what we expected and the previous evidence (DeShazo and Fermo, 2002, p.136; Caussade et al., 2005, p.631).

The band of SP design has affected the estimation precision significantly. In the function of scale parameter, band A is chosen as the reference ($\alpha_{banda} = 0$). The coefficients of dummy variables corresponding to bands B, C and D are significantly different from 0 (reference). The coefficients associated with band B and C are not statistically different (t=-1.6) from each other at the 5% level in M 8-3a, thus being combined in the HMNL model. The coefficient for band D is the lowest which indicates the largest level of variance. It tries to conclude that with the increase of the level difference between two alternatives, the variance of error term increases. Simply put, the wide range of levels among alternatives contributes to an important higher variance. It is suspected that narrow ranges place less cognitive burden on respondents, comparisons would be easier to assess, leading to a more consistent process. This agrees with previous empirical evidence (Mazzotta and Opaluch, 1995; Caussade et al., 2005, p.631) that range of attributes variation impact on the error variance.

As in this experiment, different bands are developed for representing different journey distance and attempts to conclude that shorter journey distance contributes to a higher scale parameter; this is nothing relevant to the cognitive burden. The results show that scale parameters decline (i.e. error variances increase) with increasing trip length, implying that travellers are less sensitive to the trade-off of attributes in the utility function for longer trips relative to shorter trips. This finding is consistent with the result of Koppelman and Sethi (2005, p.835), who indicate that individuals might have differences in perceptions of attributes describing the utility function for different journey lengths.

Income has shown some effects on the scale parameter. Higher income contributes a less variance of error term. This is because the level of income is connected with the education level. People with higher education are more easily able to understand the SP experiment, thus making more consistent responses in the experiment. This result is consistent with the previous empirical evidence (Caussade et al., 2005).

Cheap talk shows a positive impact on the scale parameter, which contributes to a smaller variance in the random term. The impact is not statistically significant (t-statistics = 0.65).

8.3.5 Comparison of valuations

Prior to discussing the impacts of design factors from the model estimation, this section presents the monetary values of rolling stock (VoS) obtained from the preferred MNL and HMNL models. The monetary values are obtained using the equations in section 4.4.4, taking into account the design factor impacts and individuals' socio-economic features. The variance of the monetary values is obtained using Equation 4.31 (see the discussion in section 7.4.5).

This section, firstly, presents the comparison of VoSs derived from different models (MNL & HMNL), and then the comparison of VoSs derived from different SP design within each model. The VoSs in this section are obtained from the coefficient estimation from the preferred MNL (M8-2b) and HMNL (M8-3b) models.

Comparison of VoSs obtained from different model estimation

Table 8.7 presents the comparison of VoSs with the standard error in the bracket to the right (at the 5% level). It is found that the standard error obtained from the HMNL model is generally smaller than that from the MNL model, which indicates that the values from the HMNL model are more precise than those from the MNL model.

Figure 8.3 and Figure 8.4 present the comparison. The point estimate VoSs do not show significant difference between M8-2b (MNL) and M8-3b (HMNL), with the exception of the highest income band (over £50k). For individuals who belong to this income category (over £50k), VoSs derived from M8-2b is significantly higher than that from M8-3b.

From the comparison, we conclude that the VoSs obtained from the MNL model do not show a significant difference from the HMNL model in the lower income category (<£50k). The interval of VoS is narrower from the HMNL model which indicates a more precise estimate.

Table 8.7 Comparison of VoSs obtained from different model estimation

VoS (Commuters)	M 8-2b MNL-No CT	M 8-3b HMNL-No CT	M 8-2b MNL – CT	M 8-3b HMNL-CT
< £21k	25.25 (3.85) (17.69 – 32.81)	23.02 (2.19) (18.73 - 27.31)	20.68 (3.15) (14.51 -26.86)	19.68 (1.77) (16.21 -23.15)
£21-35k	31.22 (4.83) (21.75 – 40.69)	28.69 (3.10) (22.61-34.77)	25.53 (3.85) (17.97-33.08)	23.68 (2.32) (19.13-28.23)
£36-50k	43.58 (7.38) (29.12-58.04)	33.76 (4.86) (24.23 -43.29)	32.83 (5.27) (22.50-43.16)	27.02 (3.28) (20.59-33.45)
Over £50k	59.84 (15.28) (29.90-89.78)	44.40 (8.99) (26.78-62.02)	44.79 (9.61) (25.96-63.63)	33.45 (5.29) (20.59-43.82)

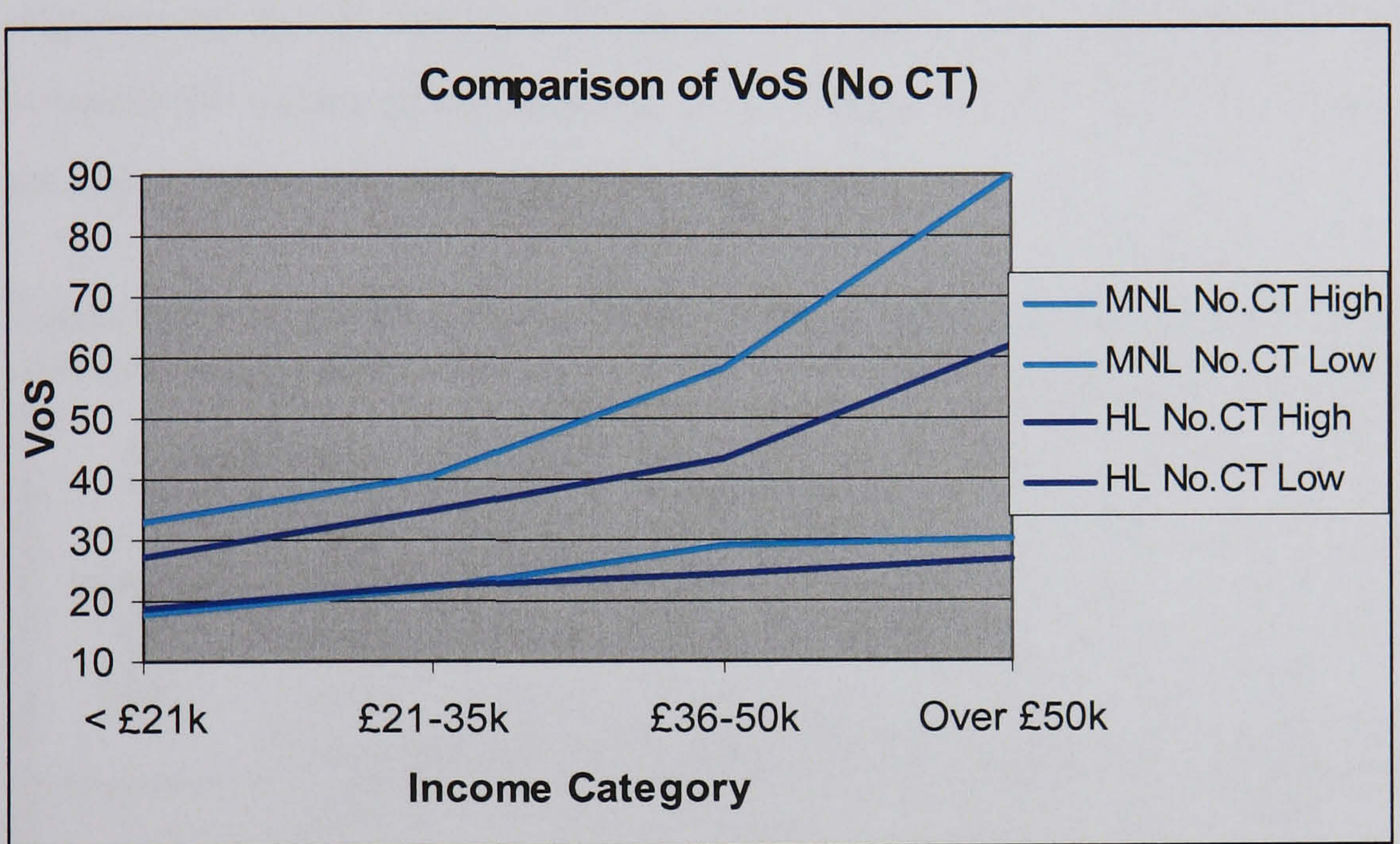


Figure 8.3 Comparison of VoSs derived from different models (No CT)

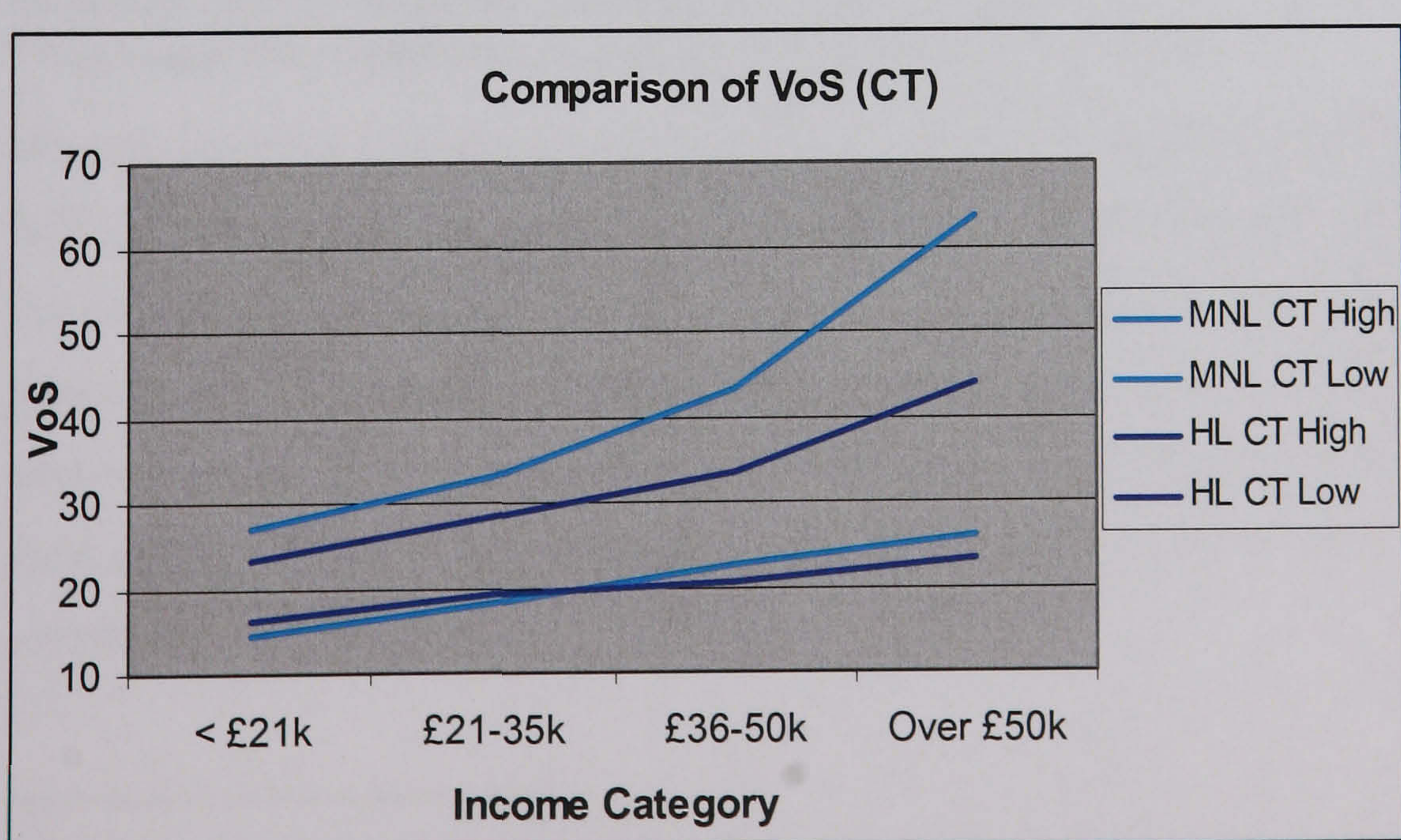


Figure 8.4 Comparison of VoSs derived from different models (With CT)

Comparison of VoSs obtained from different SP design

Table 8.8 presents the comparison of VoSs in different SP experiment. The VoSs are obtained from the preferred MNL model (M8-2b), with the standard error and t-ratio.

VoS_{CT} (VoS derived from SP responses with the CT script) is lower than VoS_{NoCT} (without the CT script) in all the income bands. However, from the t-statistic test, the difference between the values is only significant (t>1.96) in the income category (£36-50k) in the simple SP experiment, and not significant in all the income bands in the complex SP experiment.

$VoS_{Sim.}$ (VoS derived from the simple SP design) is lower than VoS_{Com} (from the complex experiment) in all the income bands. However, the t-statistic test found that the difference between the values is not significant (t>1.96) in all the income categories, no matter whether in the experiment with or without the CT script.

Table 8.8 VoS derived from MNL model (M 8-2b) (p/single journey)

VoS (Commuters)		SP Exp. No CT			SP Exp.With CT			Sig. of Difference
		value	s.e.	t-ratio	value	s.e.	t-ratio	
Simple SP Exp.	**<£20k**	23.30	3.73	6.25	19.13	2.85	6.71	*
	£21-35k	29.74	5.30	5.61	23.57	3.86	6.11	*
	£36-50k	41.32	12.30	3.36	29.44	6.37	4.62	**
	> £50k	56.53	27.36	2.07	39.81	14.34	2.78	*
	Ave. VoS	28.66	8.26	3.47	22.78	4.92	4.63	*
Complex[3] SP Exp.	**<£20k**	27.41	5.56	4.93	22.78	4.57	4.98	
	£21-35k	32.56	6.82	4.77	27.89	5.65	4.94	
	£36-50k	45.66	13.30	3.43	37.69	9.57	3.94	
	> £50k	60.74	26.15	2.32	48.37	16.55	2.92	
	Ave. VoS	33.33	9.79	3.40	27.60	6.97	3.96	

*represents that the difference is significant at the 15% level (t = 1.44)
**represents that the difference is significant at the 5% level (t = 1.96)

We also examined the impact of CT on the variation of VoS in the HMNL model estimation (M 8-3b). From the HMNL model (refers to Table 8.6), the incremental effect of complex design on the estimation is not significant, thus being removed from the model. Therefore, Table 8.9 presents the VoS derived from the HMNL model with only account of the CT impact. The comparison in Table 8.9 confirms the conclusion obtained from the MNL model estimation. Adding the CT script decreases the VoS, by on average 17%; however, this impact is only significant at the 15% level.

[3] The interaction effect of journey distance and complex design on the estimation of the ASC term is taken into account in calculating the VoS for complex experiment. For the sake of comparison, VoSs in complex design from different journey distance (short and long) are combined and presented in the table.

Table 8.9 VoS derived from HMNL model (M 8-3b) (p/single journey)

VoS (Commuters)	SP Exp. No CT			SP Exp.With CT			Sig. of Difference
	value	s.e.	t-ratio	value	s.e.	t-ratio	
<£20k	23.02	(2.19)	10.51	19.68	(1.77)	11.12	*
£21-35k	28.69	(3.10)	9.25	23.68	(2.32)	10.21	*
£36-50k	33.76	(4.86)	6.95	27.02	(3.28)	8.24	*
> £50k	44.40	(8.99)	4.94	33.45	(5.29)	6.32	*
Ave. VoS	27.21	(3.57)	7.62	22.56	(2.48)	9.10	*

*represents that the difference is significant at the 15% level (t = 1.44)

In summary, both MNL and HMNL model estimation found that the CT script had a negative effect on the magnitude of VoS. The VoS usually decreases by 20% in the SP experiment with a CT script; although the impact is not significant at the normal 5% level. Adding two more attributes does not show a significant impact on the magnitude of VoSs.

8.4 Impacts of Cheap Talk (CT) on Decision Making

8.4.1 Research hypotheses regarding to the impacts of cheap talk

Section 1.2.3 presented the research hypotheses which outlined the research scope for this study. The research hypotheses were established from the literature review in chapter 2 and 3. This section presents the examination of research hypothesis (H2), with the discussion of the research questions (see section 2.6.5) regarding the impacts and effectiveness of the CT script in the SP experiment on the valuation of improved rolling stock.

Recall the hypothesis 2 (H2), "The adding of the cheap-talk can amend individuals' incentive to strategic bias". This statement can be interpreted in two ways: the impact of the CT on the estimation of coefficients (section 8.4.2) and on the valuation of the improved rolling stock (VoS) (section 8.4.3). The null hypotheses can be stated as:

$H2_{01}: \delta_{CT} = 0$, where δ_{CT} measures the incremental effect of CT (denoted by the dummy variable d_{CT}) on the ASC. The null hypothesis is that adding CT does not affect the estimation of the ASC (refers to the preference of the improved rolling stock).

$H2_{02}: \gamma_{CT} = 0$, where γ_{CT} measures the incremental effects of the CT script on the other attributes in the utility function. The null hypothesis is that adding the CT script does not have an impact on the estimation of other attributes.

$H2_{03}: VoS_{CT} = VoS_{NoCT}$, where VoS is the monetary value of the improved rolling stock, which is obtained by the ratio of marginal utilities of the constant term (represents the preference of the improved rolling stock) and cost. VoS_{CT} refers to the monetary values obtained from the SP experiment with the CT script; whilst VoS_{NoCT} is the monetary value obtained from the SP experiment without the CT script. The null hypothesis is that adding a CT script does not have an impact on the magnitude of the VoS.

8.4.2 Impacts of cheap-talk on the estimation of coefficients

In M 8-2a (Table 8.4), adding a CT script demonstrates a negative incremental effect (δ_{CT}) on the estimation of the ASC term (preference of the improved trains). This indicates that adding CT, individuals show less preferences for the improved trains, which is same as what is expected. However, the impact is not significant at the 5% level (t = -0.34), thus removing it from the model (see M 8-2b). The same trend is found in the HMNL model (M 8-3a and M 8-3b). The null hypothesis $H2_{01}$ cannot be rejected at the 5% significance level.

Both the MNL and HMNL model estimation found significant incremental effects of adding the CT script on the estimation of time, cost and crowding coefficients. Refer to M 8-2b (Table 8.4), the absolute value of the cost coefficient increases by 0.0034 (t=2.02) in the SP experiment with the CT script, and the t-statistic test can reject $H2_{02}$ at the 5% level. This is reasonable and the same as expected. CT reminds respondents of budget constraints; therefore, it changes individuals' sensitivity to the cost in the SP choice. Higher absolute value of the cost coefficient leads to a lower VoS if all the other coefficients are the same.

The result is consistent with those of Carlsson et al. (2005) and List et al. (2006). In their application of a CT script in SC experiments, the adding of a CT script increased cost coefficient, thus decreasing the monetary value of the new product. For example, Carlsson et al. (2005) found that out of 10 attributes which characterised the good, seven were significantly less valued when a CT script was used.

When adding the CT script, respondents give more weight to the 'time' coefficient (increase the coefficient by 0.0255, t= -2.65). With the CT script, respondents give less weight to the crowding by 0.0487 (t= 2.86). The monetary values of the attributes are discussed and compared with the previous empirical evidence in the section 8.4.5.

8.4.3 Impacts of cheap-talk on the valuation of improved rolling stock

This section presents the impact of the CT script on users' valuation of the improved rolling stock. In the previous application of CT scripts, Cummings et al. (1995) found that hypothetical experiment in the CV studies is not incentive compatible, through the difference between the outcomes from actual and hypothetical experiments. CT script was found to attenuate the bias in the CV studies. However, Haab et al. (1999) debated this conclusion, contending that the variation of the attributes (caused by adding CT) may be caused by heteroskedasticity in the error variance between actual and hypothetical experiments.

The monetary values of improved rolling stock (VoSs) are compared in this section. As the VoS is obtained by the ratio of marginal utilities of the ASC term (captures the preference for the improved trains) and the cost, the impact of heteroskedasticity in the error variance will be cancelled out by taking the ratio of coefficients.

Recalls the comparisons of VoSs in section 8.3.5 (see Table 8.8 and Table 8.9), VoS_{CT} (VoS derived from the SP response with the CT script) is lower than VoS_{NoCT} (without the CT script) in all income bands, by on average 18% (or more). However, the t-statistic test of the difference between the values cannot reject the $H2_{03}$ at the 5% significance level. This leads to the conclusion that adding cheap talk script decreases the VoS by on average 20%, but the impact is not significant at the normal 5% level.

In summary, a lower VoS is obtained from SP responses with the CT script. This finding is consistent with the previous empirical evidence that adding a CT can attenuate respondents' hypothetical bias (Cummings and Taylor, 1999).

The direction of "near significance" result is consistent with findings reported in List (2001) and Aadland and Caplan (2003). The possible reason is individuals' heteroskedacity. For example, List (2001) noted that experienced individuals may not be easily swayed by the CT design as they have a well-structured preference for the good in questions. This is also be supported by some empirical evidence (such as, Murphy, 2003). The effectiveness of the CT script among different population is explored in section 8.4.4.

8.4.4 Effectiveness of cheap-talk among different population

The review of CT applications in section 2.6 found that CT script is effective in certain groups of individuals. This section investigates the relationship between respondents' characteristics and the effectiveness of the CT in our experiment. The test is conducted by adding $\delta_{CT.y} d_{CT} d_y$ to the ASC term in the utility function to identify the propensity for respondents (belongs to certain category denoted by dummy variable d_y) to choose one alternative when adding a CT.

This method is motivated by DeShazo and Fermo (2004)'s work (refer to section 4.4.5) on detecting "propensity to attend" behaviours of respondents. They incorporated an ASC term '$\delta_c d_{STALjk}$' into the utility model to detect the task complexity effect of the standard deviation among the attributes, which characterizes the cognitive cost associated with evaluating one alternative. The δ_c measures individuals' propensities to attend to that alternative, when the cognitive cost of evaluation increases.

In this study, the journey frequency and purpose are considered, as these indexes can measure if respondents are experienced or non-experienced group. Income is also examined. The CT script is a warning message of budget constraints, hence it is expected that the higher income group is less sensitive to the script, as they are less sensitive to the cost change.

Table 8.10 reports the results from a MNL model analysis of the relationship between individuals' characteristics and effectiveness of the CT script. M8-4 and M8-5 use the same model specification as the M8-2b, differs in the specification of the ASC term. $\delta_{CT.y} d_{CT} d_y$ is added to M8-4 and M8-5 to detect respondents' 'propensity to attend' the alternative (improved rolling stock), where there is a CT script in the SP experiment and respondents' characteristics belong to certain category (denoted by dummy variable d_y). For emphasis reason, only the estimates of the ASC and the incremental effects are presented. The coefficients of the other attributes are not significantly different from those of M8-2b, therefore, are not presented here.

M 8-4 includes all the interaction effects between the CT and respondents' characteristics. Non-significant variables are removed from the utility function, and some similar coefficients are combined. M 8-5 is a preferred model to M 8-4.

From the segmentation analysis of the CT by respondents' characteristics, the effectiveness of CT among different group of population can be identified. The positive sign of $\delta_{CT.y}$ can be explained as that with CT script in the SP experiment, individuals (in certain categories) show more preferences for the improved rolling stock (Super Sprinter). This indicates the CT is less effective in this category of population. If $\delta_{CT.y}$ shows negative sign, individuals show fewer preferences for Super Sprinters, which indicates the CT script is more effective.

Income shows a "U-shaped' pattern effect on the CT effectiveness. Mid-level income group is less sensitive to the CT, compared to the lowest income group. This is reasonable and the same as expected. The CT script gives a warning message of budget constraints. Higher income group is assumed to be less sensitive to the cost change than the lower income group; therefore, they are less likely to be affected by the warning message. However, this trend is reversed in the highest income group (over £50k). The coefficient $\delta_{CT.Inc5}$ is not significantly different from zero, which indicates that the CT script is more effective in this group than the mid-level group (£10-50k). This is contradictory to what we expected as higher income will be less sensitive to the cost change; therefore they would be less likely to be affected by the CT script.

There are some reasons to explain this phenomenon. Firstly, respondents in the highest income group might be more likely to treat the SP experiment as a real task, rather than a hypothetical game; therefore, they are more likely to be affected by the CT script. We will come back to this point in section 8.6.2.

Table 8.10 The relationship between individuals' characteristics and effectiveness of CT

	Model 8-4		Model 8-5	
ASC Segmentation				
Base (Commuters/PB)	0.3877	(4.75)	0.3694	(5.00)
+ Leisure/EB/School	-0.2813	(-3.49)	-0.2332	(-3.70)
+ Reimburse	0.4528	(2.93)	0.4403	(2.91)
+ Male	0.1036	(1.59)	0.1178	(2.35)
Effects of CT (cheap - talk)				
- On the ASC –interaction effects				
CT (Adding Cheap Talk)	-0.2303	(-2.27)	-0.1303	(-1.65)
+CT*JP1(Leisure)	0.0242	(0.12)		
+CT*JP2(EB- Employer's business)	-0.1327	(-0.63)		
+CT*JP3(PB – Personal business)	-0.0402	(-0.13)		
+CT*JP4(To and from school/college)	0.2433	(1.75)		
+ CT*Inc2 (Income: £10-20k)	0.2389	(3.38)	0.1686	(2.52)
+ CT*Inc3 (Income: £21-35k)	0.5230	(4.95)	0.4368	(5.39)
+ CT*Inc4 (Income: £36-50k)	0.3421	(1.86)	0.2538	(1.45)
+ CT*Inc5 (Income: over £50k)	0.2045	(0.81)		
+ CT*JFreq2 (Frequency: 2 to 4 times a week	-0.1064	(-0.8)	*-0.1888*	*(-2.20)*
+ CT*JFreq3 (Frequency: once a week)	-0.2197	(-2.72)	*-0.1888*	*(-2.20)*
+ CT*JFreq4 (Frequency: Less than once a week)	-0.2050	(-0.97)	-0.1888	(-2.20)
+ CT*Reimburse	-0.2568	(-1.23)	-0.2587	(-1.58)
+ CT*Male	0.0132	(0.14)		
Other Attributes				
...(not significantly different from that of M8-2)				
Number of observations	10885		10885	
ρ^2 (C)	0.1324		0.1321	
LL (C)	-6457.1		-6459.1	
LL test statistic	31.4		4.01	
Degree of Freedom	13(vs.M8-2a)		8 (vs. M 8-4)	
χ^2 Critical Value (5%)	22.36		15.51	

Respondents' journey frequency shows a significant impact on the effectiveness of the CT script. Less frequent travellers (who travel less than 5 or more times a week) are more sensitive to the CT message, as $\delta_{CT.FRE2-4}$ shows a negative sign. This result is consistent with previous empirical evidence that non-experienced respondents are more easily affected by the CT script compared with the experienced respondents (List, 2001; Murphy et al., 2003; List et al., 2006). The possible reason is that experienced respondents might have some well-formed preferences, so it is difficult for them to change their ideas by a simple CT script.

8.4.5 Impacts of cheap-talk on the estimation of other attributes

Table 8.11 presents the comparison of average valuations (monetary values and values relative to the in-vehicle time) of attributes for commuters, with the standard errors and t-ratios.

The values from two models (MNL and HMNL) are not significantly different from each other in each category. In our study, the MNL model estimation solved the repeated measurement problems (see discussion in section 4.3.6) using the Jack-knife function in the ALOGIT software; however, the HMNL model did not solve this problem due to the software issue. Therefore, we use the values obtained from MNL model for further examination and analysis.

