EMBEDDING DECISION HEURISTICS IN DISCRETE CHOICE MODELS:

ASSESSING THE MERITS OF MAJORITY OF CONFIRMING DIMENSIONS, EXTREMENESS AVERSION, AND REFERENCE REVISION

WAIYAN LEONG

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The Institute of Transport and Logistics Studies, The University of Sydney Business School, The University of Sydney

February 2014

ABSTRACT

Contrary to the usual assumption of fixed, well-defined preferences, it is increasingly evident that individuals are likely to approach a choice task using rules and decision heuristics that are dependent on the choice environment. More specifically, heuristics that are defined by the local choice context, such as the gains or losses of an attribute value relative to the other attributes, seem to be consistently employed. Recent empirical findings also demonstrate that previous choices and previously encountered choice tasks shown to respondents can affect the current choice outcome, indicating a form of inter-dependence across choice sets.

A number of these heuristics, namely the majority of confirming dimensions (MCD), the extremeness aversion and the reference revision heuristics, are extensively analysed in this thesis. Although these heuristics have been previously identified in the existing literature, for example, the extremeness aversion heuristic in the psychology and marketing literature, their application, using the discrete choice modelling framework, to the transportation field has only barely begun. In particular, arising from the extremeness aversion heuristic, three models are discussed. The first is a recently developed model of context dependence known as the random regret minimisation (RRM) model. The second model is a non-linear utility model that makes reference to the worst attribute level in a choice set. The third model is a "relative advantage maximisation" (RAM) model, with an updated version of the existing RAM model introduced in this thesis. All these models are compared against one another and with the standard linear-in-the-parameters random utility maximisation (RUM) model. The results strongly indicate that incorporating context dependency into existing models should be a key consideration for the practitioner.

Moreover, having identified some heuristics of especial interest, the role that multiple heuristics or decision rules can play in choice behaviour is also analysed. This can be done through models of probabilistic decision processes but interestingly, the heuristics themselves can be embedded directly into the utility functions by means of heuristic weighting functions, which weight the contribution of each heuristic to overall utility. The thesis examines the validity of such an approach.

STATEMENT OF ORIGINALITY

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in the preparation of this thesis and sources have been acknowledged.

Waiyan Leong

ACKNOWLEDGEMENTS

I would like to thank my supervisors David Hensher and John Rose for the exceptionally high level of supervision that I have received. This thesis would not be anywhere near where it is without their invaluable guidance and input from day one. Thank you, David and John, for your patience, your questions, your corrections and your affirmation of my research.

It would be remiss of me not to say a big thank you to the other academic staff at ITLS. I have benefited much from their comments and suggestions at our regular research dissemination meetings and also at our annual research student symposium.

I would also like to thank my fellow PhD students at ITLS, both past and present, for the many hours of discussions around the lunch table and in the tea room. They have challenged my assumptions and have impressed me with fabulous insights. Undoubtedly, my education at ITLS would have been so much poorer without each one of them.

Thanks to the support staff at ITLS who have made my stay at ITLS in the past three and a half years extremely smooth and enjoyable.

Many thanks also to my lovely wife Larissa who has faithfully laboured beside me throughout this journey. I cannot stress how much I am indebted to her for the many selfless sacrifices and unwavering support of me.

Last, but not least, a heartfelt gratitude to my parents. They have been so uncomplaining about my long absence away from home.

Table of Contents

ABSTRACT ... ii

STATEMENT OF ORIGINALITY ... iii

ACKNOWLEDGEMENTS ... iv

List of Tables ... ix

List of Figures ... xi

CHAPTER 1 INTRODUCTION ... 1

1.1 Preliminaries ... 1

1.2 Structure of this Thesis ... 5

CHAPTER 2 LITERATURE REVIEW ... 8

2.1 Introduction ... 8

2.2 The Interplay between Heuristics and Contextual Effects as Expressed by Choice Task Properties ... 9

2.3 Choice Task Complexity: A Review of Modelling Approaches ... 14

2.3.1 Two Stage Processes ... 14

2.3.2 ‘Soft’ Constraints ... 20

2.3.3 Mixture Models ... 22

2.3.4 Direct Representations of Complexity ... 24

2.3.5 Hierarchical Bayes Modelling ... 26

2.4 Relational Heuristics: Prospect Theory ... 27

2.5 Extremeness Aversion and Regret Minimisation ... 29

2.5.1 Empirical Demonstrations of the Extremeness Aversion Effect ... 29

2.5.2 Models of Extremeness Aversion and Regret Minimisation ... 33

2.6 Context-Free and Context Dependent Preferences ... 40

2.7 Heuristics of Choice Set Inter-Dependence: Value Learning and Strategic Misrepresentation ... 43

2.8 State Dependence and Habit Persistence ... 50

2.9 Attribute Level Editing ... 54

2.10 Other Heuristics ... 57

2.10.1 Lexicography and Elimination-by-Aspects ... 57

2.10.2 Common Metric Attribute Aggregation ... 60

2.10.3 Majority of Confirming Dimensions and Attribute Non-Attendance ... 61

2.11 Summary and Concluding Remarks ... 63

CHAPTER 3 ECONOMETRIC METHODOLOGY 68
3.1 Introduction 68
3.2 Nomenclature of Preference Data 68
3.3 Random Utility Theory (RUT) and the Multinomial Logit (MNL) Model 70
3.4 More on the Observed Component of Utility 73
3.5 The Non-Linear Random Parameters Logit Model 75
3.6 Methods of Embedding Multiple Heuristics in Choice Models 77
3.6.1 The Probabilistic Decision Process (PDP) Model 77
3.6.2 The Hybrid RRM-RUM Model 78
3.6.3 Heuristic Weighting Functions 79
3.7 Model Comparisons 80
3.7.1 Comparing Nested Models with the Likelihood Ratio Test 80
3.7.2 Comparing Non-Nested Models with the Ben-Akiva and Swait (1986) Test... 80
3.8 A Procedure for Analysing Non-linear Functions of Parameters 82
3.9 Conclusion 83
CHAPTER 4 DESCRIPTION OF THE DATA 84
4.1 Overview of the Datasets 84
4.2 Detailed Descriptions of Datasets 85
4.2.1 Australia, 2008 (Aust08) 85
4.2.2 Overview of Other Datasets 87
4.2.3 Australia, 2000 (Aust00) 88
4.2.4 New Zealand, 1999 (NZ99) 88
4.2.5 Australia, 2005 (Aust05) 89
4.2.6 Australia, 2004 (Aust04a) 89
4.2.7 Australia, 2004 (Aust04b) 90
4.2.8 New Zealand, 2007 (NZ07) 90
4.3 Summary Statistics 90
4.3.1 Summary of Key Attributes in the Reference and Hypothetical Alternatives .. 91
4.3.2 Summary of the “Worst” Attribute Levels in Choice Sets 97
4.3.3 Summary of Pair Wise Comparisons in Choice Sets 99
4.3.4 Socio-economic Characteristics 103
4.4 Some Concluding Remarks 105
CHAPTER 5 MODEL RESULTS AND ANALYSIS 106

5.1 Introduction ... 106
5.2 The Standard Context Independent MNL Model ... 107
5.3 Majority of Confirming Dimensions Heuristic as a Context Dependent Heuristic 110
5.3.1 Discussion of the *mcd* Variable ... 110
5.3.2 Generic versus Alternative Specific Parameters for the *mcd* Variable ... 112
5.3.3 A Probabilistic Decision Process Model ... 116
5.4 The Extremeness Aversion Heuristic ... 120
5.5 Models with Regret Minimisation ... 130
5.5.1 The Random Regret Minimisation (RRM) Model ... 130
5.5.2 The Hybrid RRM-RUM Model ... 133
5.6 The Relative Advantage Maximisation Model as a Description of Extremeness Aversion ... 135
5.6.1 Setting up the RAM Model ... 135
5.6.2 Some Theoretical Properties of the RAM Model ... 140
5.6.3 A Relative Disadvantage Model and its Equivalence to the RAM Model ... 144
5.6.4 Empirical Estimation of the Piecewise RAM model with Three Alternatives in Choice Set ... 145
5.6.5 Estimation of the Piecewise RAM with Two Alternatives in Choice Set ... 148
5.6.6 An Alternative Specification for the RAM Model Using the Regret Function 152
5.6.7 Relative versus Absolute Regret ... 155
5.7 Comparisons of Model Fit ... 160
5.8 The Reference Point Revision Heuristic ... 161
5.9 Heuristic Weighting Functions ... 164
5.9.1 Application of the Multiple Heuristic Approach to the RUM and NLWLR Mixture ... 166
5.9.2 Application of the Multiple Heuristic Approach to the RAM Model ... 172
5.10 Mixed Logit Estimations ... 174
5.10.1 Mixed Logit Estimation of the NLWLR Model ... 174
5.10.2 Mixed Logit Estimation of the Regret-RAM Model ... 178
5.11 Calculations of Value of Travel Time Savings (VTTS) ... 178
5.11.1 Some Preliminaries ... 178
5.11.2 The NLWLR model ... 181
5.11.3 The RRM Model and the Hybrid RRM-RUM Model ... 182
5.11.4 The RAM Model ... 183

5.11.5 Estimates and Discussion 185
5.12 Summary 188
CHAPTER 6 CONCLUSIONS 190
6.1 Main Contributions of this Thesis 190
6.1.1 Reference Revision Heuristic 190
6.1.2 Majority of Confirming Dimensions Heuristic 190
6.1.3 Models of the Extremeness Aversion Heuristic 191
6.2 Some Suggestions for Future Research 195
REFERENCES 200

List of Tables

Table 2.1: Illustration of the Allais Paradox (Kahneman and Tversky, 1979) 28
Table 2.2: Results from a Camera Purchase Experiment (Simonson and Tversky, 1992) 31
Table 2.3: Results from a Stated Choice Purchase of a Cassette Player 32
Table 2.4: An Example of the Attraction Effect (Huber *et al.*, 1982) 41
Table 2.5: An Example of the Similarity Effect (Tversky, 1972) 42
Table 2.6: Description of Possible Reference Rules 49
Table 2.7: Summary Table of Main Heuristics Discussed in Chapter 2 66
Table 4.1: Trip Attributes in Stated Choice Design 85
Table 4.2: A Summary of the Seven Tollroad Studies 92
Table 4.3: Summary Statistics of Total Trip Time and Total Trip Cost 93
Table 4.4: Summary Statistics of the Key Attributes in the Hypothetical Alternatives 95
Table 4.5: Summary of Variation of Attribute Levels around the Pivot Alternative 96
Table 4.6: Summary Statistics of $\max_{j \in s}(X_{jks})$ 98
Table 4.7: Incidence of $\max_{j \in s}(X_{jks})$ Across Alternatives 98
Table 5.1: Results from Standard MNL Estimation 109
Table 5.2: Frequency Distribution of the *mcd* Variable 111
Table 5.3: Comparison of *mcd* Parameter Assumptions Using MNL 114
Table 5.4: Estimation of Probabilistic Decision Process Models 117
Table 5.5: Results from Applying a Non-Linear Multinomial Logit Model to Aust08 122
Table 5.6: Estimates from a Non-Linear Multinomial Logit Estimation 124
Table 5.7: A Comparison of the φ_k Estimates across the Seven Toll Road Datasets (MNL Estimation) 125
Table 5.8: An Illustration of the Polarisation Effect Using Parameters Obtained from Estimation on Aust04a 129
Table 5.9: Estimation Results from the RRM (MNL) Model 132
Table 5.10: Results from Hybrid RRM-RUM (MNL) Estimation 134
Table 5.11: Comparison of Log-likelihood Statistics across the Four Models 146
Table 5.12: Estimates from the Symmetric Piecewise RAM (MNL) Model 147
Table 5.13: Comparison of Standard RUM and Piecewise RAM (MNL) Models With Two Alternatives in Choice Set 151
Table 5.14: Estimates from the Regret Form of RAM (MNL) Model (Three Alternatives in Choice Set) 154

Table 5.15: Comparison of Standard RUM, RRM, Hybrid and Regret-RAM Models (all MNL) with Three Alternatives in Choice Set 155
Table 5.16: Regret-RAM Model with Example Choice Set from Aust00 156
Table 5.17: Comparison of Log-likelihoods of RRM_1 Model (MNL) with Standard RUM (MNL) and Standard RRM (MNL) Models..157
Table 5.18: Summary and Comparison of Log-likelihood Statistics of RRM_2 (MNL) with RUM (MNL) Model.. 159
Table 5.19: Ranking of Models based on Model Fit... 160
Table 5.20: Estimation Results Embedding Reference Revision Heuristic (MNL Model) .. 163
Table 5.21: Estimation Results Embedding Reference Revision Heuristic (Error Components Logit Model).. 165
Table 5.22: Empirical Estimation of a RUM and NLWLR Mixture (MNL Model) 169
Table 5.23: Summary Statistics of W_1, W_2 and RW 170
Table 5.24: Estimation Results from a Multiple Heuristics Approach to the Piecewise RAM (MNL) Model 173
Table 5.25: Mixed Logit Estimation of the NLWLR Model.. 176
Table 5.26: Simulations of the CDFs for $\Pr(\varphi_k < 1)$ 177
Table 5.27: Estimation Results of the Mixed Logit Regret RAM Model 179
Table 5.28: Summary of VTTS Calculations for the MNL Form of the Standard RUM, NLWLR, RRM, Hybrid RRM-RUM and Regret-RAM Models... 186

List of Figures

Figure 2.1: Graph of $y = \ln(1 + \exp(x))$ 34

Figure 4.1: Illustrative Screenshot of Choice Experiment used in Aust08 86

Figure 4.2: Illustrative Screenshot of Choice Experiment Used in DS2 to DS7 87

Figure 4.3: Kernel Densities of Total Trip Time and Total Trip Cost, by Dataset.......... 94

Figure 4.4: Histograms of Age and Income Variables 105

Figure 5.1: Graph of $A_k(j, j')$ under Two Different Sign Assumptions on β 139

Figure 5.2: Piecewise RAM Model with Compromise Effect 141

Figure 5.3: Piecewise RAM Model with IIA-like property 144

Figure 5.4: Graph of the $RA(j, j')$ function 153

Figure 5.5: Kernel Density Plots of W_1, W_2 and RW from Selected Datasets 171

CHAPTER 1 INTRODUCTION

1.1 PRELIMINARIES

Human decision making typically culminates in the choice of some alternative that is picked from a larger number of alternatives that are available to the decision maker. The study of how people end up with the choices they make can be said to be the raison d'être of a significant number of researchers in the social sciences such as psychology and economics, and in many other related and applied fields as well. It is a vast and burgeoning area of research, spanning the range from theory to applied practice and fieldwork. Cutting edge techniques and methodology developed by this research are used extensively not just in the academic domain, but also in the corporate and government sectors around the world.

Discrete choice modelling is one of the most powerful tools to have emerged in the last half-century or so to understand how people make choices. At the heart of discrete choice modelling is the question of how decision processes might be expressed in a model. Frequently, the literature relies on a simplified decision making rule that is based on the use of context independent utility specifications as a representation of preferences for an alternative. By context independent, it is assumed that a person's utility for an alternative is assumed to be independent of the attributes of all other available alternatives. Hence, the presence or absence of other alternatives in the choice set does not make a difference to the respondent's subjective evaluation of any one particular alternative. Put another way, if the same alternative were to be taken from one context and placed in another, the subjective evaluation of that alternative remains the same. This assumption of context independent utilities often underpins the design of choice experiments which, in discrete choice modelling, are the main mechanism for collecting data and evidence about preferences.

The idea of context independent utility functions has a long history in economic thought. Beginning with the neoclassical economic theory of the consumer, it has been postulated that goods themselves are the direct objects of utility. In other words, while the demand for a particular good might be affected by changes in the attributes of another good such as changes in price, the utility for that good is stable and remains unaffected. This long held premise continues to be one of the basic building blocks used in standard microeconomics textbooks even till today.

It was in 1966 that Lancaster suggested an important extension to the neoclassical view of consumer behaviour. He advocated that utility is derived, not directly from the goods themselves, but from the intrinsic characteristics or attributes of the good. Even as Lancaster advanced our understanding that it is this collection of attributes of the good that is at the forefront of utility or preference ordering, no further mention was made of the possibility that utility may in part be determined by the attributes of other alternatives. Indeed, McFadden (1974) built on this notion of context independent utilities in the modelling of discrete choices by introducing linear additive models, free of contextual effects, for the non-stochastic part of the utility function.

For a variety of reasons, the idea of context independence was left largely unchallenged in microeconomic theory at that time. Context independent utilities satisfy the axioms of rational choice, and are therefore 'well-behaved' and analytically tractable, even though context independence may be inconsistent with findings from psychology. Looking back at the 1970s, Camerer (1999) suggests that economists then were not particularly concerned by the systematically false psychological assumptions of their economic agents, since their formal mathematical models of behaviour had resulted in surprisingly good predictions. Thus, the economics and psychology disciplines were kept apart for a period of time, not only by differences in methods and the ways of acquiring knowledge, but also by the reluctance of economists to delve deeper into the intricacies of human decision making.

Meanwhile, even as economists devoted themselves to deriving increasingly elegant models of consumer theory, psychologists were adopting a different approach by using results from controlled experiments to ask questions about how individual preferences might be shaped by the decision context. A few years after McFadden's seminal 1974 paper, Russo and Dosher (1983) introduced into the psychology literature the heuristic known as the majority of confirming dimensions (MCD). The motivation behind the MCD heuristic arose partly out of their research and partly out of a body of accumulating evidence from other work. Specifically, behavioural findings on processing suggest that decision makers are predominantly comparing attributes across alternatives, rather than evaluating alternatives as a whole. This idea of dimensional processing across attributes, rather than holistic processing within alternatives, led Russo and Dosher (1983) to consider how dimensional processing may proceed under various simplifying strategies.

One of these strategies turned out to be the MCD heuristic. Under the MCD, the cognitive burden of dimensional processing is reduced by assuming that decision makers are ignoring the magnitude of the differences in the attribute levels. Only the direction of change – whether an attribute is higher or lower than its counterpart in a competing alternative – is thought to matter. The MCD rule may also be said to be an example of what Tversky and Simonson (1993) would later call "context dependent preferences", since explicit consideration is made to a relative ranking in an attribute-by-attribute comparison across alternatives in the decision making process.

Continuing with this theme of context dependent preferences, around the same time of the work by Russo and Dosher (1983), Huber *et al.* (1982) demonstrated convincingly that contextual effects can influence choice in ways that can violate commonly held assumptions, such as Luce's (1959) choice axiom which states that the probability ratio between two alternatives in a choice set is not affected by the presence or absence of other alternatives. In particular, Huber *et al.* (1982) showed that choice probabilities between existing alternatives can be altered significantly when a certain class of alternatives called asymmetrically dominated or 'decoy' alternatives are introduced into the choice set. In other words, decision making is not independent of choice set composition. Shifts in perceptions arising from such choice set changes or even the use of decision strategies involving pair wise attribute-by-attribute comparisons in the spirit of Russo and Dosher (1983) were suggested as possible explanations for such behavioural anomalies.

Regret theory, articulated by Loomes and Sugden (1982) in the context of decision making under uncertainty, is yet another example that preferences may be influenced by context. Loomes and Sugden (1982) argue that decision makers will take into account the psychological experience of regret, in the sense that the pleasure derived from a course of action depends not only on the consequence of that action itself, but also on the consequences of the non-chosen action. Therefore, if the non-chosen course of action happens to lead to a more desirable consequence, decision makers might reflect on how much better their positions might have been, had they made a different choice, and this reflection may have a negative bearing on the pleasure derived from the chosen action.

The classical example of the role of contextual effects in individual decision making comes from prospect theory (Kahneman and Tversky, 1979). A key idea that emerges from the evidence is that in the editing phase of the decision, "people normally perceive outcomes as gains and losses, rather than as final states of wealth or welfare" (Kahneman and Tversky, 1979, p. 274). Gains and losses must be defined relative to some reference point. In particular, the idea that the attributes of some competing alternative can be used as the reference point by which gains and losses are defined is the hallmark of the models of the extremeness aversion heuristic that is explored in this thesis. Simply put, the extremeness aversion heuristic is characterised by the proclivity to choose a middle option that has neither very good nor very bad attributes, rather than an extreme option that has both very good and very bad attribute levels. Therefore, another objective of this thesis will be devoted to estimating and comparing models of extremeness aversion in the transportation stated choice context.

The question of whether reference points can be revised is an interesting one and has been the subject of some discussion in the literature, especially where path dependence is concerned. Whether preferences are poorly formed initially, or whether strategic considerations are at the forefront of decision making in a sequence of choice responses, initial choices made at the beginning of the sequence may come to influence subsequent decisions. Broadly speaking, a departure from the status quo and the inertia that it exerts on the decision process may mean that a non status quo alternative becomes the new point of reference for subsequent decision making.

The fundamental premise of this thesis rides on the growing trend, in discrete choice analysis, of giving greater consideration to alternative behavioural models that contrast with the standard context independent formulation. In a sense, as Camerer (1999) points out, there has recently been a reunification of psychology and economics. The field is vast, the number of heuristics potentially available to a decision maker is immense, and it is therefore necessary to consider a subset of heuristics that might be worthwhile exploring in greater detail. The study of these heuristics will take place within the transportation context. As will be fully explained in subsequent chapters, the following three alternative decision rules – the majority of confirming dimensions (MCD), the extremeness aversion heuristic, and the reference point revision heuristic – are the focus for more intense assessment in the thesis. To date, a thorough understanding of these decision rules or heuristics in the transportation discrete choice literature has been limited. This thesis will address this research gap and it is hoped that the findings will contribute to a wider discussion and debate among transportation professionals on the merits of using such alternative models for future research.

Increasingly, it is also becoming widely recognised that heterogeneity in decision making strategies is an important element to capture in formal modelling (see for example, Hensher and Greene, 2010; McNair *et al.*, 2012; Hess *et al.*, 2012; Chorus *et al.*, 2013, to name but a few). Modelling approaches typically consider differences in strategies across segments of decision makers or perhaps along the dimension of attributes. This thesis also proposes an alternative way of understanding heterogeneity by specifying that a weighted mixture of decision strategies may be integrated into one combined decision rule, with the weights of the mixture allowed to be endogenously determined through the estimation process.

Marginal willingness to pay measures constitute important outputs from a discrete choice model. Studies from behavioural economics suggest that in real life, people are probably not as 'well-behaved' as neoclassical economic theory would postulate, but the willingness to pay and consumer surplus measures grounded in microeconomic axioms are well understood, even if it almost inevitably means accepting a loss of behavioural realism. On the other side of the trade off, there may be instances where capturing more behavioural realism is preferred. Such a scenario might come about where more accurate forecasts of future demand might be required. This thesis will also examine the various implications on willingness to pay arising from the use of context dependent preferences in discrete choice models.