Table 8.11 The impact of CT on the valuation estimation

Commuters	SP Exp. without CT script						SP Exp. with CT script						Sig.
	MNL			HMNL			MNL			HMNL			
		s.e.	t		s.e.	t		s.e.	t		s.e.	t	
Monetary Values													
VoS (p/trip)	30.95	8.62	3.59	27.21	3.57	7.62	24.96	5.94	4.20	22.56	2.48	9.10	
VoT (p/min)	5.34	1.22	4.38	5.32	0.54	9.85	5.68	0.91	6.24	5.56	0.30	18.53	
VoH (p/min)	4.68	1.10	4.25	4.75	0.48	9.90	3.86	0.65	5.94	3.93	0.30	13.10	
VoP (p/min)	34.69	8.23	4.22	34.44	4.01	8.59	28.59	5.00	5.72	28.55	2.67	10.69	
VoC (p/min)	15.91	2.58	4.10	16.09	1.34	8.04	12.05	1.21	5.26	12.14	0.84	7.83	**
Values relative to the In-vehicle time													
VoS	5.80	0.91	6.37	5.11	0.72	7.10	4.39	0.80	5.49	4.06	0.51	7.96	
VoT	1			1			1			1			
VoH	0.88	0.07	12.57	0.89	0.08	11.13	0.68	0.05	13.60	0.71	0.05	14.20	
VoP	6.50	0.72	9.03	6.47	0.65	9.95	5.03	0.51	9.86	5.13	0.44	11.66	
VoC	2.98	0.26	7.62	3.02	0.23	8.78	2.12	0.18	6.22	2.18	0.13	9.08	**

**represents that the difference between the values from experiment with CT script and from that without CT script is significant at the 5% level (t = 1.96)
The VoC here is the value of the standing time, with account of the seated time.

The monetary values of attributes are generally lower in the SP experiment with the CT script. The only exception is the VoT, which is found to be slightly higher in the experiment with the CT script. The difference between the two experiments (with/without CT) is not significant at the normal 5% level, except for the VoC. We examine the values of attributes by comparing them to those obtained from previous studies. For the sake of clarification, $VoX_{No.CT}$ refers to the value of 'X' attribute obtained from SP responses without a CT script, and VoX_{CT} refers to the value obtained from SP responses with a CT script.

The VoT_{CT} is higher than the $VoT_{No.CT}$, where the former value is more consistent with the PDFH (2005) recommended value (see Table 7.13) for commuters (9.3 p/min) in the short journey distance, even though it is still on the lower side.

The VoH_{CT} is slightly lower than the $VoH_{No.CT}$ and the impact is not significant at the 5% level, again the former value is more consistent with the previous evidence. In Table 7.17, the meta-analysis (Wardman, 2004) of the British evidence suggested that the value of headway relative to the time unit is 0.57 for non-business travellers and 0.71 for business travellers for the journey distance of 10 miles and gets lower for slightly longer journey distance (50 miles). In the present study, most respondents are travellers within the Greater Manchester area (short journey distance), so that 0.68 (VoH_{CT}) is more consistent with the recommended value.

The VoP_{CT} is slightly lower than the $VoP_{No.CT}$ (not significant at the 5% level). PDFH (2005) recommended the value of punctuality relative to the time units as 2.5 to 3.0. In the present study, both VoP_{CT} and $VoP_{No.CT}$ are much higher than this value, with the VoP_{CT} more consistent to the recommended value. The large penalty of punctuality observed in this study is partly because most respondents were making short journey distance morning commuter trips with presumed severe penalties for late arrivals. Respondents gave a high penalty to delay of 10 minutes as set in our choice experiment. It is suspected that the VoP is biased upward.

The VoC significantly decreases in the SP experiment with a CT script from both models estimation. Crowding is suggested to have a strong effect on respondents' preferences of rail service. The main survey filed work and the comments provided by respondents indicated that crowding was a serious issue on commuter routes in the Greater Manchester area. It is suspected that respondents expressed their opinions of crowding and would like to have this situation changed through the SP survey. With adding of the CT script to the SP experiment, respondents were informed of budget constraints and gave a lower penalty to crowding in the experiment. The VoC_{CT} is closer to the recommended value by PDFH (2005).

Although the valuations of service attributes are found to be lower in the experiment with the CT script, no significant bias is detected. Based on the above evidence, we can conclude that the valuations obtained from the SP responses with the CT script are more consistent with PDFH (2005) recommended values.

8.4.6 Influences of adding Cheap Talk on the demand forecast

The VoS can be used to forecast demand by the introduction of improved trains. Section 3.2.3 discussed the PDFH (2005) method in forecasting demand. Using the PDFH (2005) method, we investigate the impact of adding the CT script on demand forecast. This method is to convert the improvement into an equivalent change in rail fare; and then the relevant fare elasticity is applied to calculate the expected demand increase.

Using the average VoS obtained from Table 8.11, if the average fare in the present study is set as £3.00 (see section7.5.2) per single journey, 8.3% (obtained by 24.96/300) of the fare can be obtained from the SP responses with the CT script. 10.3% (30.95/300) of the fare is obtained from those without the CT script. The following equation is applied to achieve the demand impact. Section 3.2.3 presents a introduction of this equation.

$I_F = (1-RS)^{f_t}$, where, f_t is the overall fare elasticity which is the ratio of the incremental percentage fare change with respect to an incremental percentage change in another variable. *RS* is expressed as a proportion of the base fare.

Using PDFH (2005) values that fare elasticity is -0.6 for commuter of Non London (distance < 20 miles), VoS_{CT} is 8.3% of the average fare, so that the demand change will be:

$$(1-0.083)^{-0.6} = 1.053$$

This indicates that if the train changes from Pacer to Super Sprinter and all the other factors remain the same, the passenger demand would increase by 5.3%. Table 8.12 presents demand changes caused by the improved trains using values obtained from SP responses with/without the CT script.

Table 8.12 Impact of adding the CT on demand forecast

	SP Exp. without CT	SP Exp. with CT
VoS (p/single trip)	30.95	24.96
% of average fare	10.3%	8.3%
Demand Change	6.7%	5.3%

In PDFH (2005), the most relevant value can be found is that the monetary value for 'Non air-conditioned modern sliding door South East electric multiple units replacing non air-conditioned slam door electric multiple units' is 2.5% of the fare (refer to Table 7.15). The monetary value from the present research is still higher than the officially recommended value. However it is lower than the values obtained from previous SP studies (averagely in excess of 10% of the fare) according to the review of previous studies (see chapter 3). It is concluded that adding the CT script improves the SP design in the valuation of the improved rolling stock, but this bias may remain.

8.4.7 Summary of the Cheap-talk impact on SP responses

- Decreases the coefficient of the ASC term. Individuals show less preferences for the improved train with the CT script, however, the impact is not significant at the 5% level. Adding the CT script significantly (t=2.63) increases the weight of the cost coefficient, which leads to a lower monetary value of the improved rolling stock;

- The VoS_{CT} is lower than the VoS_{NoCT} in all income bands by 18% (or more). The impact is not significant at the 5% level (Table 8.8). Similar results are found from the HMNL model (Table 8.9) estimation. We cannot reject the null hypothesis at the normal 5% significant level that adding the CT decreases the estimation of the VoS;

- The CT script is more effective in the less-frequent travellers;
- The CT script is more effective in the highest and lowest income group;
- Adding the CT script to SP survey change the values of other attributes. Compared to the recommended values (PDFH, 2005), the values obtained from SP responses with the CT are more consistent.
- Demand change caused by the improved trains is 8.3% in this study, and decreases to 5.3% with adding the CT script in the SP survey.

As explained by Cummings and Taylor (1999), as individuals become aware of the potential influence of the context of hypothetical decision on their valuation of a good, they attempt to "correct" for the hypothetical nature of the exercise. Through the internal correction process, the individuals commit cognitive efforts to retrieve a more accurate value for the good in question.

As noticed from the analysis, although adding the CT script to the SP experiment decreases the VoS, the discrepancy is not significant at the 5% level by the t-statistical test. The 'nearly significant' impact of CT can be explained by the following reasons:

Individuals' heterogeneity might contribute to the lower t – ratio of the discrepancy. It is found that experienced individuals might be less likely to be affected by the CT script as they already formed their preferences (List, 2001; Murphy, 2003; List et al., 2006). The analysis in section 8.4.3) found that frequent travellers were less sensitive to the CT script compared to the less frequent travellers. In the current study, most of the respondents are frequent travellers (59% of the whole sample).

Secondly, the CT script in the present research is a short-versioned one which is truncated from Cumming and Taylor's (1999) cheap-talk script. The literature review of the CT application found that the short version CT shows mixed evidence of the effectiveness. The reliability of CT script still needs further examination.

8.5 Impacts of Complex Design on the Choice Making

8.5.1 Research hypotheses

In the present study, two more attributes were added to some of the SP experiments. It is to test if adding more attributes to SP choices would make the research aim less transparent; therefore, respondents will be less likely to strategically overestimate the VoS. This is motivated by Wardman and Bristow (2003)'s successful exploration in valuing the aircraft noise. This study was briefly described in section 2.5.5. However, this method adds the risk of the task complexity effects to the SP experiment. The literature review of task complexity effect (section 2.7) found that the number of attributes in the SP experiment has an impact on respondents' choice behaviours: make more errors or bias their answers to simplify the decision making.

This section examines the impact of the complexity design on SP responses. In section 1.2.3, null hypotheses regarding the complex design can be stated as below:

$$H3A_{01}: \mathrm{VoS}_{\mathrm{Sim.}} = \mathrm{VoS}_{\mathrm{Com.}}$$

This hypothesis is to test if adding more attributes to SP choices can mask the research aim, thus amending respondents' incentive to strategically bias their answers. The complex design is expected to yield a lower valuation of the improved rolling stock compared with that from the simple design where the research aim is more transparent. The null hypothesis is that the VoS from the complex design is not different from that obtained from the simple design.

$$H3B_{01}: \alpha_{\text{Complex design}} = 0;$$

This hypothesis is to test if more attributes add task load to respondents, therefore yielding a larger error variance, compared to the simple design. In Equation 8.2 (section 8.3.4), $\alpha_{\text{Complex design}}$ is the parameter for the impact of complex design on the scale parameter. The null hypothesis is that adding two more attributes will not affect the consistency of choice making (i.e. does not cause more/few errors), compared to the simple SP design.

8.5.2 Impacts of complex design on the estimation of attributes

From M 8-2b (Table 8.4), there is no clear picture regarding the effect of adding two more attributes on users' preferences for the improved rolling stock (ASC term). Some interaction effects of the long journey distance and complex design are found on the estimation of the ASC term; however this impact is not significant at the 5% level. The HMNL model estimation confirms this finding.

In M8-2b, the complex design shows a negative incremental effect on the cost coefficient. The impact is not significant (t=-1.01); however, removing this variable leads to a worse model fit by the LR test. The HMNL model estimation found that this impact is not significant (t=-0.50), removing the variable does not affect the model fit significantly.

8.5.3 Impacts of complex design on the monetary values

The incremental effect of the complex design on the cost coefficient cannot be detected from the HMNL model (Table 8.6), but it can be detected by the MNL model estimation as stated in section 8.5.2. Therefore, only values obtained from the MNL model estimation are compared

(Table 8.8). It is found that: the $VoS_{Sim.}$ (VoS derived from the simple SP design) is lower than the VoS_{Com} (from the complex experiment) in every income band. However, the t-statistic test found that the difference between values is not significant, no matter whether obtained from responses with or without the CT script. Therefore, we cannot reject $H3A_{01}$ at the 5% level. The VoSs from the simple and complex SP designs are not significantly different.

Table 8.13 presents the comparison of the monetary values of attributes between the simple and complex designs, based on M8-2b. It is found that the VoS is higher, whilst values of other attributes (time and headway) are lower in the complex SP experiment. The impact is not significant for all those comparisons at the 5% level. Therefore, the values from simple and complex designs can be combined (Table 8.11). The impact of adding more attributes on demand forecast will not be discussed in the presented thesis.

Table 8.13 Impacts of complex design on the monetary values of attribute (M8-2b)

Commuters	SP Exp. No CT script						SP Exp. With CT script					
	Simple design			Complex design			Simple design			Complex design		
	value	s.e.	t	value	s.e.	t	value	s.e.	t	value	s.e.	t
VoS (p/trip)	28.66	7.17	4.00	33.42	9.94	3.36	22.75	4.90	4.64	27.47	6.94	3.96
VoT (p/min)	5.53	1.25	4.42	5.13	1.19	4.31	5.88	0.91	6.46	5.45	0.91	5.99
VoH (p/min)	4.85	1.13	4.29	4.50	1.07	4.21	4.00	0.65	6.15	3.71	0.65	5.71
VoP (p/min)	-			34.69	8.23	4.22	-			28.59	5.00	5.72
VoC (p/min)	-			15.91	2.58	4.10	-			12.05	1.21	5.26

8.5.4 Impacts of the complex design on the consistency of choice making

In the MNL model, different scale factor is incorporated into the utility function to capture the difference in the variance of error terms among different data sets. Table 8.5 shows that the scale factor for the simple design (LSS12) is larger than those of the complex design (S34/L34). The variance of SP responses in the complex design is larger than those in the simple design, which indicates the existence of the task complexity effect (Bradley and Daly, 1994). The complex SP design leads to more variances in SP responses, which implies that respondents make more errors in the complex design.

The conclusion is confirmed by the HMNL model estimation (section 8.3.4). A brief summary of findings with the hypotheses test is provided below:

- The number of attributes has a clear detrimental effect on the ability to choose, contributing to a higher error variance (section 8.3.4). The $H3B_{01}$ can be rejected at the 5% significance level.

- The range of attributes affects the precision of the estimation. The HMNL analysis (M 8-3b) found that when the difference between the levels of attributes increases, the variance of error term increases. The wide range contributes to a significant higher variance. It is suspected that narrow ranges place less cognitive burden on respondents, since trade-off among attributes tends to be similar across responses.

- As in this experiment, different bands are developed for representing different journey distances, it is attempted to conclude that shorter journey distance contributes to a higher scale parameter. This is nothing relevant to the cognitive burden.

In summary, adding two more attributes to the SP choice contributes to a larger error variance, which indicates less consistency in choice making.

8.5.5 Summary of impacts of the complex design on SP responses

The results from this study exhibit a mixed picture of the impact of adding more attributes. The findings can be summarised as below:

- The VoS obtained from the simple design is slightly lower than that from the complex design, which is contradictory to what we expected. However the difference is not significant at the 5% level (section 8.5.3).

- Values of time and headway are found to be slightly lower (not significant) in the complex experiment.

- Significant task complexity effects is found (see section 8.5.4), where adding two more attributes contributes to a higher error variance in SP responses.

8.6 Impacts of Individuals' Perceptions on SP Responses

8.6.1 Model specification and estimation

People make decisions based on their perceptions (Powe et al., 2005). Investigating influences of individuals' perceptions of the SP survey is helpful for better understanding of respondents' decision making, thus improving the validity of valuation and forecast by using SP results. In this study, three follow-up questions were presented to investigate respondents' perceptions of survey and probe their choice making processes. Respondents were asked to indicate if they experienced difficulty in the choice making (difficulty); if they could perceive the difference of trains with the help of introduction information (familiarity) and whether or not they believed the price would increase if newer trains were introduced (realism).

MNL and HMNL models are developed based on M 8-2b and M 8-3b in order to capture the influences of perceptions on SP responses:

- Impact of individuals' socio-economic features on the estimation of values of the improved rolling stock and other coefficients (same as M 7-10).
- Impact of design factors (cheap-talk (CT), complex design and longer journey distance) and their interaction effects on the valuation of improved rolling stock (same as M 8-2) and on the estimation of other attribute coefficients;
- Impact of individuals' perceptions on the estimation of the ASC and the error variance (using the HMNL model).

The perception variables are added to the ASC term in the utility function to detect their impacts on the tendencies for respondents to choose improved trains (section 4.4.5). $\delta_{perception} d_{perception}$ is added in the utility function, where $d_{perception}$ is the dummy variable denoting individuals' perceptions, defined in Table 8.14, and $\delta_{perception}$ is the coefficient for the dummy variables.

Table 8.14 Definition of dummy variables of individuals' perceptions

Perception Variables	
Cpi	The chances that the price would increase if newer trains were introduced
	Not at all (Base)
	Slightly (Cpi1)
	Moderately (Cpi2)
	Very (Cpi3)
Dich	The difficulty of making choice
	Yes, very (Base)
	Yes, quite (Dich1)
	Yes, a little (Dich 2)
	No (Dich 3)
Ditr	If respondents could perceive the difference of the trains
	Not at all (Base)
	Fairly (Ditr1)
	Very (Ditr2)
	Not sure (Ditr3)

Equation 8.3 demonstrates the function for the scale parameter in the HMNL model estimation. The model specification is same as Equation 8.2, differs in that dummy variables of respondents' perceptions (denoted by $d_{Perception}$) is incorporated into the function:

$$\lambda_s = \exp(\sum \alpha_{Design} \cdot d_{Design} + \sum \alpha_{Perception} d_{Perception} + \sum \alpha_{Socio} d_{Socio}) \qquad \textbf{Equation 8.3}$$

Table 8.15 presents the preferred MNL and HMNL models (M 8-6 and 8-7). From likelihood ratio tests, both models are significantly improved compared to the models (M8-2b and M8-3b).

Table 8.15 The impacts of individuals' perceptions on SP response

Estimation Coefficients (t-ratio)	Model 8-6 (MNL)		Model 8-7 (HMNL)	
Time (Commuters)	-0.0897	(-7.31)	-0.1184	(-6.77)
+ Leisure	0.0513	(3.21)	0.0703	(3.84)
+ EB/PB/School	-0.0243	(-1.78)	-0.0432	(-3.05)
Cost (Base)	-0.0189	(-9.10)	-0.0268	(-8.18)
+ Cost - Inc3 (£21-35k)	0.0030	(1.85)	0.0051	(3.900
+ Cost - Inc4 (£36-50k)	0.0065	(3.50)	0.0076	(3.57)
+ Cost - Inc5 (over 50k)	0.0097	(4.05)	0.0123	(4.22)
Headway (Commuters/EB/PB/School)	-0.0782	(-14.20)	-0.1087	(-8.61)
+ Leisure	0.0308	(2.99)	0.0419	(3.51)
Punctuality (Commuters)	-0.5947	(-9.06)	-0.7929	(-7.75)
+ Leisure	0.2665	(3.32)	0.3697	(3.87)
+ EB/PB/School	0.1381	(2.24)	0.1750	(2.45)
Crowding (Commuters/EB/PB/School)	-0.1803	(-7.30)	-0.2521	(-7.39)
+ Leisure	-0.0546	(-1.52)	-0.0803	(-2.33)
ASC Segmentation (Commuters/PB)	0.6979	(4.91)	1.1271	(6.25)
+ Leisure/EB/School	-0.3554	(-4.14)	-0.5148	(-4.62)
+ Reimburse	0.2822	(2.22)	0.3724	(2.87)
Design Variables on ASC				
+ CT (Adding Cheap Talk)			-0.1246	(-1.37)
+ JD*CD (Long Distance * Complexity)	0.1452	(1.27)	0.1916	(1.76)
Perceptions				
+ cpi3 (Moderately: 20%)	-0.2624	(-1.79)	-0.4089	(-2.76)
+ cpi4 (Very: 69%)	-0.3754	(-3.35)	-0.5973	(-4.29)
+ dich3 (A little bit difficult: 31%)	-0.2301	(-2.67)	-0.3359	(-4.28)
+ ditr3/4 (Distinguish trains: 64%)	0.2009	(2.30)	0.1999	(2.53)
- On the Other Attributes				
+ CT*Time	-0.0250	(-2.79)	-0.0375	(-2.56)
+ CT*Cost	-0.0038	(-2.63)	-0.0043	(-2.41)
+ CT*Crowding	0.0481	(3.04)	0.0716	(3.10)
Scale Factors				
θLS12	1			
θS34	0.6684	(8.46)		
θL34	0.5324	(8.11)		
Parameterization of Logsum				
Complex Design			-0.5558	(-7.66)
Band B/C			-0.5302	(-6.04)
Band D			-0.9117	(-7.34)
Income Group 4 & 5			0.1283	(1.70)
Adding of CT			0.0341	(0.60)
Cpi4 – Perceived the price increase be 'very likely'			0.1298	(2.14)
Dich3 – 'A little bit difficult to make the choice'			0.1238	(2.22)
Ditr3 – ' Could distinguish trains' difference			0.1358	(2.42)
ρ^2 (C)	0.1332		0.1374	
LL (C)	-6450.8		-6420.0	
LL statistics test	-43.97		-71.73	
Degree of Freedom	3 (vs. M8-2b)		10 (vs. M 8-3b)	
χ^2 Critical Value (5%)	7.81		18.31	

In the utility function, the ASC term captures the preference for the improved rolling stock and impacts of design factors and respondents' perceptions on their decision making. The coefficients associated with individuals' perceptions are significant, which indicate that perception has significant impacts on the tendency for respondents to choose the improved rolling stock. Except for the parameter of ASC term, other coefficients estimation of M 8-6 is not significantly different from that of M 8-2b at the 5% level.

In the HMNL model estimation, M 8-7 is significantly improved compared to M 8-3b. 71.73 is obtained from the LR test, which is significantly larger than the χ^2 critical value at the 5% level (18.31) with 10 degrees of freedom. Incorporation of individuals' perceptions into the utility function can better explain their decision making behaviour.

Findings from the model estimation

- Respondents' perceptions of the potential price increase have shown significant effects on valuation of the improved rolling stock. If other factors were same, the more likely respondents perceived the price would increase for the introduction of newer trains, the fewer preferences they gave to the improved rolling stock. Individuals who perceived the potential price 'very likely' to increase made fewer errors in the SP choices.

- The perceived difficulty to make the choice affects respondents' decision making. A 'U-shape' pattern is found on preference for the improved rolling stock. Respondents who perceived choice 'a little bit' difficult gave the lowest value to the ASC term, and were found to be more consistent in the choice making.

- Individuals' familiarity with the experiment alternative demonstrates a significant impact on their responses. Individuals who were familiar with the stock gave more preferences to the improved train, compared to the individuals who were not familiar with the difference between the trains. It is also found individuals, who were familiar with the stocks, were more consistent in the choice making.

Based on the results from M 8-6 and M 8-7, influences of respondents' perceptions on their responses are discussed and summarised in the following sections (8.6.2-8.6.4).

8.6.2 Influence of perceived potential price increase

Impact on the valuation

Interesting results emerged from model analyses. It is found that the higher chance respondents perceived the price would increase by the introduction of newer trains, the lower preference they gave to the new stock. For example, when respondents perceived the price 'very likely' to increase, they gave a lower weight to the ASC term by 0.3754 (t=-3.35). When the perceived chance was 'moderately', the weight was 0.2624 (t= -1.79). The same trend is found from the HMNL model (M 8-7) analysis. When the impacts are transferred to the monetary value of the improved rolling stock, individuals who perceived the potential price increase showed lower WTPs for the improved trains if the other attributes remain the same.

This finding is consistent with the theory and empirical evidence (see section 2.5). The literature review of strategic bias found that payment has a strong impact on individuals' decision making (Bohm, 1971). If respondents perceive that the payment of the new good in the SP survey is compulsory, incentive compatible can be achieved (assumed that other criteria are satisfied). In this situation, respondents give their true WTPs to the product. However, if respondents believe the payment is based on voluntary or the price change is very unlikely, they have strong incentive to strategically overestimate their WTP to increase the chance that the new product being introduced.

In this research, it is assumed that respondents' perceptions of the potential price increase reflect the level of payment coercive perceived by them. "Very likely" indicates that respondents believe that the price will increase and they will have to pay such an amount for the improved service. Therefore, respondents give a lower value to the improved rolling stock, compared to those who do not believe the price will increase.

Another possible reason is that respondents are sensitive to the cost change. If respondents believe that the price is the variable most likely to vary by the introduction of new product, reduction of the price or cost is preferred rather than increase. Wardman (2001) conducted a meta-analysis on value of time studies (from SP values) and found that there is a significant effect on the value of time if toll is the numeraire compared to other numeraires, which yield a 19% lower estimate of value of time.

Influence on the consistency of choice making

The HMNL model (M 8-7) analysis found that the coefficient corresponding to the option "the price is very likely to increase" is positive in the function of scale parameter. The impact is significant at the 5% level (t=2.26). This indicates that the error variance is smaller from individuals who believe the price would be 'very likely' to increase. Stated equally, they are found to be more consistent in the SP choice making. It is difficult to find previous evidence for comparison. The reason suspected is that individuals in this group are more likely to treat the SP experiment as reality; whilst respondents who do not believe the potential price increase are more likely to treat SP experiment as a game, therefore, might be less careful in the exercise.

Factors influencing the perception of potential price increase

The factors affecting individuals' perceptions of price increase are explored. From the qualitative analysis (section 8.2.2), no clear evidence shows that adding the CT script has a significant impact on changing individuals' perceptions of the potential price increase.

A decision model is established to examine what factors might affect respondents' perceptions of the potential price change, as shown in Equation 8.4. 'Low' represents that the perceived chance of price increase is low. This group is generated by combining options cpi1 and cpi2 (defined in Table 8.14). 'High' represents the perceived chance is high (cpi3 and cpi4). To explain the perception of potential price increase, respondents' characteristics and other perceptions (difficulty and familiarity) of SP survey together with design factors are included into the utility function by dummy variables.

$$U_{low} = \sum_a (\gamma_a \cdot d_{design}) + \sum_b (\gamma_b \cdot d_{personal}) + \sum_c (\gamma_c \cdot d_{journey}) + \sum_e (\gamma_e \cdot d_{perception}) + \sum_f (\gamma_f \cdot d_{other})$$

$$U_{high} = 0$$

Equation 8.4

Another analysis was conducted by specifying four alternatives (cpi1-cpi4) in the utility function. The $\rho^2(C)$ obtained was negative (-0.3568). Most of the coefficients were not significant at the 5% level. Therefore, four options are combined to two alternatives in the utility function as shown above.

Table 8.16 reports the preferred MNL model analysis. Various factors contribute to the perception of the potential price increase. An interpretation of results is presented below:

The CT script does not show a significant impact, which is removed from the preferred model. This indicates that adding this warning message does not affect individuals' perceptions of the potential price increase. This finding agrees with the qualitative analysis in section 8.2.2. The relationship between CT and individuals' perceptions of the price increase is not significant.

Journey frequency and purpose are not included into the function as these factors are expected to have strong correlation with ticket types. For example, commuters are expected to travel more frequent than other groups of travellers (section 6.5.2). Frequent users normally are season ticket holders. Therefore, only ticket type is included in the utility function.

Table 8.16 Factors affect respondents' perceptions of potential price increase (M8-8)

Factors	Coef.	t-ratio	Effect
Ticket Type			
Standard Day Single (Base)	0		
Saver ticket (day) • Standard Day Return • Cheap Day Return • Rail Ranger • Day Saver Ticket	-0.6703	(-2.92)	↑
Season ticket • Weekly Season Ticket • Monthly Season Ticket • Annual Season Ticket • County Card	-0.8314	(-3.64)	↑
Journey Information			
If the journey is being reimbursed	0.8847	(3.20)	↓
Personal Information			
Income Group			
+ Inc5(Over £50k)	0.6559	(2.02)	↓
If respondent's gender is male	0.4044	(2.00)	↓
Perception Questions			
A little difficult to make the choice	-0.5179	(-1.96)	↑
Not difficult to make the choice	-0.494	(-2.31)	↑
Fairly can distinguish the difference of the train	-1.4851	(-5.98)	↑
Very confident to distinguish the difference of the train	-1.3009	(-1.55)	↑
Not sure the difference between the train	-1.6745	(-1.87)	↑
$\rho^2(C)$	0.05		

The model estimation is based on 1222 observations from 1222 respondents
↑ refers to that the factor contributes to perception of potential price increase positively, and vice versa.

With reference to people who travel with "Standard Day Single" ticket, those who hold saver day tickets and season tickets are more likely to believe the price would increase by the introduction of newer trains. For example, saver ticket holders have a value of 0.6703 (t = -2.92) towards the option "high possibility"; while season ticket holders have the value as 0.8314 (t = -3.64). Frequent travellers tend to believe the price increase in the SP survey. The possible reason is that season ticket holders normally travel more frequent than day ticket holders (Table 6.14). They are assumed to have more knowledge and experiences about the rail travel than the other population.

Income levels also shows impacts on respondents' perceptions of price increase. People with the highest income (> £50k) are less likely to believe the price would change. A higher value 0.6569 (t = 2.02) is obtained, which indicates that they tend to select the "Low possibility" options, compared to other groups. Table 8.17 shows the relationship between individuals' annual income and their perceptions of the potential price increase, with the number of individuals who belong to certain category and the percentage in the bracke

Highest income group is less likely to believe the price increase in the SP survey, compared to other income groups. This finding confirms the conclusion on the effectiveness of CT script among different population (section 8.4.4), where it is found that higher income group tend to be affected by the CT script. This study attempts to conclude that individuals who believe price increase are less likely to be affected by the CT script. This is easy to understand. The CT script is a warning message of strategic bias in this study and a reminder of budget constraints. If respondents have already formed the opinion that in the SP experiment, the price will increase for the provision of new good, they are less likely to be affected by the CT message. Alternatively, if individuals do not perceive the potential price increase, they might tend to be affectedly by this warning message.

Table 8.17 Relationship between individuals' annual income and perceptions of potential price increase

Perception of the Price Increase	Not at all		Slightly		Moderately		Very likely		Total
Income group	(1)	(1/5)	(2)	(2/5)	(3)	(3/5)	(4)	(4/5)	(5=1+2+3+4)
<10k	4	(1%)	23	(8%)	68	(23%)	201	(68%)	296
£10-20k	4	(1%)	36	(9%)	70	(17%)	302	(73%)	412
£21-35k	3	(1%)	26	(8%)	72	(21%)	242	(71%)	343
£36-50k	4	(4%)	7	(7%)	21	(22%)	62	(66%)	94
> £50k	2	(3%)	12	(19%)	13	(20%)	37	(58%)	64

Males are less likely to believe the price would increase when new trains were introduced. This partly explained the higher valuation of the improved rolling stock given by males.

8.6.3 Influence of perceived difficulty in choice making

The qualitative analysis in section 8.2.3 found that perception of difficulty in choice making is closely related to the level of complexity (i.e. adding two more attributes) of SP experiment with which individuals have faced. Respondents who completed a complex SP survey were more likely to select the option that choices were difficult to make.

The impact of perceived difficulty reported by respondents is also examined. Another logit model was initially established to detect the impact of perceived difficulty on SP responses using scale factors (section 4.3.3). As scale parameter inversely relates to the variance error term, the larger scale factor refers to a smaller error variance. The logit model has the same specification as M 7-3 (section 7.3.2), only differs in the scale factor. Instead of allowing the difference in scale factors by different SP designs (simple and complex), scale factors were specified by the perceived difficulty reported by respondents. However, the model estimation found that this model was worse than the logit model specifying SP design as scale factors (M 7-3). The goodness of fit of the model was 0.1094 which was worse than that of M7-3 (0.1136).

The possible reason is that people might have different concept of difficulty. The problem can only be solved by asking individuals' personal opinion of 'difficulty' in the choice making.

In M 8-6, people who felt a little bit difficult to make the choice were found to have a lower value (-0.2301, t = - 2.67) to the improved rolling stock. The HMNL model (M8-7) confirms this finding, where the coefficient corresponding to perception of "a little bit" difficult is -0.3359. The impact is significant at the 5% level (t=-4.28).

The influence of perceived difficulty on the consistency of choice making is detected. HMNL model (M8-7) analysis found that people who selected the task was "a little bit" difficult made fewer errors in the choice making. The possible reason is the same as the 'U-Shape' task complexity effects (Keller and Staelin, 1987). If a task is simple, people would easily finish the task. When the task gets more difficult, people think it deserves more efforts to finish the task and they treat it more careful. However, when the task gets even more difficult, they will simplify their decision rules and make more errors.

8.6.4 Influence of familiarity

Respondents' familiarity with the good provided in the SP survey shows some significant effects on their decision making. From M 8-6, when respondents recognise the difference between stocks in the SP choices, they give a higher weight (0.2009, t = 2.30) to the improved train. This leads to a higher monetary value of the improved rolling stock. This finding is confirmed by the HMNL model estimation (M8-7). Respondents are willing to pay more for the improved train if they know the difference between the current and improved trains.

This finding is in contrast to finding by Wardman and Whelan (2001). In their meta-analysis on the valuation of rolling stock, they found if respondents were familiar with the rolling stocks the SP survey presented, respondents gave the improved stock a lower value (see chapter 3). A regression model showed the coefficient was 44% lower and the impact is significant at the 5% level. They concluded that unfamiliarity with improved levels of attributes will result in overestimation of the stock valuation.

Kottenhoff and Lindh (1996, p.240) suggested respondents with experience of the new rolling stock would give a higher value to the new stock (section 3.4.2). Some other evidence can be found in the SP application in residential studies. People who often visit the countryside or have lived there mostly consider it as attractive (Kaplan and Kaplan, 1989). The frequency of visit increases the chance of having rural living preference at a 1% significance level. The previous evidence demonstrates different results on the impact of familiarity. The possible reasons are:

Firstly, the measurement of familiarity is different in these cases. In the review by Wardman and Whelan (2001), familiarity impact was estimated by a regression model of the previous rolling stock studies. The familiarity was measured by the fact that respondents would be familiar with both rolling stocks. In the present research, the familiarity was measured by if respondents could perceive the difference between the trains, either by previous experience or description in the survey. The level of familiarity to the trains in these studies is different.

Secondly, the review of consumer studies provides some explanation for the difference. Johnson and Russo (1984, p.542) stated that "Familiarity with a product class could have several different results, each of which might affect consumers' information processing skills in a different way." They found that the influence of familiarity depend on individuals' decision strategy. 'Enrichment' hypothesis (Johnson and Russo, 1981) suggested that existing knowledge facilitates the learning of new information. However, 'inverted-u' effect (Bettman and Park, 1980; Johnson and Russo, 1984) suggested highly familiar consumers may search less than those who are moderately familiar. It is also found that when faced with a large number of attributes and limited processing capacity, both experienced and naïve consumers consider a subset of the available information (Bettman and Kakkar, 1977; Payne, 1976). The experienced consumers should be better able to select attributes that are predictive of product performances, which should, in turn, result in better decisions (Johnson and Russo, 1984).

The HMNL model (M8-7) reports the familiarity effect on the scale parameter. Respondents' familiarity is incorporated into the scale parameter function. The positive coefficient (0.1358, t =2.42) corresponding to the individuals who could recognise the difference between trains leads to a larger scale parameter, compared to the individuals who are not familiar with the difference. This implies that familiar individuals make fewer errors and are more consistent in the choice making, which agrees well with the previous consumer studies on the familiarity.

8.7 Interpretation and Discussion

8.7.1 Discussion of the incentive to strategic bias

The first research hypothesis (section 1.2.3) was proposed based on the literature review, that:

Hypothesis 1 (H1): *The incentive to (strategic) bias exists in the SP exercises.*

Strategic (hypothetical) bias occurs when individuals overestimate WTP to a new good/policy to increase the chance of the new product/policy being introduced, which is found in the CV and CE studies. To test the H1, the CT script and adding more attributes (complex design) to mask the research aim were introduced into the SP experiments on users' valuation of rolling stock.

The findings were discussed in sections 8.4 and 8.5. Individuals' perceptions of SP experiment were explored and presented in section 8.6.

H2 Adding Cheap Talk script

A Cheap Talk script is introduced into this study. It is found:

- The CT script significantly (at the 5% level) increased the absolute value of the cost coefficient, which led to a lower VoS in the experiment(section 8.4.2);
- The VoS_{CT} is lower than the VoS_{NoCT} in all the income bands, by 18% (or more). However, the t-statistic test of the difference between the values cannot reject the $H2_{03}$ at the 5% significance level (section 8.4.3).

In the recent research by Carlsson et al. (2005) and List et al. (2006), CT scripts were applied to SP experiments to detect the existence of hypothetical bias. Both of them found the existence of significant positive hypothetical bias in the SC experiment on valuation of private goods (food) in the mail survey/actual market place. They found that the CT can effectively reduce the stated marginal WTP in a hypothetical setting and the values are statistically indistinguishable from actual responses.

In our study, it is observed that more preferences and higher WTP are given to the improved rolling stock in the experiment without the CT script. We therefore conclude that we have demonstrated the existence of strategic bias in the context of users' valuation of improved rolling stock. The improved stock is valued on average 1.2 times higher in the conventional SP experiment than which are obtained from SP responses with a CT script.

It is difficult to obtain the magnitude of this bias, as it is difficult to compare results of this research with the real demand change by the improved rolling stock due to the data availability. Demand impact analysis (section 8.4.6) found that adding CT script, demand change due to the introduction of the improved rolling stock is decreased from 6.7% to 5.3%, using the PDFH (2005) method. The latter is closer to the recommended value by PDFH (2005) for the similar type of rolling stock. This indicated that adding CT script to the SP experiment improves the SP design in our study; however, the bias may remain.

H3A Adding more attributes to mask the research aim

More attributes were introduced to some of the SP experiments to mask the SP study aim; hence respondents would be less likely to see any single clear purpose to the experience. Therefore, it would be less likely for respondents to strategically bias their answers. This method was motivated by Wardman and Bristow (2003)'s successful empirical evidence on valuation of the aircraft noise.

Interestingly, the analysis found that the $VoS_{Sim.}$ (the VoS derived from the simple SP design) is lower than the VoS_{Com} (from the complex experiment) in each income band, although the difference between the values is not significant, no matter whether in the experiment with or without the CT script. Therefore, we cannot reject $H3A_{01}$ at the 5% significant level. The VoSs from the simple and complex SP experiments are not significantly different.

The possible reasons for the non-significant impact are: Firstly, 'Punctuality' and 'Crowding' were added to some of the SP experiments, in order to mask the study aim – users' valuation of improved rolling stock. However, it is difficult to detect whether or not adding these two attributes make the study aim less transparent. For example, the verbal and pictorial information of trains might clearly demonstrate this study is on the rolling stock valuation. In addition, adding two more attributes to the SP experiment might add the task load to respondents, therefore, respondents might have a different strategy in the choice making process. We will come back to this point later in section 8.7.2.

Secondly, interaction effects between the task complexity and other factors (journey distance) might be existed, which contributes the variation of the VoS.

Discussion on the incentive to strategic bias in SP studies

The literature review in chapter 2 suggested that the existence of the strategic bias is due to the hypothetical nature of CV and CE studies (section 2.5). Throsby and Withers (1986) categorized the incentive structure along two dimensions: the payment liability and perception of the impact of responses on the good provision. They concluded that if respondents could perceive that their responses have impact on the provision of good; the 'weak free-rider' effects might happen (Situation 4) that respondents have the incentive to strategically bias their answers. The direction/magnitude of valuation variation (over/underestimate) depends on the perceived payment liability by respondents and respondents assigned cost.

Wardman and Bristow (2003) stated that the incentive for respondents to strategically bias their answer was that respondents perceived their responses would have an impact on the provision of the new good, thus providing the biased answer for some certain better outcome. Alternatively, Carson et al. (2000) concluded that a CV/SP survey is incentive compatible when: firstly, the individual perceives responses to the survey question as potentially influencing government or company action; secondly, the individual cares about what the outcome of that action; and thirdly, information about the good, payment mechanism/vehicle and how the survey result be used in the future are provided to the respondent in some certain way. To summarise, the literature review suggested the occurrence of strategic bias can be explained by the following reasons:

- The hypothetical nature of CV/CE studies leaves the payment outside of the experiment, therefore respondents do not have any financial consequence of their statements;

- Respondents perceive their answer potentially affect the provision of new good/policy, thus having the incentive to bias their answer to increase the possibility of the introduction of the good;

Our study supports the first reason and indicates that the reminder of the payment by introduction of CT script can attenuate the incentive to strategic bias in SP responses. Again, the exploration of the individuals' perceptions of cost change in the SP experiment confirms the findings. The higher chance respondents perceived potential price increase, the lower preference they gave to the improved train (section 8.6.2). This indicates that perceptions of payment in the SP experiment play an important role on the occurrence of strategic bias. A payment reminder such as CT script can effectively reduce the bias. However, the bias still remains in our study. Further testing of the CT script on different context is needed.

From this study, no significant difference is detected in the values obtained from the simple and complex SP experiments. Therefore, no direct conclusion can be arrived.

In summary, the present study demonstrated the existence of strategic bias in the SP studies on users' valuation of rolling stock. Cheap Talk is an effective method in the SP design to reduce the bias, however, bias may remain. This study found that adding more attributes did not significantly change the magnitude of the valuation, but worse the choice consistency. We will come back to this point later. Respondents' perceptions of the potential price change in the SP experiment contribute to the variation of the new product valuation. If respondents note that price would increase by the introduction of the good, they give a lower WTP for the new good.

8.7.2 Discussion of task complexity effect

The literature review in section 2.7 suggested that the number of attributes in the SP experiment has a significant impact on the SP choice consistency. Our research hypothesis (H3B) was to test if adding more attributes lead to the higher error variance in SP responses, shows below:

H3B: An increase in the number of attributes will always increase the variance of error terms, thus affecting the valuations implicit in responses.

As stated in section 8.7.1, adding more attributes to the SP experiment did not show a significant impact on the valuations of attributes; although adding more attributes did cause the variation of some of the valuations, such as the VoT and VoH. The difference is not significant. In addition, this study found that adding more attributes to the SP choices affects the consistency in choice making. Adding more attributes contributed to a smaller scale factor, which indicated more error in responses. The H3B can be accepted at the 5% significance level.

Discussion on the variation of the valuations

In our study, 'Punctuality' and 'Crowding' were added into some of the SP experiments to test research hypotheses. The VoSs are found to be higher, whilst the VoT and VoH are found to be slightly lower in the complex SP experiment. The reasons suspected for this finding are:

Adding two attributes did not mask the research aim, but added the task complexity to the SP experiment. From the pictorial information and word description prior to SP choices, it is quite clear that the experiment is about rolling stocks as mentioned in section 8.7.1. It is suspected that when task get more complex (adding two more attributes), respondents allocate more attention to the attributes they are interested in, for example, cost and rolling stock, which leads to the variation of the valuation.

The literature review in section 2.7 found that individuals might have different strategies to process attributes in the complex experiment. Behaviour and psychology theories provided some explanation for this phenomenon. Payne et al. (1992) defined a typology of decision strategies and stated that individuals construct strategies depending on the task demands and the information they are faced with. Respondents have different information processing strategies in respect to how specific attributes are processed, in terms of exclusion and inclusion.

For example, DeShazo and Fermo (2004) proved that respondents applied a rationally-adaptive manner in choice making, and take strategies to minimize the cost and maximize the benefits of information evaluation. Previous studies do not provide a trend for impacts of complexity on the variation of the valuation and suggest the direction of change is dependent on the type of good.

Therefore, in our study, the variation of valuations can be explained as individuals might take different strategies to complete the choice tasks. We cannot identify how respondents process their choice making (for instance, ignore some attributes) in our study. This problem can only be solved by asking individuals' personal choice making strategy.

The smaller scale factor (larger error variance) found in the complex SP excitement indicates a less consistent choice making (more errors). It is suggested that SP design need to be clear and easy to understand by individuals; otherwise, the task complexity effects would happened.

8.7.3 Comparison of values of time derived from different models

In the model development, increments of model refinements are added into the utility function. Since changes in later increments could impact the decisions made in earlier stages, it would be appropriate to revisit these decision. The values of time obtained from different models are compared to examine the impact of model specification on the estimation of valuation.

In the case of linear-in-parameters utility functions, the value of time is obtained by the marginal utilities of time and cost. The equation to achieve monetary value of time and the variance of the value (Fowkes, 1991) is presented in section 4.4.4. The variance of the values is important because the estimated VoT are ratios of random variables, so they are also random.

The variance and covariance value for coefficients are generated by the ALOGIT AND GAUSS program. Table 8.18 presents the VoT with the standard deviation in the bracket on the right side. The average value is generated by the weighted mean approach (section 7.4.4). The confidence interval (with 5% level) is presented below the value and standard error (s.e.). The values of time derived from M 7-10, 8-6 and 8-7 are presented and compared. Prior to the interpretation of the values, recall the function for each model:

- M7-10 (Reference MNL): conventional MNL model, controlling the individuals' characteristics and allowing variance difference among data sets by scale factors;
- M8-6 (Joint MNL): M7-10 + effects of SP design factor (CT and complex design), as well as the impact of individual's perception on their responses;
- M8-7 (HMNL): M8-3 + impacts of design factors and perception on the precision of model estimation (by parameterization of the scale parameter);

Table 8.18 Value of time obtained from the different models (Commuter)

Value of Time Commuter	Model 7-10 (Reference MNL)	Model 8-6 (Joint MNL)	Model 8-7 (HMNL Model)
< £21k	4.90 (0.32) (4.27-5.53)	4.91 (0.36) (4.19-5.62)	4.72 (0.36) (4.02-5.41)
£21-35k	5.77 (0.42) (4.95-6.59)	5.77 (0.47) (4.85-6.69)	5.72 (0.46) (4.81-6.63)
£36-50k	7.06 (0.60) (5.88-8.24)	7.11 (1.05) (5.06-9.16)	6.41 (0.70) (5.04-7.78)
Over £50k	9.52 (1.78) (6.03-13.01)	9.28 (1.79) (5.76-12.79)	8.23 (1.39) (5.51-10.95)
Average	5.68 (0.58) (4.54-6.82)	5.67 (0.66) (4.37-6.97)	5.43 (0.55) (4.35-6.50)

Different values of time are obtained from each model. Figure 8.5 presents the comparison of the VoT with the standard error. The point estimate value of time does not show significantly difference from M7-10 (Reference MNL) with M8-6, with the exception for the highest income band (over £50k), where the value of time derived from M7-10 is slightly higher than the value derived from M 8-6.

The valuation obtained from the HMNL model (M 8-7) is smaller than that of M 7-10 and M 8-6. More specifically, the estimation of VoT from M 8-7 is lower than that from M 7-10. With increase of the income, the discrepancy is getting larger. The difference is not significant at the 5% level. The point estimates from HMNL model (M 8-7) are included within the corresponding confidence interval for the same attribute in the alternative models (MNL).

This agrees with the empirical evidence by Caussade et al. (2005, p.634) that the values achieved from HMNL model were slightly lower than that from the reference MNL model. The impact was not significant at the 5% level. DeShazo and Fermo (2002, p140) applied HMNL model in their study to investigate complexity effects on SP responses and found that the values obtained from HMNL model were significantly smaller than those from the reference MNL model. For example, controlling the heteroskedasticity among individuals, the change of one attribute was as much as 33% (lower than the reference MNL model). A discussion of the HMNL model analysis is presented in the section 8.7.4.

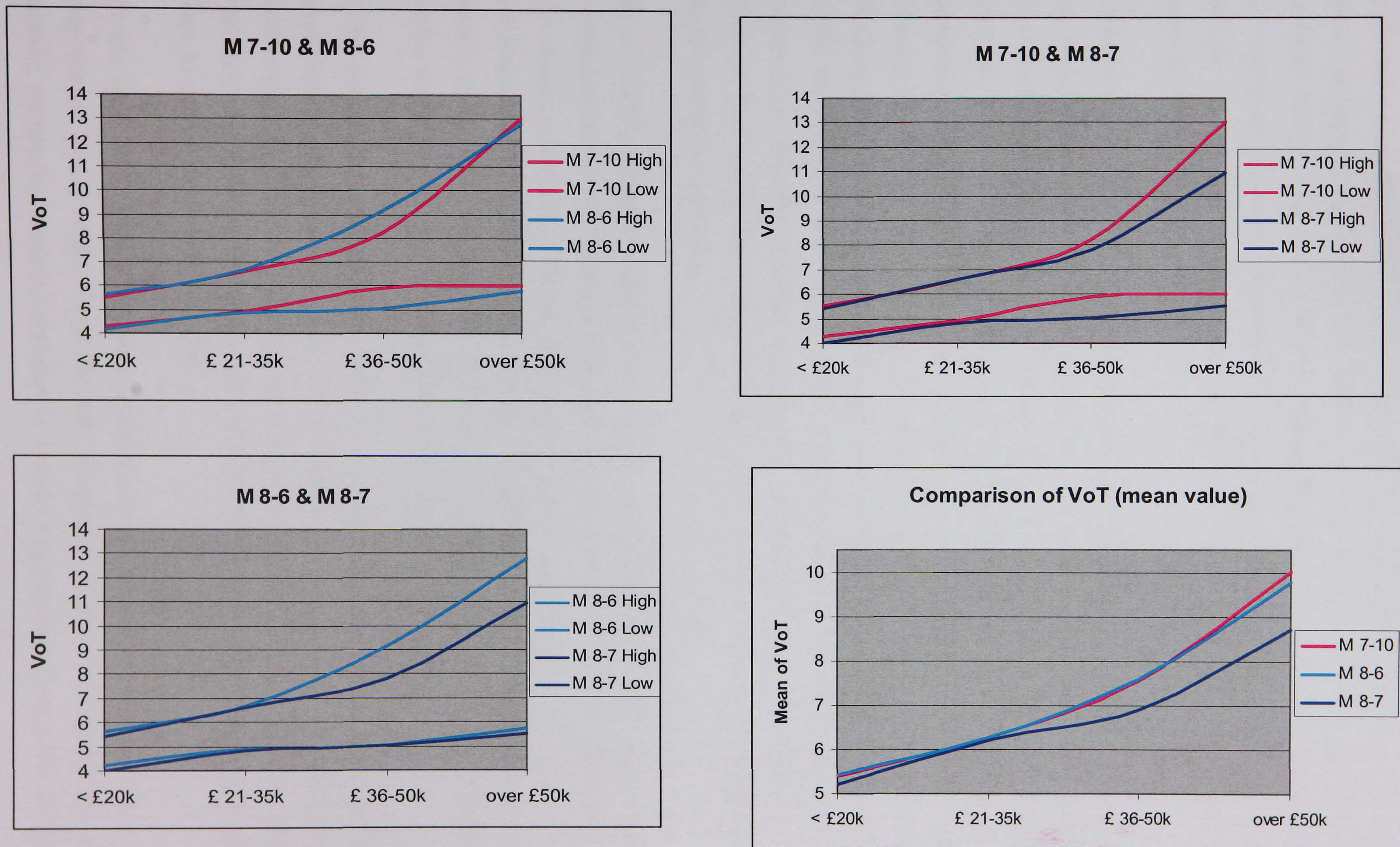


Figure 8.5 Comparison of the VoT from models analysis

8.7.4 Discussion on the HMNL model estimation

HMNL model assumes the taste parameters are equal for the same attributes in different SP experiments, but only the variance between data sources is unequal (Hensher et al., 1999; Swait and Adamowicz, 2001). This will raise the issue that if the taste parameters are not equal for the same attributes in the different SP experiments. For example, in combining the simple and complex design, if individuals have a different strategy of coping with SP choices in different SP experiments, the taste parameters for the same attribute might be different across the SP designs. This problem is found from the previous studies on the task complexity effects by applying HMNL model.

DeShazo and Fermo (2002) conducted a study to examine the task complexity effects in the SP experiment. The number of attributes in the SP experiment was chosen as a measurement of the task complexity. The strategy to vary the number of attributes consisted in selecting a few attributes from a large set in such a way that relevant information could be missing when comparing different SP designs. For example, in the experiment to value the service and infrastructure at new national park, the attributes selected in one design were Access road, Walking trials and Fee. In another SP design, the attributes selected were Access road, Walking trails, Availability of water/toilet and Fee. Significant missing-attribute problem was found by incorporating dummy variables denoting the existence of missing variables in the utility function. This implied the absence of certain attributes in the less attribute condition impacted on the coefficient estimation.

In the work by Hensher (2006b) and Caussade et al., (2005), the same measurement of task complexity (by varying the number of attributes) was examined in the SP experiments. To avoid the missing- attributes problem, some attributes were aggregated in the experiment for less attributes experiment. For example, attribute "Total time" can be disaggregated to "free flow + slowed down and stop/start time". The number of attributes in their experiments was varied in such a way that relevant information (total time /costs) was never missing.

In the present research, two more attributes punctuality and crowding were added into the SP experiments to mask the research aim and to examine if it would amend respondents' incentives to strategic bias. These two variables are selected because they are very important attributes for travellers. It is more likely to mask the aim of the SP survey (valuation of rolling stock), or at least, to make the aim less transparent.

At the same time, adding two more attributes to the experiment increase the task load to respondents, which might cause the task complexity effects. Considering the experiment context, the number of attributes cannot be varied by aggregating some of the attributes. From findings in section 8.7.2, adding two more attributes contributes to a smaller scale factor, which indicates a less consistent choice making. This proves the existence of the task complexity effect. On the other side, we also tested the variation of the valuation of different attributes. Facing more information, respondents might change their choice making strategies; therefore, they might allocate the attention to different attributes. This can only be solved by investigating individuals' personal choice making strategy.

To detect the missing attribute problem in this experiment, the term $\beta_{CD} d_{CD} X_i$ (Equation 8.1) is incorporated into the utility function. If the missing attributes problem exists, then different taste parameters ($\beta_{CD} \neq 0$) can be obtained for the same attribute (such as cost and time) in different SP experiments.

In the MNL model, the difference of error variance caused by complex design is incorporated into the utility function by allowing scale factors for different SP designs. A negative coefficient for the cost ($\gamma_{CD.Cost}$= -0.0016) is obtained from M 8-2b (Table 8.4), which implies that adding two more attributes; individuals have a different taste of cost coefficient. However, the impact is not significant, which cannot reject the null hypothesis ($\gamma_{CD} = 0$) at the 5% significance level. In the further analysis of perception impacts on SP responses (section 8.6.1), the $\gamma_{CD.Cost}$ is not significant, thus being removed from the preferred model (M 8-6).

The HMNL model applies the same model specification as M 8-2 (MNL model) in the utility function, only differs in allowing the heterogeneity across preference observations by parameterization of the scale parameter. $\gamma_{CD.Cost} X_{Cost}$ is incorporated into the utility function, as the complex design impacts on the estimation of other variables is found less significant in the MNL model analysis. However, the t –ratio for estimates of $\gamma_{CD.Cost}$ is (-0.50). Therefore, this variable is removed from the preferred HMNL model.

Louviere and Swait (1996) demonstrated that accounting for differences in variance often accounts for most of the differences in taste parameters in a number of new and published empirical preference and choice results. Hensher et al. (1999, p217) found that "differences in tastes may be less common than previous thought".

In the HMNL model, which the journey distance has been incorporated into the function of scale parameter, some significant impacts can be found in the estimation of scale parameter, but not in the coefficient estimation. The longer journey distance contributes a higher level of the error variance. When incorporated the journey distance impact into the scale parameter, the different taste parameter for coefficient is not significant at the 5% level (see section 8.3.4). It is suspected that the different valuation caused by journey distance in the present research is caused by individuals' heterogeneity.

8.7.5 Comparison between the MNL and HMNL model estimation

In the present study, the impacts of SP design and respondents' perceptions on SP responses were examined by both MNL and HMNL models. The comparisons are made between the estimation of these two models, in terms of model estimation and valuation.

<u>**Software**</u>

MNL models in this study were analysed by ALOGIT (HGC, 2000) program. It is found that the ALOGIT program is user friendly and powerful, and can be applied to most model structure of the logit family. A very attractive feature is that this package has a module to solve the repeated measurement problems (see the introduction in section 4.3.8) by using Jack-Knifed techniques. ALOGIT is very fast in the estimation, compared with other package. A limit of this software is that ALOGIT does not enable the parameterization of the scale parameter.