1.2 STRUCTURE OF THIS THESIS

This thesis is organised as follows. The following chapter, Chapter 2, discusses and summarises some of key research papers related to the modelling of context dependent preferences and heuristics in the discrete choice modelling literature. A review of the early work from behavioural decision research shows an emphasis on relating various aspects of choice task properties to the use of heuristics in decision making. Some of this work is qualitative in nature, but modelling approaches developed within the discrete choice framework tend to focus on phased decision making, with the first phase invoking some form of screening criteria to narrow down a subset of feasible alternatives from a larger universal set. In some cases, direct representations of complexity through the entropy measure are used. Following this, ‘relational’ heuristics such as extremeness aversion are then discussed. The common feature of this class of decision heuristics is the explicit use of reference points derived from attribute levels of competing alternatives within the same choice set. The final major class of heuristics reviewed in this chapter pertains to issues of choice set inter-dependence.

Chapter 3 describes the econometric methodology that will be used in the data analysis. Some basic definitions pertaining to the nomenclature of the data used are first described. These are then followed by an overview of the econometrics behind the standard linear additive random utility maximisation (RUM) model. Following this review, some discussion of the non-linear logit model is then provided, as many of the heuristics of interest result in a non-linear specification for the utility function of an alternative. The chapter concludes with a discussion of some post-estimation methods for comparing model outputs.

Chapter 4 provides an extensive description of the datasets which are used for the empirical work. The common feature across all these datasets is that each of them relates to a stated choice experiment on what route choices are made, where each alternative presented represents varying bundles of travel time and travel cost attributes. Moreover, each of these datasets makes use of a pivot design. Such a design allows the current or reference alternative to always feature in the choice set, and also allows the attribute levels of the hypothetical alternatives to be coded as variations around the reference level. The similarity of the data context enables some conclusions to be made on whether the decision rules tested are independent of respondents; in other words, whether they carry across similar data contexts.

Chapter 5 forms the empirical core of the thesis. Results from the RUM MNL model are first presented. These are then followed by a discussion on modelling the MCD heuristic. Next, models of extremeness aversion, including an extension of the contextual concavity model (Kivetz *et al.*, 2004), the random regret minimisation model (Chorus, 2010) and its variants, and the relative advantage maximisation (RAM) model (Kivetz *et al.*, 2004) are introduced and extensively tested and discussed. The last heuristic to be considered is the reference revision heuristic. As mentioned earlier, a framework for combining and integrating these various decision rules is also introduced and examined in this chapter. The chapter concludes with a section devoted to a discussion of the marginal willingness to pay measures that can be derived from the models.

Chapter 6 concludes by summarising the main findings of the thesis. Suggestions for a future research agenda on the role of heuristics, and in particular, some possible ways forward for the RAM model, are also discussed.

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

The typical approach used in much of the discrete choice modelling literature in traveller behaviour studies assumes that well-defined preferences exist for most decision tasks. Under the standard random utility theory, preferences are stable and invariant to choice tasks, and are fully known to the respondent. In the great majority of cases, the analyst writes out a utility function assuming that the respondent is cognitively indefatigable, examining all alternatives and all attributes across all choice tasks in the same fully compensatory manner. The linear weighted additive form for utility, also commonly known as the standard Random Utility Maximisation (RUM) model, has been found to be a representation which is easily tractable and which is capable of embodying all these assumptions. The RUM model has therefore become the mainstay in discrete choice modelling.

As a description of how people behave, research in the decision behaviour field, and more recently from the choice experiment literature, has cast doubt that the weighted additive function, assumed independent of contextual effects, comes close to being an accurate representation of the actual processes used in the majority of decision tasks (see for example, McFadden, 1999; Gigerenzer *et al.*, 1999). Empirical research from the psychology literature has shown that preferences for an alternative are influenced by the choice context itself; in other words, by factors that are beyond the immediate attributes of the alternative under consideration. Earlier work focused on how choice task properties, such as the number of alternatives and attributes, impact decisions in terms of how decision rules are selected and applied (e.g., Payne *et al.*, 1993). Decisions may also be made according to some reference point selected by the respondent (see for example, Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Li and Hensher, 2011). This reference may have something to do with the other alternatives in the same choice set (Simonson and Tversky, 1992; Kivetz *et al.*, 2004) or even across previously encountered choice sets (McNair *et al.*, 2011; 2012). Moreover, different respondents may be attending to various subsets of attributes, and such heterogeneity may be masked if it is assumed that preference weights are the same across all individuals in the entire dataset. Outside the immediate environment of a choice experiment, the social context, through an intermediate construct of comparative happiness, may also influence utility (Abou-Zeid and Ben-Akiva, 2011). Compared to the context independent, linear weighted additive, form of utility, these studies suggest that other representations of utility which better approximate the realism of real-life decision making can lead to an improved goodness of fit and more plausible model estimates and outputs. Swait *et al.* (2002) best summarise this position when they write that

> *"In choice modelling, we believe a shift in emphasis to mapping the psychological processes of an individual in comprehending and responding to the decision task at hand could bring some significant gains in terms of a richer insight into individual preferences, improved accuracy in predicting individual behavioural responses to economic changes or policy interventions, and a more realistic assessment of the impact of policy interventions on individual welfare."* (Swait *et al.*, 2002, p. 197).

The purpose of this chapter is to review the decision heuristics that have been modelled in the discrete choice literature, drawing from extant work in the fields of transport, environment, marketing and psychology. The examples cited in this chapter primarily deal with stated choice data, although the relevance of applying decision heuristics to revealed preference data cannot be ruled out, even if examples from revealed preference data are much fewer in comparison. Heuristics prompted by contextual effects in the form of choice task properties are first discussed, followed by a review of the various methodological approaches to embed such heuristics into choice models. Contextual effects embodied in relational heuristics linked to the idea of referencing are then discussed. The chapter concludes by identifying some research gaps in the literature, especially as they relate to discrete choice models in the transportation literature, and lays out a roadmap for the rest of this thesis.

2.2 THE INTERPLAY BETWEEN HEURISTICS AND CONTEXTUAL EFFECTS AS EXPRESSED BY CHOICE TASK PROPERTIES

The psychology literature has amassed a wealth of evidence to suggest that humans rely on the use of quick mental processing rules, known as heuristics, to manage the vast number of decisions that must be made in everyday life. While the fully compensatory weighted additive rule is commonly used for the purposes of modelling, psychologists have argued that this rule, if followed strictly to the letter, is cognitively demanding and time consuming (Payne *et al.*, 1993). It moreover implies an assumption of stable, well-articulated preferences which appears to hold only under conditions where the choice task is familiar or when the respondent has experience with the various alternatives that are presented. In many instances, it is argued that these conditions fail to apply, and preferences are not determined in advance of the choice situation, but are instead constructed in response to the contextual effects which vary according to the properties of the choice task. As described by Payne *et al.* (1999, p. 245), the construction process involves an interaction between "the properties of the human information processing system and the properties of the choice task."

Therefore, rather than static decision processes which are repeatedly applied to different choice contexts, the conclusion drawn by behavioural decision research is that "individuals have a repertoire of decision strategies for solving decision problems" (Bettman *et al.*, 1998, p. 194). Some decision strategies that might be a part of this repertoire include satisficing (Simon, 1955), lexicography (Tversky, 1969), elimination-by-aspects (EBA) (Tversky, 1972) and the majority of confirming dimensions (Russo and Dosher, 1983). Brief descriptions of these heuristics follow below:

(1) *Satisficing* (Simon, 1955): Under this heuristic, an attribute may be associated with a pre-defined cut-off and the first alternative with all attribute levels satisfying the cut-off criteria is chosen. If none of the alternatives meets the cut-off criteria, the cut-offs may be relaxed and the process repeated; or a random choice is made.
(2) *Lexicography* (Tversky, 1969): The respondent evaluates all alternatives based on what is deemed to be the most important attribute. The alternative with the best value on that attribute is chosen. If there is a tie, or if the difference between the best levels of the attribute is not noticeable, then the remaining alternatives are evaluated on the next most important attribute, and so on.
(3) *EBA* (Tversky, 1972): The respondent identifies the most important attribute (either deterministically or probabilistically) and its associated cut-off threshold. An alternative is eliminated if its attribute fails to satisfy the cut-off. This process is repeated with the second most important attribute and so on until one alternative remains.
(4) *Majority of confirming dimensions* (Russo and Dosher, 1983): The first two alternatives are compared and the one with the larger number of winning attributes is retained. The retained alternative is compared with the next alternative and so on until

all alternatives have been evaluated. The alternative with the highest number of winning attributes is selected.

Payne *et al.* (1993) argue that heuristics are used to manage situations of high choice task complexity. Choice complexity is largely determined by the choice context, which in this case are choice task properties such as the number of alternatives in the choice set, the number of attributes in each alternative, and the correlation between attribute levels across multiple alternatives. This idea of complexity can be contrasted with Hensher's (2006) notion of relevancy, which pertains to providing more complete descriptions of attributes in the choice task and allowing respondents to form their own processing rules with regards to relevancy. Hence, a choice task that disaggregates say a time attribute into its various components such as free-flow time, slowed down time and stop-start time may be more relevant to a respondent, even though this task would be "more complex" from Payne *et al.*'s (1993) perspective.

An effort-accuracy trade-off framework has been proposed as one possible mechanism by which individuals select the decision strategy or combination of decision strategies out of the repertoire available to them (Payne *et al.*, 1993). This framework is not a formal choice model, but it still postulates a relationship between decision strategies and choice task complexity. Essentially, the decision maker's choice of heuristic is thought to be the outcome of trading-off between two conflicting goals: maximising the accuracy of a decision, with the weighted linear additive rule defining the normative benchmark level of accuracy, and minimising the cognitive effort required to reach that decision. One way of measuring cognitive effort for each decision strategy is to consider the number and type of elementary information processes (EIPs) required. For example, in a linear additive representation of utility, an alternative is assessed through EIPs such as MULTIPLYING weights with attribute levels and then ADDING up part-utilities. An alternative is ELIMINATED if the sum of its part-utilities is less than that from a competing alternative. The cognitive processes of multiplying, adding, comparing, eliminating and choosing are collectively known as EIPs.

Using qualitative evidence such as verbal protocols and process tracing, Payne *et al.* (1993) suggest that heuristics such as lexicography and EBA are more commonly used when choice tasks become more complex. Their understanding of complexity essentially relates to the dimensions of the choice task and the quantity of information provided, such as the number of alternatives and the number of attributes per alternative. A choice task is therefore said to be more complex when the information load given to respondents increases. Hensher (2006) takes issue with this particular definition of complexity, arguing that it is 'relevancy' that matters, but following Payne *et al.*'s (1993) line of reasoning for the moment, it has been argued that in more complex choice tasks, heuristics are more likely to be called upon to aid decision making because they require much less cognitive resources but at the same time, do not require much sacrifice in terms of accuracy, where accuracy is defined as the ability of a decision heuristic to replicate the choice outcome from the weighted additive rule. In this sense, Hensher (2006) and Payne *et al.* (1993) agree that when the dimensionality of a choice experiment is high, individual respondents are more likely to use 'coping' strategies (Hensher's terminology) or decision heuristics (Payne *et al.*'s terminology). Of course, the choice of heuristic is not necessarily only as a strict consequence of the dimensionality of a choice experiment, but can be associated with previous experience in adopting specific heuristics, and hence it is important to recognise and account for both potential sources of influence on choice making and selection of a heuristic.

As a corollary of the effort-accuracy trade-off framework, a phased decision strategy may also be employed under some circumstances (Stevenson *et al.*, 1990). In this strategy, respondents are thought to initially rely on some non fully-compensatory heuristic to reduce choice task complexity before using a fully compensatory strategy to evaluate the reduced number of alternatives and/or attributes. For example, Payne (1976) has observed that respondents demonstrate the use of attribute-based non-compensatory strategies like EBA early in the decision process, to reduce the number of alternatives before using an alternative-based strategy such as additive utility, to arrive at the final outcome. This raises an interesting possibility in terms of whether a mixture of heuristics can be modelled sequentially, instead of simultaneously, a feature which can be incorporated into models of two stage processes, reviewed below.

Conceptually, the effort-accuracy framework requires the decision maker to be cognisant of the costs and benefits of each strategy as applied to the choice task under consideration. The realism of such an assumption is debatable, in view of the maintained hypothesis that cognitive effort is a scarce resource. Cost-benefit approaches potentially lead to the infinite regress issue. At the empirical level, as pointed out by Cameron and DeShazo (2010), a significant challenge remains in terms of identifying and quantifying the cognitive effort associated with each heuristic. Nonetheless, the effort-accuracy framework describes broad conditions under which non-compensatory heuristics like lexicography or EBA are more likely to be used, thus providing some guidance in terms of model specification. For example, Young (1984) assumes an EBA model of choice because choosing a residential location among many attributes and alternatives is thought to be a complex choice task. Therein also lies a potential limitation of the effort-accuracy framework, for it does not explain the strong empirical evidence for cognitively more demanding relational heuristics (Section 2.4 and Section 2.5) that require the use of various reference points.

Discrete choice models of information processing strategies such as attribute non-attendance (see for example, Hensher, 2010; Hensher and Greene, 2010; Scarpa *et al.*, 2009) may be interpreted as a "bottom-up" or "data-driven" view of preference construction (Payne *et al.*, 1993, p. 171). The "bottom-up" view differs from a top-down view of strategy selection where the decision maker selects the best strategy from his/her repertoire of strategies, on the basis of some effort-accuracy trade-off. The bottom-up view to preference construction allows respondents to shape or change decision strategies on the spot by exploiting previously encountered problem structures, which can be a reflection of accumulated overt experiences. As people learn more about dimensions of the decision problem, processing strategies may change to reflect the peculiarities of the decision problem. In other words, processing is opportunistic. In bottom up processing, decision problems may be subsequently restructured as an intermediate step, making them more amenable to analysis using certain heuristics. Information in choice tasks might be transformed through rounding or through calculations to standardise values in a common metric. Information might also be rearranged or further simplified by deeming certain attributes irrelevant. It is argued that such restructuring serves to make difficult decision tasks more manageable by reducing the perceived complexity of the choice task and making later processing more efficient (Payne and Bettman, 1992).

2.3 CHOICE TASK COMPLEXITY: A REVIEW OF MODELLING APPROACHES

2.3.1 Two Stage Processes

Modelling decisions as following a two stage decision process is one way of incorporating a phased decision strategy into a discrete choice model. A general form of a two-stage model, attributed to Manski (1977), is given in Equation (2.1):

$$P_j = \sum_{C \in G} P(j \mid C) P(C), \tag{2.1}$$

where P_j is the unconditional probability that alternative j is chosen, $P(j \mid C)$ is the probability of choosing alternative j given the reduced choice set C, and $P(C)$ is the probability that the reduced choice set is C, among all the non-empty subsets of a master choice set G.

In the first stage of decision making, respondents are assumed to invoke screening rules in selecting a subset of alternatives from a larger universal set. The final choice is made from the reduced set. In these two-stage models, Swait and Ben-Akiva (1987) have observed that the first stage, which is a form of choice set generation, establishes the set of feasible alternatives and in many instances, resembles the "considered subset" in Simon's (1955) satisficing model. To generate the feasible choice set, the analyst must identify appropriate screening rules or constraints. These rules could be based on the history of past choices of the respondent, or on the attribute levels of alternatives in the current choice situation. Swait and Ben-Akiva (1987) argue that this process "must not only consider constraints such as income and transport infrastructure, but must also account for informational, psychological, cultural and social restrictions" (Swait and Ben-Akiva, 1987, p. 91). Because constraints are external to the choice process, and are not often observed by the analyst, they may be thought of as random or probabilistic, especially if the degree of confidence in the constraint is low. To operationalise these constraints probabilistically, Swait and Ben-Akiva (1987) suggest an appeal to random thresholds, for example, a travel alternative is included in the feasible choice set if its cost is below a threshold T_1, with T_1 distributed according to an unknown (but estimable) mean $\overline{T_1}$ and variance $\sigma^2(T_1)$.

Expanding on the Manski (1977) equation, Cantillo and Ortuzar (2005) assume a first stage elimination involving the use of a rejection mechanism based on individual-specific thresholds of attribute levels. Alternatives which survive the first stage screening are then evaluated in the usual compensatory manner within the random utility framework. Cantillo and Ortuzar (2005) suggest that the threshold might be determined by "the most favourable value among those that the attribute can take for the set of potential alternatives; it could also be the value that the attribute takes for the chosen alternative or simply any reference value" (Cantillo and Ortuzar, 2005, p. 644). The distribution of thresholds can even be a function of systematic influences. Hence, there is a great amount of flexibility as to how the individual-specific thresholds are modelled.

Following Swait and Ben-Akiva's (1987) suggestion, Cantillo and Ortuzar (2005) assume that the vector of such thresholds is distributed across all individuals with a certain mean and variance-covariance structure, since they do not have other data sources on what constitutes acceptable threshold levels,. Hence, $\mathbf{T}_n$, may be specified as a $m \times 1$ threshold vector, where m is the number of attributes subject to threshold considerations, satisfying $0 \leq m \leq K$. K is the total number of attributes in the alternative. $\mathbf{T}_n$ is assumed to be a random vector distributed according to a joint density function $\Omega(\delta)$ with mean $E[\mathbf{T}_n] = \overline{\mathbf{T}}$ and a variance-covariance matrix $\mathrm{var}[\mathbf{T}_n] = \mathbf{\Sigma}$. The vector of means of the distribution $\overline{\mathbf{T}}$ can also be made a function of individual specific characteristics. Hence, an alternative j is included in the second stage consideration if $\mathbf{X}_{nj} \leq \mathbf{T}_n$, for all the m attributes which are assumed to be threshold constrained.

Applying this model to a route choice stated preference experiment of possible car trips in which alternatives are described by travel time, toll charge and the number of accidents per year, Cantillo and Ortuzar (2005) do not find evidence of a threshold effect for the time and cost attributes. For the accident rate attribute however, there is evidence for an age varying threshold effect, with men below thirty years old having a larger threshold for accidents. Interestingly, once the threshold effect for the accident variable is explicitly modelled, the accident attribute no longer contributes to indirect utility in the compensatory decision making stage and its effect appears entirely in the non-compensatory screening stage. whereas cost and travel time continue to remain relevant in the compensatory stage in all model specifications. It is also interesting that in the accident threshold model, Cantillo and

Ortuzar (2005) find that the value of travel time is much higher compared to the MNL model. What Cantillo and Ortuzar (2005) do not address in their paper, however, is the situation where a very low attribute level for cost or time might invoke a believability threshold, for example, in the work on attribute attendance and processing (Hensher, 2010).

Another example of a two-step model comes from Suzuki's (2007) model of airline and airport choice. In this model, the probability of an alternative surviving the first stage is assumed to be negatively related to its likelihood of violating some minimum acceptable standard or threshold. The probability of an alternative *j* belonging to a reduced choice set C is given in Equation (2.2) by a negative exponential function:

$$\Pr(j \in C) = \frac{1}{\exp(\delta_1 L)}, \tag{2.2}$$

where L is a penalty variable that increases as thresholds (minimum standards) are violated and δ_1 is a parameter to be estimated.

Like Cantillo and Ortuzar (2005), Suzuki (2007) does not collect respondent data on thresholds. Since the thresholds are unknown for each respondent, they are arbitrarily approximated. However, unlike Cantillo and Ortuzar (2005), Suzuki determines these acceptable standards by appealing to contextual effects using a decision rule reminiscent of lexicography and majority of confirming dimensions. Various specifications for these standards, such as whether an attribute is the 'best' performer, 'best'/'second-best' performer or 'best'/'second-best'/'third-best' performer in the choice set, are tested for goodness-of-fit. If the 'best value' definition is used to specify the threshold, then only the alternative that has the most attractive value of a given attribute is considered as meeting the respondent's acceptable standard for that attribute.

Alternatives are then evaluated according to a disjunctive or conjunctive rule via penalty variables denoted by L. Under the disjunctive rule, an alternative simply needs to meet the respondent's acceptable standards in at least one attribute to have a 100 percent chance of being considered in the second stage; in this case, where acceptable standards are met, the penalty variable satisfies $L = 0$. If the respondent's acceptable standards are not satisfied, then $L = 1$ and the probability of second stage consideration will be lower. On the other hand, the conjunctive rule explicitly takes into account the number of attributes meeting the respondent's acceptable standards. An alternative with a larger number of attributes meeting these acceptable standards is granted a higher probability of survival into the second stage. Under the conjunctive rule, $L = 0$ for the alternative(s) having the maximum number of attributes meeting the acceptable standard, from among all alternatives in the choice set. All other alternatives will have a non-zero L variable, with the magnitude of L increasing with the alternative that has fewer attributes meeting the acceptable standard.

The Suzuki (2007) model specifies a traveller's airport-airline choice probability as specified in Equation (2.3):

$$\Pr(h, j) = \Pr(h)\Pr(j \mid h), \tag{2.3}$$

where $\Pr(h, j)$ is the joint probability of choosing departure airport h and airline j.

If a two-step process of airline choice is assumed, then the conditional probability $P(j \mid h)$ is given by Equation (2.4):

$$\begin{aligned} P(j \mid h) &= \frac{\exp(u_{hj}).\Pr(j \in C_{hj})}{\sum_{j' \in J_h} \exp(u_{hj'}).\Pr(j' \in C_{hj})} \\ &= \frac{\exp(u_{hj} - \delta_l L_{hj})}{\sum_{j' \in J_h} \exp(u_{hj'} - \delta_l L_{hj'})}, \end{aligned} \tag{2.4}$$

where u_{hj} is the observed component of the utility for airline j given airport h,

J_h is the airline consideration set given airport h and

C_{hj} is the airline choice set given airport h.

Turning to airport choice, if a one-step process is assumed, the marginal probability $P(h)$ satisfies Equation (2.5):

$$\Pr(h) = \Pr\{\max_{j \in J_h}(v_h + u_{hj} - \delta_l L_{hj}) \geq \max_{j \in J_{h'}}(v_{h'} + u_{h'j} - \delta_l L_{h'j})\} \forall h' \in H, \tag{2.5}$$

where H is the consideration set of airports available and

v_h is the observable component of the utility for airport h.

Equation (2.5) can be converted into Equation (2.6):

$$\Pr(h) = \frac{\exp(\lambda_h v_h + \lambda_h IV_h)}{\sum_{h' \in H} \exp(\lambda_{h'} v_{h'} + \lambda_{h'} IV_{h'})}, \tag{2.6}$$

where $IV_h = \ln \sum_{j \in J_h} \exp(u_{hj} - \delta_l L_{hj})$ and λ is an inclusive value parameter.

Equation (2.6), which is used to model the one-step process of airport choice, may be modified to take account of a two-step process of airport choice, as in Equation (2.7):

$$\Pr(h) = \frac{\exp(\lambda_h v_h + \lambda_h IV_h).\Pr(h \in D)}{\sum_{h' \in H} \exp(\lambda_{h'} v_{h'} + \lambda_{h'} IV_{h'}).\Pr(h' \in D)}, \tag{2.7}$$

where D is the airport choice set and $\Pr(h \in D) = \dfrac{1}{\exp(\delta_p L_h)}$. The parameters may be estimated simultaneously by maximising the log-likelihood function associated with $\mathrm{P}(h, j)$.