To estimate the HMNL model, a code was written in GAUSS (Aptech Systems, 1997), using the MAXLIK routine to maximize the values of the log likelihood function. Gauss program is not as user friendly as ALOGIT, but enables more flexible model specification. In estimation of the HMNL model, it took normally very long time (24 hours or even longer time) to achieve convergence. In the thesis, the repeated measurement problem was not solved in the estimation of HMNL model by using GAUSS program.

<u>**Model estimation and valuation**</u>

In the present study, different SP designs were developed to investigate the impacts of design on responses. MNL and HMNL models were applied to interpret individuals' decision behaviour. Recall the analysis in sections 8.3.4 and 8.6.1, it is found that HMNL is better than MNL estimation from the approximate log likelihood ratio tests.

By comparing the valuation obtained from both models (Table 8.7), it is found that the values from HMNL model are not significantly different from those from the reference MNL model. The standard error for the monetary value from HMNL model is smaller than those from the MNL model. For example, it is found that the VoT obtained from the HMNL model is slightly lower than that of the MNL model (section 8.7.3), although the difference is not significant.

8.7.6 Fatigue effects in SP experiments

Previous SP studies found when the task becomes complex, respondents would simply make more errors or bias their results by using some heuristics (see the review in section 2.7). The increase in the task load of the choice set is correlated with increase in the variance of random component (i.e. error) of logit models. As the ranking become lower and as the number of choices completed becomes greater, the amount of unexplained variance is shown to increase. This is defined as fatigue effect (Bradley and Daly, 1994).

Bradley and Daly (1994), applying the scaling approach, detected 'fatigue' effect (higher unexplained variance) in approximately the 5th experiments and to become much stronger by the time of 12th experiments. The fatigue effects were not existed in the simulated data sets. They also found fatigue effect did not significantly change the relative magnitude of the model coefficients. They concluded that "any adverse influence of the order-related fatigue effect may have been 'randomised out' of the data." Recent empirical evidence also found the existence of fatigue effect (Koppelman and Sethi, 2005, p387) in the SP experiments. Scale approach (see sections 4.3.3 and 4.3.4) was normally used to investigate the variances of choices within a single SP experiment. This section explores if the fatigue effects is existed in this study.

Data and scaling approach

There are two types of SP design (complex /simple) in the present research. There are 5 attributes in each binary choice in the complex design, and 3 attributes in the simple design. As discussed before (section 7.3), the datasets can be combined by three groups – allowing two scale factors relative to the reference data set, which are the simple design for both short and long journey distance, complex design for short journey distance and complex design for long journey distance. The scale factor of the complex design for both short and long journey distance is not quite different from each other.

In both designs, there are 9 binary choices. The choices are presented randomly among all the different questionnaires to avoid the order effect. To obtain more information, there is no dominant comparison in the SP design. In the analysis, the first choice is set as the reference, which the scale is fixed to 1.0. A hierarchical logit model with 8 scale factors is built. Each scale factor represents the impact of the choice number on the error variance. Recall that scale factor is inversely related to the error variance, which large scale factor implies a small level of error variance. The process of the scaling approach is discussed in detail in section 4.3.4. The 'artificial tree' is presented in the Figure 8.6.

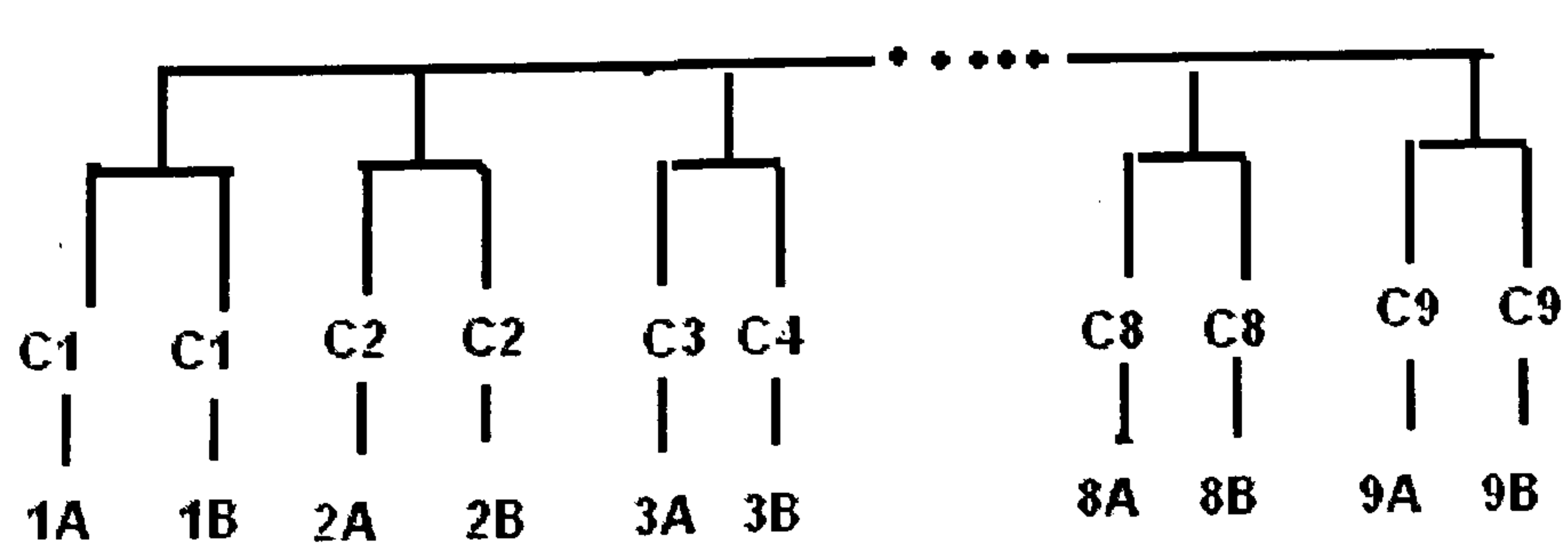


Figure 8.6 Artificial tree for obtaining the scale factors

Model estimation

Initially a joint model of the effects of different design and tests for fatigue effects was set and estimated (extension from M8-6), however, the model cannot arrive the convergence in the estimation. Therefore separate analyses for simple and complex design are conducted in the following. As we are interested in demonstrating the scale factors for different choice sets, a simple model is set without account for individuals' taste variation. Table 8.19 reports the results from the simple and complex design separately, and the combined data sets.

The scaled models accommodates extensive variance scaling by including scale factors for each choice scenario to allow for respondent fatigue and/or learning effects in making choice. The scaled models reject the reference models at a significance level of 0.05, for simple, complex and joint data sets. The results are illustrated in Figure 8.7.

Within the simple design, the result shows a gradual decline in the magnitude of the scale factor with increasing question number initially. This suggests that respondents are less careful in making decision about their second /third choice than about their first choice. This result is consistent with the findings of Bradley and Daly (1994) in which they observed declining precision in parameter estimates with increasing rank of the SP experiment.

However, within the complex design, the results show a steadily increase in the magnitude of the scale factor with increasing question number initially. Differences in the complexity of SP design may have contributed to theses differences between complex and simple design, the individual who faced with the complex SP design may learn with each choice scenario.

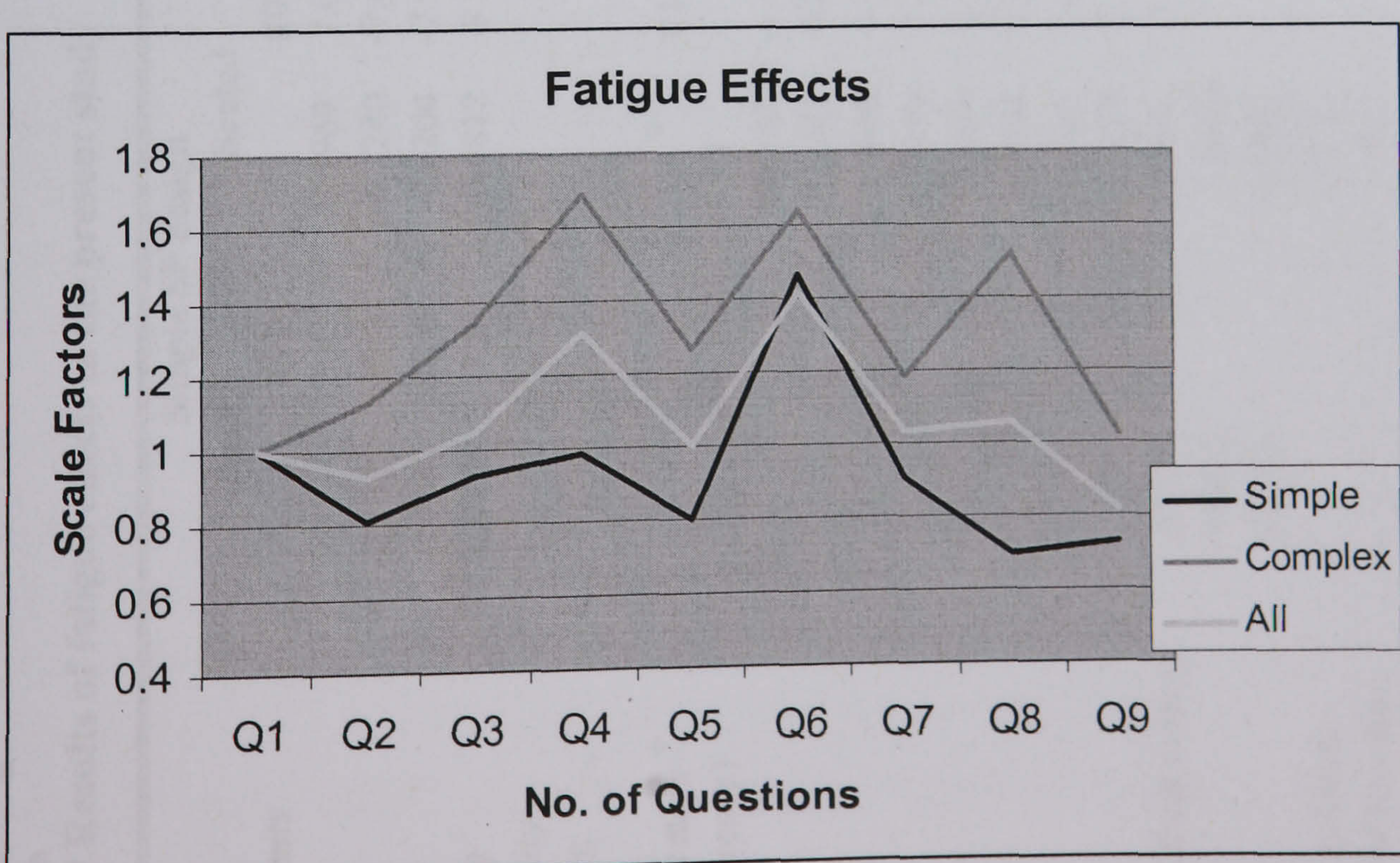


Figure 8.7 Fatigue effects in the present SP study

Table 8.19 Results of fatigue effects in the present study

Model	**Simple SP design**				**Complex SP design**				**ALL**			
	Base		**Scaled**		**Base**		**Scaled**		**Base**		**Scaled**	
Coefficients		t(0)		t(0)		t(0)		t(0)		t(0)		t(0)
ASC	0.3534	3.4	0.3969	7.0	0.3304	6.1	0.2428	4.6	0.3493	11.4	0.3286	8.3
Time	-0.1023	-17.1	-0.1240	-7.4	-0.0505	-5.3	-0.0403	-4.8	-0.0705	-12.8	-0.0676	-8.8
Cost	-0.0184	-13.0	-0.0204	-7.9	-0.0115	-8.0	-0.0089	-5.4	-0.0138	-18.9	-0.0128	-9.8
Headway	-0.0735	-7.8	-0.0812	-8.5	-0.0451	-6.8	-0.0352	-5.5	-0.0555	-28.6	-0.0523	-10.3
Punctuality					-0.3138	-16.1	-0.2390	-5.4	-0.3289	-24.2	-0.3033	-8.7
Crowding					-0.0951	-13.9	-0.0734	-5.2	-0.1018	-20.9	-0.0959	-8.0
Scale Factors				t(1)				t(1)				t(1)
Q1 (reference)			1				1				1	
Q2			0.8060	1.4			1.1320	0.5			0.9310	-0.6
Q3			0.9242	0.4			1.3430	1.3			1.0470	0.3
Q4			0.9848	0.1			1.6870	2.0			1.3126	1.8
Q5			0.7970	1.2			1.2620	1.0			0.9995	0.0
Q6			1.4624	-1.8			1.6410	1.9			1.4153	2.3
Q7			0.9065	0.6			1.1810	0.7			1.0288	0.2
Q8			0.6929	1.4			1.5200	1.6			1.0532	0.3
Q9			0.7287	2.1			1.0270	0.1			0.8179	-1.2
No. of Observations	5559		5559		5326		5326		10885		10885	
LL (C)	-3576.0		-3564.9		-3022.4		-3009.2		-6621.8		-6605.4	
ρ^2 **(C)**	0.0372		0.0402		0.1599		0.1635		0.1043		0.1066	
LL test statistics			22.2				26.4				32.8	
Degree of Freedom			8				8				8	
χ^2 Critical Value (5%)			15.5				15.5				15.5	

t(1) refers to the t-statistic vs. 1.0

After the 6th choice, scale factors of SP choice numbers from both complex and simple design start to decrease. With the simple design, the scale factors of the choice from both design fluctuated around 1 with a decreased scale factor in the end. For the complex design, the scale factors for each choice scenario fluctuated around 1.4.

This result can be interpreted by the theory proposed by Keller and Staelin (1987), that complexity of choice experiments may hold a U-Shape relationship with decision effectiveness. That is, as the situations becomes more complex, individuals initially exert additional efforts and become more effective, until a point is reached at which their effectiveness begins to deteriorate. This is also can be explained by the learning effects. In the beginning part of the SP experiments, respondents are not familiar with the 'game', with the increase of choices, thus making more mistakes (gives a higher level of variances). With the increase of the choices, they begin to understand the task and make less mistakes till the point they get bored and tired of the 'game', the variance are increased again.

8.8 Summary of Design Impacts on SP Responses

This section presented influences of design factors and respondents' perceptions on their choice making. The qualitative analysis was conducted prior to the model estimation on the influence of design factors on respondents' perceptions of the SP survey.

MNL and HMNL models were developed to interpret impacts of design factors on SP responses. This study found that the CT script decreased the VoS in the SP experiment; however the impact was not significant. We demonstrated the existence of strategic bias in the context of users' valuation of rolling stock by detecting the difference between the VoSs obtained from experiments with and without the CT script. With the introduction of the CT, valuations of other attributes in this SP study were more consistent with the PDFH (2005) value, which indicated no additional bias was found. This indicated that adding a CT script is an effective method to reduce the bias; however, in this study bias may remain. The effectiveness of the CT script varies among different population.

Adding two more attributes to mask the research aim was introduced to this study as a second method to amend the incentive to strategic bias. It was assumed that respondents would be less likely to perceive the study aim, thus amending their incentives to strategic bias. This study found that the impact of complex design on the valuation of rolling stock was not significant, although some positive interaction effects could be found in the longer journey distance group with the complex design. It was found that complex design contributed to a higher variance of random term in the estimation, which indicated less consistency in the choice making.

Respondents' perceptions affected their decision making. Respondents' perceptions of the potential price increase affected their valuation of improved stocks. If other factors were same, the more likely respondents believed the price would increase, the less value they gave to the improved stock, and they were found to be more consistent in the choice making. Various factors could explain individuals' perceptions of the potential price change. Among them, ticket types, gender, income and some perceptions of the SP experiments (the difficulty and familiarity) were found to have significant impacts.

From the qualitative analysis, the perception of difficulty was found to be closely related with the SP experiments with which individuals experienced. Individuals would more likely to perceive the choice making was difficult if they faced with the complex SP survey. The perception of difficulty of choice making also had some effect on the coefficient estimation. If all else equal, a "u-shaped" pattern was emerged in the impact of perceived difficulty on the valuation of the improved rolling stock and also the estimation of the error variance.

Familiarity with the rolling stock types demonstrated a significant positive effect on the ability to choose, contributing to a less error variance in the choice making.

In the model development, increments of model refinements were added into the utility function. Since changes in later increments could impact the decisions made in earlier stages, it would be appropriate to revisit these decision. In this case, examination of the final model indicated that earlier decisions about which variables to include in the utility function and the structure of the model have demonstrated high degree of consistency of those parameters and their levels of significances across models.

Chapter 9

Conclusions

9.1 Summary of Research

This final chapter aims to provide a summary and conclusion for the research which has been demonstrated in the previous eight chapters. Chapter 1 presented the research objective and framework of the study. Chapter 2 reviewed biases in the previous SP studies. A typology of the biases in SP studies has been developed from the review. Among them, incentives to strategic bias and task complexity effects have been reviewed. Also methods to amend the incentive to bias were discussed in this chapter, which led to the requirements for the study. Chapter 3 reviewed the previous studies on the valuation of rolling stock.

Chapter 4 provided the methodologies used in the study to achieve the objective. Chapters 5 and 6 presented the development of a series of SP experiment for research hypothesis testing.

Chapters 7 and 8 reported results from data analyses. Chapter 7 established a base model for users' valuation of the improved rolling stock. The base model controlled several factors (i.e. income and journey purpose) which cause the variation of valuations to avoid their potential confounding effects. Chapter 8 presented the research hypothesis tests and explored the effects of design factors (cheap-talk and complex design) on SP responses from the base model. In addition, influences of individuals' perceptions on their choice making were explored.

This section provides a summary of research objectives and methodology. Section 9.2 summarises the findings from the research and their implications. Finally, section 9.3 suggests the direction for future research.

9.1.1 Research objectives

SP methods have been used extensively in transport research and elsewhere both for demand forecasting purposes and to value the importance attached to different product features and travel attributes. Alongside the broader acceptance and wider application of SP methods, some practitioners (Bates, 1998; Ampt et al., 2000; Wardman and Shires, 2001) have argued for greater openness in discussing what they see as significant concerns surrounding SP. The present study is motivated by the desire to analyse and reduce biases in the SP application, specifically addressing the issue of the strategic biasing of SP responses.

The aim of this study was to examine the existence and consequence of incentives to strategic bias in the SP application, and the effects of SP design on responses, in the context of users' valuation of the improved rolling stock.

This study reviewed and summarised the concerns surrounding the extent to which the response to hypothetical questions reliably reflect individuals' true preferences when there is an incentive to bias responses. The discussion was illustrated with examples from research in the field of transport, environmental science and market research.

Based on the literature review, this study examined two methods to amend the incentive to bias. This involved the impact of the Cheap Talk (CT) script which was proved to be effective in amending the strategic bias in CV and SP applications in environment studies. This also involved the effect of masking research aim by introducing more attributes, which was motivated by a successful exploration in transport field.

The experiment context was selected as users' valuation of rail passenger improved rolling stock. A review of past studies (chapter 3) found that the values obtained from SP experiments sometimes are much higher than that obtained from other evidence (such as RP method or demand analysis using ticket sales data). The suspected reason was strategic bias in SP responses. A suite of SP experiments was designed. These two methods (the CT script and adding more attributes) were introduced into the SP experiments.

In addition, this research tested the effects of SP design on the estimation, as the above methods might add to the task load for respondents. A review of task complexity effect (section 2.7) found that the context (task complexity) of the SP experiment has a significant impact on individuals' choice making. Therefore, this study examined impacts of adding more attributes to the SP design (which logically increases task complexity) and adding a CT script on respondents' choice making.

Furthermore, this study examined the influence of individuals' perceptions on their choice making. The perceptions of difficulty in the choice making, familiarity of alternatives provided in the SP experiment and perceptions of potential price increase were explored for better understanding respondents' choice making process.

Finally, it provided a study on users' valuation of improved rolling stock in Greater Manchester. The study also investigated the values of service attributes, such as value of time, headway, punctuality and crowding.

9.1.2 Methodology

The aim of this research was to examine the incentive compatibility of SP experiment. The experiment context was selected as users' valuation of the improved rolling stock (Super Sprinters vs. Pacers) in Greater Manchester. To test research hypotheses, a series of SP experiments were developed, with two methods to eliminate the strategic bias.

There were four types of questionnaires combing the addition of a CT script and adding more attributes to mask the research aim (complex design). Three questions were provided to identify influence of respondents' perceptions on their choice making.

The design and development of the paper-based SP questionnaire (chapters 5 and 6) was carried out through two pilot surveys between April and August 2005. The main survey was conducted between October and December 2005 in Greater Manchester in 14 railway stations. The four types of questionnaires were distributed to the rail passengers in rotation. There were 1222 respondents in the sample, which generated about 11000 preference observations.

The analysis technique was based on the random utility theory, in which individuals were assumed to maximise their utility by choosing the option with the highest preference or utility to them. This was used to formulate the multinomial-logit model of preference for rolling stock. Prior to the data analysis, different sources (SP design) of data were combined using simultaneous estimation. A Hierarchical Logit model was established to allow different scale factors for data sets in order to pool the data. A segmentation model was used to identify taste variation of respondents. A Heteroskedastic Multinomial Logit (HMNL) model was later used to examine the impacts of design factors, individuals' socio-economic features and perceptions on their different abilities to cope with the task load in choice making through a parameterization of the scale parameter.

9.2 Main Findings and Implications

9.2.1 Research hypotheses testing

The main findings of this research relate to the examination of SP design impacts on biases in responses and suggestions on how to reduce bias in the SP design and application. Three research hypotheses were tested, which involved the examination of existence and consequence of incentives to strategic bias in the SP experiment in the context of users' valuation of improved rolling stock (from Pacers to Super Sprinters), two methods to amend incentives to strategic bias and the exploration of possible task complexity effects on responses.

Table 9.1 summarises the testing of research hypotheses.

Table 9.1 Research hypotheses testing

Research Hypotheses	Testing Results
H1: The incentive to (strategic) bias exists in the SP exercises. Respondents will overestimate the utility/valuation of the service improvement to increase the likelihood of its introduction.	Detected the strategic bias in the context of users' valuation of rolling stock.
H2: The adding of the cheap-talk can amend individuals' incentive to strategic bias, where $VoS_{CT} < VoS_{NoCT}$	Adding a CT script decreased the estimation of VoS, but the impact was not significant at the 5% level (rejected)
H3A: Masking the research aim (by introducing more attributes) can amend the incentive to strategic bias, where $VoS_{Sim.} > VoS_{Com}$	Adding more attributes did not significantly change the VoS (rejected)
H3B: An increase in the number of attributes will always increase the variance of error terms, thus affecting the valuations implicit in responses.	Adding more attributes showed a significant impact on the estimation of the scale factor (accepted at the 5% level), which indicated higher error variance in the complex SP design.

Following sections present the main findings from research hypotheses testing and implications /insights this study has brought to the experience and knowledge gathered in previous studies.

9.2.2 Sources of bias in the SP application

Based on a desk study of biases observed in the past SP studies, this study developed a typology of biases. Sources of bias can be roughly categorized to three main types: incentive to strategic bias (caused by the hypothetical nature of SP experiments), task complexity effects (caused by the multiple choices presented by SP experiments) and unrealistic design. Our study provided a contribution in the examination of incentive compatibility of SP responses, relating to two methods (adding cheap-talk script and more complexity to SP choices) in the questionnaire. This study reviewed and summarised concerns surrounding the extent to which SP responses to hypothetical questions reliably reflect individuals' true preferences when there is an incentive to bias responses. The discussion was illustrated with examples from research in transport field, environment science and marketing.

9.2.3 Impacts of cheap talk on SP responses

This study introduced a Cheap Talk (CT) script in the SP experiment to investigate its impact on amending the incentive to strategic bias in SP responses. As far as the author noted, this is the first application of the CT script in the transport related context. In previous CT application, respondents were normally university students, where the sample was believed to be homogenous. Respondents in this study were rail passengers in Greater Manchester area. The findings of CT application added the knowledge of understanding how respondents cope with anti-bias information in the real market. The findings of this study are summarised as follows:

- Adding the CT script significantly (at the 5% level) increased respondents' sensitivities to the cost attribute in the SP experiment.

- The VoS_{CT} is lower than the VoS_{NoCT} in all income bands by 18% (or more). The impact is not significant at the normal 5% level.

- According to the segmentation analysis of the CT impact on the ASC term, it was found the CT script is more effective in the less-frequent travellers and more effective in the highest and lowest income group.

- Adding the CT script to SP surveys changed the relative values of other attributes. The values obtained from the SP experiment with a CT script were found to be more consistent to the PDFH (2005) recommended values. The comparison of relative values with the previous evidence in the similar context showed that adding the CT script did not introduce other biases.

- Using PDFH (2005) method, the improved train increased the passengers' demand by 8.3% from the SP experiment without a CT script and by 5.3% from the SP survey with the CT script. The latter is closer to the PDFH (2005) recommended value. However it is still higher which indicates the strategic bias still remains.

- According to the HMNL model analysis, CT did not show a significant impact on the scale parameter. This indicated that adding the CT script did not lead individuals to make more or less errors, thus affecting the consistency of choice making.

- According to the qualitative and model analyses, adding CT script did not show a significant impact on changing individuals' perceptions of the potential price increase.

In summary, adding the CT script in the SP survey corrected the higher value of the improved rolling stock, although the impact was not significant at the 5% level which indicated that strategic bias may remain in our study. CT script provided a cost-efficient and easily implemented way to examine the incentive compatibility in the SP experiment.

9.2.4 Impacts of task complexity on SP responses

This study examined that if adding two attributes to the SP experiment can make the research aim less transparent to respondents, thus amending the incentive to strategic bias. The thesis does not provide innovative theory in this area, but presents findings that contribute to the discussion of task complexity effects in the SP design. Findings are summarised as follows:

- It was found that the VoS obtained from the simple design was slightly lower than that from the complex design. This was contradictory to what we expected. However the difference was not significant at the 5% level.

- Values of time and headway were found to be slightly lower in the complex experiment. The impact was not significant at the 5% level in this study. The possible reason is that when choice task gets complex, respondents allocate more attention to the attributes they are interested in.

- Significant task complexity effect was found in the model estimation. Adding two more attributes contributed to a higher error variance in SP responses, which could not be detected from the simulation test. This indicated that the choice making process is less consistent in the experiment with more attributes.

- According to the qualitative analysis, the relationship between individuals' perceived difficulty in choice making and the SP design (complex) was found to be significant. Respondents who completed the SP experiment with two more attributes were more likely to perceive the choice making was difficult.

- In the study, the range of attribute levels was developed for representing different level of journey distance. It was found with the increase of the attribute range between alternatives, individuals were less consistent in their choice making.

In summary, no significant impact of complex design on the monetary valuation was detected. This indicated with the increase of the number of attributes, individuals made more errors in the choice making but did not bias their answers.

To test research hypotheses (see chapter 1), two more attributes punctuality and crowding were added into the complex SP experiments. These two attributes are very important to rail passengers in their travel. This leads to an open discussion that in the SP experiment with more attributes whether or not respondents would change their decision strategies (for instance, using heuristics) in their choice making, thus causing the variation of relative values and the change of the scale parameter. It is interesting to probe how respondents make SP choices, for instance, asking their personal opinions by focus group.

9.2.5 Influence of perceptions on the SP responses

In the present study, three follow-up questions were provided to investigate impacts of respondents' perceptions on their choice making. Our findings in this aspect contribute to the debate on the more general issue of how to interpret individuals' choice behaviour and improve the SP design. Some significant results were found, shown below.

Impacts of perceptions on difficulty to make the choice

Individuals' perceived difficulty in their choice making has shown a significant impact on their responses. When respondents felt the task was 'a little bit difficult' to finish, they gave less preferences to the improved rolling stock and were shown to be more consistent (make fewer errors) in the choice making.

Impacts of familiarity

It was found that if respondents could recognise the difference between rolling stocks in the choice experiment, they gave more preferences to the improved rolling stock. This indicated that respondents would be willing to pay more for the introduction of the improved rolling stock if they knew the difference of trains. In addition, this study found that individuals who were familiar with the experiment alternatives made fewer errors and were more consistent in the choice making.

Impacts of perceptions on price increase in the SP experiment

The study found that if respondents perceived the potential price increase by the introduction of newer trains, they gave a lower value to the new stock. It was also found that individuals, who believed the potential price change, were more consistent in the choice making.

Various factors contributed to respondents' perceptions of the potential price increase. As demonstrated in the analysis, the CT script did not show a significant impact on respondents' perceptions of the potential price increase due to the introduction of improved trains. Frequent travellers were more likely to believe the potential price increase due to the introduction of the improved trains. The study also found that males were less likely to believe the potential price increase.

9.2.6 Users' valuation of the improved rolling stock

In the PDFH (2005), most of the rolling stock valuation studies were conducted in London or South London area. Our study was based on approximately 11,000 preference observations obtained from 1222 rail passengers in Greater Manchester area. Besides the contribution to the theories of biases in SP study, our study provided an empirical evidence for the valuation of the rail travel attributes.

The standard analysis in chapter 7 concluded that valuations obtained from the study were sensible and generally in line with the PDFH (2005) recommended values and previous evidence: business travellers would be willing to pay the highest (about 11.2% of single fare) to improve the train, followed by commuters (8.8%) and then other groups. Business travellers have the highest value of time (6.56p/min) followed by commuters (5.68p/min) and then other groups of people (2.6p/min). The journey frequency was valued at averagely 0.7 minute of the in-vehicle time, varied by journey purposes. Leisure travellers have the highest time unit value of punctuality and crowding.

Furthermore, according to the model analysis of design factors, we found that attribute values (for commuters) obtained from experiment with the CT script were more consistent with the PDFH (2005) recommended values.

- Commuters are willing to pay 8.3% of the average fare to improve the rolling stock from Pacers to Super Sprinters.
- They are willing to pay 6.68 pence for one minute travel time saving.
- The value of headway is 0.68 times relative to the value of in-vehicle time, which is consistent with the PDFH recommended value.
- The value of punctuality is 5.03 times relative to the value of in-vehicle time, which is suspected to be biased upward.
- The value of crowding is 2.12 times relative to the value of in-vehicle time, which is consistent with the recommended value.
- Using PDFH (2005) method, we derived the demand impact due to the introduction of improved rolling stock at 5.3% (with CT). It is still higher than the recommended value which indicates bias still remain.

9.2.7 Model estimation and comparison

In addition to the traditional MNL model, this study introduced Heteroskedastic Multinomial logit models into the data analysis. By parameterization of the scale parameter, the HMNL model incorporated the impacts of SP design, individuals' characteristics and perceptions of the SP survey into the variance of the error term in the utility function. This study found that applying HMNL models in data analysis, the goodness of fit index improved, which indicated the HMNL model can better explain the travel behaviour and respondents' choice making. The valuations obtained from different models were compared. It was found that the valuation obtained from the HMNL model was not significantly different from those derived from the reference MNL model. However, smaller standard errors were obtained for the monetary values derived from the HMNL model estimation which indicated a more accurate estimate.

9.2.8 Suggestions for the SP design

The reviews, experiments and discussions described in this thesis tackles various issues regarding strategic bias behaviour in SP responses and suggested the method to amend the bias. To have an optimal SP design, in addition to the consideration of statistics property of the SP design, the incentive compatibility and task complexity effects need to be taken into account.

In brief, this research suggests that researchers should consider the incentive compatibility in the SP questionnaire design, especially in the experiment that a new/improved product will be introduced. Due to the hypothetical nature of the SP experiment, the payment is left out of the process. Respondents do not have any financial consequence of their decisions, thus having the incentive to bias their results for a better outcome, for instance, the introduction of the new good. To amend respondents' incentives to strategic bias, Cheap Talk (CT) script provides a cost-efficient and easily implemented way in the hypothetical survey. A CT script includes a discussion of the previous bias and a reminder of the payment. With the help of the CT script in the SP survey, respondents are reminded the payment of the new product and the hypothetical nature of the exercise would be "corrected". With the help of the CT script, individuals also become aware of the potential influence of the context of hypothetical decisions on their valuation of a good. Through the internal correction process, individuals commit cognitive effort to retrieve a more accurate value for the good in question.

Many researchers suggest that SP experiment design should take into account of individuals' cognitive ability in the choice making. This study has detected task complexity effects on SP responses. To better explain individuals' travelling behaviour, more information is suggested to put in the choice experiment to better represent the reality. Nevertheless, individuals are found to be subjected to the cognitive ability in choice making. In this study, respondents were found to make more errors and to be less consistent in the choice making in the SP experiment with more attributes. It was suspected that adding more attributes to the SP experiment would change individuals' strategy to make choices, such as using decision heuristics. The SP design should balance the task load and provision of sufficient information. On the other hand, individuals' ability to cope with the choices should be taken into account in the model analysis for better interpretation of choice behaviour.

The study has found that individuals who are familiar with the experiment are more consistent in the choice making. Therefore, this study suggests, same as many researchers, that SP experiment presentation should be easily understood by individuals and in the way that the difference between alternatives is clear to respondents.

9.3 Suggestions for Further Research

Many of the issues that this thesis raises deserve a more thorough and focused investigation. Some of the issues are here listed as potential themes for further research.

Firstly, the impact of the CT script was found to be nearly significant in this study in amending the incentive to strategic bias in the valuation of improved rolling stocks. Further research can be conducted to vary the context of the CT script and examine its impact. For example, the CT script applied by Cummings and Taylor (1999) explicitly discussed the previous bias in terms of the reasons and magnitude. Some researchers found that a CT script without explaining the direction and magnitude of bias cannot eliminate the hypothetical bias (Poe, 2002; Aadland and Caplan, 2006). It would be interesting to test if in this CT script, the magnitude of the bias is explicitly discussed, for example "the valuation is found to be three times higher than from other evidence", how respondents would cope with this warning message.

This study provided a method for amending the incentive to strategic bias in the SP experiment. A limit to this study is that it was conducted by the paper-based questionnaire; therefore, we could not explore how individuals cope with the CT script in their choice making. From the exploration of respondents' perceptions of the potential price increase, no clear evidence shows that the CT script introduced other types of bias. Further research can be conducted to explore how respondents cope with this information in their choice making by a focus group. Only by doing that, the effectiveness and robustness of the CT script can be observed.

In addition, the reliability of the CT script on similar goods and other products needs further exploration. It is worthwhile to investigate the effectiveness of CT on the other format of SP survey such as the computer based survey. It is suspected that individuals would pay more attention in such format, compared to paper-based mailed-back questionnaires.

Secondly, further exploration of impacts of design factors on the valuation of the improved rolling stock is needed. In the present study, due to the limit of the model specification, some important interaction effects of complex design and journey distance cannot be discerned. In future research, more advanced models can be applied to investigate these impacts. For example, mixed logit model where the attribute parameters are randomly distributed in the population, can be applied to detect the interaction impacts of the design factors and attributes in the utility function. Another extension could be to consider a heteroskedastic-mixed logit specification, where attribute parameters are randomly distributed in the population and the Gumbel scale parameter is the function of design features, respondents' characteristics and their perceptions.

Thirdly, future research could be done on the specific information strategy followed by each individual. This would enrich our understanding on how people process information in a choice experiment.

Hensher et al. (2005, p.204) stated that "It is also important to recognise that complexity is not strictly defined by the quantity of information to process, with more information suggesting greater complexity. Designs, with a small number of attributes and alternatives may, for some individuals, be "complex" if an individual expects more information that they know is relevant in making such a choice in a real market setting".

It is worthwhile to understand how individuals employ a pre-choice algorithm. Theoretical evidence from behaviour science and psychology will provide more insights in this aspect.

Fourthly, in the present research, impacts of individuals' perceptions of the experiment on SP responses were examined. The perception was incorporated into the model to interpret individuals' behaviour. It was found including the perceptions significantly improves the model estimation and could better explain the choice behaviour. Further research can be conducted to incorporate these factors into the SP design and results interpretation. For example, by knowing individuals' familiarity of the experiment, a specific SP design can be developed to tailor individuals' characteristics. On the other side, by detecting the impact of individuals' perceptions of experiment on responses, adjustment can be made to the estimation results by their specific perceptions for better interpretation and forecast. Impacts of individuals' perceptions of the experiments on the choice making still remain unexplored.

References

Aadland, D. and A. J. Caplan (2003), "Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias", American Journal of Agricultural Economics, 85(2), 492-502.

Aadland, D. and A. J. Caplan (2006), "Cheap-talk reconsidered: new evidence from CVM", Journal of Economics Behaviour and Organization, 60, 562-578.

Aarts. H and A. Dijksterhuis (2000), "Habits as knowledge structures: Automaticity in goal-directed behaviour", Journal of Personality and Social Psychology, 78, 53–63.

Abde-Aty, M.A., R. Kitamura, and P.P. Jovanis (1995), "Investigating effective travel time variability on route choice using repeated measurement stated preference data", Transportation Research Record 1493, 39-45.

Accent Marketing and Research (1994), InterCity Rolling Stock Refurbishment Phase 3 Report, Prepared for InterCity, British Railways Board.

Accent Marketing and Research (2006), The Effects of New Rolling Stock on Rail Demand, final report, Prepared for ATOC, UK.

Adamowicz, W.L., J. Louviere, and J. Swait (1998), Introduction to Attribute-Based Stated Choice Method, Submitted to Resource Valuation Branch Damage Assessment Center, NOAA (National Oceanic and Atmospheric Administration), US Department of Commerce.

Adamowicz, W.L., P.C. Boxall, J.J. Louviere, and J. Swait (1999) "Stated Preference methods for valuing environmental amenities" in Bateman, I.J., and K.G. Willis (eds), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Methods in the US, EU and Developing Countries", Chapter 13, 460-482, Oxford University Press.

Ampt, E., J. Swanson, and D. Pearmain (2000), "Stated Preference: Too much Deference?" Steer Davies Gleave Ltd, Stated Preference Modelling Techniques, PTRC, 191-201.

Aptech Systems (1997), GAUSS User's Manual, Mapple Valley, WA.

Arentze, T., A. Borgers, H. Timmermans, and R. DelMistro (2003), "Transport Stated Choice responses: effects of task complexity, presentation format and literacy", Transportation Research Part E, 39, 229-244.

Armstrong, P.M., R.A. Garrido, and J.de D. Ortuzar (2001), "Confidence intervals to bound the value of time", Transportation Research 37E, 143–161.

Arrow, K., R. Solow, P. Portney, E. E. Leamer, R. Radner, and H. Schuman (1993), Report of the NOAA Panel on Contingent Valuation, Federal Register, 58(10), 4602-4614.

ATOC (2002, 2005), Public Transport Forecast Handbook (PDFH 4.0)

Axhausen, M.V. (2002), The Impact of Tilting Trains in Switzerland: a Route Choice Model of Regional and Long Distance Public Transport Trips, paper submitted to 82nd Annual Meeting of the Transportation Research Board.

Babtie Traffic and Transportation (1993), Suffolk Rail Study, Prepared for Suffolk County Council.

Bates, J.J. (1988), "Econometric issues in Stated Preference analysis", Journal of Transport Economics and Policy, pp59-69

Bates, J. (1998), "Reflections on Stated Preference: theory and practice", in J de D. Ortúzar, D.A. Hensher and S. Jara-Diaz (eds), Travel Behaviour Research: Updating the State of Play, Chapter 6, 89-103, Pergamon, UK.

Bates, J. and G. Terzis (1997), "Stated preference and the 'ecological fallacy'", Proceedings 25th European Transport Forum: Seminar F, 155-169.

Ben-Akiva, M. and S.R. Lerman (1985), Discrete Choice Analysis: Theory and Applications to Travel Demand, Cambridge: The MIT Press.

Benshoof, J.A. (1970), "Characteristics of drivers' route selection behaviour", Traffic Engineering and Control 11(12), 604-10.

Bettman, J.R. and P.Kakkar (1977), "Effects of information presentation format on consumer information acquisition strategies", Journal of Consumer Research, 3, 233-240.

Bettman, J.R. and W.Park (1980), "Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis", Journal of Consumer Research, 7, 234-248.

Bhat, C (1995), "A heteroscedastic extreme value model of intercity choice" Transportation Research B, 29

Bohm, P. (1971), "An approach to the problem of estimating demand for public goods", Swedish Journal of Economics, 73 (1), 51-66.

Bohm, P. (1972), "Estimating demand for public goods: an experiment", European Economic Review, 3, 55-66.

Bohm, P. (1979), "Estimating willingness to pay: why and how?" Scandinavian Journal of Economics, 81, 142-153.

Bohm, P. (1984), "Revealing demand for an actual public good", Journal of Public Economics, 24, 135-151.

Bonsall, P.W. (1983), "Transfer price data - its use and abuse", Proceedings of 11th PTRC Summer Conference, 70-81.

Bonsall, P.W. (1986), "Transfer price data – its definition, collection and use", in E. Ampt, W. Brog and A.J. Richardson (eds), Selected Proceedings of Second International Conference on New Survey Methods in Transport, 63-76. VNU Science Press, Utrecht.

Bradley, M. (1988), "Realism and adaptation in designing hypothetical travel choice concepts", Journal of Transport Economics and Policy, 22 (1), 121-137.

Bradley, M. and A.J. Daly (1991), "Estimation of logit choice models using mixed Stated Preference and Revealed Preference information", Proceedings of 6th International Conference on Travel Behaviour, Quebec, Vol 1, 117-133.

Bradley, M. and A.J. Daly (1993), "New analysis issues in stated preference research." Proceedings 21st PTRC Summer Annual Meeting PTRC, UK

Bradley, M. and A.J. Daly (1994), "Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data", Transportation 21, 167-184.

Brown, T. C., I. Ajzen, and D. Hrubes (2003), "Further tests of entreaties to avoid hypothetical bias in referendum Contingent Valuation", Journal of Environmental Economics and Management, 46(2), 353-361.

Brubaker, E. (1975), "Free rider, free revelation, or golden rule", Journal of Law and Economics, 18, 147-161.

Bulte, E., S. Gerking, J.A. List, and A. de Zeeuw (2005), "The effect of varying the causes of environmental problems on stated values: evidence from a field study", Journal of Environmental Economics and Management, 49, 330-342.

Bryman, A. and D. Cramer (2005), Quantitative Data Analysis, Routledge, UK.

Cambridge Dictionaries online (2007), <http://dictionary.cambridge.org/>

Cameron, T., G. Poe, and W. Schulze (2002), "Alternative non-market value-elicitation methods: are revealed and stated preference the same?", Journal of Environmental Economics and Management, 44, 391-421.

Carlsson, F. and P. Martinsson (2001), "Do hypothetical and actual marginal willingness to pay differ in choice experiments?", Journal of Environmental Economics and Management, 27, 179-192.

Carlsson, F., P. Frykblom, and C.J. Lagerkvist (2005), "Using Cheap Talk as a test of validity in choice experiments", Economics Letter, 89, 147-152.

Carson, R.T., T. Groves, and M.J. Machina (2000), Incentive and Informational Properties of Preference Questions, paper presented at the Kobe Conference on Theory and Application of Environment Valuation. Kobe: Kobe University, January.

Caussade, S., J.D. Ortuzar, L.I. Rizzi, and D.A. Hensher (2005), "Assessing the influence of design dimensions on stated choice experiment estimates", Transportation Research Part B, 39, 621-640

Chase, W.G. and H.A. Simon (1973), "Perception in Chess", Cognitive Psychology, 4, 55-81

Chatterjee, A., F. J. Wegmann, and M.A. McAdams (1983), "Non-Commitment bias in public opinion on transit usage", Transportation 11, 347-360.

Cho, H.J. (1998), Route Choice Responses to Various Road Users Charges and Traffic Information, PhD Thesis, Institute for Transport Studies, University of Leeds, unpublished.

Cirillo, C., A. Daly, and K. Lindveld (2000), "Eliminating bias due to the measurement in problem in SP data", in Ortuzar, J.de D.(ed.), Stated Preference Modelling Techniques, PRTC, UK

Clarke, E. H. (1971), "Multipart pricing of public goods", Public Choice 8, 19–33.

Cooper, J. T. Ryley, A. Smyth and A. Alayo (2001), "The interaction between consumer response and urban design: Empirical results from Belfast", Environment and Planning A, 33, 1265-1278.

Cummings, R. G., G. W. Harrison, and L. L. Osborne (1995), "Can the bias of Contingent Valuation surveys be reduced?", Economics working paper, Columbia, SC: Division of Research, College of Business Administration, Univ. of South Carolina.

Cummings, R. G., and L. O. Taylor (1999), "Unbiased value estimates for environmental goods: A Cheap Talk design for the Contingent Valuation method", The American Economic Review.

89(3),649 - 665.

Dargay, J. M. (1993), "Demand elasticities: a comment", Journal of Transport Economics and Policy 27, 87-90.

Dellaert, B.G.C., J. D. Brazell, and J.J. Louviere (1999), "The effect of attribute variation on consumer choice consistency", Marketing Letters 10, 139-147.

DePalma, A., G.M. Meyers, and Y. Y. Papageorgiou (1994), "Rational choice under an imperfect ability to choose", American Economics Review 84, 419-440.

DeShazo, J.R. and G. Fermo (2001), Rational Choice with Cognitive Limitations: Empirical Support for Partial-Information Processing Strategies, Unpublished working paper, Department of Economics, UCLA.

DeShazo, J.R. and G. Fermo (2002), "Designing choice sets for Stated Preference methods: The effects of complexity on choice consistency", Journal of Environmental Economics and Management, 44, 123-143.

DeShazo, J.R. and G. Fermo (2004), "Implications of Rationally-adaptive Pre-choice Behaviour for the Design and Estimation of Choice Models", working paper, School of Public Policy and Social Research, University of California at Los Angeles, August.

Dhar, R. (1997), "Consumer preference for a no-choice option", Journal of Consumer Research, 24, 215–231.

Diamond, P.A. and J. Hausman (1994), "Contingent Valuation: Is some number better than no number?", Journal of Economics Perspectives 8, 45-64.

Erikson, R. S. (1988), "The puzzle of midterm loss", Journal of Politics 50, 1011–1129.

Fischhoff, B. (2002), "Cognitive processes in Stated Preference methods." Forthcoming in the Handbook of Environmental Economics, Karl-Göran Mäler and Jeffrey Vincent (eds.), Elsevier, North-Holland.

Festinger, L. (1957), A Theory of Cognitive Dissonance, Stanford, CA: Stanford University Press.

Fowkes, A.S. (1985), Checking Equi-utility VoTs for SP Design, unpublished Value of Time Working Paper, 25/2/85, UK DoT Value of Time Project.

Fowkes, A.S. (1991), "Recent developments in Stated Preference techniques in transport research", PTRC conference, reprinted in Ortuzar J. de D. (2000), Stated Preference Modelling Techniques, 37-52, PTRC, London.

Fowkes, A.S. (1995), "The influence of modelling error on the shapes of estimated demand functions", PTRC Conference, published as Transportation Planning Methods, 49-60, PTRC, London.

Fowkes, A.S. (1998), The Development of Stated Preference Techniques in Transport Planning, working paper, Institute for Transport Studies, University of Leeds.

Fowkes, A.S., I. Johnson, and P. Marks (1985), Long Distance Business Travel and Mode Choice: The Results of Two Surveys of Business Travellers, Working papers, Institute for Transport Studies, University of Leeds.

Fowkes, A.S., C. A. Nash, and A. E. Whiteing (1985), "Understanding trends in InterCity Rail traffic in Great Britain" Transportation Planning and Technology 10, 65-80.

Fowkes, A.S. and M. Wardman (1988), "Design of Stated Preference travel choice experiments with special reference to taste variation", Journal of Transport Economics and Policy, 22(1), 27-44.

Fowkes, A.S. and C. A. Nash (1991), Analysing Demand for Rail Travel, Avesbury, Aldershot,

Fowkes, A.S., and J. Preston (1991), "Novel approaches to forecasting the demand for new local rail services", Transportation Research A, 25 (4), 209-218.

Fox, J., J. Shogren, D. Hayes, and J. Kleibenstein, (1999), "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." American Journal of Agricultural Economics, 80(3), 455- 65.

Gibbard, A. (1973), "Manipulation of voting schemes: A general result", Econometica 41, 587-601.

Green, J. R. and J.J. Laffont (1978), "A sampling approach to the free rider problem", in Agnar Sandmo, ed., Essays in Public Economics (Lexington, MA: Lexington Books).

Green, P.E. and V. Srinivasen (1978), "Conjoint analysis in consumer research issue and outlook", Journal of Consumer Research, 5, 103-123.

Green, P.E. and V. Srinivasan (1990), "Conjoint analysis in marketing: new developments with implications for research and practice", Journal of Marketing, vol. 54, no. 4, 3-19.

Groves, T. (1973), "Incentive in teams," Econometrica, 41, 617-631.

Groves, T., R. Radner, and S. Reiter, eds. (1987), "Information, Incentives, and Economics Mechanisms: Essays in Honour of Leonid Hurwicz", 48-111, Minneapolis: University of Minnesota Press.

Gunn, H. (2001), "Spatial and temporal transferability of relationships between travel demand, trip cost and travel time", Transportation Research Part E 37, 163-189.

Haab, T.C., J.C.Huang and J.C.Whitehead (1999), "Are hypothetical referenda incentive compatible? A comment." Journal of Polit. Economy 107, p. 186-196

Hague Consulting Group (2000), ALOGIT 4.0EC.

Hague Consulting Group, Accent Marketing and Research and Department of the Environment, Transport and Regions (1999), The Value of Travel Time on UK Roads. The Hague, Netherlands.

Hanemann, M. (1984), "Welfare evaluations in contingent valuation experiments with discrete responses", American Journal of Agricultural Economics 66, 332–341.

Hastings, N. and J.B. Peacock (1975), Statistical Distributions, New York: John Wiley and Sons.

Heiner, R. A. (1983), "The origin of predictable behaviour", American Economic Review 73, 662-676.

Hensher, D. (1994), "Stated Preference analysis of travel choices: The state of practice", Transportation 21 (1994), 107-133.

Hensher, D. A. (1998), "Intercity rail services: a nested logit stated choice analysis of pricing options", Journal of Advanced Transportation, 32, 130-151.

Hensher, D. A.(2006a), "Revealing differences in willingness to pay due to the dimensionality of Stated Choice designs: an initial assessment", Environmental and Resource Economics34, 7-44

Hensher, D. A.(2006b), "Towards a practical method to establish comparable values of travel time savings from Stated Choice experiments with differing design dimensions", Transportation Research Part A 40, 829-840.

Hensher, D.A. and J.J. Louviere (1983), "Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event", Journal of Consumer Research, Vol. 10(3), 348-361.

Hensher, D.A. and M. Bradley (1993), "Using stated response data to enrich revealed preference discrete choice models", Marketing Letters 4 (2), 139-152.

Hensher, D.A., J.J. Louviere, and J. Swait (1999), "Combining sources of preference data", Journal of Econometrics, vol. 89, no. 1-2, 197-221.

Hensher, D.A, P.R. Stopher, and J.J. Louviere (2001), "An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice applications", Journal of Air Transport Management 7, 373-379.

Hensher, D.A., J. Rose, and W.H. Greene (2005a), Applied Choice Analysis: A Primer, Cambridge University Press, Cambridge.

Hensher, D., J. Rose and W.H. Greene (2005b), "The implication on willingness to pay of respondents ignoring specific attributes", Transportation, vol. 32 (3), 1-19.

Hensher, D., J. Rose and T. Bertoia (2007), "The implications of willingness to pay of a stochastic treatment of attribute processing in Stated Choice studies", Transportation Research Part E 43, 73-89.

Holden, D. G. (1992), Design Procedures for Stated Preference Experiments, PhD thesis, Institute for Transport Studies, University of Leeds.

Hurwicz, L. (1972), "On informationally decentralized systems", in Radner, R. and C. B. McGuire (Eds.), Decision and Organization: A Volume in Honor of Jacob Marschak, 297–336. Amsterdam: North-Holland.

Hurwicz, L. (1986), "Incentive aspects of decentralization", in K.J.Arrow and M. D. Intrilligator, eds, Handbook of Mathematical Economics, vol. 3, Amsterdam: North-Holland.

Johnson, E.J. and J.E. Russo (1981), "Product familiarity and learning new information", in Advances in Consumer Research, Vol. 8 ed. Kent B. Monroe, Ann Arbor, M1:Association for Consumer Research.

Johnson, E.J. and J.E. Russo (1984), "Product familiarity and learning new information", Journal of Consumer Research, 11, 542-551.

Jones, P. (1997), "Addressing the 'packaging' problem in stated preference studies". In Proceedings of Seminar D, PTRC European Transport Forum.

Kahneman, D. and A. Tversky (1979), "Prospect theory: an analysis of decision under risk", Econometrica 47, 263-291

Kaplan, R. & S. Kaplan (1989), The Experience of Nature: A Psychological Perspective, Cambridge: Cambridge University Press.

Keller. K. L. and R. Staelin (1987), "Effects of quality and quantity of information on decision and effectiveness", Journal of Consumer Research, 15, 411-421.

Kroes, E. P. and R. J. Sheldon (1988), "Stated Preference methods – an Introduction", Journal of Transport Economics and Policy, vol. XXII, no1, 11-25.

Kocur, G., T. Adeler, W.Hyman and B.Aunet (1982), Guide to Forecasting Travel Demand with Direct Utility Assessment, Technique Report, US Department of Transportation, USA.

Kottenhoff, K. (1999), Evaluation of Passenger Train Concepts - Methods and Results of Measuring Travellers' Preferences in Relation to Costs, KTH, Division of Traffic and Transport Planning, Stockholm.

Kottenhoff, K. and C. Lindh (1996), "The value and effects of introducing high standard train and bus concepts in Blekinge, Sweden", Transport Policy 2 (4), 235–241.

Koppelman, F.S. and V. Sethi (2005), "Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modelling long distance travel choice behaviour", Transportation Research Part B 39, (9), 825–853.

Kuhfeld, W. F., R.D. Tobias, and M. Garatt (1994), Efficient Experimental Design with Marketing Research Applications, SAS Institute, TS-650C.

Lawrence, H (1990), The Effect of Intangible Product Attributes on Rail Passenger Demand with Special Reference to Ride Quality, PhD thesis, Cranfield Institute of Technology, Centre for Transport Studies.

Lerman, S. R., and J.J. Louviere (1978), "On the use of direct utility assessment to identify functional from in utility and destination choice models", Transport Research Record, 673, 78-86.

List, J.A. (2001), "Do explicit warning eliminates the hypothetical bias in elicitation Procedures? Evidence from Field auction experiments", American Economic Review, 91(5), 1498-1507.

List, J.A. and C.A.Gallet (2001), "What experimental protocol influence disparities between actual and hypothetical stated values?", Environmental and Resource Economics 20, 241-54.

List, J.A., P. Sinha, and M.H. Taylor (2006), "Using choice experiments to value non-market goods and services: Evidence from field experiments", The B.E. Journal of Economic Analysis and Policy, Vol 6(2).

Loomis, J. B., T. Brown, B. Lucero and G. Peterson (1996), "Improving validity experiments of Contingent Valuation methods: Results of efforts to reduce the disparity of hypothetical and actual willingness to pay", Land Economics, 72(4), 450- 461.

Louviere, J. J. (1988), Analyzing Decision Making: Metric Conjoint Analysis, Newbury Park; London.

Louviere, J. J. and D. A. Hensher (1983), "Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event", Journal of Consumer Research, 10, 348-361.

Louviere, J. J. and J. D. Swait (1996), "Searching for regularities in choice processes, or the little constant that could", working paper, Department of Marketing, University of Florida, Gainesville.