The overall conclusion from Suzuki's (2007) modelling results is that the goodness of fit measures support the assumption of a two-step process in airline choice (but a one-step process in airport choice) over a conventional utility model specification which assumes no such thresholds. In general, the better model fit is obtained using the conjunctive rule and a stricter definition of acceptable standards; hence, the best model fit is achieved with the conjunctive method and by assuming that the 'best' performing attribute defines the acceptable standard for respondents. Further tests of Suzuki's two-step model also show that it outperforms the standard MNL and nested logit models in terms of forecasting and prediction.

Another application of two-stage models lies in modelling endogenous attribute attendance and attribute non-attendance (Hole, 2011; Hensher and Greene, 2010). In the first stage, rather than deciding on the composition of the considered subset of alternatives, respondents are instead assumed to decide which subset of attributes to take into account when making their choice. Subsequently, in the second stage, all available alternatives in the choice task are evaluated conditioned on the subset of attributes chosen in the first stage. The probability that decision maker *n* takes attribute *k* into account in the first stage is assumed to follow a logistic function, which is Equation (2.8):

$$P_k = \frac{\exp(\delta_k' \mathrm{z}_\mathrm{n})}{1 + \exp(\delta_k' \mathrm{z}_\mathrm{n})} \tag{2.8}$$

In this equation, z_n is a vector of individual-level observed characteristics and δ_k is a vector of parameters to be estimated. Comparing the results of the endogenous attribute attendance (EAA) model to the standard logit model, Hole (2011) finds that the both models are qualitatively similar in terms of the sign and the statistical significance of the parameters, with the only exception being an attribute that is significant in the logit model but not in the EAA model. Comparison of model goodness-of-fit using a likelihood ratio test allows Hole to conclusively reject the standard fully compensatory logit model in favour of the two stage model. In terms of attribute attendance, Hole (2011) finds that a substantial share of respondents is not attending to one or more attributes when making their choices. For example, the most frequently ignored attribute was the cost attribute which was only attended to by about 30 percent of respondents.

There are also noticeable differences in the willingness-to-pay (WTP) measures between the logit model and the EAA model. In most cases, the standard logit estimates are considerably higher than the EAA model estimates. Hole (2011) concludes that imposing the assumption that all respondents trade off all attributes in a fully compensatory manner might lead to biased WTP estimates. He cautions however that allowing for endogenous attribute attendance/non-attendance results is a different challenge as it may not be possible to infer anything about WTP for the (relatively large) segment of respondents who seem to ignore cost. While Hole (2011) does not give a reason for the high incidence of costs non-attendance, Hensher *et al.* (2012b) suggest that such this result might have come about as a consequence of the design of the choice experiment. For example, some respondents may not have seen much merit in trading off costs with the other attributes of the alternatives given the ranges and levels presented in the experiment.

2.3.2 'Soft' Constraints

Instead of assuming a 'hard' inviolable cut-off constraint and then modelling the probability of whether an alternative is dropped from consideration or not, another class of models allows cut-offs to be possibly violated, and then penalises any violation by directly adding a penalty cost into the utility function. This is the approach taken by Swait (2001). Hence, an alternative whose attributes violate the cut-off can still be chosen provided sufficient compensation in the other attributes is available to outweigh the dis-benefit of violating the cut-off. The modelled component of the utility function for alternative *j* is written in Equation (2.9) as follows:

$$V_j = \sum_k \beta_{jk} X_{jk} + \sum_k \left(\omega_k \lambda_{jk} + \upsilon_k \kappa_{jk} \right) \tag{2.9}$$

In this model, λ_{jk} and κ_{jk} are the respective penalties of violating the lower bound and upper bound constraints on attribute *k*. Therefore, the parameters ω_k and υ_k may be interpreted as the marginal disutilities of violating the lower and upper cut-offs. Let the lower bound cut-off threshold and the upper bound cut-off threshold for attribute *k* be denoted by c_k and d_k respectively, where c_k and d_k may be allowed to vary across individuals. λ_{jk} and κ_{jk} may then be defined in terms of c_k and d_k as in Equation (2.10):

$$\lambda_{jk} = \begin{cases} 0 \text{ if } c_k \text{ does not exist} \\ \max(0, c_k - X_{jk}) \end{cases} \qquad \kappa_{jk} = \begin{cases} 0 \text{ if } d_k \text{ does not exist} \\ \max(0, X_{jk} - d_k) \end{cases} \tag{2.10}$$

Incorporating these cut-offs creates piecewise linear utility functions. To estimate the model, Swait (2001) uses self-reported cut-off information from respondents to determine c_k and d_k. Estimation results show that the inclusion of penalty parameters is able to significantly improve the goodness-of-fit of the model. However, for empirical identification of the penalty parameters, sufficient choices must be made under conditions where cut-offs are violated, and this may be easier to achieve with experimentally manipulated stated choice data rather than revealed preference data.

In a subsequent application of a ‘soft’ constraint model, Hensher and Rose (2012) extend Swait’s (2001) model by allowing the entire utility function, including the penalty function, to be conditioned on individual specific perceptions, such as the acceptability of the alternative, the certainty of the choice response and the incidence of attribute levels in the perceived attribute threshold rejection region. More specifically, an index A_{jn} defining the acceptability of alternative *j* for the individual *n* is constructed and by allowing the observed component of utility to be specified as $A_{jn}V_{jn}$, A_{jn} conditions the utility expression multiplicatively. The formulation of A_{jn} recognises that individual-specific perceptions, proxied by statements on relevance of attributes defining each alternative, condition the marginal (dis)utility of each observed attribute associated with alternative *j* in a pre-defined choice set. An example of how A_{jn} may be constructed is by means of Equation (2.11):

$$A_{jn} = 1 + \delta_j \left(AC_{jn} + \sum_{k=1}^{K} \gamma_k R_{kn} \right) \qquad (2.11)$$

In Equation (2.11), AC_{jn} is a variable denoting whether an alternative is perceived to be acceptable or not by the n^{th} individual, R_{kn} is a dummy variable indicating whether the level of attribute *k* is in a perceived attribute threshold rejection region or not for individual *n*, and δ_j and γ_k are estimated parameters. The inclusion of R_{kn} recognises that the perception of alternative acceptability is fundamentally determined by the attributes, in particular the role of attribute thresholds.

In terms of the data requirements to identify the model, respondents are asked to indicate, for each alternative in each choice task, whether it was acceptable or not. Prior to the commencement of the actual choice experiments, respondents are also asked to reveal their lower and upper thresholds for attributes. Hensher and Rose (2012) find that besides improved fit and in sample prediction success, conditioning on these perceptions, especially on alternative acceptability, leads to noticeable differences in mean direct elasticities compared to a model without such conditioning.

2.3.3 Mixture Models

Another approach, used by Swait (2009), to formalise decision process heterogeneity is to consider a mixed model of random utility, where an alternative may be evaluated in one of several discrete states, with each state corresponding to a different decision rule or cognitive process. One of these states pertains to the usual utility maximising, fully compensatory condition, while other states may represent a more extreme version of attractiveness or unattractiveness, which aims to capture the possible use of a non-compensatory strategy, context dependence and/or attribute independence. Equation (2.12) illustrates this model for a simple two-condition scenario, where alternative j is assigned to either the first state which represents the trade-off condition in the usual sense, or another state representing a rejection condition, where the utility for alternative j is not defined over attribute values.

$$U_j \begin{cases} = V_j + \varepsilon_j \text{ with probability } p_j \\ = -\infty \text{ with probability } q_j \end{cases} \tag{2.12}$$

Swait's (2009) model can be set up to embed the EBA heuristic as part of choice set formation, by allowing q_j, the probability of an alternative being in the rejection condition, to be written as a function of a disjunctive screening rule: it takes just one attribute to fail the threshold cut-off before the alternative is eliminated. Conversely, p_j, which is the probability that an alternative is in the usual random utility maximising, fully compensatory trade-off condition, is written in the conjunctive sense: it is the probability of all attributes satisfying the threshold criteria before fully compensatory processing takes place.

In the model, for each attribute of interest, individual-specific thresholds can be assumed to be randomly distributed across the population, according to say, a normal distribution with mean $\bar{\tau}_k$ and variance σ_k^2. Consider an example where the EBA heuristic is applied to one aspect, for example, departure time. Then Equation (2.13) is obtained, where τ_k takes on a lower bound threshold (i.e., departure time no earlier than τ_k):

$$
\begin{aligned}
p_j &= \Pr(\tau_k < X_{jk}) = \Pr\left(Z < \frac{X_{jk} - \bar{\tau}_k}{\sigma_k} \right) = \Phi\left(\frac{X_{jk} - \bar{\tau}_k}{\sigma_k} \right) \\
q_j &= 1 - p_j = 1 - \Phi\left(\frac{X_{jk} - \bar{\tau}_k}{\sigma_k} \right)
\end{aligned}
\tag{2.13}
$$

Equation (2.13) can be generalised to elimination by *m* aspects, $1 \le m \le K$, in which case a joint density function for the vector of thresholds, $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_m)'$ is required. Other parameterisations of p_j are also possible, for example, as a logistic function of attributes or person characteristics. Using such a logistic specification, Swait (2009) concludes that these mixed models are preferred over the standard linear additive model despite the loss of degrees of freedom in estimating an increased number of parameters.

Despite sharing some similarities to two-stage models, Swait's (2009) model departs from the strict two-stage model in the sense that there is a non-zero probability of all alternatives being in the rejection condition. If this happens, the alternatives are simply selected according to a random choice rule. This contrasts with the typical two-stage model, where the set of all possible choice subsets in the second stage does not include the null, thereby excluding the possibility that all alternatives are 'rejected'. In a choice experiment with status quo alternatives, a modification of this decision rule might require the status quo alternative to be always in the trade-off condition. Even more complex hybrid rules that depend on alternative, person and/or choice context characteristics can also be considered in extensions of this model. For example, the majority of confirming dimensions heuristic might be invoked by allowing p_j to depend on the number of 'best' attributes that an alternative possesses.

Andersen *et al.* (2007) use a conceptually similar mixture model to represent dual latent decision processes in decision making under uncertainty. In their model, they assume that decision making by any individual is based on two criteria. One criterion is simply a form of expected utility maximisation. The other criterion is based on an aspiration-type process that incorporates a minimum income threshold. The weights of the criteria in decision making are *a priori* unknown and are therefore inferred in the estimation process. Their results indicate that respondents appear to assign a nearly two-third probability weight to the aspiration-type process, and assigning only a one-third weight to the expected utility process.

2.3.4 Direct Representations of Complexity

In a more direct approach of embedding complexity into discrete choice models, Swait and Adamowicz (2001) suggest that a formal relationship might exist between the error variance or conversely, scale in preferences, and entropy, which can be seen as a measure of complexity. Entropy is defined in Equation (2.14) with π_{js} denoting the probability of choosing alternative *j* in choice situation *s*. π_{js} is obtained by *a priori* estimating a basic MNL model:

$$H_s = -\sum_{j \in s} \pi_{js} \log(\pi_{js}) \tag{2.14}$$

As preferences among alternatives become more indistinguishable, π_{js} approaches 1/J and a high level of entropy (and complexity) is obtained. Entropy is related to scale by noting that at both low and high levels of entropy, the scale is high as decision-making is relatively easy in the former case and alternatives are all approximately similar in utility terms in the latter. At moderate levels of complexity however, more preference inconsistency (lower scale) may be evident as respondents resort to using simplifying heuristics. Hence, the scale of choice task *s*, μ_s, may be related to H_s through a quadratic form to account for these non-linear effects, as in Equation (2.15):

$$\mu_s = \exp(\theta_1 H_s + \theta_2 H_s^{\,2}) \tag{2.15}$$

According to the preceding hypothesis, it would be expected that $\theta_1 < 0$ and $\theta_2 > 0$. These theoretical predictions are confirmed in eight of the ten cases of consumer choice analysed by Swait and Adamowicz (2001). In each of these cases, which involve either revealed preference or stated preference data, the null hypothesis of a homoscedastic MNL assumption is strongly rejected at the five percent level. The estimated parameters θ_1 and θ_2 all have the expected sign, implying that scale is a function of entropy. However, Swait and Adamowicz (2001) do not test if entropy affects preferences directly.

By using contextual attributes to condition the preference and scale functions of a discrete choice model, Zhang and Adamowicz (2011) test if such conditioning can control for the choice format effect. Essentially, the choice format effect describes the different model outcomes, for example parameter estimates and WTP measures, that have been observed purely as a consequence of how the choice experiment has been structured, for example, whether the choice experiment has two alternatives (a status quo alternative plus one other experimentally designed alternative) or three alternatives (status quo plus two alternatives). Zhang and Adamowicz (2011) postulate that the emergence of a choice format effect may be a result of the model misspecification which arises due to the application of the standard RUM assumptions without regard to the psychology of choice decision making. They reason that if the choice effect format comes about as a result of preferences that are context dependent, then the choice format effect may disappear once cognitive components are incorporated into the standard RUM model.

In terms of the contextual variables that are tested in the model, Zhang and Adamowicz (2011) consider, besides entropy, the number of attributes whose levels differ across alternatives (*numdiff*), the dispersion of attribute levels within each alternative, and the dispersion of the standard deviation across alternatives. These are variables which can be used to capture the structure of information (DeShazo and Fermo, 2002; Boxall *et al.*, 2009). The order of a choice task (*order*) is also considered as another candidate context variable (Holmes and Boyle, 2005). Zhang and Adamowicz (2011) consider the following representations in Equations (2.11a to 2.11c) for the modelled component of utility, where $\mathbf{Z_s}$, which is the vector of choice task specific context variables, is allowed to condition preferences through interaction terms with the attributes of the alternatives in the choice task. In Equation (2.16a), $\boldsymbol{Z_s}$ enters the preference function alone.

$$V_{js}^{P} = \boldsymbol{\beta}\mathbf{X}_{js} + \boldsymbol{\gamma}\mathbf{X}_{\mathbf{js}} * \mathbf{Z}_{\mathbf{s}} \tag{2.16a}$$

If $\boldsymbol{Z_s}$ enters the scale function, then Equation (2.16b) follows.

$$V_{js}^{S} = \mu(\mathbf{Z}_{\mathbf{s}}) * \boldsymbol{\beta}\mathbf{X}_{js} \tag{2.16b}$$

Finally, if the context variables are assumed to enter both the preference and scale functions, then Equation (2.16c) is obtained.

$$V_{js}^{P,S} = \mu(\mathbf{Z_s})V_{js}^{P} \tag{2.16c}$$

From their dataset, Zhang and Adamowicz (2011) find that *numdiff*, entropy and *order* are the contextual effects that are most significant in the various model specifications, with *numdiff* appearing to have the strongest effect, followed by entropy and *order*. The best fit model is the random parameter logit model where these contextual effects are fully interacted with all main variables, with many of the interaction terms highly significant. Hence, they conclude that the choice format affects preferences and that it is necessary for contextual effects to at least enter the preference function in order to control for the effects of different choice formats on preference elicitation.

Zhang and Adamowicz (2011) also find that these contextual effects do not control for the choice format effect when entered into the scale function only. In particular, the lack of significance of the entropy variable in the scale function, contrary to what Swait and Adamowicz (2001) have found, leads Zhang and Adamowicz (2011) to postulate that entropy might be more suitable for capturing continuous relationships between the scale function and the level of choice complexity (such as the number of alternatives) in a more complex choice set rather than the situation they analyse, which involves a discrete change from only two alternatives to three alternatives.

2.3.5 Hierarchical Bayes Modelling

Hierarchical Bayes modelling has also proven useful when estimating heterogeneity in cases where the decision sequence, constraints and thresholds are latent. Gilbride and Allenby (2004) model the two stage decision processing strategy by assuming that screening rules exist to restrict a larger choice set into a smaller subset of alternatives for final evaluation. The screening rules considered include (i) a compensatory screening rule, where the deterministic portion of utility in the traditional compensatory sense must exceed a threshold; (ii) a conjunctive screening rule, where all attribute values must be acceptable and (iii) a disjunctive rule, where at least one attribute value needs to be acceptable. These thresholds are determined endogenously and are allowed to vary by respondents, not unlike the approach taken by Swait (2009) and Cantillo and Ortuzar (2005). In conclusion, Gilbride and Allenby (2004) find that the conjunctive screening model explains the data best. Further extensions of this work include modelling the EBA processing rule and an economic screening rule which stops all processing of an alternative whenever an undesirable attribute level is present (Gilbride and Allenby, 2006).

2.4 RELATIONAL HEURISTICS: PROSPECT THEORY

Another broad category of heuristics might be called "relational" heuristics. Unlike the models discussed earlier, choice task properties and complexity are not the focus of attention here. Instead, these heuristics emphasise the use of some reference point(s), or in some specific applications, a comparison of ratings of one alternative against another. Prospect theory (Kahneman and Tversky, 1979) is perhaps one of the most well-known behavioural models in this category. Some brief remarks about prospect theory (PT) follow, but for a more comprehensive treatment, see Van de Kaa (2010a; 2010b) and Li and Hensher (2011).

Essentially, prospect theory assumes that respondents frame alternatives and attributes relative to a reference state, that respondents are loss averse, and that there is diminishing sensitivity to gains and losses. If evaluating outcomes which are uncertain, laboratory experiments have shown that relative to certain outcomes, outcomes with low probabilities are overweighted and outcomes with high probabilities are underweighted. Hence, PT assumes that each individual evaluates the expected probabilities according to an inverse S-shaped weighted probability function.

Despite numerous examples demonstrating that an individual's reference state is updated in successive choice settings, the original version of PT as proposed by Kahneman and Tversky (1979) does not discuss this issue. Neither is PT concerned with heterogeneity in choice behaviour strategies. To address these gaps in the theory, Van de Kaa (2010a) proposes an extension of prospect theory, aptly named extended prospect theory (EPT), suggesting that "attributes of alternatives are framed as context-dependent changes (gains and losses) relative to an updated reference state, and most individuals value losses higher than gains of equivalent size." (Van de Kaa, 2010a, p. 773)

A review of empirical evidence in the transportation literature in support of PT/EPT was undertaken by Van de Kaa (2010a), whose meta-analysis finds strong support for the key PT/EPT assumption of reference-dependent, loss-aversive choice behaviour in close to 100 studies reviewed. In these studies, the evidence shows that loss-neutral utility maximisation appears to be more of an exception rather than a rule. As for the other assumptions of PT/EPT, some studies also demonstrate that respondents display diminishing sensitivity to gains and losses and that PT's inverse S-shaped weighted probability function offers a better descriptive ability than the expected utility concept.

In a series of experiments inspired by Kahneman and Tversky (1979), Avineri and Prashker (2004) show that traveller behaviour in the light of travel time uncertainties violate the assumptions of expected utility theory. For example, respondents are found to make decisions according to the Allais paradox. The Allais paradox represents a violation of one of the axioms of expected utility theory as it has been shown empirically that the addition of an independent event influences choice behaviour. The following well-known example of the Allais Paradox in Table 2.1 is due to Kahneman and Tversky (1979):

Table 2.1: Illustration of the Allais Paradox (Kahneman and Tversky, 1979)

Problem I: choose between	A	B
	33% chance of winning 2,500 66% chance of winning 2,400 1% chance of winning nothing	2,400 for sure
Problem II: choose between	**C**	**D**
	33% chance of winning 2,500 67% chance of winning nothing	34% chance of winning 2,400 66% chance of winning nothing

Kahneman and Tversky (1979) find that a majority of respondents when confronted with Problem I choose B over A. In Problem II however, the majority would choose C over D. This finding violates expected utility theory since the only difference in going from Problem I to Problem II is the elimination of a 66 percent chance of winning 2,400 across both alternatives.

There is also evidence suggesting that small probabilities are inflated. In another experiment, Avineri and Prashker (2004) show that when there are substantial probabilities of a gain, as measured by travel time savings, people prefer the option with larger probabilities, but when the probabilities of travel time savings are very close to zero, preferences are switched and the option with the larger time savings is preferred. In summary, the evidence gathered by Avineri and Prashker (2004) seems to point to respondents as possibly prospect maximisers instead of utility maximisers. Likewise, consistent with prospect theory, Senbil and Kitamura (2004) observe reference dependency in the decision frames relating to departure time decisions. At the same time, Hess *et al.* (2008) demonstrate, in the context of stated choice data, the existence of referencing around a recently experienced alternative and that preference formation may better relate to differences in respondent specific reference points, rather than the absolute values of the attributes shown in stated choice experiments.

On a more cautionary note however, Timmermans (2010) and Avineri and Prashker (2004) raise some questions regarding the application of prospect theory to travel behaviour under uncertainty, especially in dynamic situations where there is feedback. In particular, Avineri and Prashker (2004) highlight that their results are obtained in the context of one-shot experiments and do not explicitly capture learning and day-to-day effects.

A review of the evidence suggests that PT and its variants hold a significant amount of promise as an alternative theory of traveller behaviour. However, until further research convincingly demonstrates its applicability in what is a fairly complex environment of decision making - multi-day, multi-choice situations where dynamic feedback and learning is highly possible – the literature suggests that PT should be embraced more cautiously in travel applications.

2.5 EXTREMENESS AVERSION AND REGRET MINIMISATION

2.5.1 Empirical Demonstrations of the Extremeness Aversion Effect

More specific examples of relational heuristics abound. The discussion in this section focuses on models of extremeness aversion (Simonson and Tversky, 1992; Tversky and Simonson, 1993) and regret minimisation (Chorus *et al.*, 2008; Chorus, 2010; Chorus, 2012).

The extremeness aversion heuristic has been put forward as a possible explanation for the fairly robust empirical findings of the so-called "compromise effect" (Simonson, 1989). The compromise effect states that respondents prefer an in-between alternative when extreme alternatives are available in the choice set. In this particular context, extreme alternatives are defined as those which perform best on some attributes, but worst on others. Loss aversion may explain the compromise effect; in that the disadvantages of an alternative (defined relative to the other alternatives in the choice set) are weighted more heavily than its advantages (Tversky and Kahneman, 1991; Simonson and Tversky, 1992). Hence, the in-between alternative, with its smaller advantages and disadvantages, is more highly favoured compared to the extreme alternative, which has larger advantages, but also larger disadvantages. Louviere and Myer (2008) also show that this effect can arise from preference uncertainty among risk averse individuals. It might also be pointed out that much of the early literature on the compromise effect focuses on analysing alternatives with only two attributes; hence, with the notable exception of the random regret minimisation model reviewed later, many of the examples cited here relate to the case of two attributes.

Under some circumstances, extremeness aversion may violate the principle of regularity (Tversky and Simonson, 1993). Regularity states that the market share of an alternative cannot be increased by enlarging the choice set. Formally, for any item that is a part of choice set A where A is, in turn, a subset of B, the probability of choosing an element j from A must not be less than choosing j from B, i.e., for all $j \in A \subseteq B, Pr(j;A) \geq Pr(j;B)$. Regularity is a very weak condition and violations are generally not expected (Luce, 1977). It is the minimum condition for most existing choice models (Huber *et al.*, 1982) and it is implied by the standard linear additive representation of utility. However, through a series of experiments, such as the one reported in Table 2.2, Simonson and Tversky (1992) demonstrate the existence of the extremeness aversion heuristic in several instances of decision making, which leads to unexpected violations of regularity[1].

[1] The addition of asymmetrically dominated alternatives to a choice set can also result in violations of regularity (Huber *et al.*, 1982).