Louviere, J J., D. A. Hensher, and J. Swait (2000) Stated Choice Methods: Analysis and Application, Cambridge : Cambridge University Press.

Louviere, J. J. (2001), "Choice Experiments: an overview of Concepts and Issues" in The choice Modelling Approach to Environmental Valuation, J. Bennett & R. Blamey, eds., Edward Elgar Publishing Limited, Cheltenham, UK, 13-36.

Luce, R.D. (1959), Individual Choice Behaviour. New York: Wiley.

Luce, R.D. and J.W. Tukey (1964), "Simultaneous conjoint measurement". Journal of Mathematical Psychology, 1, 1-27.

Lust, J.L. and T.C. Schroeder (2004), "Are choice experiments incentive compatible? A test with quality differentiated beefsteaks", American Journal of Agricultural Economics 86, 467-482.

Mackie, P.J., M. Wardman, A.S. Fowkes, G. Whelan, J. Nellthorp, and J. Bates (2003), "Values of Travel Time Savings in the UK", Report prepared for Department of Transport. Institute for Transport Studies, University of Leeds.

Malhota, N.K. (1982), "Information load and consumer decision making", Journal of Consumer Research 8, 419-430.

Mazzotta, M.J. and J.J. Opaluch (1995), "Decision making when choices are complex: A test of Heiner's hypothesis", Land Economics 71, 500-515.

Manski, C. (1977), "The structure of random utility models", Theory and Decision, 8, 229-54.

Margolis, H.M. (1981), "A new model of rational choice", Ethics, 91, 265-279.

McFadden, D. (1973), "Conditional logit analysis of qualitative choice behaviour", In Frontiers of Econometrics, ed. P.Zarembka. New York Academic Press.

McFadden, D. (1981), "Econometric models of probabilistic choice". In: Manski, C., McFadden, D., (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA 198-272.

McFadden, D. (1986), "The choice theory approach to market research", Marketing Science, 5, 275-297.

McFadden, D. (2001), "Economic choices", American Economic Review, vol. 91, no. 3, 351-378.

Miller, D.T., and D. McFarland (1987), "Pluralistic ignorance: when similarity is interpreted as dissimilarity", Journal of Personality and Social Psychology, 53(2), 298-305.

Mitchell, R.C. and R.T. Carson (1989), Using Surveys to Value Public Goods: The Contingent Valuation Method, Resources for the Future, Washington DC.

Morikawa, T. (1989), Incorporating Stated Preference Data in Travel Demand Analysis, Ph.D. Dissertation, Department of Civil Engineering, MIT.

Murphy, J. J., T. H. Stevens, P. G. Allen, and D. Weatherhead (2003), A meta-analysis of hypothetical bias in Stated Preference valuation, working paper, Amherst, MA: Univ. of Massachusetts, Dept. of Resource Economics.

MVA Consultancy (1985), InterCity Rolling Stock Market Research: Report on Phase One, Prepared for British Railways Board.

MVA Consultancy, ITS University of Leeds, TSU University of Oxford (1987), The Value of Travel Time Savings, Policy Journals, Newbury, Berks.

MVA Consultancy (1991), Market Research for ORCATS Model, Prepared for InterCity, British Railways Board

MVA Consultancy (1992), InterCity 225 Evaluation, Prepared for InterCity, British Railways Board.

MVA Consultancy (1993), "Passenger Priorities Research Stage 1", Prepared for London Underground Limited.

National Oceanographic and Atmospheric Administration (NOAA) (1994), "Proposed Rules for Valuing Environmental Damages", Federal Register 59, 1062-1191.

Neil, H. R. (1995), "The context for substitutes in CVM Studies: some empirical observations", Journal of Environmental Economics and Management, 29(3), 393-397.

Opaluch, J. J. and K. Segerson (1989), "Rational roots of 'irrational' behaviour: New theories of economic decision-making", Northeastern Journal of Agriculture and Resource Economics, Vol. 18 No. 2.

Oscar Faber TPA (1994), Rolling Stock Refurbishment Benefit Cost Study, Prepared for North London Railways.

Ortúzar, J. de D.and R.A. Garrido (1991), "Rank, rate or choice? An evaluation of SP methods in Santiago", Proceedings 19th PRTC Summer Annual Meeting: Seminar G, 301-312.

Ortúzar, J. de D.and A. G. Rodrigo (1994), "On the semantic scale problems in stated preference rating experiments", Transportation 21, 185-201.

Ortúzar, J de D., D.A.Roncagliolo and U.C. Velarde (1997), Interactions and independence in stated preference modelling, Proceedings of 24th European Transport Forum: Seminar F. PTRC, UK, 143-154

Ortúzar, J. de D. and L. I. Rizzi (2001), "Valuation of road fatalities: A stated preference approach", Travel Behaviour Research: The Leading Edge. D. Hensher Ed., Elsevier Science Ltd, Oxford, 855-867.

Ortúzar, J. de D. and G. Rodriguez (2002), "Valuing reductions in environmental pollution in a residential location context", Transportation Research D 7 (6), 407-427.

Ortúzar, J. de D. and L.G. Willumsen (2002), Modelling Transport, John Wiley and Sons, Third edition.

Osterlind, S. J. (1976), Test Item Bias, a SAGE UNIVERSITY PAPER.

Ouwersloot, H. and P. Rietveld (1996), "Stated choice experiments with repeated observations", Journal of Transport Economics and Policy 30(2), 203-212.

Payne, J.W. (1976), "Task complexity and contingent processing in decision making: an information search and protocol analysis", Organizational Bahaviour and Human Performance, 16, 366-380.

Payne, J., J. Bettman, E. Coupey, and E. Johnson (1992), "Behavioural decision research: A constructive processing perspective", Annual Review of Psychology, 43, 87-131.

Pearmain, D. and E. Kroes (1990), Stated Preference Techniques A Guide to Practice, Steer Davies Gleave Ltd, Richmond, Surrey, U.K.

Pearmain, D. and J. Swanson (1991), Stated Preference Techniques: A guide to Practice, (Second Edition), Steer Davies Gleave.

Poe, G. L., J. E. Clark, D. Rondeau, and W. D. Schulze (2002), "Provision point mechanisms and field validity tests of Contingent Valuation", Environmental and Resource Economics, 23, 105-131.

Powe, N.A., G.D. Garrod, and P.L. McMahon (2005), "Mixing methods within stated preference environmental valuation: choice experiments and post-questionnaire qualitative analysis", Ecological Economics 52, 513– 526.

Saelensminde, K. (2001), "Inconsistent choices in Stated Choice data: use of logit scaling approach to handle resulting variance increases", Transportation 28, 269-296.

Samuelson, P. (1954), "The pure theory of public expenditure", The Review of Economics and Statistics, 36(4), 387-389.

Samuelson, W. and R. Zeckhauser (1988), "Status quo bias in decision making", Journal of Risk and Uncertainty 1 March, 7-59.

Satterthwaite, M. (1975), "Strategy-proofness and arrow conditions: existence and correspondence theorems for voting procedures and welfare functions", Journal of Economic Theory 10, 187-217.

Schlag, B. and J.Schade (2000), "Public acceptability of traffic demand management and pricing measures in Europe", Transport Engineering and Control, 41(8), 314-318.

Schuman, H. and S. Presser (1981), Questions and Answers in Attitude Surveys, New York: Academic Press.

Simon, H. (1955), "A behavioural model of rational choice", Quarterly Journal of Economics, 69, 99-118.

Swait, J. (2000), "Distinguishing taste variation from error structure in discrete choice data", Transportation Research Part B 34, 1-15.

Swait, J. and J.J. Louviere (1993), "The role of the scale parameter in the estimation and comparison of multinomial logit model", Journal of Marketing Research, vol. 30(3), 305-314.

Swait, J. and W. Adamowicz (2001), "Choice environment, market complexity, and consumer behaviour: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice", Organizational Behaviour and Human Decision Processes, Vol. 86 No.2, 141-167.

Train, K., D. Revelt and P. Ruud (1999), Mixed Logit Estimation Routine for Panel Data, *<http://elsa.berkeley.edu/Software/abstracts/train0296.html>*,

Train, K. (2003), Discrete Choice Methods with Simulation, Cambridge University Press, U.K.

Throsby, C.D. and G.A.Withers (1986), "Strategic bias and demand for public goods: Theory and an application to the Arts", Journal of Public Economics 31 307-327.

Tudela, A. (2000), "Simulations as a necessary step in the design of stated preference experiments", Proceedings of European Transport Conference: Seminar F, 81-91, PTRC, UK.

Tversky, A. (1972), "Elimination by aspects: a theory of choice", Psychological Review 79, 281-299.

Tversky, A. and E. Shafir (1992), "Choice under conflict: the dynamics of deferred decision," Psychological Science, 6, 358–361.

Varian, H. (1992), Microeconomics Analysis, 3rd ed. (New York: Norton).

Vickrey, W. (1961), "Counterspeculation, auctions, and competitive sealed tenders", Journal of Finance 16, 8–37.

Wardman, M.R. (1986), Route choice and the value of motorists' travel time: theoretical and methodological issue, working paper, Institute for Transport Studies, University of Leeds.

Wardman, M.R. (1987), An Evaluation of the Use of Stated Preference and Transfer Price Data in Forecasting the Demand for Travel, PhD thesis, Institute for Transport Studies, University of Leeds.

Wardman, M.R. (1988), "A comparison of revealed and stated preference models of travel behaviour" Journal of Transport Economics and Policy 22, 71-91.

Wardman, M.R. (1998), "The value of travel time: A review of the British evidence". Journal of Transport Economics and Policy, 32(3), 285-316.

Wardman, M.R. (2001), "A review of British evidence on time and service quality valuations" Transportation Research Part E: Logistics and Transportation Review 37, 107-128.

Wardman, M.R. (2003), Reliability of Value Obtained from Stated Preference Methods, working paper, Institute for Transport Studies, University of Leeds.

Wardman, M.R. (2004), "Public transport values of time", Transport Policy 11, 363-377.

Wardman, M.R and G. Whelan (1998), Rolling Stock Quality Improvements and User Willingness to Pay, Working Paper 523, Institute for Transport Studies, University of Leeds.

Wardman, M.R and G. Whelan (2001), "Valuation of improved railway rolling stock: a review of the literature and new evidence", Transport Review, Vol 21(4), 415-447.

Wardman, M.R and J. Shires (2003), Review of Fare Elasticities in Great Britain, Working Paper 573. Institute for Transport Studies, University of Leeds.

Wardman, M. R. and A. L. Bristow (2003), "Valuation of Aircraft noise using Stated Preference Methods within a Broader Quality of Life Dimension" ETC 2003, Strasbourg FRANCE.

Wardman, M.R. and A.L. Bristow (2004), "Traffic related noise and air quality valuations: evidence from Stated Preference residential choice models", Transportation Research Part D: Transport & Environment, 9, 1-27.

Wardman, M.R. and A.L.Bristow (2005), Incentive to Bias in Stated Preference Valuation of Aircraft Noise, Institute for Transport Studies, University of Leeds, unpublished.

Wegener, D. and R. Petty (1995), "Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias", Journal of Personality and Social Psychology, 68(1), 36-51.

Whelan, G. (2003), Modelling Car Ownership in Great Britain, PhD thesis, Institute for Transport Studies, University of Leeds.

Appendix A

An Example of the SP Questionnaires (Simple Design)

RAIL TRAVEL SURVEY

ITS

Dear Passenger

Thank you for agreeing to answer this short questionnaire about your journey today. This survey is being undertaken as part of research into Rail Travel at the Institute for Transport Studies, University of Leeds. The information you provide will be treated confidentially.

This questionnaire contains four parts. Once completed, please return the questionnaire in the FREEPOST envelope. If you have any queries about this study or how to complete this form, please contact Lucy on 0113-343-7325. It will take you about 10 minutes to complete this questionnaire. Thank you for your help.

Part 1 – About your Journey

Q1 Please list the stations where you get on and off the train on the current leg of your journey.

Starting Station: ☐

Interchange station: ☐

Final Station: ☐

Q2 Is your ticket (Please tick one box)?

Standard Day Single	☐	Standard Day Return	☐	Cheap Day Return	☐
Rail Ranger	☐	Day Saver Ticket	☐	Weekly Season Ticket	☐
Monthly Season Ticket	☐	Annual Season Ticket	☐	County Card	☐
Other-please specify	☐	____________			

Q3 How much did your ticket cost? £________

Q4 Is your ticket paid for by others? (E.g. reimbursed by your employer) Yes ☐ No ☐

Q5 What is the purpose of your journey?

Commuting to/from work	☐	Employer's business	☐	Personal Business	☐
To/from School /College	☐	Visiting friends/relatives	☐	Sport/Entertainment	☐
Shopping	☐	Other – please specify	☐	____________	

Q6 According to the timetable, how long will the train journey take? ______*hrs*_____*mins* *Don't Know* ☐

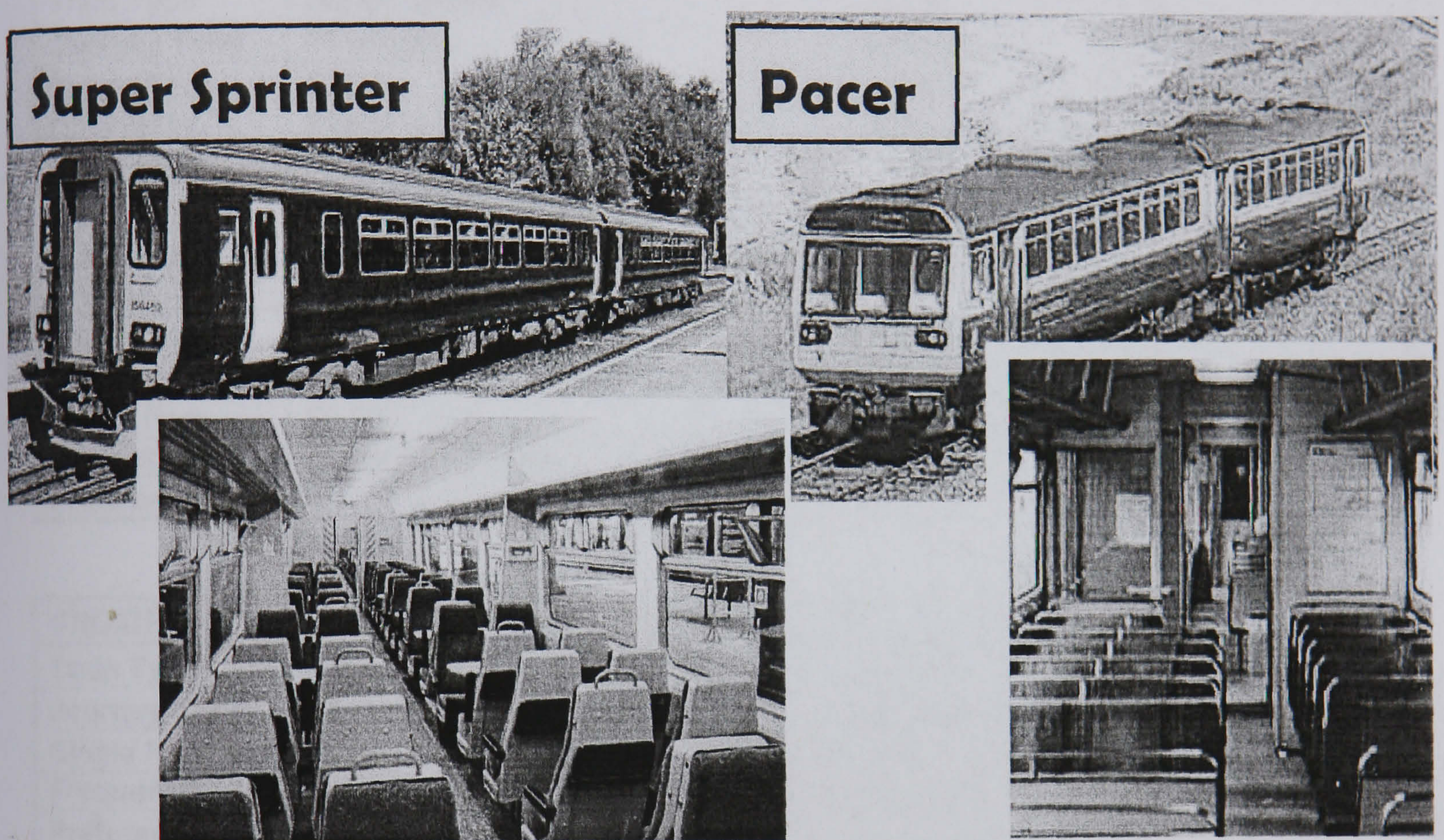
Q7 How often do you make the journey in Question 1?

5 or more times a week	☐	2 to 4 times a week	☐	Once a week	☐
Once every two weeks	☐	Once a month	☐	Less Frequent/First Time	☐

Please Turn Over

Part 2 –Indicate your preference on the train services

One way of improving a train service is introducing newer trains. We would like to know how you react to such a change by presenting you with sets of fictional options. Imagine you have the choice between two types of train: Super Sprinters (New) and Pacers (Old). Super Sprinters are air-conditioned and Pacers have the alternative of opening the window. The trains' pictures are presented below.



Each situation is described in terms of Type of Train, Journey Time (the amount of time spent on the train), Single Fare and Frequency of the service. You should assume that everything else is the same for the two options. **Now please consider each of the 9 choices set out below and in each case, tick your preferred option.**

CHOICE 1	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	25 minutes	30 minutes
Single Fare	£3.00	£1.80
Frequency	Every 15 minutes	Every 30 minutes
Preference	☐	☐

CHOICE 2	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	20 minutes	30 minutes
Single Fare	£2.50	£2.00
Frequency	Every 15 minutes	Every 20 minutes
Preference	☐	☐

CHOICE 3	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	25 minutes	25 minutes
Single Fare	£2.00	£2.50
Frequency	Every 20 minutes	Every 10 minutes
Preference	☐	☐

CHOICE 4	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	15 minutes	25 minutes
Single Fare	£2.00	£1.80
Frequency	Every 15 minutes	Every 20 minutes
Preference	☐	☐

CHOICE 5	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	20 minutes	20 minutes
Single Fare	£2.00	£1.80
Frequency	Every 20 minutes	Every 10 minutes
Preference	☐	☐

CHOICE 6	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	15 minutes	30 minutes
Single Fare	£3.00	£2.00
Frequency	Every 20 minutes	Every 10 minutes
Preference	☐	☐

CHOICE 7	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	20 minutes	25 minutes
Single Fare	£3.00	£2.00
Frequency	Every 15 minutes	Every 30 minutes
Preference	☐	☐

CHOICE 8	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	25 minutes	20 minutes
Single Fare	£2.20	£2.00
Frequency	Every 15 minutes	Every 20 minutes
Preference	☐	☐

CHOICE 9	Option A	Option B
Train Type	Super Sprinter	Pacer
Journey Time	15 minutes	20 minutes
Single Fare	£2.50	£2.00
Frequency	Every 15 minutes	Every 30 minutes
Preference	☐	☐

Please Turn Over

Part 3 – About You

This information helps us to determine whether the sample in our survey is representative.

Q8 Are you male ☐ or female☐?

Q9 In which of the following age groups does your own age fall?

Under 18	☐	18-25	☐	26-35	☐
36-50	☐	51-59	☐	60 and Over	☐

Q10 What is your annual income?

Less than £10k	☐	£10k-£20k	☐	£21k-£35k	☐
£36k-£50k	☐	Over £50k	☐	Do not want to say	☐

Part 4 – About the Survey

Q11 Did you find it difficult in making the choices?

Yes, very ☐ Yes, quite ☐ Yes, a little ☐ No ☐

Q12 Did you feel confident that you could distinguish Super Sprinters from Pacers with the help of the information provided?

Not at all ☐ Fairly ☐ Very ☐ Not sure ☐

Q13 How likely do you think it is that fares would increase if new trains would be introduced?

Not at all ☐ Slightly ☐ Moderately ☐ Very ☐

Your comments are welcomed and greatly appreciated!

Thank you for your co-operation!

End

This research is not commissioned by Northern Rail Ltd and is conducted for academic purposes only. Northern Rail Ltd will not necessarily be using the results from this survey. Northern Rail Ltd has authorised the survey, but has no involvement with the survey content and distances itself from any suggestions made in the survey.

Appendix B

An Example of the SP Questionnaires (Complex Design)

RAIL TRAVEL SURVEY

ITS

Dear Passenger

Thank you for agreeing to answer this short questionnaire about your journey today. This survey is being undertaken as part of research into Rail Travel at the Institute for Transport Studies, University of Leeds. The information you provide will be treated confidentially.

This questionnaire contains four parts. Once completed, please return the questionnaire in the FREEPOST envelope. If you have any queries about this study or how to complete this form, please contact Lucy on 0113-343-7325. It will take you about 10 minutes to complete this questionnaire. Thank you for your help.

Part 1 – About your Journey

Q1 Please list the stations where you get on and off the train on the current leg of your journey.

Starting Station: ☐

Interchange station: ☐

Final Station: ☐

Q2 Is your ticket (Please tick one box)?

Standard Day Single	☐	Standard Day Return	☐	Cheap Day Return	☐
Rail Ranger	☐	Day Saver Ticket	☐	Weekly Season Ticket	☐
Monthly Season Ticket	☐	Annual Season Ticket	☐	County Card	☐
Other-please specify	☐	______________			

Q3 How much did your ticket cost? £________

Q4 Is your ticket paid for by others? (E.g. reimbursed by your employer) Yes ☐ No ☐

Q5 What is the purpose of your journey?

Commuting to/from work	☐	Employer's business	☐	Personal Business	☐
To/from School /College	☐	Visiting friends/relatives	☐	Sport/Entertainment	☐
Shopping	☐	Other – please specify	☐	______________	

Q6 According to the timetable, how long will the train journey take? ______*hrs*_____*mins* *Don't Know* ☐

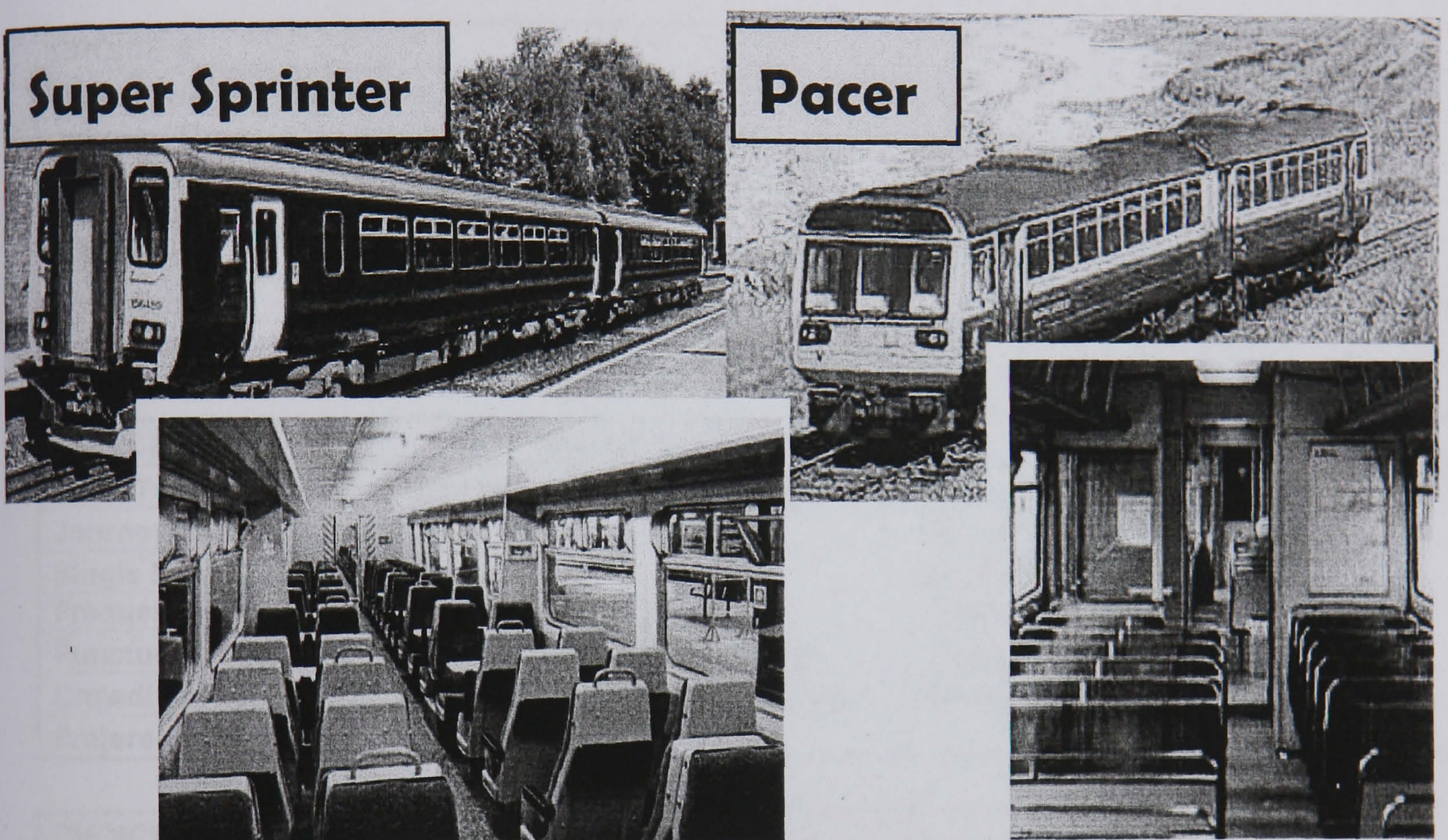
Q7 How often do you make the journey in Question 1?

5 or more times a week	☐	2 to 4 times a week	☐	Once a week	☐
Once every two weeks	☐	Once a month	☐	Less Frequent/First Time	☐

Please Turn Over

Part 2 –Indicate your preference on the train services

One way of improving a train service is introducing newer trains. We would like to know how you react to such a change by presenting you with sets of fictional options. Imagine you have the choice between two types of train: Super Sprinters (New) and Pacers (Old). Super Sprinters are air-conditioned and Pacers have the alternative of opening the window. The trains' pictures are presented below.



Each situation is described in terms of Type of Train, Journey Time (the amount of time spent on the train), Single Fare, Frequency of the service, Punctuality and Crowding. You should assume that everything else is the same for the two options. **Now please consider each of the 9 choices set out below and in each case, tick your preferred option.**

CHOICE 1	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	15 minutes	20 minutes
Single Fare	£2.00	£1.80
Frequency	Every 15 minutes	Every 20 minutes
Punctuality	1 out of 5 times delay for 10 minutes	2 out of 5 times delay for 10 minutes
Crowding	1 out of 5 times stand for whole journey	Enough seats
Preference	☐	☐

CHOICE 2	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	20 minutes	25 minutes
Single Fare	£2.50	£2.00
Frequency	Every 15 minutes	Every 20 minutes
Punctuality	Always on time	1 out of 5 times delay for 10 minutes
Crowding	1 out of 5 times stand for whole journey	Enough seats
Preference	☐	☐

CHOICE 3	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	20 minutes	20 minutes
Single Fare	£3.00	£2.00
Frequency	Every 15 minutes	Every 30 minutes
Punctuality	1 out of 5 times delay for 10 minutes	2 out of 5 times delay for 10 minutes
Crowding	2 out of 5 times stand for whole journey	1 out of 5 times stand for whole journey
Preference	☐	☐

CHOICE 4	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	15 minutes	25 minutes
Single Fare	£3.00	£2.00
Frequency	Every 20 minutes	Every 10 minutes
Punctuality	Always on time	1 out of 5 times delay for 10 minutes
Crowding	Enough seats	2 out of 5 times stand for whole journey
Preference	☐	☐

CHOICE 5	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	25 minutes	20 minutes
Single Fare	£2.00	£2.50
Frequency	Every 20 minutes	Every 10 minutes
Punctuality	2 out of 5 times delay for 10 minutes	Always on time
Crowding	1 out of 5 times stand for whole journey	Enough seats
Preference	☐	☐

CHOICE 6	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	25 minutes	25 minutes
Single Fare	£2.00	£1.80
Frequency	Every 15 minutes	Every 30 minutes
Punctuality	1 out of 5 times delay for 10 minutes	2 out of 5 times delay for 10 minutes
Crowding	Enough seats	2 out of 5 times stand for whole journey
Preference	☐	☐

CHOICE 7	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	15 minutes	30 minutes
Single Fare	£2.50	£2.00
Frequency	Every 15 minutes	Every 30 minutes
Punctuality	2 out of 5 times delay for 10 minutes	Always on time
Crowding	2 out of 5 times stand for whole journey	1 out of 5 times stand for whole journey
Preference	☐	☐

CHOICE 8	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	20 minutes	30 minutes
Single Fare	£2.00	£1.80
Frequency	Every 20 minutes	Every 10 minutes
Punctuality	2 out of 5 times delay for 10 minutes	Always on time
Crowding	Enough seats	2 out of 5 times stand for whole journey
Preference	☐	☐

Please Turn Over

CHOICE 9	Option A	Option B
Train Type	Super Sprinters	Pacers
Journey Time	25 minutes	30 minutes
Single Fare	£3.00	£2.00
Frequency	Every 15 minutes	Every 20 minutes
Punctuality	Always on time	1 out of 5 times delay for 10 minutes
Crowding	2 out of 5 times stand for whole journey	1 out of 5 times stand for whole journey
Preference	☐	☐

Part 3 – About You

This information helps us to determine whether the sample in our survey is representative.

Q8 Are you male ☐ or female ☐?

Q9 In which of the following age groups does your own age fall?

Under 18	☐	18-25	☐	26-35	☐
36-50	☐	51-59	☐	60 and Over	☐

Q10 What is your annual income?

Less than £10k	☐	£10k-£20k	☐	£21k-£35k	☐
£36k-£50k	☐	Over £50k	☐	Do not want to say	☐

Part 4 – About the Survey

Q11 Did you find it difficult in making the choices?

Yes, very ☐ Yes, quite ☐ Yes, a little ☐ No ☐

Q12 Did you feel confident that you could distinguish Super Sprinters from Pacers with the help of the information provided?

Not at all ☐ Fairly ☐ Very ☐ Not sure ☐

Q13 How likely do you think it is that fares would increase if new trains would be introduced?

Not at all ☐ Slightly ☐ Moderately ☐ Very ☐

Your comments are welcomed and greatly appreciated!

Thank you for your co-operation!

End

This research is not commissioned by Northern Rail Ltd and is conducted for academic purposes only. Northern Rail Ltd will not necessarily be using the results from this survey. Northern Rail Ltd has authorised the survey, but has no involvement with the survey content and distances itself from any suggestions made in the survey.

Appendix C

Simulation for other SP designs

Simulation tests of SP design (Band A) – Model Specification 1

Input Values			Estimated Coefficients								Estimated Values				
VoT	VoH	VoS	Time	(t)	Cost	(t)	Headway	(t)	ASC	(t)	VoT	VoH	VoS	ρ^2(c)	SD
8	5	20	-0.2172	-8.3	-0.0255	-5.1	-0.1230	-6.8	0.5310	5.8	8.53	4.83	20.86	0.1126	50
8	5	20	-0.2238	-8.5	-0.0309	-6.0	-0.1545	-7.8	0.6567	7.0	7.25	5.00	21.26	0.1233	50
											7.89	**4.92**	**21.06**		
5	3	20	-0.1924	-7.6	-0.0440	-8.3	-0.1169	-6.7	0.9448	9.0	4.37	2.66	21.47	0.0820	30
5	3	20	-0.2428	-9.0	-0.0496	-9.0	-0.1338	-7.2	0.9353	9.0	4.90	2.70	18.88	0.1095	30
											4.64	**2.68**	**20.17**		
6	4	10	-0.2253	-9.0	-0.0404	-8.0	-0.1826	-9.6	0.4386	4.7	5.57	4.52	10.85	0.1311	30
6	4	10	-0.2956	-10.9	-0.0469	-9.0	-0.1703	-9.3	0.2752	2.9	6.31	3.63	5.87	0.1606	30
											5.94	**4.07**	**8.36**		

Note: The value of the improved rolling stock transferred to be positive

Input Values					Estimated Coefficients						Estimated Values						
VoT	VoH	VoS	VoP	VoC	Time	Cost	Headway	ASC	Punctuality	Crowding	VoT	VoH	VoS	VoP	VoC	ρ^2(c)	SD
6	4	20	20	10	-0.0833 -6.8	-0.0138 -4.7	-0.0367 -3.9	0.3401 4.3	-0.2584 -13.5	-0.1264 -7.5	6.03	2.66	24.64	18.72	9.16	0.1405	100
6	4	20	20	10	-0.0807 -6.5	-0.0125 -4.3	-0.0311 -3.4	0.2737 3.4	-0.2717 -14.2	-0.1178 -7.1	6.44	2.48	21.84	21.68	9.40	0.1446	100
											6.24	**2.57**	**23.24**	**20.20**	**9.28**		
8	5	20	20	10	-0.0946 -7.4	-0.0071 -2.4	-0.0375 -3.9	0.1807 2.3	-0.2874 -14.6	-0.1242 -7.2	13.35	5.29	25.50	40.56	17.53	0.1634	100
8	5	20	20	10	-0.1335 -10.3	-0.0144 -4.9	-0.0444 -4.5	0.2630 3.3	-0.2733 -13.8	-0.1260 -7.1	9.26	3.08	18.24	18.95	8.74	0.1728	100
											11.30	**4.18**	**21.87**	**29.76**	**13.13**		
10	8	20	20	10	-0.1367 -10.6	-0.0121 -4.0	-0.0677 -6.6	0.2660 3.3	-0.2489 -12.5	-0.1353 -7.3	11.33	5.62	22.06	20.64	11.22	0.1631	100
10	8	20	20	10	-0.1100 -10.6	-0.0095 -3.2	-0.0539 -5.4	0.2812 3.5	-0.2697 -13.8	-0.1175 -6.6	11.60	5.68	29.64	28.43	12.39	0.1544	100.0
											11.47	**5.65**	**25.85**	**24.53**	**11.80**		

Simulation tests of SP design (Band A) – Model Specification 2

Input Values			Estimated Coefficients								Estimated Values				
VoTA	VoH	VoTB	TimeA	(t)	Cost	(t)	Headway	(t)	TimeB	(t)	VoTA	VoH	VoTB	ρ^2 (c)	SD
7	5	9	-0.1827	-6.4	-0.0282	-5.6	-0.1363	-7.3	-0.2337	-8.9	6.49	4.84	8.30	0.1172	50
7	5	9	-0.1952	-6.9	-0.0306	-6.1	-0.1435	-7.8	-0.2377	-9.3	6.38	4.69	7.77	0.1165	50
											6.44	**4.77**	**8.04**		
5	3	7	-0.2191	-6.9	-0.0486	-8.4	-0.1572	-7.4	-0.3107	-10.3	4.51	3.24	6.39	0.1552	30
5	3	7	-0.1286	-4.1	-0.0319	-5.7	-0.1029	-5.2	-0.2332	-8.0	4.03	3.22	7.31	0.0927	30
											4.27	**3.23**	**6.85**		
12	7	15	-0.2526	-7.3	-0.0279	-4.9	-0.1493	-6.3	-0.3269	-9.8	9.07	5.36	11.74	0.1988	80
12	7	15	-0.2233	-6.5	-0.0141	-2.6	-0.0809	-4.1	-0.287	-8.8	15.79	5.72	20.30	0.1755	80
											12.43	**5.54**	**16.02**		
8	5	8	-0.1921	-7.3	-0.0222	-4.7	-0.1201	-7.5	-0.1978	-8.4	8.67	5.42	8.92	0.108	50
8	5	8	-0.1699	-6.6	-0.0207	-4.5	-0.1207	-7.8	-0.164	-7.3	8.21	5.83	7.93	0.0889	50
											8.44	**5.63**	**8.42**		

Input Values					Estimated Coefficients						Estimated Values						
VoTA	VoH	VoTB	VoP	VoC	TimeA	Cost	Headway	TimeB	Punctuality	Crowding	VoTA	VoH	VoTB	VoP	VoC	ρ^2(c)	SD
7	5	9	20	10	-0.0930 -6.4	-0.0152 -5.1	-0.0522 -5.0	-0.1268 -9.7	-0.2853 -14.3	-0.1674 -8.8	6.12	3.44	8.35	18.78	11.02	0.1831	100
7	5	9	20	10	-0.0946 -6.8	-0.0108 -3.9	-0.0314 -3.4	-0.1161 -9.5	-0.2337 -12.3	-0.1050 -6.2	8.78	2.92	10.78	21.70	9.75	0.1325	100
											7.45	**3.18**	**9.56**	**20.24**	**10.38**		
5	3	7	30	10	-0.0715 -4.9	-0.0165 -5.8	-0.0100 -1.1	-0.0941 -7.6	-0.2993 -15.2	-0.1232 -7.5	4.34	0.61	5.71	18.17	7.48	0.1571	120
5	3	7	30	10	-0.0823 -5.7	-0.0129 -4.6	-0.0147 -1.6	-0.0933 -7.6	-0.3096 -15.6	-0.1152 -7.0	6.38	1.14	7.24	24.00	8.93	0.1722	120
											5.36	**0.87**	**6.47**	**21.09**	**8.21**		
12	7	15	20	10	-0.1409 -9.7	-0.0108 -3.7	-0.0477 -4.6	-0.1697 -12.7	-0.2309 -11.5	-0.1069 -5.7	13.05	4.41	15.71	21.38	9.90	0.1648	120
12	7	15	20	10	-0.1350 -9.2	-0.0092 -3.1	-0.0507 -4.8	-0.1677 -12.4	-0.2067 -10.3	-0.1175 -6.0	14.67	5.51	18.22	22.46	12.77	0.1601	120
											13.86	**4.96**	**16.97**	**21.92**	**11.33**		

Simulation tests of SP design (Band C) – Model Specification 1

Input Values			Estimated Coefficients								Estimated Values				
VoT	VoH	VoS	Time	(t)	Cost	(t)	Headway	(t)	ASC	(t)	VoT	VoH	VoS	ρ^2(c)	SD
8	5	20	-0.188	-9.4	-0.026	-7.7	-0.118	-9.9	0.493	4.1	7.34	4.59	19.26	0.1640	60
8	5	20	-0.156	-8.3	-0.019	-6.1	-0.099	-9	0.455	3.9	8.26	5.25	24.03	0.1342	60
											7.80	**4.92**	**21.65**		
6	4	10	-0.2503	-7.6	-0.0426	-8.0	-0.1703	-9.9	0.7282	7.1	5.87	3.99	17.08	0.1093	30
6	4	10	-0.2913	-8.4	-0.0455	-8.2	-0.1918	-10.6	0.7112	6.8	6.40	4.21	15.62	0.1421	30
											6.14	**4.10**	**16.35**		

Note: The value of the improved rolling stock transferred to be positive

Input Values					Estimated Coefficients						Estimated Values						
VoT	VoH	VoS	VoP	VoC	Time	Cost	Headway	ASC	Punctuality	Crowding	VoT	VoH	VoS	VoP	VoC	ρ^2(c)	SD
6	4	20	20	10	-0.0413 -3.6	-0.0096 -6.3	-0.0348 -6.2	0.2423 3.2	-0.2641 -13.6	-0.0909 -11.4	4.31	3.63	25.33	27.61	9.50	0.1521	150
6	4	20	20	10	-0.0353 -3.1	-0.0080 -5.3	-0.0291 -5.3	0.2863 3.9	-0.2269 -12.0	-0.0914 -11.7	4.41	3.63	35.78	28.36	11.43	0.1339	150
											4.36	**3.63**	**30.56**	**27.98**	**10.46**		
8	5	20	20	10	-0.0861 -7.6	-0.0085 -5.7	-0.0344 -6.2	0.2036 2.8	-0.1482 -8.0	-0.0903 -11.3	10.15	4.05	24.00	17.47	10.64	0.118	150
8	5	20	20	10	-0.0881 -7.7	-0.0114 -7.6	-0.0500 -9.0	0.2450 3.3	-0.1251 -6.8	-0.0945 -11.5	7.71	4.38	21.45	10.95	8.28	0.1203	150
											8.93	**4.22**	**22.72**	**14.21**	**9.46**		
10	8	20	20	10	-0.1142 -9.2	-0.0118 -7.2	-0.0880 -13.4	0.1977 2.5	-0.2306 -11.4	-0.1144 -11.4	9.69	7.47	16.78	19.58	9.71	0.2053	120
10	8	20	20	10	-0.0958 -7.8	-0.0097 -6.0	-0.0842 -13.2	0.1888 2.4	-0.1803 -9.2	-0.1131 -11.6	9.91	8.71	19.53	18.65	11.70	0.1834	120
											9.80	**8.09**	**18.16**	**19.11**	**10.70**		

Simulation tests of SP design (Band C) – Model Specification 2

Input Values			Estimated Coefficients								Estimated Values				
VoTA	VoH	VoTB	TimeA	(t)	Cost	(t)	Headway	(t)	TimeB	(t)	VoTA	VoH	VoTB	ρ^2 (c)	SD
7	5	9	-0.150	-6.2	-0.023	-5.8	-0.098	-7.3	-0.195	-8.4	6.464	4.207	8.387	0.1330	60
7	5	9	-0.122	-4.8	-0.019	-4.4	-0.111	-8.1	-0.170	-7.1	6.493	5.910	9.034	0.1569	60
											6.479	**5.059**	**8.711**		
12	5	15	-0.142	-5.7	-0.010	-2.7	-0.061	-5	-0.174	-7.5	14.575	6.257	17.952	0.1252	100
12	5	15	-0.211	-7.8	-0.020	-5.5	-0.071	-5.6	-0.242	-9.4	10.614	3.582	12.183	0.1837	100
											12.595	**4.920**	**15.067**		
5	5	5	-0.139	-7.7	-0.029	-9.9	-0.137	-12.9	-0.141	-8.8	4.708	4.644	4.780	0.1883	50
5	5	5	-0.131	-7.4	-0.026	-9.1	-0.129	-12.6	-0.131	-8.3	5.077	4.988	5.050	0.1708	50
											4.893	**4.816**	**4.915**		
8	5	8	-0.168	-8.6	-0.023	-7.7	-0.116	-10.7	-0.171	-9.7	7.299	5.046	7.434	0.1631	60
8	5	8	-0.181	-9.2	-0.022	-7.6	-0.119	-10.9	-0.176	-10	8.078	5.306	7.881	0.1741	60
											7.688	**5.176**	**7.657**		

Input Values					Estimated Coefficients						Estimated Values						
VoTA	VoH	VoTB	VoP	VoC	TimeA	Cost	Headway	TimeB	Punctuality	Crowding	VoTA	VoH	VoTB	VoP	VoC	ρ^2(c)	SD
7	5	9	20	10	-0.104 -7.8	-0.014 -7.9	-0.056 -8.2	-0.128 -9.7	-0.257 -9.8	-0.132 -12.6	7.54	4.06	9.30	18.70	9.65	0.1946	100
7	5	9	20	10	-0.1012 -7.8	-0.0120 -7.1	-0.0572 -8.7	-0.1220 -9.6	-0.2307 -11.6	-0.1233 -11.8	8.43	4.76	10.17	19.23	10.28	0.1781	100
											7.99	**4.41**	**9.73**	**18.96**	**9.96**		
5	3	7	30	10	-0.0742 -5.7	-0.0114 -6.6	-0.0282 -4.4	-0.0960 -7.6	-0.3286 -15.8	-0.1138 -12.0	6.49	2.47	8.40	28.75	9.96	0.2100	120
5	3	7	30	10	-0.0536 -4.3	-0.0093 -5.7	-0.0201 -3.4	-0.0737 -6.2	-0.2908 -14.8	-0.0961 -11.2	5.77	2.16	7.93	31.31	10.35	0.1739	120
											6.13	**2.31**	**8.17**	**30.03**	**10.15**		
12	7	15	20	10	-0.1160 -8.3	-0.0107 -5.9	-0.0804 -10.6	-0.1493 -10.6	-0.2306 -10.9	-0.1008 -8.5	10.84	7.51	13.95	21.55	9.42	0.1735	120
12	7	15	20	10	-0.1300 -9.2	-0.0108 -6.1	-0.0688 -9.4	-0.1598 -11.3	-0.1684 -8.1	-0.1077 -9.2	12.01	6.36	14.77	15.56	9.95	0.1601	120
											11.43	**6.93**	**14.36**	**18.56**	**9.69**		

Simulation tests of SP design (Band D) – Model Specification 1

Input Values			Estimated Coefficients								Estimated Values				
VoT	VoH	VoS	Time	(t)	Cost	(t)	Headway	(t)	ASC	(t)	VoT	VoH	VoS	ρ^2(c)	SD
8	5	20	-0.1289	-8.1	-0.0160	-9.4	-0.0797	-9.9	0.4602	4.6	8.036	4.968	28.691	0.1102	80
8	5	20	-0.1248	-8	-0.0165	-9.7	-0.0897	-10.9	0.2611	2.7	7.545	5.423	15.786	0.1285	80
											7.791	**5.195**	**22.238**		
5	3	20	-0.0663	-4.7	-0.0159	-9.5	-0.0525	-7.5	0.3621	3.4	4.181	3.314	22.845	0.0975	80
5	3	20	-0.0878	-6.2	-0.0173	-10.1	-0.0424	-6.2	0.4344	3.8	5.088	2.454	25.168	0.1095	80
											4.635	**2.884**	**24.007**		

Note: The value of the improved rolling stock transferred to be positive

Input Values					Estimated Coefficients						Estimated Values						
VoT	VoH	VoS	VoP	VoC	Time	Cost	Headway	ASC	Punctuality	Crowding	VoT	VoH	VoS	VoP	VoC	ρ^2(c)	SD
6	4	20	20	10	-0.0534 -6.3	-0.0099 -9.8	-0.0278 -6.8	0.3068 3.7	-0.2467 -12.2	-0.0934 -14.7	5.41	2.81	31.06	24.97	9.45	0.199	150
6	4	20	20	10	-0.0482 -5.7	-0.0092 -9.4	-0.0355 -8.6	0.2010 2.5	-0.2664 -13.0	-0.0883 -14.0	5.23	3.84	21.78	28.87	9.56	0.202	150
											5.32	**3.33**	**26.42**	**26.92**	**9.51**		
8	5	20	20	10	-0.0671 -7.9	-0.0089 -9.2	-0.0421 -10.2	0.2247 2.9	-0.1803 -9.5	-0.0830 -12.9	7.51	4.72	25.16	20.19	9.29	0.17	150
8	5	20	20	10	-0.0675 -7.9	-0.0086 -8.9	-0.0462 -11.0	0.2098 2.7	-0.1669 -8.8	-0.0860 -13.0	7.81	5.34	24.27	19.31	9.95	0.173	150
											7.66	**5.03**	**24.71**	**19.75**	**9.62**		
10	8	20	20	10	-0.0782 -8.5	-0.0080 -7.9	-0.0732 -15.6	0.1863 2.3	-0.2020 -10.2	-0.0809 -10.9	9.75	9.13	23.23	25.18	10.09	0.223	150
10	8	20	20	10	-0.0841 -9.3	-0.0081 -8.1	-0.0650 -14.3	0.1883 2.3	-0.1881 -9.7	-0.0784 -10.9	10.35	8.00	23.18	23.15	9.65	0.2026	150
											10.05	**8.56**	**23.20**	**24.17**	**9.87**		

Simulation tests of SP design (Band D) – Model Specification 2

Input Values			Estimated Coefficients								Estimated Values				
VoTA	VoH	VoTB	TimeA	(t)	Cost	(t)	Headway	(t)	TimeB	(t)	VoTA	VoH	VoTB	ρ^2 (c)	SD
7	5	9	-0.1194	-5.7	-0.0166	-8.2	-0.0881	-9.2	-0.1505	-7.5	7.215	5.326	9.094	0.1233	80
7	5	9	-0.1076	-5.5	-0.0162	-8.2	-0.0752	-8.3	-0.1393	-7.5	6.663	4.653	8.625	0.1073	80
											6.939	**4.989**	**8.860**		
5	3	7	-0.0711	-4.3	-0.0160	-8.5	-0.0486	-6.3	-0.1032	-6.7	4.435	3.029	6.438	0.0888	80
5	3	7	-0.0930	-5.4	-0.0177	-9.2	-0.0623	-7.7	-0.1246	-7.6	5.244	3.514	7.024	0.1100	80
											4.839	**3.271**	**6.731**		
12	7	15	-0.1675	-4.9	-0.0146	-5.4	-0.1030	-7.2	-0.2098	-6.3	11.457	7.045	14.350	0.1426	100
12	7	15	-0.1074	-4.3	-0.0088	-4.1	-0.0643	-6	-0.1420	-5.9	12.221	7.312	16.158	0.0839	100
											11.839	**7.179**	**15.254**		
8	5	8	-0.1422	-8.7	-0.0170	-10.2	-0.0893	-11.1	-0.1405	-9.3	8.345	5.238	8.245	0.1337	80
8	5	8	-0.1063	-6.8	-0.0154	-9.4	-0.0809	-10.4	-0.1080	-7.4	6.912	5.262	7.022	0.1112	80
											7.628	**5.250**	**7.634**		

Input Values					Estimated Coefficients						Estimated Values						
VoTA	VoH	VoTB	VoP	VoC	TimeA	Cost	Headway	TimeB	Punctuality	Crowding	VoTA	VoH	VoTB	VoP	VoC	ρ^2(c)	SD
7	5	9	20	10	-0.056 -6.2	-0.0078 -7.9	-0.0382 -8.9	-0.0727 -8.4	-0.1824 -9.7	-0.0734 -10.9	7.21	4.92	9.36	23.47	9.45	0.1460	150
7	5	9	20	10	-0.0539 -5.8	-0.0062 -6.3	-0.0377 -8.6	-0.0680 -7.7	-0.1657 -8.8	-0.0823 -11.7	8.65	6.04	10.91	26.57	13.20	0.1476	150
											7.93	**5.48**	**10.13**	**25.02**	**11.32**		
5	3	7	30	10	-0.0503 -5.5	-0.0090 -8.8	-0.0257 -6.0	-0.0684 -7.9	-0.2278 -11.8	-0.0895 -13.2	5.56	2.84	7.57	25.22	9.91	0.1839	150
5	3	7	30	10	-0.0364 -4.0	-0.0056 -5.7	-0.0192 -4.5	-0.0491 -5.8	-0.2765 -14.1	-0.0791 -12.5	6.45	3.40	8.71	49.03	14.03	0.1867	150
											6.01	**3.12**	**8.14**	**37.13**	**11.97**		
12	7	15	20	10	-0.1036 -8.6	-0.0109 -8.9	-0.0614 -9.7	-0.1350 -10.9	-0.1704 -8.3	-0.1044 -9.8	9.50	5.63	12.37	15.62	9.57	0.1744	150
12	7	15	20	10	-0.1065 -8.7	-0.0087 -7.1	-0.0657 -10.2	-0.1328 -10.6	-0.1745 -8.5	-0.1012 -9.4	12.18	7.51	15.18	19.95	11.57	0.1848	150
											10.84	**6.57**	**13.78**	**17.79**	**10.57**		

Appendix C-2: An Example of the Simulation Tests

An Example of a Program File to Create the Responses

(Source: Fowkes, internal use)

```
   DOUBLE PRECISION E1,E2,RANDOM
   INTEGER COSTA(30),COSTB(30),TIMEA(30),TIMEB(40),HEADA(40), &
    HEADB(40),PUNCA(30),PUNCB(30),CROWDA(30),CROWDB(30)
   OPEN (UNIT=11,FILE='DESIGN.txt')
   OPEN (UNIT=12,FILE='DESIGN.DAT')
   CALL DATE_TIME_SEED@
   ASC=0.0
   VOT=-8.0
   VOH=-5.0
   VOP=-10.0
   VOC=-10.0
   VOC=-1.0
   nnn=0
   STDEV=400.0
   DENOM=1.28/STDEV
! SP CALCULATIONS
   sum1=0
   sum2=0
   tot=0
   DO 10 I=1,18
   READ (11,*) timeA(I),timeB(I),costA(I),costB(I),HEADA(I),HEADB(I), &
    PUNCA(I),PUNCB(I),CROWDA(I),CROWDB(I)
 10 CONTINUE
!  NEXT NUMBER IS NUMBER OF PEOPLE IN SIMULATION
  DO 15 I=1,100
   DO 20 J=1,18
   E1=RANDOM()
   E1=-DLOG(E1)
   E1=-DLOG(E1)
   E1=E1/DENOM
   E2=RANDOM()
   E2=-DLOG(E2)
   E2=-DLOG(E2)
   E2=E2/DENOM
   UA=(VOC*COSTA(J))+(VOT*TIMEA(J))+(VOH*HEADA(J))    &
   +  (VOP*PUNCA(J))+(VOC*CROWDA(J))+E1+ASC
   UB=(VOC*COSTB(J))+(VOT*TIMEB(J))+(VOH*HEADB(J))   &
   +   (VOP*PUNCB(J))+(VOC*CROWDB(J))+E2

   IF(UA.GE.UB) THEN
   IC=1
   ELSE
   IC=2
   ENDIF
   WRITE (12,77) timeA(J),timeB(J),costA(J),costB(J),HEADA(J),HEADB(J),  &
    PUNCA(J),PUNCB(J),CROWDA(J),CROWDB(J),IC
 77 FORMAT (10I6,I3)
 20 CONTINUE
 15 CONTINUE
   STOP
   END
```

Appendix C-3: An Example of Created Responses

15	20	200	180	15	20	2	4	3	0	1
20	25	250	200	15	20	0	2	4	0	2
20	20	300	200	15	30	2	4	8	4	2
15	25	300	200	20	10	0	2	0	10	2
25	20	200	250	20	10	4	0	5	0	1
25	25	200	180	15	30	2	4	0	10	1
15	30	250	200	15	30	4	0	6	6	1
20	30	200	180	20	10	4	0	0	12	1
25	30	300	200	15	20	0	2	10	6	1
15	25	200	180	15	20	4	0	6	5	1
25	20	220	200	15	20	4	0	0	8	2
20	25	300	200	15	30	4	0	4	0	2
15	30	300	200	20	10	2	4	3	0	2
20	20	200	180	20	10	0	2	8	4	1
25	25	200	250	20	10	2	4	10	5	1
20	30	250	200	15	20	2	4	0	12	1
15	20	250	200	15	30	0	2	0	8	2
25	30	300	180	15	30	0	2	5	0	1
15	20	200	180	15	20	2	4	3	0	2
20	25	250	200	15	20	0	2	4	0	1
20	20	300	200	15	30	2	4	8	4	2
15	25	300	200	20	10	0	2	0	10	1
25	20	200	250	20	10	4	0	5	0	2
25	25	200	180	15	30	2	4	0	10	2
15	30	250	200	15	30	4	0	6	6	2
20	30	200	180	20	10	4	0	0	12	2
25	30	300	200	15	20	0	2	10	6	2
15	25	200	180	15	20	4	0	6	5	1
25	20	220	200	15	20	4	0	0	8	2
20	25	300	200	15	30	4	0	4	0	1
15	30	300	200	20	10	2	4	3	0	2
20	20	200	180	20	10	0	2	8	4	2
25	25	200	250	20	10	2	4	10	5	2

Appendix D An Example of CT script

Cummings and Taylor, 1999, (p.651):

...in a recent study, several different groups of people voted on a referendum just like the one you are about to vote on. Payment was hypothetical for these groups, as it will be for you. No one had to pay money if the referendum passed. The results of these studies were that on average, across the groups, 38 percent of them voted "yes." With another set of groups with similar people voting on the *same* referendum as you will. Vote on here, *but* where payment was real and people *really* did have to pay money if the referendum passed, the results on average across the groups were that 25 percent voted yes. That's quite a difference, isn't it?