Table 2.2: Results from a Camera Purchase Experiment (Simonson and Tversky, 1992)

Category: 35mm camera		Share (%)	
Quality of Camera	Price ($)	Version 1	Version 2
Alt 1: Low-end	169.99	50	22
Alt 2: Middle range	239.99	50	57
Alt 3: Top quality	469.99	-	21

In the first version of the experiment, Simonson and Tversky (1992) asked respondents to choose between two alternatives: a low-end camera (*Alt 1*) and a middle-of-the-range camera (*Alt 2*). In this experiment, both low-end and middle-of-the-range alternatives proved to be equally popular. When an extreme alternative (*Alt 3*), a top-of-the-range camera, was added to the choice set (Version 2 of the experiment), the middle alternative *Alt 2* became more popular relative to the low-end camera *Alt 1*, in both relative and absolute terms. Simonson and Tversky (1992) found that the difference in the market shares of *Alt 2* relative to *Alt 1* between Version 2 and Version 1 of the experiment was statistically significant at the five percent level, allowing them to conclude that regularity had been violated.

Extremeness aversion also violates a betweenness inequality which is a stronger condition than regularity (Tversky and Simonson, 1993). To understand the betweenness inequality, first define *Alt 2* to be between *Alt 1* and *Alt 3* if for all attributes *k*, either $x_{1,k} \le x_{2,k} \le x_{3,k}$ or $x_{1,k} \ge x_{2,k} \ge x_{3,k}$ is true. Under a ranking condition which occurs with empirical regularity (Tversky and Simonson, 1993), if *Alt 2* lies between *Alt 1* and *Alt 3*, then the ranking *Alt 3 > Alt 2 > Alt 1* should be relatively more common than the ranking *Alt 3 > Alt 1 > Alt 2*. Putting the ranking condition together with the value maximisation condition according to classical utility theory, the betweenness inequality essentially states that the middle alternative *Alt 2* should lose relatively more than the existing extreme alternative *Alt 1* when an extreme alternative such as *Alt 3* is introduced into the choice set. Intuitively, this is because *Alt 3* is more similar to *Alt 2* than to *Alt* 1.

To see the violation of the betweenness inequality in Table 2.2, observe that in Version 1 of the experiment, the choice share of *Alt 2* relative to *Alt 1* is equal to $\frac{0.5}{0.5+0.5} = 0.5$, while in

Version 2 of the experiment, the choice share of *Alt 2* relative to *Alt 1* is equal to

$\frac{0.57}{0.57+0.22}=0.72$. Hence, the introduction of *Alt 3* has taken relatively more from *Alt 1* than from *Alt 2*.

The compromise effect is a symmetric form of extremeness aversion, occurring when disadvantages loom larger than advantages on all attributes. The in-between alternative becomes the compromise, and its share is enhanced relative to both extreme alternatives. A non-symmetric form of extremeness aversion, known as polarisation, may occur when the introduction of an in-between alternative substantially reduces the relative share of one of the extreme alternatives but makes the other extreme alternative even more favoured. This happens when disadvantages loom larger than advantages on only some of the attributes of the alternatives, but not on others. To illustrate polarisation, another example from Simonson and Tversky (1992) would be helpful. This example is reported in Table 2.3.

In Choice Set 4, the relative choice share of *Alt 2* with respect to *Alt 3* in the presence of *Alt 1* is $\frac{0.48}{0.48+0.43}=0.53$. In Choice Set 2, the relative choice share of *Alt 2* with respect to *Alt 3* in the absence of *Alt 1* is 0.4. In this case, Simonson and Tversky (1992) find that the increase

Table 2.3: Results from a Stated Choice Purchase of a Cassette Player (Simonson and Tversky, 1992)

Option (AM/FM cassette player)		Share (%)			
Brand	Price ($)	Choice Set 1	Choice Set 2	Choice Set 3	Choice Set 4
Alt 1: Low quality Emerson	39.99	45	-	51	9
Alt 2: Mid quality Sony	69.99	55	40	-	48
Alt 3: Top quality Sony	149.99	-	60	49	43

in *Alt 2*'s relative choice share in Choice Set 4 compared to Choice Set 2 ($0.53-0.4=0.13$) is not statistically significant, indicating that the inclusion of *Alt 1* into the choice set does not lead to a noticeable extremeness aversion effect away from *Alt 3* in favour of the in-between alternative *Alt 2*.

However, Simonson and Tversky (1992) found that if the market share of *Alt 2* relative to *Alt 1* in Choice Set 1 is compared to its counterpart in Choice Set 4, there is a statistically significant increase of $\frac{0.48}{0.48+0.09} - 0.55 = 0.29$. This result suggests that the inclusion of extreme alternative *Alt 3* leads to extremeness aversion away from *Alt 1* towards *Alt 2*. This example illustrates the point that extremeness aversion works only in one direction, which is an aversion away from *Alt 1*, rather than an aversion away from *Alt 3*. Moreover, a comparison between Choice Sets 3 and 4 shows that the inclusion of the middle alternative *Alt 2* disproportionately shifts market share away from *Alt 1* towards *Alt 3*. Simonson and Tversky (1992) conclude from these results that there is a polarisation effect in favour of quality.

It would be interesting to speculate on the differences in circumstances that might lead to the emergence of the compromise effect and the polarisation effect, but empirical observations aside, there does not seem to be much further discussion in the literature on why the compromise effect is observed in some cases and the polarisation effect in others. This may in fact be a potential area of upstream research with interesting implications for the modelling techniques used in discrete choice modelling.

2.5.2 Models of Extremeness Aversion and Regret Minimisation

The random regret minimisation (RRM) model, first proposed in Chorus *et al.* (2008) and subsequently refined in Chorus (2010) has been shown to be able to accommodate the extremeness aversion heuristic, in particular, the compromise effect. In the RRM model, regret is said to occur when a non-chosen alternative leads to a more desirable outcome, for example, when a foregone alternative performs better on a certain attribute compared to the chosen alternative. This notion of regret should be distinguished from the early representation of heuristics encapsulated by regret theory (Bell, 1982; Loomes and Sugden, 1982), which focuses on the study of risky choices in single attribute (usually monetary) alternatives. The RRM model has been primarily developed for the analyses of riskless multi-attribute choice such as those commonly found in travel demand modelling, although it is possible for the RRM to be extended to the study of risky choices as well (Chorus, 2012).

Under the RRM model, respondents are assumed to engage in regret avoidance behaviour by choosing the alternative which minimises regret. The systematic regret for any considered alternative j, denoted $Reg(j)$, is the sum of all binary regrets associated with the bilateral comparisons of alternative j over all non-considered alternatives j' in choice set s. Specifically, the binary regret function $Reg(j,j')$ may be written in the form of Equation (2.17a), with the subscript k denoting the k^{th} attribute of the alternative, X_{jk} denoting the level of attribute k in alternative j and β_{jk} denoting the taste parameter associated with attribute k of alternative j:

$$Reg(j,j') = \sum_k \ln\left(1+\exp\left[\beta_{j'k}X_{j'k} - \beta_{jk}X_{jk}\right]\right) \tag{2.17a}$$

The systematic regret function $Reg(j)$ is given by Equation (2.17b):

$$Reg(j) = \sum_{\substack{j' \neq j,\\ j' \in s}} Reg(j,j') \tag{2.17b}$$

In the limit, as the regret term $[\beta_{j'k}X_{j'k} - \beta_{jk}X_{jk}]$ becomes sufficiently negative, $Reg(j,j')$ with respect to attribute k falls towards zero. Likewise, if $[\beta_{j'k}X_{j'k} - \beta_{jk}X_{jk}]$ becomes sufficiently large, $Reg(j,j')$ with respect to attribute k approaches $[\beta_{j'k}X_{j'k} - \beta_{jk}X_{jk}]$. As illustrated by Chorus (2010) and reproduced here, a graph of the function $y=\ln(1+\exp(x))$ is provided in Figure 2.1. It is readily apparent that the regret function is a convex function.

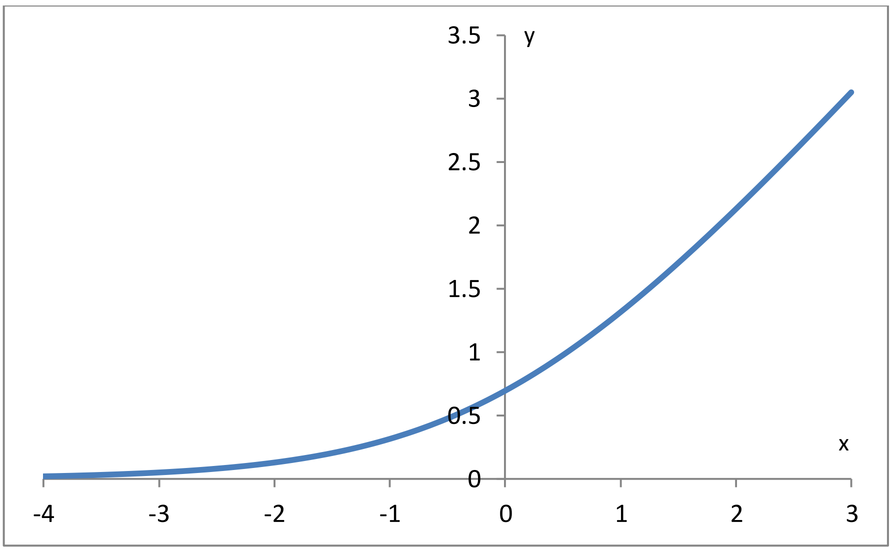


Figure 2.1: Graph of $y=\ln(1+\exp(x))$

As the regret function is convex, the extent to which an improvement in an attribute can compensate for the deterioration in the value of another attribute depends very much on its value relative to the other alternatives. For example, an improvement in an attribute that is already far superior to its counterparts in the other alternatives leads to a minimal reduction in regret, while a worsening of another attribute which was comparing poorly to begin with can lead to a substantial increase in regret. This implies that the RRM model is able to capture semi-compensatory behaviour (Chorus, 2010). Semi-compensatory behaviour occurs when a disproportionately large improvement in an attribute is required to offset a given decline in the performance of another attribute. If the decline in the performance of the latter attribute (that is, an increase in regret) is large enough, no amount of improvement in the first attribute will compensate sufficiently.

It is the convexity/semi-compensatory property of the RRM model that allows it to also account for the compromise effect (Chorus, 2010). Put another way, when an attribute of an alternative performs well relative to other alternatives in the choice set, such as in the case of an extreme alternative, an improvement generates only a small decrease in regret; whereas on another attribute in which the extreme alternative fares poorly, a small amount of deterioration in that attribute will generate a large amount of regret. Consequently, in the RRM model, the 'in-between' alternative is favoured over the extreme alternative.

Other properties of the RRM model are also worth highlighting. For example, it might also be observed that the RRM model is just as parsimonious as the standard RUM model, unlike other models of contextual effects which typically require the estimation of additional parameters (Chorus, 2010). As attributes of the non-considered alternatives enter into the regret function, the RRM model also does not display the property of independence from irrelevant alternatives (IIA), even if independent and identically distributed (*i.i.d.*) error terms are assumed for the unobserved part of the regret function. As for parameter interpretation, Chorus (2012) observes that the parameters "reflect the *upper bound* of the extent to which a unit increase or decrease in *relative* performance on an attribute influences the level of regret that is associated with a comparison with another alternative." (Chorus, 2012, p. 79, emphasis his)

Despite these desirable properties of the model, empirical support for the RRM, in terms of goodness of fit measures, appears to be mixed, at least as far as stated preference data are concerned (Chorus, 2012). This may be because the level of anticipated regret engendered in stated choice data is less prescient than in revealed preference data which reflect real life choices and trade-offs. However, even where the RRM model does not fit the data as well as the standard linear additive RUM model, the context dependency engendered in the RRM model means that there are different implications for marginal willingness to pay measures and mean elasticities (Hensher *et al.*, 2011). From a behavioural perspective, it may be more desirable to consider the outputs from the RRM model if it is believed that context dependency, regret and referencing matter in decision making.

Embedding the compromise effect into discrete choice models is also possible through the proposed specifications offered by Kivetz *et al.* (2004) in the marketing literature. Their contextual models incorporate the use of reference points and also account for either loss aversion or concavity in gains in the context of comparisons against competing alternatives in the current choice set. Like the RRM model, instead of a reference alternative, reference attribute levels are used. In their loss aversion model, the reference point is taken to be the mid-point of the attribute range of the alternatives in the local choice set *s*, which is not necessarily equal to the attribute levels in the existing status-quo choice option. The value function for the loss aversion model is defined in Equation (2.18) as:

$$\begin{aligned} V_j = & \sum_k \left[v_{jk}(X_{jk}) - v_{rk}(X_{rk}) \right] \times 1\left(v_{jk}(X_{jk}) \geq v_{rk}(X_{rk}) \right) \\ & + \sum_k \lambda_k \left[v_{jk}(X_{jk}) - v_{rk}(X_{rk}) \right] \times 1\left(v_{jk}(X_{jk}) < v_{rk}(X_{rk}) \right) \end{aligned} \tag{2.18}$$

In Equation (2.18), V_j is the value of alternative *j* (given a choice set *s*), $v_{jk}(X_{jk})$ is the utility of attribute *k* of alternative *j*, λ_k is the loss aversion parameter for attribute *k* and X_{rk} indicates the reference value of attribute *k* in choice set *s*. If respondents display loss aversive tendencies, λ_k will be greater than one.

On the other hand, Kivetz *et al.*'s (2004) contextual concavity model takes the attribute value with the lowest part-utility as the reference point and codes the utility of other attribute values as gains against the reference. This model specification is shown in Equation (2.19):

$$V_j = \sum_k \left(v_{jk}(X_{jk}) - v_{rk}(X_{rk}) \right)^{c_k} \tag{2.19}$$

As its name suggests, the contextual concavity model assumes that the utility gains are concave relative to the reference, as an outcome predicted by prospect theory. Hence, c_k is introduced as a concavity parameter for attribute *k* and would typically take a value between zero and one. X_{rk} in this case is the attribute value that gives the lowest utility on attribute *k* across all alternatives in the choice set. The reference is context specific and will change from choice task to choice task. More generally, the concavity parameter implies diminishing marginal sensitivity to gains; thus, the in-between alternative with its moderate gains on the attributes benefits more compared to the extreme alternatives.

As an application to the transportation literature, Chorus and Bierlaire (2013) compare the contextual concavity model with the RRM model in a stated route choice experiment and find that in terms of model fit, the contextual concavity model has a "slight (but statistically significant) edge" over the RRM model, even after correcting for the additional number of parameters (Chorus and Bierlaire, 2013, p. 561). They also find that the concavity parameters associated with the attributes of travel time and travel time variability are not statistically different from one at the usual levels of significance (implying absence of concavity), but the associated concavity parameters for travel cost and percentage of travel time spent in congestion are statistically less than one. In terms of predictive ability for out of sample forecasts, Chorus and Bierlaire (2013) find that the differences between the RRM and the contextual concavity model are small. There is no clear 'winner' as such, since the RRM performs better on certain metrics and the contextual concavity model is better on others. Chorus and Bierlaire (2013) also explore another more explicit way of postulating that respondents are focusing on a compromise alternative by constructing a compromise variable to measure the 'in-betweenness' score of an alternative which is then entered into a conventional linear-in-the-parameters logit model. However, they find that their empirical results favour either the contextual concavity model or the RRM over the linear + compromise model.

Another model, attributable first to Tversky and Simonson (1993) as a componential context model, is another candidate for explaining the compromise effect. This model was later empirically estimated as a relative advantage model by Kivetz *et al.* (2004). The componential context model or relative advantage model is shown in Equation (2.20):

$$V_j = \sum_k v_k(X_{jk}) + \lambda \sum_{j' \in s} RA(j, j') \tag{2.20}$$

In Equation (2.20), $RA(j, j')$ denotes the relative advantage of alternative j over alternative j', and λ is the weight given to the relative advantage component of the model. $RA(j, j')$ may be defined, according to Tversky and Simonson (1993), as follows: First, for a pair of alternatives (j, j') in choice set s, consider the advantage of j over j' with respect to an attribute k, denoted in Equation (2.21) by:

$$A_k(j, j') = \begin{cases} v_k(X_{jk}) - v_k(X_{j'k}) \text{ if } v_k(X_{jk}) \geq v_k(X_{j'k}), \\ 0 \text{ otherwise.} \end{cases} \tag{2.21}$$

Define the disadvantage of j over j' with respect to an attribute k as an increasing convex function $\delta_k(.)$ of the corresponding advantage function $A_k(j', j)$, that is, $D_k(j, j') = \delta_k(A_k(j', j))$. $\delta_k(.)$ is assumed to be convex as a consequence of loss aversion in evaluating a disadvantage. A functional form for $D_k(j, j')$, due to Kivetz *et al.* (2004), is suggested in Equation (2.22):

$$D_k(j, j') = A_k(j', j) + L_k A_k(j', j)^{\psi_k} \tag{2.22}$$

Here, L_k is a loss aversion parameter (*a priori* expected to be greater than zero) and ψ_k is a power parameter (*a priori* expected to be greater than one). The relative advantage of j over j' is then defined in Equation (2.23) as:

$$RA(j, j') = \frac{\sum_k A_k(j, j')}{\sum_k A_k(j, j') + \sum_k D_k(j, j')} = \frac{A(j, j')}{A(j, j') + D(j, j')} \tag{2.23}$$

Like the RRM model, the relative advantage model assumes a bilateral comparison of each alternative against all other alternatives in the choice set. Moreover, since the componential context/relative advantage model was primarily motivated by the need to explain the compromise effect, Tversky and Simonson (1993) supposed that $RA(j, j') = 0$ if choice set s contains two or less elements, as by definition, the compromise effect cannot occur in such a choice situation. Empirically however, this assumption is unnecessarily restrictive. Chapter 5 discusses and estimates a relative advantage model on binary choice data.

Testing the various model forms for the compromise effect, Kivetz *et al.* (2004) find that the parameter estimates for all models are consistent with *a priori* expectations from theory. However, the contextual concavity model and the loss aversion model have superior measures of fit and predictive validity compared to the relative advantage model, with the latter performing no better than the standard model in a handful of cases. The poorer performance of the relative advantage model may be a consequence of the way the model was estimated. Since the model is highly non-linear, Kivetz *et al.* (2004) resorted to *a priori* imposing, without further testing, the restriction of a common ψ_k across all k attributes of the alternatives and then using a grid search to find the optimal value for ψ [2]. It is also noteworthy that Kivetz *et al.* (2004) reported that in all their empirical applications, none of the loss aversion parameters L_k turned out to be statistically different from zero.

Although less widely documented than extremeness aversion, the opposite effect, known as extremeness seeking, may also be true, especially when choice sets are "non-alignable" (Gourville and Soman, 2007). A non-alignable choice set entails alternatives "that vary along discrete, non-compensatory attributes, such that one alternative may possess one set of desirable features, while a second alternative may possess a different set of desirable features" (Gourville and Soman, 2007, p. 10). For example, in having to choose among

[2] In their empirical applications, the optimal value of ψ is reported to be either 5 or 10. It is not clearly stated in the paper how the grid search was conducted, but one suspects that given these reported optimal values for ψ, the search was only conducted over integer values of ψ.

multiple car models, with say one model having rear seat DVD entertainment but no sun roof, and another model having the sun roof, but no rear seat entertainment, the trade-off across attributes is discrete, such that by choosing one alternative, the desirable features of another may have to be given up completely. In such cases, Gourville and Soman (2007) found that respondents displayed an increased tendency to either of the extreme alternatives. Hence, Gourville and Soman (2007) proposed that extremeness aversion occurs under more specific circumstances when the attributes are alignable, i.e., when attributes can be traded off incrementally. For example, a choice involving a low-priced, low processing speed computer model and a medium-priced, medium processing speed model is alignable, and the introduction of an extreme high priced, high processing speed option causes the market share of the intermediate option to go up.

If the hypothesis suggested by Gourville and Soman (2007) is valid, then extremeness seeking behaviour should be taken into consideration when modelling choice data where alternatives are described by discrete attributes. In the more likely case where alternatives have a mix of discrete and continuous attributes, a mixture of extremeness seeking and extremeness aversion may even be possible. To the best of this author's knowledge, while the compromise effect has been modelled for continuous attributes, the case of discrete attributes has not received any empirical attention in the discrete choice literature.

2.6 CONTEXT-FREE AND CONTEXT DEPENDENT PREFERENCES

Arising from the utility specification found in the relative advantage model, Simonson (2008) highlights a distinction between "inherent" and "constructed" preferences. He argues that there may be a more important role for stable or inherent preference values to play in shaping decisions than has been recognised in the behavioural decision literature, which has mainly been focused on demonstrating that preferences are largely constructed. Such inherent preferences or dispositions (to like or dislike) may even exist for objects that people have not yet experienced. The relative advantage model described in Equation (2.20) thus represents the notion that preferences are separable into inherent and constructed components (Kivetz *et al.*, 2008). The first term on the right hand side of Equation (2.20) represents value maximisation independent of contextual effects while the second term is context dependent utility. The parameter λ can be taken as an indication of the strength of the choice context in determining preferences and represents the weight given to constructed preference vis-à-vis inherent preferences. Additionally, as in explored in Chapter 5, the parameter λ may be expressed as a function of respondent and choice task characteristics, allowing the weight of constructed preferences to vary across respondents and choice tasks.

Rooderkerk *et al.*'s (2011) model essentially follows this line of thinking by decomposing utility into additively separable context-free and context dependent components. Their key innovation is to consider a contextual model of utility where the context dependent component is written as a linear combination of three contextual effects that have been well-documented in the marketing and psychology literature: the compromise effect, discussed in Section 2.5, the attraction effect (Huber *et al.*, 1982), and the similarity effect (Tversky, 1972). Both the attraction and similarity effects are described below.

The attraction effect occurs when an asymmetrically dominated alternative is included in the choice set. An asymmetrically dominated alternative is one which is dominated by one alternative in the choice set, but not by others. The inclusion of such a decoy alternative increases the relative market share of the dominating alternative, leading to a violation of the IIA assumption. Table 2.4 below, obtained from Huber *et al.* (1982), provides an illustration of the attraction effect. What Huber *et al.* (1982) found was that in a binary choice between Beer *A* and Beer *B*, 43 percent of respondents chose Beer *A* when the decoy was not present. However, when the asymmetrically dominated alternative (the decoy) was added to the choice set, the market share of Beer *A* increased to 63 percent, violating the regularity condition.

Table 2.4: An Example of the Attraction Effect (Huber *et al.*, 1982)

Six-Pack Beer	Beer A	Beer B	Decoy
Price ($)	2.60	1.80	3.00
Quality rating	70	50	70
Market share of Beer A (binary choice, no decoy present)	43%		
Market share of Beer A (with decoy present)	63%		

The similarity effect states that the inclusion of an alternative takes away more market share from a similar alternative than from a dissimilar alternative. The following example presented in Table 2.5 is taken from Tversky (1972).

Table 2.5: An Example of the Similarity Effect (Tversky, 1972)

Potential college applicants	A	B	C
Intelligence score	78	60	75
Motivation score	25	90	35
Share of A relative to B in binary choice {A,B}	56%		
Share of A relative to B in triple {A,B,C}	46%		
Share of C relative to B in binary choice {B,C}		47%	
Share of C relative to B in triple {A,B,C}	39%		

In this experiment, respondents were asked to choose the most promising college applicant among profiles with different intelligence and motivation scores. The alternatives were designed so that applicants A and C were more similar to each other than either of them was to applicant B. Tversky (1972) found that the inclusion of alternative C into the choice pair {A, B} led to a fall in the popularity of A relative to B, from 56 percent to 46 percent. Likewise, the inclusion of A into the choice pair {B, C} led to the similar alternative C losing relatively more than the dissimilar alternative B. The difference in market shares of A and C between the three-alternative choice set and the two-alternative choice set was statistically significant at the ten percent level.

Returning back to Rooderkerk *et al.* (2011)'s model, the impact of these contextual effects are expressed as direct utility gains or losses in the value function. The compromise, attraction and similarity effects are modelled using measures of Euclidean distance between attributes. In particular, the compromise effect is modelled using a different perspective from the RRM model of Chorus (2010), the compromise variable of Chorus and Bierlaire (2013) and the models of Kivetz *et al.* (2004). For Rooderkerk *et al.* (2011), a compromise alternative, which may be a virtual alternative that does not exist in the choice set, is first constructed. The attribute levels of such an alternative are defined by the mid-points of the range of attribute values in the choice set, as in Equation (2.24):

$$x_{ck} = \frac{x_{\min,k} + x_{\max,k}}{2} \tag{2.24}$$

Instead of using a non-linear specification for the compromise effect (Chorus, 2010; Kivetz *et al.*, 2004), Rooderkerk *et al.* (2011) measure the strength of the compromise effect by a linear representation of the distance between alternatives *j* and the (possibly virtual) alternative *c*. The hypothesis advocated by Rooderkerk *et al.* (2011) is that the closer the attribute values of alternative *j* are to the compromise alternative *c*, the more likely *j* will benefit from being like the compromise and the larger its choice share. As for the attraction effect, it is modelled by assuming that it is only present when there are pairs of dominating/dominated alternatives in the choice set. Whenever an alternative is dominating, the attraction effect, which makes the dominating alternative more attractive in the presence of a dominated alternative, becomes stronger as the distance between it and the dominated alternative increases, with the attraction effect adding to utility; while the same attraction effect reduces utility for the dominated alternative. Finally, the similarity effect is measured by the distance between an alternative and its closest neighbour. The higher this distance becomes, the more dissimilar is the alternative to others in the choice set and this should add to utility. As an extension to the model, an interaction between the similarity and attraction variables is also included.

Compared to a RUM model without contextual effects and also to the main effects contextual model without any interaction terms, Rooderkerk *et al.* (2011) find that the extended contextual model (which includes the interaction term) leads to an improvement in most of the descriptive and prescriptive fit statistics. In the extended model, all contextual parameters are statistically significant and of the correct sign. In conclusion, they suggest that context effects can be quite easily embedded into commonly used choice models by treating them as extra components in the utility function. This approach sits well with the MCD and reference revision heuristics examined in this thesis.

2.7 HEURISTICS OF CHOICE SET INTER-DEPENDENCE: VALUE LEARNING AND STRATEGIC MISREPRESENTATION

The notion of "relational" can be extended to allow preceding choice tasks or choice outcomes to impact current choice. As noted by Simonson and Tversky (1992), "in deciding whether or not to select a particular option, people commonly compare it to other alternatives that are currently available as well as with relevant alternatives that have been encountered in the past" (Simonson and Tversky, 1992, p. 282). This notion is not entirely dissimilar to case-based decision theory advocated by Gilboa and Schmeidler (1995), where respondents are thought to recollect similar choice problems and the associated choice outcomes encountered previously. As most stated choice experiments require respondents to answer a series of choice tasks, the implication arising from Simonson and Tversky (1992) and Gilboa and Schmeidler (1995) is that preferences over attributes are not necessarily independent across choice sets.

There is now a significant body of evidence indicating that what respondents encounter in previous choice sets matters in the current decision making context. "Ordering anomalies", where choice is biased by the sequence of attribute values observed in the preceding choice set(s), are not uncommon, according to Day and Prades (2010). For example, if a price attribute of one alternative is seen to increase from one choice set to another, the proportion of respondents choosing that alternative in the second choice set is smaller than if the choice sets are reversed. A proposed explanation for this observation may be found in a 'good deal / bad deal' heuristic (Bateman *et al.*, 2008) whereby 'good deals' in the current choice task, relative to the price-attribute combinations encountered in previous choice tasks, are chosen more frequently than relative 'bad deals'. A trade-off contrast (Simonson and Tversky, 1992) by which current preferences are revised on the basis of previous price or cost attributes may also explain ordering effects. Such ordering anomalies may be considered to be specific examples of a more general phenomenon of preference reversal (Tversky *et al.*, 1990).

Strategic misrepresentation has also been invoked as one justification for incorporating the attributes of some previously chosen non status-quo alternative as a reference point in the current choice set. The argument from a public goods provision context is that people aim to increase the likelihood of their most preferred alternative being implemented by deliberately withholding the truth about their preferences in the current choice task if chosen alternatives in previous choice tasks have better attribute values (such as lower cost) than those in the current choice task. Strategic misrepresentation assumes that the respondents have stable and well formed preferences, but that a discrepancy exists between stated preferences and underlying true preferences. A weaker version of strategic misrepresentation allows respondents to consider the likelihood that the good would not be provided if they do not reveal their true preferences, and hence to only reject truth-telling probabilistically (McNair *et al.*, 2012). Other papers have also concluded that the data are consistent with respondents using previously encountered information and past choices and that estimated parameters are sensitive to how choice sets are ordered. Strategic behaviour in respondents, particularly the weaker version of strategic behaviour, is not rejected (Scheufele and Bennett 2010; 2013; McNair *et al.*, 2011).

Another explanation for considering features of previously encountered choice sets into the current choice set involves a value learning heuristic, which assumes truth telling, but poorly formed initial preferences. Value learning involves the discovery of preferences and taste parameters may change according to the attribute levels presented to the respondent. Hence, preferences can be influenced by the starting point and subsequent attribute values (McNair *et al.*, 2011), with the 'good-deal / bad-deal' heuristic being a specific case in point. The 'good-deal/bad-deal' heuristic posits that an alternative is less (more) likely to be chosen if its attributes are relatively inferior (superior) to those of the preceding alternatives.

McNair *et al.* (2011) show that responses to a sequence of binary choice tasks involving the provision of an underground electricity network are consistent with both a weak form of strategic misrepresentation and with a 'good deal/bad deal' heuristic. The relationship between cost sensitivity and the positioning of the cost level relative to the levels presented in previous choice tasks may be ascertained by interacting the cost variable with variables that indicate the relative cost position. In any given choice task, the cost level presented must satisfy one of the following four conditions:

1. both the minimum and the maximum level presented in the sequence to that point;
2. the minimum, but not the maximum level presented in the sequence to that point;
3. the maximum, but not the minimum level presented in the sequence to that point ; or
4. neither the minimum nor the maximum level presented in the sequence to that point.

Using these interaction terms, McNair *et al.* (2011) reject the standard null hypothesis that the relative cost position in the choice set does not matter. They find that the cost sensitivity is closest to zero and WTP the highest when the cost level is the minimum, but not the maximum level presented up to that point. Conversely, cost sensitivity is highest (and WTP lowest) when cost is the maximum, but not the minimum level presented in the sequence to that point. These results imply that an order with an increasing sequence of cost levels will underestimate WTP, while an order with a declining sequence of cost levels might overestimate WTP. With this perspective in mind, it would be prudent to account for such ordering effects in the data, whether by means of experiment design or by modelling or both.

Decision process heterogeneity is another dimension that can or should be explicitly considered as well. It is possible that in a group of respondents, several heuristics are at work and no one heuristic dominates. One way of testing this hypothesis is to use equality constrained models of probabilistic decision processes (PDP), which are essentially latent class models of decision rules, to represent heterogeneity in decision behaviour across respondents. For example, strategic misrepresentation and value learning can be modelled as distinct classes of heuristics for sub-groups of respondents, in addition to the standard assumption about utility (McNair *et al.*, 2012). In that paper, the authors particularly emphasised the role of cost levels in the value learning heuristic as cost levels are thought to be a significant factor in how the value learning process shapes preferences. The impact of cost levels might be especially important in modelling stated choice data in which similar goods are offered at very different prices over the course of a sequence of surveys. For the value learning heuristic, utility functions were specified to allow the alternative-specific preference to vary with the average of cost levels observed in the sequence up to and including the current choice task. Therefore, while the status quo alternative was specified in the usual way, the utility functions of the hypothetical/experimentally designed alternatives included a variable that measures the average cost level observed up to that point in the experiment (Equation (2.25)):

$$\begin{aligned} V_{SQ} &= \sum_k \beta_k X_{SQ,k} \\ V_{alt} &= \sum_k \beta_k X_{alt,k} + \beta_{k+1}(z^0 - \breve{z}) \end{aligned} \tag{2.25}$$

In Equation (2.25), z^0 is the average of the cost levels observed up to and including the current choice task. $\breve{z}$ is the average of cost levels in the whole sample (averaged across all choice tasks and all respondents). The purpose of $\breve{z}$ is to normalise the average observed cost variable so that its sample mean is approximately zero.

To model the strategic misrepresentation heuristic, McNair *et al.* (2012) specify utility functions which have the following two features. The first feature requires respondents to compare alternatives to those accepted in previous choice tasks. Defining a reference alternative as the highest cost alternative that was previously accepted in the sequence, the strategic misrepresentation heuristic suggests that the choice of the status quo option not only occurs when the status quo is preferred to the alternatives in the current choice set, but, potentially, also when a previously accepted alternative is preferred to the hypothetical alternatives currently on offer. When this latter condition occurs, respondents are also assumed to replace the status quo with the reference alternative. The second feature of the strategic misrepresentation heuristic is that respondents consider the probability of provision. When a similar good is offered at very different cost levels over the course of a sequence of choice tasks, it would be reasonable for respondents to assume that the higher-cost alternatives are more likely to be provided where the respondents' stated WTP is higher. McNair *et al.* (2012) assume that the perceived probability of project provision is equal to the ratio of the maximum cost level previously accepted by the respondent and the maximum cost level observed in all intervening choice sets up to the current choice task. Under strategic misrepresentation, the utility expressions for the status quo and experimental alternatives may be expressed by Equation (2.26):

$$\begin{aligned} V_{SQ} &= p_{sq}\left(\sum_k \beta_k X_k^a\right) \\ V_{alt} &= p_{alt}\left(\sum_k \beta_k X_{alt,k}\right) \end{aligned} \tag{2.26}$$

The variables in Equation (2.26) are defined as follows. X_k^a is the level of attribute k in the highest-cost alternative accepted in a previous choice task; $p_{sq} = \frac{X_{\cos t}^a}{X_{\cos t}^0}$, $X_{\cos t}^0$ being the maximum cost level observed up to and including the current choice task; and $p_{alt} = \max(p_{sq}, \frac{X_{alt,\cos t}}{X_{\cos t}^0})$.

In terms of estimation results, using the PDP/latent class approach, McNair *et al.* (2012) find that the project attributes have the expected sign where they are significant. On the class membership probabilities of the various decision processes modelled, they find that about 40 percent of respondents behave according to the value learning heuristic. Some 30 to 45 percent of respondents are estimated to be choosing according to the strategic misrepresentation heuristic. The class with the lowest membership probability is the decision process based on the standard assumptions of truthful response and stable preferences. Within the value learning heuristic, the estimated parameter β_{k+1} is positive, which agrees with prior expectations that a higher cost level observed in a previous choice task leads to a higher probability that the hypothetical alternative is chosen in the current choice task, all else equal. However, McNair *et al.* (2012) do not find any significant differences in mean prior WTP calculations across this proposed probabilistic decision process model and a standard MNL model.

Underpinning value learning and strategic misrepresentation is the notion of reference point revision (DeShazo, 2002). In experiments which include the status quo as one of the alternatives, the oft-observed status quo bias (Samuelson and Zeckhauser, 1988; Fernandez and Rodrik, 1991) may mean that the status quo itself simply ends as the reference point, as respondents continue to stick disproportionately to the status quo. Samuelson and Zeckhauser (1988) argue that

> *"A decision maker in the real world may have a considerable commitment to, or psychological investment in, the status quo option. The individual may retain the status quo out of convenience, habit or inertia, policy (company or government) or custom, because of fear or innate conservatism, or through simple rationalization."* (Samuelson and Zeckhauser, 1988, p. 10).

Hensher and Collins (2011) test whether reference points are shifted when non status quo alternatives are chosen and find that if a non-reference (i.e., non status quo) alternative is chosen in the preceding choice set $s-1$, the reference in the current choice set s is revised and the utility of the non status quo alternatives increases. This suggests a shift in the value function around a new reference point.

In the marketing literature, Briesch *et al.* (1997) have suggested that a relatively large proportion of consumers are fairly accurate in their recollection of prices and may therefore rely on their memories for past prices when evaluating current purchases across various brands. When previously encountered attributes or alternatives are used as a reference, judgements are assumed to be memory-based because information is retrieved from memory and then compared to what is currently available in the choice set. Memory based judgements "are likely to occur when consumers are able and are motivated to recall past prices from memory and use this information for the task at hand." (Briesch *et al.*, 1997, p. 204). Thus, a vastly superior and dominant alternative encountered in a choice set would create favourable conditions for memory-based judgments to take place and it would be likely for such an alternative to be held in memory as a reference point in future choice sets. By contrast, when the current attributes of another alternative are used as the reference, judgement is said to be stimulus based. Concerning high frequency purchases of consumer goods, Briesch *et al.* (1997) evaluate various econometric specifications of references involving memory-based or stimulus-based prices. Several possible candidate specifications considered by Briesch *et al.* (1997) are described in Table 2.6.

In their specification of the utility function, Briesch *et al.* (1997) consider different sensitivity to gains and losses. They find that a specification assuming memory based judgements involving brand specific past prices provides the best model of reference price across all the consumer product categories they analysed. Respondents appear to consider up to around six

Table 2.6: Description of Possible Reference Rules

Memory-based prices	Stimulus-based prices
Price of all previously chosen brands: This reference specification is motivated by a body of findings that consumers have a stronger memory for an actual chosen brand than for a non-chosen brand. Prices of previously chosen brands are more readily accessible, with more recently encountered prices given a greater weight.	**Current price of a random brand**: An extreme case where the consumer has no knowledge of brand prices and is unable to determine which of the current brands is appropriate as a reference. Hence, a random brand is selected and its price used as the reference.
Brand-specific past prices: Consumers use the past price history of each brand as a reference price to evaluate the respective brands in the current choice set. The reference price is unique for each brand.	**Current price of previously chosen brand**: Assumes that consumers do not remember the price paid previously, but do recall the brand they chose. The current price of this reference brand becomes the reference price.
Brand-specific past prices and other information: The reference used is not only the past prices specific to each brand, but also the brand's price trend and frequency of deals. This rule places greatest demand on memory.	

time periods when choosing the length of the price history to include in the reference price. Briesch *et al.* (1997) caution however that the results are sensitive to the specification of reference price. For example, the use of past prices of all previously chosen brands as a reference does not lead to any improvement in the fit over a model that does not contain a reference price term. Briesch *et al.* (1997) conclude that this finding is significant because it demonstrates that misspecifications of the reference price can lead to wrong conclusions about the absence of the reference price effect, even when it may actually exist.

2.8 STATE DEPENDENCE AND HABIT PERSISTENCE

With a sufficient number of choice sets per respondent, such as when decisions of a panel of respondents are recorded over a period of time, it may be possible to test whether intervening choice sets matter. Heckman (1981) suggested the idea of modelling the dynamics of past influences on current decisions through a general model of structural state dependence and habit persistence. Structural state dependence occurs when previous choice outcomes affect current utility, whereas habit persistence allows for previous utility evaluations to affect current utility. Structural state dependence is not dissimilar to what Gilboa and Schmeidler (1995) would call “case-based decision theory”. In a model of dynamic vehicle choice, Hensher *et al.* (1992) capture the dependence of current choice behaviour on past behaviour through the use of an expectations and an experience effect. In particular, the proposed experience effect measures the influence of past behaviour on current choice. It is given by the difference, in absolute terms, in the current attribute level with the same attribute level in some past instance, weighted by a time discount factor. From the results, Hensher *et al.* (1992) find that the strong statistical significance, with the correct sign, points to the important role that habit plays in influencing current vehicle-type choice[3].

Heckman’s (1981) model is also picked up by Swait *et al.* (2004) in the context of choice behaviour among recreational anglers, who have to repeat the choice of one of several fishing

---

[3] Hensher *et al.* (1992) also allowed for serial correlation across time periods for a dynamic model of vehicle use.

sites multiple times over the season. In the model specification for habit persistence, the current utility at time or choice set s of alternative j is defined through a meta-utility function as shown in Equation (2.27):

$$\hat{V}_{js} = \prod_{i=0}^{i=s} \alpha_{j,s-i} \exp(V_{j,s-i}) \tag{2.27}$$

Meta-utility $\hat{V}_{js}$ is dependent on all past (static) utilities $V_{j,s-i}$ which is itself dependent only on the attributes in the period $s - i$. The link between current utility and historical observed utilities is achieved through a path-dependence parameter $\alpha_{j,s-i}$, where $\alpha_{j,s-i}$ might also be interpreted as the weights associated with the previous periods. $\alpha_{j,s-i}$ satisfies the conditions $0 \leq \alpha_{j,s-i} \leq 1$; $\alpha_{js} = 1$. Taking logs to obtain a linear additive form and adding past and contemporaneous error terms results in Equation (2.28):

$$\ln(\hat{V}_{js}) = \sum_{i=0}^{i=s} V_{j,s-i} + \sum_{i=0}^{i=s} \ln(\alpha_{j,s-i}) + \sum_{i=0}^{i=s} \varepsilon_{j,s-i} \tag{2.28}$$

As the first right hand side term of Equation (2.28) contains all past attribute levels, this equation can also be seen to link "current utility to historical observed attribute levels in a fashion that is consistent with learning about attributes or updating." (Swait *et al.*, 2004, p. 98). Attribute levels in previous periods are therefore combined with current attribute levels in a form of temporal averaging. Incorporating state dependence into the model involves using a dummy variable that equals 1 for alternative j in current choice set s if the same alternative had been chosen in choice set $s - 1$. The variance structure of the disturbance term can be allowed to vary over time, providing a form of temporal heteroscedasticity. With single respondents answering repeated choice experiments, this model provides another way of investigating the role of the value learning heuristic.

Heckman's (1981) idea of state dependence implies that the utility for alternatives in the current state or choice set is directly modified by previous choices or previously encountered attributes. One difficulty with this approach is the potential endogeneity that might arise as a result of the error term being correlated with some of the co-variates. The use of lagged endogenous variables to model habits is avoided in Adamowicz and Swait (2013) through a two stage decision process. In their model, the higher level decision involves a choice from one of various decision strategies. These decision strategies correspond to: (i) randomly choosing any alternative that is different from the last alternative chosen (pure variety seeking heuristic with minimum effort); (ii) repeating the choice of the same alternative again in order to minimise cognitive effort (habitual behaviour); or (iii) conducting a full evaluation of all products based on attributes and prices (utility maximisation in a fully compensatory sense). In their approach, tastes, as reflected in the utility function, are constant over time and utility is stable, in the sense that the evaluation of an alternative is not influenced by simple cumulative experience with an alternative.

The second stage of the decision process is the evaluation of alternatives conditioned on the decision strategy of the first stage. Past choice behaviour influences the higher level decision on which decision strategy to adopt, but does not directly affect the evaluation of the alternatives themselves. Consider first the fully evaluative strategy. The expected value of the maximum utility from a set of alternatives can be summarised as the log-sum or inclusive value $I_s = \ln \sum_{j \in s} e^{V_j}$ (Ben-Akiva and Lerman, 1985). Whilst the benefit of pursuing the full evaluation strategy in the current choice task is a function of its log-sum value, Adamowicz and Swait (2013) suggest that because decision makers are cognitive misers, they do not use I_s in deciding whether the full evaluation strategy is worthwhile or not, as the proper assessment of this benefit requires the full evaluation of all alternatives. Instead, it is assumed that people use I_{s-1} which is the log-sum value from the prior choice task to naively forecast I_s.

Given that the full evaluation strategy is likely to require some cognitive resources, Adamowicz and Swait (2013) assume that the utility of the full evaluation strategy takes the form $V_{FE} = I_{s-1} - \theta$, where θ represents the net effect of the cognitive processing costs of undertaking full evaluation and the utility of knowing that the "best" option has been chosen and that all available alternatives have been fully evaluated. The use of I_{s-1} is not without its difficulties however, as the computation of I_{s-1} assumes that respondents make a full evaluation of all alternatives even if the decision strategy used previously was something other than full evaluation.

The utility of the pure habitual decision making strategy is given by the utility of the previous choice, so that $V_H = V_{s-1}$. V_H does not include a component for processing cost because by assumption, the pure habitual strategy requires low cognitive effort.

For the variety seeking strategy, Adamowicz and Swait (2013) suggest that a simplifying heuristic that can be used to approximate this decision rule is for respondents to make a random choice from all alternatives in the current choice set but excluding the alternative chosen in the previous period. This strategy eventually provides a diversity of choice outcomes which satisfies the desire for variety. The variety seeking strategy is also differentiated from the full evaluation strategy due to the very parsimonious use of information, needing only the prior choice to be known. Nevertheless, in terms of modelling the utility of the variety seeking strategy, certain aspects of the model are borrowed from the full evaluation strategy model, primarily, that the benefit of adopting pure variety-seeking is the maximum expected utility from the previous set of alternatives excluding the alternative that was chosen.

More formally, for the variety seeking strategy, $I'_{s-1} = \ln(\sum_{j \neq j*} \exp(V_j))$, where $j*$ denotes the alternative that was previously chosen. Because choice is assumed to be random, all the alternatives $j \neq j*$ appear to have equal utility from the decision maker's perspective, hence it may be assumed that $V_j = \tilde{V}$ and $I'_{s-1} = \ln(\sum_{j \neq j*} \exp(\tilde{V}))$. This assumption of equal utility reduces the informational processing load that is required of respondents. Finally, a constant γ is added to reflect the net effect between a "variety premium" and the disutility of knowing that the random choice may generate an undesirable outcome. In summary, the observed utility in the variety seeking strategy is $V_{VS} = I'_{s-1} + \gamma$. Since I'_{s-1} collapses to a constant because of the constant term $\tilde{V}$, Adamowicz and Swait (2013) are only able to identify one overall term that captures the net utility of variety seeking relative to the other decision strategies.