We call this a "hypothetical bias." Hypothetical bias is the difference that we continually see in the way people respond to hypothetical referenda as compared to real referenda...

Now can we get people to think about their vote in a hypothetical referendum like they think in a real referendum, where if enough people vote "yes," they'll really have to pay money? How do we get them to think about what it means to really dig into their pocket and pay money, if in fact they really aren't going to have to do it?

Let me tell you why I think that we continually see this hypothetical bias, why people behave differently in a hypothetical referendum than they do when the referendum is real. I think that when we hear about a referendum that involves doing something that is basically good - helping people in need, improving environmental quality, or anything else - our basic reaction in a hypothetical situation is to think: *sure*, I would do this. *I really would vote "yes"* to spend the money...

But when the referendum is real, and we would actually have to spend our money if it passes, we think a different way. We basically still would like to see good things happen, but when we are faced with the possibility of having to spend money, *we think about our options*: if I spend money on this, that's money I don't have to spend on other things...we vote in a way that takes into account the limited amount of money we have...This is just my opinion, of course, but it's what I think may be going on in hypothetical referenda.

So if I were in your shoes...I would ask myself: if this was a real referendum, and I had to pay \$10.00 if the referendum passed: do I *really want* to spend my money this way? If I really did, I would vote yes; if I didn't, I would vote no...

In any case, I ask you to vote just exactly as you would vote if you were really going to face the consequences of your vote: which is to pay money if the proposition passes. Please keep this in mind in our referendum.

Appendix E Mixed Logit Model Analysis

E.1 Introduction

This report presents the data analysis and research hypotheses testing using Mixed Logit (MXL) models. Chapter 7 presented a base standard logit model for analysing users' valuation of the improved rolling stock. The base model controlled several factors (i.e. income and journey purpose) which might cause the variation of valuations, to avoid the potential confounding effects. Chapter 8 explored the effects of design factors (cheap-talk and complex design) on the SP responses from the base model. Standard MNL model and HMNL model were applied in the data analysis and research hypotheses testing. This report starts from the point reached at the end of the chapter 7 and adds to the analysis and discussion of chapter 8, using other techniques.

The second section presents a brief introduction of the MXL. The third section presents the model estimation. The fourth section discusses the valuations compared with those obtained from the other model (MNL and HMNL) estimations in the thesis. The fifth section discusses the research hypotheses testing based on the MXL model estimation. The last section summarises the findings from MXL model analysis.

As before, the main research objective is to examine the existence and consequence of the incentive to strategic bias and discuss the impacts of two methods to reduce the bias in the SP experiment. The possible task complexity effect is examined. In this report, we only apply the MXL model to analyse the impacts of adding the Cheap Talk (CT) script and adding more attributes to mask the research aim (Complex Design – CD) on SP responses. The impact of individuals' perceptions on SP responses will be examined using this technique in the future.

E.2 Literature review of Mixed Logit Model

E.2.1 Introduction

Train (2003, p.138) states that "Mixed Logit is a highly flexible model that can approximate any random utility model (MaFadden and Train, 2000). It obviates the three limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time."

In our data analysis, travellers' taste variation in valuation of the improved rolling stock (as well as the other attributes) is accounted by segmentation analysis (see section 4.3.6) according to their income and journey purpose factors.

One of the strengths of MXL models is that the model estimation allows for variation in the variables (and thus in the monetary valuation) that results from those travellers' characteristics whose explicit information is not known. This feature makes the MXL model estimation a better technique to obtain a more accurate valuation of variables.

Rich literature documented the development of MXL model estimation (Brown, Bunch and Train, 2001; Batley et al., 2001; Hensher and Greene, 2003; Train, 2003 and many others). The following presents a brief description of the general form.

E.2.2 Model description

Following Train (2003, p.148), the utility derived from alternative i for individual n is specified:

$$U_{ni} = \beta'_n x_{ni} + \varepsilon_{ni}$$

Where coefficients β'_n are distributed with density $f(\beta|\theta)$, where θ refers collectively to the parameter of this distribution (such as mean and covariance of β). ε_{ni} is a random term with zero mean that is IID across alternatives.

The researcher specifies the function form $f(\beta|\theta)$ and wants to estimate the parameters of θ.

The choice probabilities are:

$$P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta$$

where, $L_{ni} = \dfrac{\exp(V_{ni}(\beta))}{\sum\limits_{j \in J} \exp(V_{nj}(\beta))}$

In the standard logit model, the mixed distribution $f(\beta)$ is generated at fixed parameters b: $f(\beta) = 1$ for $\beta = b$ and 0 for $\beta \neq b$. For example, the density of β can be normal with mean b and covariance W. Then the choice probability under this density becomes (Train, 2003, p.141):

$$P_{ni} = \int \left(\frac{\exp(\beta' x_{ni})}{\sum_j \exp(\beta' x_{nj})}\right) \phi(\beta|b, W) d\beta$$

Where $\phi(\beta|b, W)$ is the normal density with mean b and covariance W. The distribution can also be lognormal, uniform and triangular.

In the estimation of MXL models, researchers need to select the random variables and specify the distribution for the random variable as the input to estimate the mean and covariance of the variable. This is one of the difficulties of this technique. For example, a large standard deviation would lead the parameter to be positive (especially when the mean of parameter estimate is very close to zero), which leads to a negative WTP. Recently, several researchers have tackled the problems relating to the selection of the random variance and choice of distribution (although this not always the main focus of their discussion), for instance, the possibility of positive time-related values. Hess et al. (2005), based on the synthetic data set, found the positive parameter values of some of the population, which lead to a negative WTP. They addressed that the negative value of in-vehicle time should be treated with suspicion as it is contradictory to the economic theory of time valuation. They recommended using bounded distribution where the bounds are estimated from the data. Similarly, Batley et al. (2001) found a positive tail in the distribution of a mean lateness parameter, and decided as a result not to allow any distribution of this parameter. This characteristic of MXL model leads to an open discussion. We will examine this issue in the later model estimation.

E.3 Model Specification

E.3.1 Introduction

One property of MXL model is that it allows for the random variation among individuals in the value of parameters that represent the unobserved factors/characteristics; therefore, the model estimate is more accurate. This section presents the attempts to specify and estimate some Mixed Multinomial Logit models, using the same model specification as the standard MNL model (M8-2), but not account for the scale factor effects as this requires further tests and consideration of the program. The impacts of individuals' socio-economic information (i.e. income and journey purpose) are taken into account. The impacts of design factors (adding a cheap-talk script and adding more attributes to mask the research aim, namely complex design) are explored in the model estimation.

E.3.2 Selection of random variables

We assume the cost coefficient to be fixed in the MXL model estimation, since this variable will be needed to derive the monetary values of other variables. This is a common practice in the MXL estimation. If the cost variable is selected as the random variable, it might lead to an unbelievably large monetary value of the other variable when the value of cost coefficient gets very small.

The other variables (ASC, time, headway, punctuality and crowding) in the choice are selected as random variables initially. Impacts of design factors on the estimation of other coefficients (incremental effects) are also selected as random variables. This is to test if the impacts of the design factors are heterogeneous among different individuals in the values of parameters.

E.3.3 Selection of the distribution

As stated in the second section, we need to specify the distribution for the selected random variables. The model established here used some of the most common distribution for the random variables, namely normal and triangular distributions. We did not include the lognormal distribution in our analysis, as the literature reported of the difficulty with lognormal parameters (Small, 2005). We will leave the examination of the statistical performance of this distribution as future work.

E.3.4 Software for MXL estimation and other issues in the estimation

To estimate the MXL model, a code was written in GAUSS (Aptech Systems, 1997) program based on the code by Kenneth Train (see Train et al., 1999). Test models were estimated using GAUSS and BIOGEME. The same specification led to the consistent results from different software packages. Therefore, in this study, all the MMNL models were estimated by GAUSS package.

Train (2003) presents a discussion on the impact of number of Halton draws on the estimation. Researchers (Train, 1999) found that the simulation variance in the estimated parameters was lower using lower number (for instance, 100) of draws than larger number of draws (for example, 1000), however, the estimation is faster in the lower number draws. In this study, we examined the estimation from 500 and 1000 Halton draws. It normally takes about an hour to reach the convergence for 500 Halton draws and much longer for 1000 Halton draws.

E.3.5 Findings from MXL model estimation

After several attempts, some not significant variables are removed or combined with the variables which have similar values. Table E. 1 presents the estimation result from the preferred MXL models.

MXL1 is a MMNL model with normally-distributed parameters. The number of draws is 500. MXL2 is same as MXL1, but the number of Halton draws is 1000. MXL2 has a slightly better log-likelihood compared with MXL1. The variances of the estimated parameter from these two models are not significantly different from each other, except for that of the standard deviation of variable (CT*crowding). MXL3 is a MMNL model with triangularly-distributed parameters. The number of draws is 500. The log-likely hood of MXL3 is not as good as MXL1 and MXL2. Therefore, the normally-distributed model (MXL2) is selected as the preferred model for the subsequent analysis and discussion.

Table E.2 MXL model estimation results

Estimation Coefficients (t-ratio)	MXL1 (normal distribution)		MXL2 (normal distribution)		MXL3 (t-distribution)	
Time						
Time (Commuters)	-0.1326	(-9.86)	-0.1351	(-9.91)	-0.1326	(-9.82)
Time. standard deviation/spread	0.1530	(13.08)	0.1555	(13.24)	0.3719	(13.49)
+ Leisure	0.1050	(4.34)	0.1035	(4.33)	0.1057	(4.32)
+ EB/PB/School	-0.0408	(-2.24)	-0.0368	(-1.98)	-0.0396	(-2.16)
Cost						
Cost (Base)	-0.0275	(-16.96)	-0.0279	(-17.04)	-0.0274	(-17.03)
+ Cost - Inc3 (£21-35k)	0.0039	(2.35)	0.0039	(2.35)	0.0038	(2.32)
+ Cost - Inc4 (£36-50k)	0.0088	(3.04)	0.0094	(3.19)	0.0089	(3.07)
+ Cost - Inc5 (over 50k)	0.0112	(3.16)	0.0115	(3.29)	0.0111	(3.16)
Headway						
Headway (Commuters/EB/PB/School)	-0.1215	(-19.54)	-0.1219	(-19.41)	-0.1222	(-19.59)
Headway. standard deviation/spread	0.1123	(19.05)	0.1132	(19.20)	0.2708	(19.74)
+ Leisure	0.0462	(3.09)	0.0480	(3.16)	0.0453	(3.03)
Punctuality						
Punctuality (Commuters)	-0.7649	(-18.22)	-0.7872	(-18.18)	-0.7755	(-18.25)
Punctuality. Standard deviation/spread	0.5346	(14.59)	0.5404	(14.57)	1.2918	(15.37)
+ Leisure	0.3282	(3.55)	0.3448	(3.68)	0.3225	(3.46)
+ EB/PB/School	0.1650	(2.26)	0.1938	(2.53)	0.1774	(2.46)
Crowding						
Crowding (Commuters/EB/PB/School)	-0.2540	(-13.96)	-0.2599	(-13.82)	-0.2540	(-13.95)
Crowding. standard deviation/spread	0.1664	(10.54)	0.1769	(11.94)	0.4086	(10.95)
+ Leisure	-0.0614	(-1.77)	-0.0599	(-1.75)	-0.0578	(-1.71)
ASC Segmentation						
Commuters/PB	0.7298	(10.05)	0.7349	(10.05)	0.7502	(10.33)
ASC. standard deviation/spread	1.5163	(21.22)	1.5363	(21.33)	3.6148	(21.69)
+ Leisure/EB/School	-0.3800	(-2.85)	-0.3650	(-2.72)	-0.3684	(-2.78)
On the Other Attributes						
+ CT*Time	-0.0390	(-2.26)	-0.0382	(-2.20)	-0.0380	(-2.20)
CT*Time. standard deviation/spread	0.1035	(4.14)	0.0931	(3.45)	0.2527	(4.07)
+ CT*Cost	-0.0064	(-3.54)	-0.0062	(-3.42)	-0.0064	(-3.56)
+ CT*Crowding	0.0553	(2.43)	0.0573	(2.49)	0.0523	(2.31)
CT*Crowding. standard deviation/spread	0.0715	(2.50)	-0.0191	(-0.34)	0.1611	(1.80)
ρ^2 **(C)**	0.2296		0.2303		0.2294	
LL (C)	-5733.7		-5728.12		-5735.1	
No. of draws	500		1000		500	

The follow findings can be obtained from the model estimation (MXL2):

The change in the maximum likelihood, compared to the standard MNL model (M8-2), proves that the MXL model is better than MNL model in explaining individuals' behaviour.

The sign and trend of the variable estimation is the same as obtained from M8-2. The valuations obtained from MXL2 are compared with those obtained from the previous models, and will be presented in the fourth section.

The random variables and the standard deviation of the variables are significant, except for variable (CT*crowding) which measures the incremental effect of adding the Cheap Talk script on the crowding attribute in the choice set. The t-ratio for this variable is getting small after taking more Halton draws, which indicates that the randomness of this variable is not significant. The significant standard deviation of the random variables indicates respondents have a wider tastes, for instances, the time attribute.

Table E.2 gives the percentage of the population attach to a positive parameter estimates from the MXL2 estimation.

Table E. 2 Percentage of the population attach to a positive parameter

Variable	Mean	Standard deviation	Percentage
Time	-0.1351	0.1555	19%
Headway	-0.1219	0.1132	14%
Punctuality	-0.7872	0.5404	7%
Crowding	-0.2599	0.1769	7%
ASC	0.7349	1.5363	32%
CT*Time	-0.0382	0.0931	34%

Note: for the estimation of ASC (preference of the improved rolling stock), the value is the negative of the parameter attached to the percentage of the population.

We find that 19% of the population in the sample has a positive valuation of the time parameter. A person with a positive time parameter is one that prefers a longer in-vehicle time.

This can be explained by the fact that some of the respondents do not value time savings or would rather extend the journey (Cirillo and Axhausen, 2004). We decided to accept the model. In a model by Bhat and Sardesai (2005), positive travel time parameters are attached to 27% of the population, and the authors do not see this as a reason to reject the model. Further testing on the log-normal distribution where a negative time coefficient can be obtained is needed in the future research.

14% of the respondents in the sample have a negative valuation of the headway. The small percentage is acceptable, as most of the respondents travel frequently and they might have a good knowledge of the timetable. For those who travel in the off-peak time, they might do not mind to extend the headway of the services as they would not wait long in the rail station if they know the timetable.

The percentage of the population who have a positive value of punctuality and crowding is very low (7%). This can be accepted. For example, many working places operates flexible working times, it is likely that some travellers might wish to delay. The off-peak travellers would be less likely to have the crowding problem during their journey; therefore, they might not value the crowding in their choices.

The percentage of the respondents who have a negative preference of the improved rolling stock is 32%, which accounts for a big proportion of population. Various reasons can explain why some respondents prefer the old rolling stock (Pacers). One reason suspected is that some of the travellers might prefer Pacers as it provides more capacity as described in chapter 5.

E.4 Comparisons of valuations obtained from different models

For comparison reason, we did not calculate the distribution of the monetary values, but derive the values as the ratio of parameters of target variable and cost coefficient. As stated by Hensher and Greene (2003), the monetary values of the variables are another big challenge of the MXL model estimation as it is obtained from the ratios of the random parameters. In our MXL model estimation, we assumed the cost coefficient as the fixed variable, while the other variables in the choice set are assumed to be normal distributed. That makes the problem slightly easier.

To test the research hypotheses, monetary values from our study need to be compared with the official values (most importantly, valuation of improved rolling stock) from past studies where average values with consideration of different journey purpose and income effects are commonly available (for instance, values in PDFH). Therefore, the average value is our interest. The value of variable is obtained by the ratio of the mean variable coefficient and cost coefficient. Tables E. 3/4 report the comparison of values obtained from the model estimation in the thesis.

In the thesis, we presented lots of discussion of comparisons between the values obtained from our studies and the recommended values (see sections 7.5 and 8.4.5 for more discussion). Therefore, we will only briefly discuss the comparison with the previous evidence, but focus on the comparisons between the values obtained from different models.

Table E.3 Comparison of VoSs obtained from different model estimation

VoS (Commuters)	SP Exp. without CT script			SP Exp. with CT script		
	MNL (M8-2)	HMNL(M8-3)	MXL2	MNL (M8-2)	HMNL (M8-3)	MXL2
< £21k	25.25 (3.85) (17.69 – 32.81)	23.02 (2.19) (18.73 - 27.31)	26.34(2.80) (20.85 – 31.83)	20.68 (3.15) (14.51 -26.86)	19.68 (1.77) (16.21 -23.15)	21.55(2.18) (17.28-25.82)
£21-35k	31.22 (4.83) (21.75 – 40.69)	28.69 (3.10) (22.61-34.77)	30.62(3.58) (23.60-37.64)	25.53 (3.85) (17.97-33.08)	23.68 (2.32) (19.13-28.23)	24.33(2.61) (19.21-29.45)
£36-50k	43.58 (7.38) (29.12-58.04)	33.76 (4.86) (24.23 -43.29)	39.72(7.52) (24.98-54.46)	32.83 (5.27) (22.50-43.16)	27.02 (3.28) (20.59-33.45)	29.75(4.47) (20.99-38.51)
Over £50k	59.84 (15.28) (29.90-89.78)	44.40 (8.99) (26.78-62.02)	44.81(10.47) (24.29-65.33)	44.79 (9.61) (25.96-63.63)	33.45 (5.29) (20.59-43.82)	32.52(5.81) (21.13-43.91)

Table E. 4 The impact of CT on the valuation estimation

Commuters	SP Exp. without CT script									SP Exp. with CT script									Sig.
	MNL (M8-2)			HMNL(M8-3)			MXL2			MNL (M8-2)			HMNL (M8-3)			MXL2			
	value	s.e.	t	value	s.e.	t	value	s.e.	t	value	s.e.	t	value	s.e.	t	value	s.e.	t	
Monetary Values																			
VoS (p/trip)	30.95	8.62	3.59	27.21	3.57	7.62	30.11	4.47	6.73	24.96	5.94	4.20	22.56	2.48	9.10	23.94	2.95	8.12	
VoT (p/min)	5.34	1.22	4.38	5.32	0.54	9.85	5.53	0.66	8.41	5.68	0.91	6.24	5.56	0.30	18.53	5.65	0.43	13.17	
VoH (p/min)	4.68	1.10	4.25	4.75	0.48	9.90	5.00	0.70	7.18	3.86	0.65	5.94	3.93	0.30	13.10	3.97	0.35	11.23	
VoP (p/min)	34.69	8.23	4.22	34.44	4.01	8.59	32.26	4.05	7.96	28.59	5.00	5.72	28.55	2.67	10.69	25.65	2.42	10.62	
VoC (p/min)	15.91	2.58	4.10	16.09	1.34	8.04	16.19	1.41	7.53	12.05	1.21	5.26	12.14	0.84	7.83	12.25	0.77	8.62	**
Values relative to the In-vehicle time																			
VoS	5.80	0.91	6.37	5.11	0.72	7.10	5.44	0.79	6.89	4.39	0.80	5.49	4.06	0.51	7.96	4.24	0.56	7.57	
VoT	1.00			1.00			1.00			1.00			1.00			1.00			
VoH	0.88	0.07	12.57	0.89	0.08	11.13	0.90	0.09	10.03	0.68	0.05	13.60	0.71	0.05	14.20	0.70	0.06	11.72	
VoP	6.50	0.72	9.03	6.47	0.65	9.95	5.83	0.62	9.40	5.03	0.51	9.86	5.13	0.44	11.66	4.54	0.41	11.08	
VoC	2.98	0.26	7.62	3.02	0.23	8.78	2.92	0.22	8.75	2.12	0.18	6.22	2.18	0.13	9.08	2.17	0.13	8.99	**

E.4.1 Comparisons of VoSs obtained from different model estimation

Table 3 presents the comparison of monetary values of improved rolling stock (VoSs) with the standard error in the bracket on the right side (at the 5% level) and the t-ratio.

For the SP experiment both with and without the Cheap Talk script, the values obtained from the MXL model are generally consistent with those from MNL and HMNL models in most of the income bands. In the highest income band (over £50k), the VoSs obtained from the MXL model are lower than those from the MNL model (the impact is not significant at the 5% level); however, they are more consistent with those obtained from HMNL model.

It is found that the standard error obtained from the HMNL model is generally smaller than that from the MXL and MNL models, which indicates the values from HMNL model are more precise. The reason suspected is that HMNL model controls the heterogeneity of individuals through a parameterization of the scale factor. Another possible reason is that the estimation of HMNL model did not solve the repeated measurement problems as stated in section 8.3.4. Previous evidence has found that the repeated measurement problem lead to a smaller standard error in the parameter estimates which might partly explains the smaller standard error of values obtained from the HMNL model. A possible future work is to extend our MXL model estimation to incorporate the scale factor effects, for instance, a mixed HMNL model.

The standard error of the values obtained from the MXL model is much smaller than MNL model, which indicates a more precise estimate.

Again, the values obtained from the MXL model supports our finding that the VoS_{CT} (VoS derived from the SP response with CT script) is lower than VoS_{NoCT} (from the experiment without CT script) in all the income bands. However, from the t-statistic test, the difference between the values is not significant (t=1.63) for the average value.

As the incremental effect of complex design (adding two more attributes – CD) on the estimation of other attributes are all not significant in our MXL model estimation, we did not detect a significant impact of CD on the magnitude of the monetary values. This supports our finding in section 8.5.3.

E.4.2 Comparisons of impacts of the CT on the values of other variables

Table E. 4 presents the impact of CT script on the monetary values of variables and the values relative to the in-vehicle time in our SP experiment.

The monetary values obtained from the MXL model are generally consistent, except for some variables.

For the monetary values, the VoT_{CT} obtained from MXL is slightly higher the VoT_{CT} obtained from MNL and HMNL models (the impact is not significant), while the $VoT_{No.CT}$ obtained from MXL model is consistent with those from other models.

The VoP_{CT} obtained from MXL model is slightly lower than that obtained from the other two models (not significant), which lead to a lower time units value (value relative to the IVT). The value obtained from MXL model is closer to the PDFH (2005) recommended value of punctuality relative to the time units (2.5 to 3.0). However, it is suspected that the VoP is biased upward (see discussion in section 8.4.5).

In addition, we detected nearly significant difference in VoHs (t=1.85) and VoPs (t = 1.93) obtained from the MXL model between SP experiments with and without the CT script. The possible reason is that respondents become more sensitive to the cost attribute in the SP experiment with a CT script. Again, the values obtained from the experiment with CT are more consistent with the PDFH (2005) values (see the discussion in section 8.4.5).

E.5 Research Hypotheses Testing

This section reports the research hypothesis testing using the estimation from MXL models. Recall the research hypothesis in chapter 1, the findings from MXL model can be summarised as follows:

- Adding the cheap talk script shows a significant impact on the cost coefficient (refer to Table 1), the t-ratio is (-3.65). This indicates that with the CT script in the SP experiment, respondents are more sensitive to the cost in the hypothetical choices.

- VoS_{CT} is lower than VoS_{NoCT} in all the income bands. The impact is not statistically significant (t=1.63) at the normal 5% level for the average value of rolling stock. We cannot reject the null hypothesis at the normal 5% significant level that adding the CT decreases the estimation of VoS;

- The incremental effect of complex design (adding more attributes) on the estimation of other coefficients are all not significant, thus being removed from the preferred model. Therefore, adding more attributes does not change the magnitude of the values in our experiment.

- The scale factor effect is not incorporated into our MXL model estimation; therefore, the impact of complex design on the variance of the error term is not being explored. We cannot draw any conclusion on that.

E.6 Summary and Future Work

This report presents the data analysis and research hypotheses testing using Mixed Logit (MXL) model. The much smaller likelihood obtained from the MXL model estimation compared with the MNL and HMNL models indicates that MXL model estimation is a more accurate estimation.

The above sections presents the comparisons of model estimation and valuations obtained from the MXL model with other models (MNL and HMNL) in the thesis. The comparisons indicate that results from MXL model estimation are consistent with the findings from other models, which supports our conclusions on the research hypotheses testing.

The MXL model estimation provides us with significant insight in the future work. As discussed in this report, we only tested some simple and common forms of the distribution (normal and triangular) on the random parameters. The model estimation found that 19% of the respondents have negative values of journey time. Further testing of lognormal (and other) distributed random parameter is needed. However, this distribution might increase the difficulty of obtaining the average values of other variables as the parameter with a lognormal distribution has a very long tail (Hensher and Greene, 2003).

Another suggestion is that bounded distribution and constrained distribution can be applied to avoid large portion of negative VoT. Hensher and Greene (2003) suggested to impose constraints on a distribution by making the standard deviation /spread of each random parameter a function of the mean. For instance, the standard deviation can be constrained by a factor to make it smaller than the mean of the parameter, as a standard deviation greater than the mean estimate "typically result in behaviourally unacceptable parameter estimates" (p.147). It is important to investigate whether or not these tests would be able to lead to more reliable and accurate estimate.

As mentioned in section E.4.1, another possible future work is to extend the MXL model estimation to incorporate the scale factor effects, for instance, a mixed HMNL (MHMNL) model. The advantage of HMNL model is that it can parameterize the scale factor by a function of factors such as individuals' socio-economic features and the SP design features. By a MHMNL model, attribute parameters are randomly distributed in the population and the Gumbel scale parameters are the functions of the design factors, respondents' characteristics and perceptions.

References

Batley, R., A.S. Fowkes and G. Whelan (2001), Models for Choice of Departure Time, paper presented at the European Transport Conference, Homerton College, Cambridge.

Bhat, C.R. and R. Sardesai (2005), On examining the impact of stop-making and travel time reliability on commute mode choice: an application to predict commuter rail transit mode for Austin, TX, Proceedings of the 84th TRB annual meeting, Washington, D.C.

Hensher, D.A. and W.H. Greene (2005), The mixed logit model: the state of practice, Transportation, No. 30 p. 133-176

Hess, S. M. Bierlaire and J.W. Polak (2005), Estimating of value of travel-time savings using mixed logit models, Transportation Research A, Vol. 39, p221-236

Small, K.A., C. Winston and J. Yan (2005), "Uncovering the distribution of motorists' preference for travel time and reliability", Econometrica, Vol. 73, No4, p1367-1382

Train, K., D. Revelt and P. Ruud (1999), Mixed Logit Estimation Routine for Panel Data, *<http://elsa.berkeley.edu/Software/abstracts/train0296.html>*,

Train, K. (2003), Discrete Choice Methods with Simulation, Cambridge University Press, U.K.

Appendix F

Lists of Papers Presented in Conference

H.Lu (2005), "Incentives for Respondents to bias their answer in the Stated Preference Application", 37th UTSG Annual Conference, Bristol, UK

H. Lu, A. Fowkes and M. Wardman (2006), "The influence of SP design on Incentive to Bias in Response," presented in European Transport Conference 2006, Strasbourg, France

H. Lu (2007), "The impact of perception on the incentive to bias in SP responses" 39th UTSG Annual Conference, Harrogate, UK, won the second best prize (NYCC prize) for best paper and presentation

H. Lu, A. Fowkes and M. Wardman (2008), "Amending the incentive for strategic bias in Stated Preference studies: A case study in the users' valuation of rolling stock." presented at the 87th Annual Meeting of the Transport Research Board, Washington DC, January 2008. Recommended for publication in the TRB journal.