When applied to a dataset of routine and repeated purchases of catsup and yoghurt over time, Adamowicz and Swait (2013) find that their proposed decision strategy model is preferred over a state dependence model on the basis of information criteria (AIC and BIC) and also on the plausibility of direct price elasticities. Among the three decision strategies, the highest propensity is reserved for the fully evaluative model, but the propensity for the habitual strategy is not insignificant either. The purchase behaviour of catsup seems to favour the habitual model quite highly, possibly due to catsup being consumed and replenished at a lower rate and also due to there being less flavour variations in catsup and therefore, a smaller choice set to choose from. This means that price has very little effect on catsup choice. By not correcting for habit persistence where it is more likely, Adamowicz and Swait (2013) argue that the standard state dependence model leads to an over-prediction of price elasticity relative to their proposed model.

2.9 ATTRIBUTE LEVEL EDITING

Another way of linking the attribute levels in the preceding choice set to the attributes in the current choice set involves a just noticeable difference heuristic (Cantillo *et al.*, 2006). This heuristic may also be interpreted as an avenue for respondents to edit the attribute levels presented to them. A change in the attribute level from choice task $s-1$ to choice task s is assumed to be perceptible to the respondent if the magnitude of the change in the attribute levels of attribute k exceeds a certain threshold, i.e., $\left|\Delta X_{nk,s}\right| = \left|X_{nk,s} - X_{nk,s-1}\right| \geq \delta_{nk}$, for non-negative threshold values δ_{nk} of attribute k for respondent n. Like several of the threshold formulations described earlier, thresholds can be assumed to be individual-specific, randomly distributed across the population, and may also depend on socio-demographic characteristics.

Cantillo *et al.* (2006) hypothesise that respondents only perceive the part of the attribute level change that is bigger than the threshold, as in Equation (2.29) below:

$$X_{jnks} = X_{jnk,s-1} + sgn(\Delta X_{jnks}) \max(\left|\Delta X_{jnks}\right| - \delta_{nk}, 0) \tag{2.29}$$

If m out of the K attributes are associated with a perception threshold, the modelled component of utility can be written in Equation (2.30) as:

$$
\begin{aligned}
V_{njs} &= \sum_{k=1}^{m} \beta_{jk} X_{njks} + \sum_{k=m+1}^{K} \beta_{jk} X_{njks} \\
&= \sum_{k=1}^{m} \beta_{jk} \left[X_{njk,s-1} + \Delta X_{njks} \left(1 - \frac{\delta_{nk}}{\left| \Delta X_{njks} \right|} \right) I_{njk} \right] + \sum_{k=m+1}^{K} \beta_{jk} X_{njks}
\end{aligned}
\tag{2.30}
$$

$$
\text{where } I_{njk} = \begin{cases} 1 \text{ if } \left| \Delta X_{njks} \right| \geq \delta_{nk} \\ 0 \text{ otherwise.} \end{cases}
$$

To complete the model, Cantillo *et al.* (2006) assume that δ_{nk} can be expressed as a proportion of $X_{njk,s-1}$ so that $\delta_{nk} = X_{njk,s-1}(\bar{\delta}_k + \eta_{nk})$, where $\bar{\delta}_k$ is the expected perceived value that is allowed to vary by individual characteristics and η_{nk} represents individual deviations from the perceived value following a probability density function with mean zero and variance σ_η^2. If multiple perception thresholds are considered in the model, then a joint density function is required. In their empirical application, since there is no theory to determine which attribute is likely to be threshold constrained, Cantillo *et al.* (2006) test every attribute modelled in the utility function against a threshold constrained assumption until the best model fit is obtained. Applying this model to a route-choice stated preference survey for car trips, and for the thresholds, Cantillo *et al.* (2006) conclude that a threshold exists for the travel time attribute but not for the cost or variability attribute. A threshold model for travel time significantly improves the log-likelihood of the model compared to the reference MNL model, with substantial increases in both the coefficients for travel time and cost. These parameter estimates imply a range of values of travel time savings which are on average lower compared to the travel time benefits obtained from a traditional MNL model.

Cantillo *et al.*(2006)'s just noticeable difference heuristic provides one way of allowing respondents to modify the attribute values presented to them in a particular choice task, thereby relaxing the frequently maintained assumption in most choice models that respondents take the attribute levels as given. In applications where variability matters, for example, in transport where both travel times and variability of travel times are important determinants of choice, the travel time and the probability attributes may themselves be changed or edited by the respondent, with the magnitude of the edit possibly depending on the variability attribute and any associated threshold. In the rank dependent utility model of Hensher and Li (2012), where one of three possible outcomes (early, on time or late arrival) in a risky choice situation will occur, the decision weight associated with outcome *m* of is not solely based on the objective probability of occurrence itself, but is also in part determined by some non-linear transformation of the cumulative probability; the latter being dependent on the rank ordering of all possible outcomes. This transformation, which is another form of editing, allows low objective probabilities to be overweighted and high objective probabilities to be underweighted in decision making. In this model, the utility for an alternative is assumed to be dependent on the transformed travel time that is associated with each of these *m* outcomes.

More specifically, Hensher and Li (2012) assume that the non linear transformation of the objective probabilities follows a probability weighting function proposed by Tversky and Kahneman (1992). This function, called $w(p)$, is given by Equation (2.31):

$$w(p) = \frac{p_m^{\gamma}}{\left[p_m^{\gamma} + (1 - p_m)^{\gamma}\right]^{\frac{1}{\gamma}}} \tag{2.31}$$

Ranking outcomes from the least desired to the most desired, the decision weights $\pi(p_m)$ are given by Equation (2.32):

$$\begin{gathered} \pi(p_m) = w(p_m + p_{m+1} + \ldots + p_{M-1}) - w(p_{m+1} + \ldots + p_{M-1}) \text{ for } m = 1, \ldots, M-1 \text{ and} \\ \pi(p_M) = w(p_M) \end{gathered} \tag{2.32}$$

The part-utility for an attribute *k* which can take on *m* different outcomes is then represented by the following Equation (2.33):

$$U(X_k) = \sum_{m=1}^{M} \pi(p_m) U(X_k^m) \tag{2.33}$$

For the assumed functional specification of *w(p)*, Hensher and Li (2012) find that the estimated curvature parameter γ leads to a convex probability weighting curve. This means that the decision weight of the most desired outcome, which is on time arrival, is smaller compared to its objective probability. The decision weights of the less desired outcomes (early or late arrivals) are higher than their associated objective probabilities. Respondents may therefore be described as conservative in terms of their beliefs regarding the likelihood of on time arrivals.

2.10 OTHER HEURISTICS

Heterogeneity in decision processes may be inferred from observed choice outcomes by directly embedding various heuristics into the modelled component of utility functions in latent class or probabilistic decision process models (Hensher and Collins, 2011; Hess *et al.*, 2012; Hensher and Greene, 2010; McNair *et al.*, 2012; Scarpa *et al.*, 2009). In transport applications, these models have been used to test the heuristics of common-metric attribute aggregation, attribute non-attendance and decision rules like majority of confirming dimensions (which can be considered a form of editing) to explain choice in the context of a toll road/non toll road alternative (Hensher, 2010; Hensher and Greene, 2010; Hensher and Collins, 2011).

2.10.1 Lexicography and Elimination-by-Aspects

By means of the latent class approach, Hess *et al.* (2012) model the lexicographic and EBA decision rules. This represents a starting point that differs from the approach of Swait (2009). In Swait (2009), all respondents are assumed to be homogenous in the sense that the same decision rule is used in evaluating alternatives; heterogeneity in the treatment of alternatives comes about as a result of individual-specific thresholds. In contrast, the latent class model or the probabilistic decision process model assumes heterogeneity in the decision processes of the sample of respondents and hence potentially very different model structures in each latent class.

Hess *et al.* (2012) exploit the knowledge that (apparent) lexicographic behaviour is easier to spot in a two attribute choice task (travel time and cost) and assume that respondents who always choose the fastest travel time are lexicographic on travel time and those who always choose the lowest cost alternative are lexicographic on cost. Hess *et al.* (2012) find that including the two lexicographic models lead to an improvement in fit compared to the simple MNL and MMNL models. More significantly, they find that including the lexicographic classes reduces the random taste heterogeneity that is retrieved from a pure MMNL model. It appears that non-trading behaviour and extreme sensitivities are best modelled by explicitly specifying a lexicographic model structure rather than by requiring a random taste parameter to explain non trading effects through an extreme tail on its distribution.

In modelling the EBA heuristic, Hess *et al.* (2012) encounter the usual challenge that in the absence of data to specify the kind of EBA decision rule that is being used, some plausible EBA rules have to be assumed. In a choice experiment involving rail travel behaviour, with the alternatives described on the basis of two continuous variables (travel time and fare) and three other attributes that are described in terms of being present/absent, Hess *et al.* (2012) assume the following four EBA rules:

EBA1: Eliminate the worst (for the considered attribute) of any remaining alternatives at a given stage;

EBA2: Eliminate all but the best (for the considered attribute), equating to a dominance based approach;

EBA3: Eliminate all options that are 10 minutes slower than the reference trip, or \$0.50 more expensive when using fare (depending on which attribute is used as a basis for elimination); and

EBA4: Eliminate an alternative if the time or fare is worse than that for the reference trip (again depending on which attribute is used as the basis for elimination).

Hess *et al.* (2012) consider a choice experiment with five different aspects $\{k_1, k_2, k_3, k_4, k_5\}$, such that under an EBA heuristic, if each aspect is associated with weight $w_1, w_2, ..., w_5$, then the probability that an elimination ordering is $\{k_1, k_2, k_3, k_4, k_5\}$ is equal to:

$$p_{1,2,3,4,5} = \frac{w_1}{\sum_{i=1}^{i=5} w_i} \frac{w_2}{\sum_{i=2}^{i=5} w_i} \frac{w_3}{\sum_{i=3}^{i=5} w_i} \frac{w_4}{\sum_{i=4}^{i=5} w_i} \qquad (2.34)$$

With a given elimination ordering $\{k_1, k_2, k_3, k_4, k_5\}$, the EBA heuristic states that alternatives which do not possess aspect 1 are first eliminated. Remaining alternatives are eliminated on the basis of aspect 2 and so on until one alternative remains. If this remaining alternative coincides with the actual choice made by the respondent, then $P_{ns} = 1$, otherwise $P_{ns} = 0$. In cases where more than one alternative remain after all aspects have been cycled through, the remaining alternatives are given an equal probability of being chosen. The unconditional probability of observing the actual sequence of choices is the weighted average across all possible elimination orderings:

$$P_n = \sum_{a=1}^{5} \sum_{b \neq a} \sum_{c \neq b,a} \sum_{d \neq c,b,a} \sum_{e \neq d,c,b,a} p_{a,b,c,d,e} \prod_{s} P_{ns} \tag{2.35}$$

By estimating four different latent class models, with each model comprising a dual class structure of an MNL rule and one of the four EBA rules, Hess *et al.* (2012) find that each of the latent class models embedding the EBA rule outperforms the standard MNL model in terms of goodness of fit, despite the MNL part of the model still accounting for the majority share of the class probabilities. The EBA2 rule provides the best model fit. Interestingly, EBA2 coincides with one of Suzuki's (2007) assumed decision rules on how an alternative may be retained for further consideration in a two-step model.

In a latent class model of MMNL and EBA2, Hess *et al.* (2012) find an across-the-board increase in the mean values of each of the four WTP measures, compared to the pure MMNL model, and a reduction in the degree of heterogeneity in the WTP measures. This result indicates that the EBA2 rule has a role to play in explaining a portion of heterogeneity that was previously ascribed to random taste variance in the simple MMNL model. Interestingly, Hensher *et al.* (2012a) also come to a similar conclusion when they observe that using a random parameter approach might confound with attribute processing, and so it might be preferable to account for attribute processing using fixed parameters in discrete latent classes rather than to use continuously distributed random parameters and ignore attribute processing altogether. The overall conclusion from the research is that accounting for decision process heterogeneity in latent classes – that is, allowing heuristic use to vary by subgroups of respondents (up to a probability) – leads to improvements in model fits compared to the standard multinomial logit model, and where assessed, the mixed logit model based on a standard utility specification.

2.10.2 Common Metric Attribute Aggregation

To model the incidence of common metric attribute aggregation, Hensher (2010) assumes that a threshold exists such that attribute aggregation occurs if the distance between common metric attributes is smaller than the threshold. More formally, assuming that X_1 and X_2 are the attributes with a common metric, then V_j may be conditioned in Equation (2.36) as follows:

$$\begin{aligned} V_j &= \beta_1 X_1 + \beta_2 X_2 + \sum_{k=3}^{K} \beta_k X_k \text{ if } \left(X_1 - X_2\right)^2 \geq \alpha \\ &= \beta_{12}(X_1 + X_2) + \sum_{k=3}^{K} \beta_k X_k \text{ if } \left(X_1 - X_2\right)^2 < \alpha \end{aligned} \tag{2.36}$$

The parameter α, which is assumed to take a random distribution in the population, is the threshold distance that determines if common metric attributes are aggregated or not. Recalling the earlier discussion on the similarity effect in Section 2.6, this formulation for V_j intuitively captures the notion of similarity, so that the heuristic in this case is the rule that common metric attributes whose values are close enough to each other are indistinguishable. Hensher (2010) assumes an exponential distribution for α, such that:

$$\begin{aligned} &\Pr(\mathit{attribute\ disaggregation}) = \Pr(\alpha \leq (X_1 - X_2)^2) = 1 - e^{-\lambda(X_1 - X_2)^2} \text{ and} \\ &\Pr(\mathit{attribute\ aggregation}) = \Pr(\alpha > (X_1 - X_2)^2) = e^{-\lambda(X_1 - X_2)^2} \end{aligned} \tag{2.37}$$

Hensher (2010) argues that the exponential distribution is a good choice for this model as it has a large mass near zero which allows for the sample to behave as standard optimisers in a large fraction of the choice sets while the tail allows for the relatively less frequent occurrence of attribute aggregation. Recalling the earlier discussion in Section 2.2 on how the properties of the choice task can influence the choice of heuristic, note that the probability of attribute aggregation/disaggregation is a function of a choice task property, which is the squared difference in attribute levels X_1 and X_2, and it therefore follows that this probability will vary within the same respondent as long as $(X_1 - X_2)^2$ is not constant. The unconditional utility is simply given by Equation (2.38):

$$V_j = (V_j \mid \alpha \le (X_1 - X_2)^2) Pr(\alpha \le (X_1 - X_2)^2) + (V_j \mid \alpha > (X_1 - X_2)^2) Pr(\alpha > (X_1 - X_2)^2) \qquad (2.38)$$

Estimating this model, Hensher (2010) finds that when the common metric attribute aggregation rule is applied, the mean value of travel time savings is higher than in a traditional MNL model.

2.10.3 Majority of Confirming Dimensions and Attribute Non-Attendance

Hensher and Collins (2011) operationalise the majority of confirming dimensions rule by identifying, for each attribute, the attribute level that is the best in the choice set (no ties allowed) and then taking the count of the number of 'best' attributes for each alternative. This variable is entered as an additional component in the utility function. Hensher and Collins (2011) find that the parameter on this variable is highly significant and positive in sign, so that as the number of best attributes increases in the alternative, that alternative is more likely to be chosen. However, on the value of travel time estimates, no statistical difference is observed between the standard MNL model and the MNL model plus the majority of confirming dimensions rule.

However, Hensher and Collins (2011) find a significant improvement in model fit between a model that takes respondent stated attribute non-attendance into account versus a model that ignores non-attendance completely (i.e., assumes all attributes are attended to, regardless of what respondents say). A further improvement is obtained when the majority of confirming dimensions rule is re-interpreted to account only for those attributes that are attended to. Where the estimates for the value of travel time saving are concerned, there is a significant difference between models that embed attribute non-attendance and models that do not. This suggests that for policy applications which require the use of these outputs, it would be preferable to account for potential attribute non-attendance in the model.

Fundamentally, attribute non attendance implies the existence of a subset of attributes for which the marginal rate of substitution is not computable, as there is no trade-off between attributes in the neoclassical economic sense. For non market valuation, no relative implicit price can be obtained for respondents who do not make such trade-offs. There is a danger that biased WTP estimates are obtained when pooling respondents who fully attend to all attributes with respondents who ignore some attributes. Accordingly, Scarpa *et al.* (2009) advocate the use of equality constrained latent class models to operationalise attribute non-attendance, which is one way of inferring and accounting for attribute non-attendance when data on self reported attribute non attendance are not available. In these models, latent classes are specified with zero utility weights for selected attributes assumed not attended to, while the taste parameters for the attributes assumed attended to are required take the same value across classes. Scarpa *et al.* (2009) report evidence showing that the incidence of respondents attending to all attributes is very low – in the order of less than 0.1 – and that most respondents seem to ignore at least two attributes when making their decisions. In more recent applications, such as Hensher *et al.* (2012a), the equality constraint assumption is relaxed and parameter heterogeneity within and across various types of attribute processing rules may be observed.

In instances where supplementary questions are a part of the survey instrument, directly embedding self-stated responses to questions of whether certain attributes are ignored has further improved the explanatory power of the models and the efficiency of marginal WTP estimates. For example, Scarpa *et al.* (2010) measure attribute attendance and non attendance at the level of the choice task by asking respondents at the end of each choice task whether an attribute was attended to or not. A measure of serial non attendance can be constructed if there is consistent non-attendance for a particular attribute throughout the sequence of choice tasks. From the data collected, Scarpa *et al.* (2010) observe that choice task attribute non attendance is far more prevalent than serial non-attendance, indicating to the importance of monitoring such behaviour at the choice task level for multi-attribute choices. If there is within-respondent variability in the incidence of attribute attendance and non-attendance during the course of the choice experiment, asking respondents one question on attendance/non attendance at the end of the survey, as is typically the case, may lead to over (or under) generalisation of the issue.

Indeed, Scarpa *et al.* (2010) find that assuming only serial attendance/non attendance and disregarding intra-respondent variation in non-attendance leads to a worsening of fit compared to the conventional model which assumes all attributes are attended to. The best-fitting specifications are provided by the model which accounts for choice task attribute non-attendance statements. They also find that accounting for choice task attribute non-attendance tends to lead to smaller WTP estimates than either assuming serial non-attendance or assuming full attendance.

2.11 SUMMARY AND CONCLUDING REMARKS

One of the main purposes of this chapter is to convincingly demonstrate that a significant amount of research has been undertaken in the discrete choice modelling literature towards modelling the contribution of decision heuristics and the influence of contextual effects in explaining choice behaviour. Much of this work is very recent and undoubtedly, a lot more work on the role of heuristics in explaining choice behaviour is currently in progress or is being planned.

The empirical evidence consistently shows that embedding heuristic rules into the modelled component of utility leads to improved measures of fit. What this review chapter also shows is that in many cases, single heuristics are considered and compared against the standard RUM model. However, given the plethora of heuristics that can be called upon by respondents to aid their decision making, it may now be an opportune time to consider testing how multiple heuristics may be embedded in choice models. Modelling certain aspects of heterogeneity in decision rules might well complement the current research focus of modelling preference heterogeneity through more advanced discrete choice models such as the mixed multinomial logit.

As a practical way forward, subsequent chapters of this thesis will advocate the use of mixture models, where multiple heuristics are weighted in a utility function, using weighting functions that depend on the socio-economic characteristics of the respondent and other choice context variables, including individual specific perceptions data, where available. These mixture models thus allow a certain degree of heterogeneity in the use of decision rules. Such an approach is especially relevant in testing cases such as the relative advantage model, where respondents' utility for a certain alternative are assumed to be a function of both inherent preference (predispositions) and constructed preferences which depend on the choice context. Rooderkerk *et al.* (2011) have pointed out that context dependent components may be added to the context free or context independent RUM component of the utility function, and a mixture model of heuristics can easily be seen as a natural extension of this work.

A related observation is that in contrast to prospect theory, there appears to be comparatively much less research done on the application of models of extremeness aversion (the compromise effect) and other choice set contextual components within the discrete choice literature in transportation. Only very recently has there been a nascent attempt at this (Chorus and Bierlaire, 2013). With the notable exception of existing work done using the RRM model, this lack of research into the extremeness aversion heuristic is all the more remarkable considering that authors in the management science and marketing fields have consistently identified these effects as important determinants of choice. Perhaps this heuristic has not attracted much research attention because of the limitations in discrete choice modelling software in estimating utility functions which are non-(additively) linear in the parameters and attributes. Fortunately, testing these heuristics is highly feasible even with existing datasets and with recently developed advances in modelling software. Another advantage of estimating extremeness aversion models is that compared to models which incorporates thresholds, no additional information on thresholds needs to be collected from respondents, nor do rules on thresholds need to be assumed and tested. Therefore, another objective of this thesis will be devoted to estimating and comparing models of extremeness aversion in the transportation stated choice context.

In addition, much research in this area involves using the same dataset to estimate various models with decision heuristics embedded and comparing these to a standard model which is usually specified to be MNL, fully compensatory, linear-in-the-parameters and linear-in-the-attributes. Again, there is comparatively less work done on testing the independence of heuristics across multiple datasets, in other words, testing if the heuristics that purportedly explain the data better in one dataset can do so likewise for other datasets. This line of inquiry would be especially pertinent for models which rely on additional data for identification (such as models using self-reported threshold data). Considering how some of these heuristics have a non-linear functional representation, it will be useful to have a better understanding of how general the use of such heuristics are, before starting on a data collection process or an estimation process that might be resource intensive.

The heuristics that have been reviewed in this chapter have been developed and assessed within specific discrete choice modelling frameworks; however while the evidence is limited to static models, it does signal a need to consider incorporating the findings into other behavioural frameworks that incorporate choice model components such as activity-based models and even model systems that allow for dynamic factors (learning, adaptation, habitual behaviour), and mental processes that have an important role in heuristic decision making. Wherever a utility expression is included in a model system, the evidence herein suggests that value might be added if new or alternative functional forms that capture a range of decision making rules or heuristics are specified.

In a similar vein, the literature has observed that model outputs such as welfare estimates and marginal willingness to pay can be substantially different when the model departs from the standard assumptions about decision making. However, there is mixed evidence as to the direction of change to willingness to pay measures when heuristics are embedded into choice models. For example, some papers suggest an increase in the value of travel time savings when heuristics are modelled (e.g., Hensher, 2010), while others come to the opposite conclusion (e.g., Cantillo and Ortuzar, 2005). This is another research question that will be addressed in Chapter 5.

To conclude this chapter, the main heuristics that have been discussed are summarised in Table 2.7. More importantly, Table 2.7 also lays out a roadmap for the empirical analysis in Chapter 5 in terms of the types of heuristics that are examined in this thesis. The three heuristics that will be discussed in greater detail are the majority of confirming dimensions rule, the extremeness aversion heuristic and the reference point revision. Before that takes place, some discussion of the methodology used in this thesis is presented in Chapter 3.

Table 2.7: Summary Table of Main Heuristics Discussed in Chapter 2

Heuristic	Brief Description	Key References	Examined in thesis?
Satisficing	Choose the first alternative whose every attribute meets an acceptability threshold	Simon (1955)	No
Majority of Confirming Dimensions (MCD)	Select the alternative that has the highest count of 'best' attributes	Russo and Dosher (1983)	Yes
Elimination-by-Aspects (EBA)	Select an attribute based on some function of importance weight. Eliminate all alternatives whose attribute level fail to meet the threshold.	Tversky (1972)	No
Lexicography	Identify the most important attribute and select alternative with the 'best' level in that attribute.	Tversky (1969)	No
Extremeness Aversion Heuristic	Extreme alternatives – those with very good and very bad attributes – tend to be avoided.	Tversky and Simonson (1992); Simonson and Tversky (1993)	Yes
Compromise Effect	The compromise or in-between alternative has a higher probability of being selected. Postulated to be a result of loss aversion.		Yes
Polarisation Effect	One of the extreme alternatives tends to be selected over another extreme alternative. Can be considered to be an asymmetric form of extremeness aversion where respondent exhibits aversion to only a subset of attributes.		Yes
Extremeness Seeking Heuristic	Extreme alternatives are preferred to the compromise.	Gourville and Soman (2007)	Yes, in relation to extremeness aversion
Random Regret Minimisation	Select an alternative to minimise negative emotion, where negative emotion is defined as the feeling of loss when a non-chosen alternative does better than the chosen alternative on some attribute.	Chorus *et al.* (2008); Chorus (2010); Chorus (2012)	Yes

Heuristic	Brief Description	Key References	Examined in thesis?
Prospect Theory	Attributes are framed as gains and losses relative to a reference. People are aversive to losses.	Kahneman and Tversky (1979)	Not directly, but related to extremeness aversion.
Reference Revision Heuristic	The choice of a non status-quo alternative shifts the reference point and increases the probability of another non status-quo alternative being chosen in subsequent choice sets.	DeShazo (2002); Hensher and Collins (2011)	Yes
Strategic Misrepresentation	Withholding true preferences in the hope that the most preferred alternative, which was encountered in some previous choice set, will be implemented.	McNair *et al.* (2012)	No
Value Learning	Preferences are 'discovered' and may be influenced by both the starting point and intervening attribute values.	Bateman *et al.* (2008); McNair *et al.* (2011)	No
Attribute non-attendance (ANA)	Some attributes of an alternative are ignored in the decision process.	Hensher (2010); Scarpa *et al.* (2009); Scarpa *et al.* (2010).	No, but models discussed in Chapter 5 can be extended to include ANA.

CHAPTER 3 ECONOMETRIC METHODOLOGY

3.1 INTRODUCTION

This chapter discusses the methodology that will be used to test how the heuristics of interest identified in Chapter 2 perform as alternative decision rules of choice behaviour. The general data setting is first discussed in Section 3.2. Random utility theory and the multinomial logit (MNL) model are then introduced in Section 3.3, before more advanced logit models are discussed in Sections 3.4 and 3.5. Section 3.6 explains several methods of embedding multiple heuristics into discrete choice models, including the approach of using heuristic weighting functions. The final sections in this chapter are concerned with methods of model comparisons, in terms of model fit as well as post-estimation analyses of model output.

3.2 NOMENCLATURE OF PREFERENCE DATA

Broadly speaking, there are two classes of data which can be used in a study of individual preferences. Revealed preference (RP) data are collected from responses observed in an actual market setting, while stated preference (SP) data come from responses observed in an experimental setting. RP data collection might, for example, be interested in observing how a person actually travels from home to work, given knowledge of existing real world alternatives and their attribute levels. SP data on the other hand require respondents to state their preferences in response to some hypothetical scenarios. Regardless of whether the data is RP or SP, it is important to note that actual preferences are never directly obtained, but are inferred from information that is provided by either RP or SP data.

The experimental setting in which SP data collection proceeds can be both an advantage and a disadvantage for the modeller. While hypothetical biases and predictive validity are always a concern, SP data can improve understanding of preferences, especially when RP data are inadequate. For example, SP data are particularly useful in analysing the response behaviour of a person who is confronted with both existing and new alternatives, since the latter are not present in RP data. SP data are also useful in cases where new attributes or new attribute levels of an existing alternative need to be considered.

This thesis is primarily concerned with the analysis of SP data. Even within the realm of SP data, several different preference elicitation methods may be used[4]. These methods include matching methods and conjoint analysis. Matching methods are frequently used in contingent valuations where respondents are asked to provide a number that will make them indifferent in some sense, for example, they may be asked to state their willingness to pay (i.e., give up income) in exchange for some good that they do not have. In conjoint analysis, respondents may be asked to rate or provide a complete ranking of the various treatment combinations of multi-attribute alternatives shown to them. These combinations will most likely involve different attribute levels so that the analyst can make meaningful statements on how respondents are trading off among attributes. However, Louviere *et al.* (2010) observe that in conjoint analysis, data are collected

> *"in ways that cannot be analyzed to be consistent with neoclassical economic theory because ratings and attribute importance measures do not readily translate into choice or matching (e.g., direct expression of WTP) data, the primitives on which utility theory is based."* (Louviere *et al.*, 2010, p. 61).

Rather than conjoint analysis, the focus of the thesis will be on stated choice analysis (of SP data). In other words, the preference elicitation method used is the discrete choice experiment (DCE). Implied by its name, decision makers in a DCE are required to make some type of choice. A typical DCE always involves some hypothetical choice scenarios and researcher specified attributes and attribute levels which are systematically varied within and/or between respondents by means of an experimental design. Moreover, choices are discrete in the sense that there is a finite, countable set of alternatives from which the decision maker must make a choice. A DCE immediately assumes that the respondent will face at least two alternatives in all choice situations, so that choice is meaningful. According to Louviere *et al.* (2010), DCEs should not be viewed as a special case of conjoint analysis, but rather "DCEs are based on a long-standing, well-tested theory of choice behaviour that can take inter-linked behaviours

[4] Carson and Louviere (2011) describe the nomenclature of various SP elicitation methods in great detail.

into account. The theory … is called random utility theory (RUT)" (Louviere *et al.*, 2010, p. 62).

Even within the genre of DCE itself, preference elicitation may proceed along several different pathways. For example, respondents may be asked to simply state their most preferred alternative from among all available alternatives on offer (used in the majority of studies), or their least preferred alternative, or what they consider the best and worst outcomes to be, the latter approach being applicable to DCEs where there are three or more alternatives in the choice set. The preference elicitation approach used in all the datasets analysed in this thesis is the one that requires each respondent to state only the best alternative for him/her, among all the alternatives available in the choice set. The analysis of such 'best only' data is supported by a large body of well developed theory, described in Section 3.3 and Section 3.4.

The DCE used in the empirical application can be labelled or unlabelled. In a labelled choice experiment, the title of the alternative conveys some information to the respondent, such as 'car', 'bus' and 'train'. In an unlabelled experiment, the title of an alternative, such as 'alternative A' or 'alternative B' does not convey any such information, except perhaps the order in which the collection of attributes occur. The datasets in this thesis all relate to unlabelled experiments. A fuller description of the data used in the thesis can be found in Chapter 4.

3.3 RANDOM UTILITY THEORY (RUT) AND THE MULTINOMIAL LOGIT (MNL) MODEL

Respondents in a DCE will encounter sets of alternatives, known as choice sets, from which some choice must be made. These choice sets will contain a fixed number of mutually exclusive alternatives. Random utility theory (RUT) assumes that for each of these choice alternatives, there is some latent construct known as "utility" that is known to the decision maker, but not to the researcher. The overall utility of an alternative j may be denoted by U_j.

Technically, U_j may vary by respondent n and choice set s; however, for the time being, the n and s subscripts are suppressed for notational simplicity. RUT further assumes that this overall level of utility may be decomposed into an 'observable' and an 'unobservable' or 'random' component. The observed component of utility may be denoted by V_j and the unobserved component by ε_j. Louviere *et al.* (2010) describe V_j as comprising "attributes [that explain] differences in choice alternatives and covariates [that explain] differences in individuals' choices." (Louviere *et al.*, 2010, p. 62) On the other hand, ε_j may

> "*comprise all unidentified factors that impact choices. Psychologists further assume that individuals are imperfect measurement devices; so, random components also can include factors reflecting variability and differences in choices associated with individuals and not choice options per se.*" (Louviere *et al.*, 2010, p. 62)

Typically, it is assumed that V_j and ε_j are independent and additive, such that $U_j = V_j + \varepsilon_j$. This assumption has a long history in discrete choice modelling (see for example, McFadden (1974)) and all the indirect utility functions estimated in this thesis will assume this independent and additive form between V_j and ε_j.

To the researcher at least, since utilities are inherently a stochastic construct because of the unobserved component ε_j, it is meaningful only to speak of the probability that a respondent chooses alternative *j* in choice set *s*. Since the 'best outcome' question is the one on view, it is reasonable to conclude, under random utility maximisation based on the axioms of choice theory, that the alternative chosen provides the highest utility of all alternatives in the choice set. Denoting alternatives in the choice set by *j'*, $P(j)$, which is the probability that a particular alternative *j* is chosen, is given by Equation (3.1):

$$P(j) = P[U_j > \max_{j' \neq j}(U_{j'})] = P[V_j + \varepsilon_j > \max_{j' \neq j}(V_{j'} + \varepsilon_{j'})] \quad (3.1)$$

RUT gives rise to probabilistic discrete choice models. Since ε_j is unobserved, any number of assumptions can be made about its distribution, but a commonly used assumption is that ε_j is independent and identically distributed (*i.i.d.*) standard Extreme Value Type I. Hence, ε_j has a cumulative density function given by Equation (3.2):

$$F(\varepsilon_j) = \exp(-e^{-\varepsilon_j}) \tag{3.2}$$

The *i.i.d.* Extreme Value Type I distribution is convenient since it results in a closed form solution for $P(j)$. Following Louviere *et al.* (2000, pp. 45 - 46), $P(j)$ may be obtained as follows. From Equation (3.1),

$$P(j) = P[\varepsilon_{j'} < \varepsilon_j + V_j - V_{j'}] \quad \forall j' \in s, j' \neq j \tag{3.3}$$

For a given value of ε_j, say $\varepsilon_j = b$ and using the *i.i.d.* extreme value type I assumption for ε_j,

$$P(j \mid \varepsilon_j = b) = \prod_{\substack{j' \in s \\ j' \neq j}} \exp(-e^{-(b+V_j-V_{j'})}) = \exp(-\sum_{\substack{j' \in s \\ j' \neq j}} e^{-(b+V_j-V_{j'})}) \tag{3.4}$$

The probability density function of ε_j is given by:

$$f(\varepsilon) = e^{-\varepsilon} \exp(-e^{-\varepsilon}) \tag{3.5}$$

Hence, the unconditional probability $P(j)$ is given by the integral of $P(j \mid \varepsilon_j = b)$ weighted by its density:

$$\begin{aligned} P(j) &= \int_{-\infty}^{\infty} \exp(-\sum_{\substack{j' \in s \\ j' \neq j}} e^{-(b+V_j-V_{j'})}) f(b)\,db \\ &= \int_{-\infty}^{\infty} \exp(-\sum_{j' \in s} e^{-(b+V_j-V_{j'})}) \exp(-b)\,db \\ &= \int_{-\infty}^{\infty} \exp(-e^{-b} \sum_{j' \in s} e^{(V_{j'}-V_j)}) \exp(-b)\,db \end{aligned} \tag{3.6}$$

Let $z = \exp(-b)$ and $a = \sum_{j' \in s} e^{(V_{j'}-V_j)}$. Therefore, $b = -\ln(z)$ and $db = -\frac{1}{z}dz$

Applying the substitution to the integrand and changing the limits of integration,

$$
\begin{aligned}
P(j) &= \int_{\infty}^{0} z e^{-za} \left(-\frac{1}{z}\right) dz \\
&= \int_{0}^{\infty} e^{-za} dz \\
&= -\frac{1}{a} e^{-za} \Big|_{0}^{\infty} = \frac{1}{a}
\end{aligned}
\tag{3.7}
$$

By substituting $a = \sum_{j' \in s} e^{(V_{j'} - V_j)}$ back into the equation and re-arranging, the multinomial logit (MNL) model is obtained in Equation (3.8):

$$
P(j) = \frac{\exp(V_j)}{\sum_{j' \in s} \exp(V_{j'})} \tag{3.8}
$$

3.4 MORE ON THE OBSERVED COMPONENT OF UTILITY

This section elaborates on some ways to model and estimate V_j, which is the observable component of utility. The notion that an alternative is essentially a bundle of attributes is a useful starting point and hence, following Lancaster (1966), the link from the attributes and the attribute levels to the utility of an alternative is often made. As alluded to in Chapter 2, the alternative decision rules or heuristics that are modelled in V_j will imply a different specification for V_j. A more thorough comparison and discussion of the various candidate functional definitions of V_j is left aside until Chapter 5.

A choice alternative in a DCE is described by any number of attributes. For example, attributes that describe a public transport alternative might include its one way fare, its average travel time, its crowding level and its frequency. Now, assume that each alternative is described by *k* attributes. Let the attribute level of attribute *k* in alternative *j* be denoted by $\underline{X_{jk}}$. Covariates that may be entered into the observed component of utility include data such as the socio economic characteristics of the respondent. The value of each of the *l* covariates may be represented by a vector $\mathbf{z}_l$. Hence, V_j can be represented by some function of the attribute levels and covariates, so that

$$V_j = h(\mathbf{X}_{jk}, \mathbf{z}_l; \boldsymbol{\beta}_{jk}, \boldsymbol{\delta}_l)$$

where

$\boldsymbol{\beta}_{jk}$ is a vector of coefficients associated with the attribute vector $\mathbf{X}_{jk}$

$\boldsymbol{\delta}_l$ is a vector of coefficients associated with the covariate vector $\mathbf{z}_l$

The function *h* need not be linear, although in practice, V_j is commonly parameterised by some linear additive function such as in Equation (3.9):

$$V_j = \sum_k \beta_{jk} X_{jk} + \sum_l \delta_{jl} Z_l \tag{3.9}$$

The parameters of the model may be obtained by maximum likelihood estimation (MLE). Simply put, MLE searches for values of the parameters for which the observed sample data are most likely to have occurred. Let the parameters to be estimated be denoted generically by $\boldsymbol{\theta}$. The likelihood function in the case of the MNL is then

$$L(\boldsymbol{\theta}) = \prod_n \prod_s \prod_j P_{nsj}(\mathbf{X}_{nsj,} \mathbf{z}_n \mid \boldsymbol{\theta})^{y_{nsj}}, \tag{3.10}$$

where $y_{nsj} = 1$ if alternative *j* in choice set *s* was selected by respondent *n* and 0 otherwise.

It is usual to work with the log-likelihood function since the value of $\boldsymbol{\theta}$ that maximises the likelihood function will also maximise the log-likelihood function. Hence, Equation (3.11) follows:

$$LL(\boldsymbol{\theta}) = \sum_n \sum_s \sum_j y_{nsj} \ln[P_{nsj}(\mathbf{x}_{nsj,} \mathbf{z}_n \mid \boldsymbol{\theta})] \tag{3.11}$$

To maximise the value of the log-likelihood function, it is necessary for $\hat{\boldsymbol{\theta}}_{MLE}$ to satisfy the following first order conditions:

$$\left. \frac{\partial LL(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\hat{\theta}_{MLE}} = \mathbf{0} \tag{3.12}$$

For $LL(\boldsymbol{\theta})$ to be at a local maximum rather than a local minimum when $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{MLE}$, the second order conditions must also be satisfied, that is, the Hessian matrix must be negative semidefinite when evaluated at $\hat{\boldsymbol{\theta}}_{MLE}$. The MNL model with a linear representation for V_j leads to a globally concave log-likelihood function and the existence and uniqueness of the MLE estimator is therefore guaranteed (McFadden, 1974).

3.5 THE NON-LINEAR RANDOM PARAMETERS LOGIT MODEL

In many cases, the heuristics tested in this thesis will require V_j to be specified as a non-linear function in the attributes and respondent characteristics, hence necessitating the use of a non linear logit model. Some of these functional forms have been alluded to earlier in Chapter 2, and will be more extensively discussed in the relevant empirical sections of Chapter 5.

The non-linear model may also be extended to account for random parameters. The subscript *n* for each respondent is now introduced which means that the set of parameters $\boldsymbol{\theta}$ can potentially vary across respondents. Assuming each respondent *n* is faced with S_n choice sets. For any respondent *n* in each choice situation *s*, the more general model for the conditional probability, assuming non linear utility functions, is given by Equation (3.13):

$$P_{ns}(j \mid v_n) = \frac{\exp[h(\mathbf{x}_{nsj}, \mathbf{z}_{\mathbf{n}}, \boldsymbol{\theta}_n^{'})]}{\sum_{j \in s} \exp[h(\mathbf{x}_{nsj}, \mathbf{z}_{\mathbf{n}}, \boldsymbol{\theta}_n^{'})]} \tag{3.13}$$

where

$$\boldsymbol{\theta}_n = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{pmatrix}_n$$

$\boldsymbol{\beta}_n = \boldsymbol{\beta} + \Psi \mathbf{z}_n + \boldsymbol{\Gamma} \mathbf{v}_n = \boldsymbol{\beta} + \Psi \mathbf{z}_n + \boldsymbol{\eta}_n$

$\boldsymbol{\delta}_n = \boldsymbol{\delta} + \boldsymbol{\Gamma} \mathbf{v}_n = \boldsymbol{\delta} + \boldsymbol{\eta}_n$

Ψ is the matrix of structural parameters on the observed heterogeneity

$\mathbf{v}_n \sim$ with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$

$\boldsymbol{\Gamma}$ is a full unrestricted lower triangular matrix

The probabilities $P_{ns}(j \mid v_n)$ have to be obtained by simulated replications. Let the joint conditional probability of any replication *r* be given by

$$P_n(j;r) = \prod_{s=1}^{S_n} P_{ns}[j \mid \mathbf{v}_n(r)] \tag{3.14}$$

The unconditional choice probability $P_n(j)$ is obtained by simulating the conditional probabilities $P_{ns}(j;r)$ *r* times over the distribution $f(\boldsymbol{\eta}_n \mid \mathbf{z}_n, \boldsymbol{\Psi}, \boldsymbol{\Gamma}, \boldsymbol{\theta})$:

$$P_n(j) = \frac{1}{R}\sum_r P_n(j;r) \tag{3.15}$$

so that the simulated log-likelihood function is given by

$$LL_S(\boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Gamma} \mid \mathbf{x}, \mathbf{z}) = \sum_n \log P_n(j) \tag{3.16}$$

Simulation methods vary but one common method is to use intelligent or pseudo random draws such as Halton sequences to ensure more uniform coverage over the distribution (Train, 1999).

Denote by vec($\boldsymbol{\psi}$) and vec($\boldsymbol{\Gamma}$) the column vectors formed by stacking the rows of $\boldsymbol{\psi}$ and $\boldsymbol{\Gamma}$, respectively. Then,

$$\frac{\partial \log LL_S(\boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Gamma} \mid \mathbf{X}, \mathbf{y}, \mathbf{z})}{\partial \begin{pmatrix} \boldsymbol{\theta} \\ vec(\boldsymbol{\Psi}) \\ vec(\boldsymbol{\Gamma}) \end{pmatrix}} = \sum_n \frac{1}{P_n(j)} \frac{1}{R} \sum_{r=1}^{R} [P_n(j;r)] \sum_{s=1}^{S_n} \begin{pmatrix} \left(\mathbf{g}_j[n,s,(r)] - \bar{\mathbf{g}}[n,s,(r)]\right) \\ \left(\mathbf{g}_j[n,s,(r)] - \bar{\mathbf{g}}[n,s,(r)]\right) \otimes \ \mathbf{z}_n \\ \left(\mathbf{g}_j[n,s,(r)] - \bar{\mathbf{g}}[n,s,(r)]\right) \otimes \ \mathbf{v}_n \end{pmatrix} \tag{3.17}$$

where

$$\mathbf{g}_j[n,s,(r)] \ = \ \frac{\partial h_j[\mathbf{x}_{nsj}, \mathbf{z}_n, \boldsymbol{\theta}_n(r)]}{\partial \boldsymbol{\theta}_n(r)}$$

and

$$\bar{\mathbf{g}}[n,s,(r)] = \sum_{j=1}^{J_{ns}} P_{ns}[j \mid \mathbf{v}_n(r)] \ \mathbf{g}_j[n,s,(r)].$$

For optimisation, the derivatives of the simulated log likelihood function must be simulated as well since the heterogeneity in $\boldsymbol{\theta}$ needs to be averaged out.

3.6 METHODS OF EMBEDDING MULTIPLE HEURISTICS IN CHOICE MODELS

3.6.1 The Probabilistic Decision Process (PDP) Model

If it is believed that there is heterogeneity in decision processes, i.e., different heuristics are employed by different respondents, one possible approach is to appeal to probabilistic decision process (PDP) models where the functional form of the heuristic under consideration is expressed through the utility expressions in each class. The PDP model is essentially a special case of a latent class model. Work by Hensher and Collins (2011), Hess *et al.* (2012) and McNair *et al.* (2012) fall into this category. Typically in such models, each latent class is thought to represent one heuristic. As a matter of interpretation, the modeller is implicitly assuming that each respondent is relying only on one heuristic in the decision process. However, what that heuristic might be for each individual can only be known up to a probability, which is the estimated class membership.

It will be appropriate to describe the PDP model in some detail as there will be scope to apply the PDP model to a study of multiple heuristics in Chapter 5. As mentioned, in the PDP (latent class) model, heterogeneity in the decision rules may be handled via discrete distributions of the parameters where each class represents a decision process. Within each of these classes, parameters and choice probabilities are assumed to be generated by MNL models. Conditional on a decision process *d*,

$U_{nsj|d} = V_{nsj|d} + \varepsilon_{nsj|d}$, where $\varepsilon_{nsj|d} \sim$ *i.i.d.* EV type I.

Hence, $P_{nsj|d}$ follows the logit probability defined earlier.

$$P_{nsj|d} = \frac{\exp(V_{nsj|d})}{\sum_{j' \in s} \exp(V_{nsj'|d})} \tag{3.18}$$

Since the decision process is not directly observed, decision processes may only be specified up to a probability. This probability specification makes use of the MNL logit form.

$$\Pr(\text{decision process} = d) = \frac{\exp(\alpha_d)}{\sum_{d=1}^{D} \exp(\alpha_d)} \tag{3.19}$$

In Equation (3.19), the decision process-specific probabilities are specified as a set of fixed constants, with one of these constants set to zero. However, if required, the decision process probabilities themselves may be considered as functions of respondent specific characteristics.

For a given respondent *n*, the probability of choosing alternative *j* in choice set *s* is the expected value of $P_{nsj|d}$ over all the decision process probabilities, that is,

$$\begin{aligned} P_{ns}(j) &= \sum_{d=1}^{D} \Pr(\text{decision process} = \text{d}) P_{nsj|d} \\ &= \sum_{d=1}^{D} \frac{\exp(\alpha_d)}{\sum_{d=1}^{D} \exp(\alpha_d)} P_{nsj|d} \end{aligned} \tag{3.20}$$

In a panel dataset where multiple observations are captured per respondent, the probability of observing a sequence of choices for some alternative *j* in choice sets $s = 1$ through to $s = S_n$ is

$$\begin{aligned} P_n(j_{(s=1)}, j_{(s=2)}, \ldots, j_{(s=S_n)}) &= \sum_{d=1}^{D} \left[\Pr(\text{decision process} = \text{d}) \prod_{s} P_{nsj|d} \right] \\ &= \sum_{d=1}^{D} \frac{\exp(\alpha_d)}{\sum_{d=1}^{D} \exp(\alpha_d)} \prod_{s} P_{nsj|d} \end{aligned} \tag{3.21}$$

3.6.2 The Hybrid RRM-RUM Model

Another approach of invoking multiple heuristics in a utility function is to assume heterogeneity in the way subsets of attributes of an alternative are processed by the respondents. The hybrid RRM-RUM model advocated by Chorus *et al.* (2013) is an example of this approach. In these hybrid RRM/linear additive RUM models, respondents are assumed to process a subset of attributes according to RRM, and the remaining attributes of the alternatives according to a linear additive processing rule. If it is assumed that attributes *1,...,* *m* of alternative *j* are processed according to linear additive RUM and attributes *m+1, ..., K* are processed according to RRM, then the observed component of utility can be described by Equation (3.22):

$$V_j = \beta_{0,j} + \sum_{k=1,\dots,m} \beta_{jk} X_{jk} - \sum_{\substack{j' \in s,\\ j' \neq j}} \sum_{k=m+1,\dots,K} \ln(1 + \exp(\beta_{j'k} X_{j'k} - \beta_{jk} X_{jk})) \tag{3.22}$$

3.6.3 Heuristic Weighting Functions

Finally, another suggested alternative to the latent class model approach that will be advocated in this thesis is to weight each heuristic directly in the utility function. Each heuristic, which can be given its own functional specification, can be embedded within the utility function for each alternative. By weighting each heuristic, this approach allocates the proportional contribution of each heuristic to overall utility, with the possibility of linking this share outcome to the characteristics of respondents and other possible contextual influences.

In a model with a total of *M* heuristics, the weights of each heuristic, denoted by W_m, $m=1,2,\dots,M$ can be given by means of an exponential function shown in Equation (3.23):

$$W_m = \exp(\sum_l \delta_{lm} z_l) \tag{3.23}$$

z_l has been defined earlier and it denotes the value of variable *l* which is typically a socio-economic variable or a variable describing context characteristics. δ_{lm} is a parameter weight that is allowed to vary according to each of the *l* variables and each of the *m* heuristics. To ensure identification of the model, it will be necessary to normalise, for every variable *l*, one δ_{lm}.

This multiple heuristics approach will be examined in Chapter 5 as a 'mixture' of the standard fully compensatory, linear-in-the-parameters RUM decision rule in conjunction with the models for the extremeness aversion heuristic, such as the RAM model described in Chapter 2.

3.7 MODEL COMPARISONS

A key research question addressed by the empirical analysis in Chapter 5 relates to whether models of context dependent heuristics and models of multiple heuristics perform better than the standard linear additive, context independent utility specification. Therefore, statistical tests to distinguish between models are necessary and these are described in this section.

3.7.1 Comparing Nested Models with the Likelihood Ratio Test

When deciding which of two discrete choice models provide the better fit for the data, the likelihood ratio test may be used in cases where one of the models is nested in the other, that is, when one model is a restricted version of the other.

Suppose there are two models, $m = 1, 2$, to be compared, which each model using K_m parameters to explain the same choices. Model 1 is assumed to be nested in Model 2. Let LL_m be the sample log-likelihood of model m at convergence. Under the null hypothesis that Model 1 is the correct specification, then the statistic $-2(LL_1 - LL_2)$ follows a χ^2 distribution with degrees of freedom given by $K_2 - K_1$. If this statistic is "large enough", as defined by the five percent critical cut-off value, then the null hypothesis may be rejected and the unrestricted Model 2 can be said to be preferred to the restricted Model 1.

3.7.2 Comparing Non-Nested Models with the Ben-Akiva and Swait (1986) Test

Several statistical tests are available in cases where one of the models is not a restricted form of the other, that is, when model specifications are non-nested. In the case of several models of heuristics estimated in Chapter 5, functional specifications of the context dependent heuristics in the indirect utility function are not nests of each other. For non-nested models, two widely used tests are due to Ben-Akiva and Swait (1986) and Vuong (1989). However, in a series of Monte Carlo simulations, Strazzera *et al.* (2013) find that the Vuong test has much lower power compared to the Ben-Akiva and Swait test, implying that the Vuong test too often fails to reject the null of no differences in model when the alternative hypothesis is true.

With this in mind, the Ben-Akiva and Swait test is therefore used as the primary method of comparing non-nested models in this thesis.

The Ben-Akiva and Swait test may be described as follows.

Again, suppose there are two models, $m = 1, 2$, to be compared, with the difference this time being that Model 1 is not nested in Model 2. As before, denote by K_m the number of parameters used by each model to explain the same choices. Define $LL(0)$ as the log-likelihood of the equal probabilities naïve model, that is, assuming that all parameters are zero and therefore choices are random. Again, let LL_m be the sample log-likelihood of model m at convergence. Now define a goodness of fit measure, as in Equation (3.24):

$$\bar{\rho}_m^2 = 1 - \frac{LL_m - K_m}{LL(0)} \tag{3.24}$$

Assuming that model 2 has the better goodness-of-fit, i.e., $\bar{\rho}_2^2 > \bar{\rho}_1^2$, then Ben Akiva and Swait (1986) show that under the null hypothesis that Model 1 is the correct specification, the following asymptotic bound holds:

$$\Pr(\bar{\rho}_2^2 - \bar{\rho}_1^2 \geq Z) \leq \Phi\left(-\sqrt{-2 \times Z \times LL(0) + (K_2 - K_1)}\right) \tag{3.25}$$

In Equation (3.25), Z is the difference in the goodness-of fitness measures between Model 2 and Model 1 and it is assumed to be greater than zero. Φ is the standard normal cumulative distribution function.

Equation (3.25) may be interpreted as describing an upper bound for the probability of erroneously choosing Model 2 when in fact Model 1 is the true specification. Using the conventional five percent significance level, the statistic $-\sqrt{-2 \times Z \times LL(0) + (K_2 - K_1)}$ needs to be smaller than –1.64 in order to justify the rejection of the null hypothesis, that is, for Model 2 to be preferred to Model 1.

3.8 A PROCEDURE FOR ANALYSING NON-LINEAR FUNCTIONS OF PARAMETERS

Typically, outputs of interest that are derived from discrete choice models include willingness to pay measures and calculations of elasticity. As the models will be estimated in utility space, these outputs are non-linear functions of the parameters and in the case of the extremeness aversion heuristic, where reference is made to the other attributes of competing alternatives in the choice set, outputs are non-linear in the sample data as well. As is well-known, even in a simple linear additive model where the utility of an alternative, say a transport mode, is given by a price and time attribute such as $V_j = \beta_1 \times price + \beta_2 \times time$, the marginal willingness to pay for a unit of travel time saved is non-linear and is equal to $\frac{\beta_2}{\beta_1}$.

The delta method described below is a post-estimation procedure that may be used to analyse such non-linear functions, for example, to obtain standard errors of the functions of the parameters, where the parameters are estimated with some uncertainty. This procedure is used for where parameters are assumed to be fixed in the model.

Suppose that *i* non-linear functions of the form $\omega_i(\boldsymbol{\beta})$ are required. Let $\omega(\boldsymbol{\beta})$ denote the vector of these non-linear functions. Furthermore, let $c_i(\boldsymbol{b})$ be the sample estimate of $\omega_i(\boldsymbol{\beta})$. Denote by $\boldsymbol{\mu}_\mathrm{i}$ the set of partial derivatives

$$\frac{\partial \omega_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \boldsymbol{\mu}_\mathrm{i}$$

$\boldsymbol{\mu}_\mathrm{i}$ may be estimated by inserting the parameter estimates $\boldsymbol{b}$ into the function defined by $\boldsymbol{\mu}_\mathrm{i}$. Denote such estimates of $\boldsymbol{\mu}_\mathrm{i}$ by $\boldsymbol{m}_i$. Denoting by $\hat{\Sigma}$ the estimated asymptotic variance-covariance matrix of $\boldsymbol{\beta}$, then the estimated asymptotic variance-covariance matrix of $\omega(\boldsymbol{\beta})$ is given by Equation (3.26):

$$\mathbf{G} = \mathbf{M} \times \hat{\boldsymbol{\Sigma}} \times \mathbf{M}', \tag{3.26}$$

where the rows of $\mathbf{M}$ are given by $\boldsymbol{m}_i$.

Where the non-linear functions involve sample data, such that $\omega_i(\boldsymbol{\beta}, x)$, the functions themselves can either be evaluated at the means of the data or they can be averaged over the sample observations. In the first case, the estimate of $\omega_i(\boldsymbol{\beta}, x)$ is given by $c_i(\mathbf{b}, \bar{x})$ and in the second case, the function $\omega_i(\boldsymbol{\beta}, x)$ is given by $\frac{1}{N}\sum_{n=1}^{N} c_i(\mathbf{b}, x_n)$.

3.9 CONCLUSION

This chapter has reviewed some of the econometric theory of discrete choice modelling that can be called upon to investigate the role of various decision heuristics and rules in choice behaviour. This body of theory is well established and is frequently referenced in the literature. Having described the methodology that will be used to analyse the data, the next chapter will introduce the datasets themselves.

CHAPTER 4 DESCRIPTION OF THE DATA

4.1 OVERVIEW OF THE DATASETS

This chapter describes the datasets that will be used for the empirical application in Chapter 5. Seven stated choice (SC) datasets have been selected for analysis. These datasets are from five Australian and two New Zealand toll road studies that were conducted over a ten year period between 1999 and 2008. They are very similar to one another in the sense that the centrepiece of each dataset comprises a SC experiment described by three unlabelled alternatives in the context of a sample of commuters choosing among bundles of trip time, trip variability and cost attributes associated with tolled and non-tolled routes. Each of the SC experiments required every sampled respondent to complete 16 choice tasks by choosing what they considered to be the best alternative from the set of alternatives on offer.

In each of these datasets, respondents were asked to provide a profile of a recent trip, in terms of the attribute levels for the attributes of interest. Attribute levels of the other two hypothetical alternatives in the choice task were then generated by pivoting off the values of the recent (or reference) commuting trip profile provided by the respondent. Pivoting is said to offer more realism in the SC experiment since hypothetical alternatives are defined relative to the reference alternative (status quo), giving better specificity in the context of the choice task (Train and Wilson, 2008). Since the attributes of the recent trip are based on actual market data, the inclusion of the recent trip profile as one of the alternatives in the choice set ensures a sufficiently desirable or at the very least, a no worse-off option, should the hypothetical alternatives, for whatever reason, be completely unappealing to the respondent. All seven surveys were conducted as computer aided personal interviews (CAPI).

With the exception of NZ99, which was collected in 1999 and thereby making it the oldest among the seven datasets, all the other SC experiments also asked a second choice question which required the respondent to assume that the recent trip was unavailable and to therefore make a choice between the two hypothetical alternatives only.

All datasets include running costs and toll costs, as well as free flow time; however non-free flow time may be defined as either a single attribute ‘congested time’ or two separate attributes ‘slowed down time’ and ‘stop/start/crawling time’. Given that the sampled New Zealand’s car commuters had no tolling experience before they were interviewed, self reported toll costs are only available for Australian studies. The labelling of the datasets follows the convention *countryyear*, where *country* is either Australia (“Aust”) or New Zealand (“NZ”) and *year* is the year of collection, which can be 1999 (“99”), 2000 (“00”), 2004 (“04”), 2005 (“05”), 2007 (“07”) or 2008 (“08”). Hence, a dataset which was collected in Australia in 2008 is labelled Aust08.

Each dataset is described more extensively below in Section 4.2. Further details of the toll road datasets may also be found in Hensher *et al.* (2012c). Section 4.3 discusses some summary statistics of each of the seven datasets and Section 4.4 provides some concluding remarks.

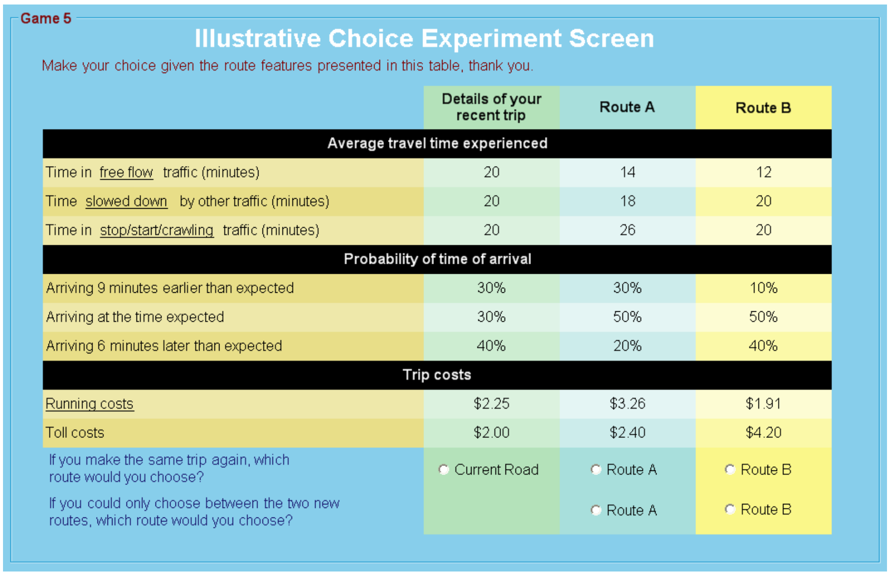
4.2 DETAILED DESCRIPTIONS OF DATASETS

4.2.1 Australia, 2008 (Aust08)

Undertaken in 2008, this study is the most recent of all the seven toll road datasets. The data are drawn from a commuter study undertaken in a metropolitan area in Australia in the context of toll versus free roads. The three alternatives shown in each choice set, that is, the recent trip and two hypothetical alternatives (i.e., route *A* and route *B*), were described in terms of free flow time, slowed down time, stop/start/crawling time, running cost, toll cost, and travel time variability (see Table 4.1). Figure 4.1 provides an illustrative screen capture of the experiment.

Table 4.1: Trip Attributes in Stated Choice Design

Routes *A* and *B*
Free flow travel time
Slowed down travel time
Stop/start/crawling travel time
Arriving *x* minutes earlier than expected
Arriving *y* minutes later than expected
Probability of arriving earlier than expected
Probability of arriving at the time expected
Probability of arriving later than expected
Running cost
Toll Cost



Game 5

Illustrative Choice Experiment Screen

Make your choice given the route features presented in this table, thank you.

	Details of your recent trip	Route A	Route B
Average travel time experienced			
Time in free flow traffic (minutes)	20	14	12
Time slowed down by other traffic (minutes)	20	18	20
Time in stop/start/crawling traffic (minutes)	20	26	20
Probability of time of arrival			
Arriving 9 minutes earlier than expected	30%	30%	10%
Arriving at the time expected	30%	50%	50%
Arriving 6 minutes later than expected	40%	20%	40%
Trip costs			
Running costs	$2.25	$3.26	$1.91
Toll costs	$2.00	$2.40	$4.20
If you make the same trip again, which route would you choose?	Current Road	Route A	Route B
If you could only choose between the two new routes, which route would you choose?		Route A	Route B

Figure 4.1: Illustrative Screenshot of Choice Experiment used in Aust08

Compared to the other toll road datasets, Aust08 is unique in terms of how travel time variability is portrayed, where each alternative has three travel scenarios - 'arriving *x* minutes earlier than expected', 'arriving *y* minutes later than expected', and 'arriving at the time expected'. Each time is associated with a corresponding probability[5] of occurrence to indicate that travel time is not fixed but can vary stochastically.

Aust08 used a D-efficient experimental design structured to increase the statistical performance of models with relatively smaller samples than are required for other less-efficient (statistically) designs such as orthogonal designs (see e.g., Rose *et al.* 2008). For all attributes except the toll cost, minutes arriving early and late, and the probabilities of arriving on-time, early or late, attribute levels of the hypothetical alternatives are obtained by pivoting around the knowledge base of travellers, that is, variations around the values for the current trip. Given the lack of exposure to tolls for many travellers in the study catchment area, the

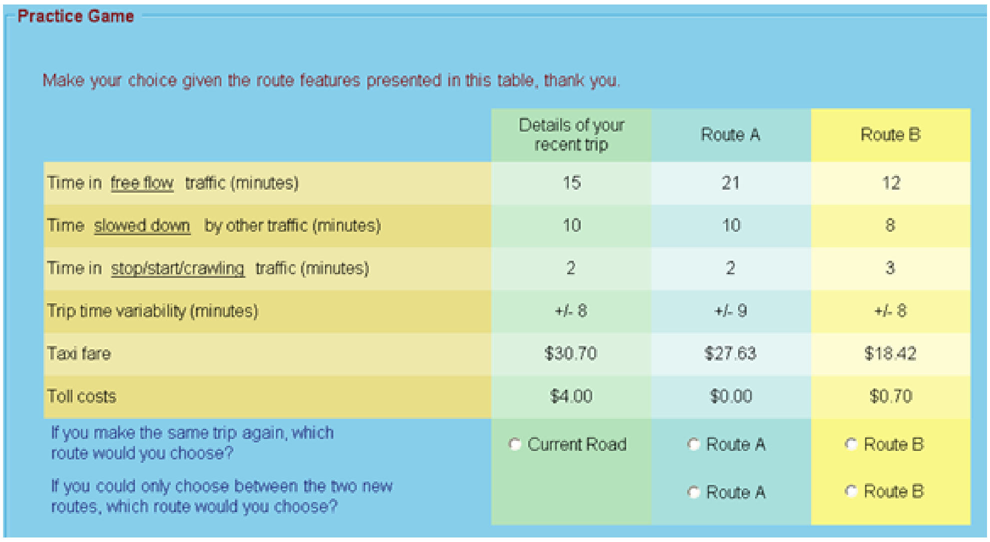
[5] The probabilities are designed and hence exogenously induced from each respondent's perspective, similar to other travel time reliability studies.

toll levels are fixed over a range, varying from no toll to $4.20, with the upper limit determined by the trip length of the sampled trip.

A telephone call was used to establish eligible participants from households. During the telephone call, a time and location were agreed for a face-to-face CAPI. In total, 280 commuters (with less than 120 minutes' trip length) were sampled for this study, each responding to 16 choice tasks, resulting in 4,480 observations for model estimation.

4.2.2 Overview of Other Datasets

The surveys used in the other datasets (Australia00, NZ99, Australia05, Australia04a, Australia04b and NZ07 are similar to that shown in Figure 4.2. An orthogonal design was used in Aust00 and NZ99 and a D-efficient design was used for Aust04a, Aust04b and NZ07. All studies allowed the disaggregation of trip cost into the running cost and the toll cost. In terms of travel time, respondents in NZ99 and Aust05 were shown the three time components of free flow time, slowed down time, and stop/start/crawling time; while respondents in Aust00, Aust04a, Aust04b and NZ07 were given the two time components of free flow time and congestion time for consideration.


Practice Game

Make your choice given the route features presented in this table, thank you.

	Details of your recent trip	Route A	Route B
Time in free flow traffic (minutes)	15	21	12
Time slowed down by other traffic (minutes)	10	10	8
Time in stop/start/crawling traffic (minutes)	2	2	3
Trip time variability (minutes)	+/- 8	+/- 9	+/- 8
Taxi fare	$30.70	$27.63	$18.42
Toll costs	$4.00	$0.00	$0.70
If you make the same trip again, which route would you choose?	Current Road	Route A	Route B
If you could only choose between the two new routes, which route would you choose?		Route A	Route B

Figure 4.2: Illustrative Screenshot of Choice Experiment Used in Other Datasets

In contrast to Aust08, where three arrival scenarios along with their probabilities of occurrence for a trip were presented in the choice experiments, the other six studies defined the trip time variability attribute as *plus* or *minus* a level of trip time associated with a trip.

Given previous evidence using these other datasets that the variability attribute was poorly specified and often not statistically significant, trip time variability in all model estimation is excluded, despite the innovation in the trip variability attribute in Aust08 (Hensher *et al.*, 2012c).

4.2.3 Australia, 2000 (Aust00)

The data in this dataset come from a study undertaken in Australia in 2000. The sample of 147 effective interviews, each responding to 16 choice sets, resulted in 2,352 observations for model estimation. The attributes included in the choice experiment are free flow time, congestion time, trip time variability, running cost, toll cost and toll payment options (cash, Electronic/Tag, and Electronic/Licence plate recognition (no tag required)). Attributes except for toll payment of the hypothetical alternatives are based on the values for the current trip in terms of travel times and cost (including tolls if a toll was paid). In the design of the choice experiment, important considerations were given to the range of the toll costs (between $0 and $16) and to logical outcomes. Examples of such considerations include requiring longer trip lengths to be associated with higher toll costs, and for hypothetical alternatives involving a toll to be mostly faster than the current trip, in cases where the current trip is not tolled. It is also assumed that a faster total travel time correlates with a higher toll cost; a lower running cost, and lower free-flow time, congestion time and trip time variability.

4.2.4 New Zealand, 1999 (NZ99)

This study was conducted in late June and early July of 1999, sampling residents of seven cities and regional centres in New Zealand. Given a sample size of 152 car commuters evaluating 16 choice sets each, 2,432 choice observations were obtained.

The design is based on two unlabelled alternatives each defined by six attributes each of four levels (i.e., 4^{12}): free flow travel time, slowed down travel time, stop/start travel time, uncertainty of travel time, running cost and toll costs. Except for toll costs, the levels are *proportions* relative to those associated with a current trip identified prior to the application of the SC experiment. Including the current alternative, described by the exact same six attributes as the two hypothetical alternatives, the design starts with six columns of zeros for the last trip attributes followed by six attributes for alternative *A* and then six attributes for alternative *B*. The six attributes for alternative *A* are orthogonal to the six columns for alternative *B*, allowing for the estimation of models with complex structures for the random components of the utility expression associated with each of the alternatives (Louviere and Hensher, 2001). The levels of the attributes for both hypothetical alternatives were rotated to ensure that neither alternative *A* nor alternative *B* would dominate the reference trip, and to ensure that alternatives *A* and *B* would not dominate each other. The fractional factorial design has 64 rows. Each respondent was randomly allocated one of the four blocks of 16 choice tasks.

4.2.5 Australia, 2005 (Aust05)

This survey, conducted in 2005, sampled 304 car commuters resident in an Australian Metropolitan Area, resulting in 4,864 observations over the entire sample. The trip attributes associated with each alternative are: free flow time, slowed down time, and stop/start/crawling time, travel time variability, toll cost and running cost. For all attributes except the toll cost, the values for the hypothetical alternatives are variations around the values for the current trip. Given the lack of exposure to tolls for many travellers in the study catchment area, the toll levels are fixed over a range, varying from no toll to $8, with the upper limit determined by the trip length of the sampled trip. A D-efficient design was used for the experiment.

4.2.6 Australia, 2004 (Aust04a)

The data collected in this study is from a study undertaken in Australia in 2004. Each alternative is described in terms of free flow time, congestion time, trip time variability, running cost and toll cost. With the exception of trip time variability, the attribute values for the hypothetical alternatives are variations around the counterpart values for the current trip. A D-efficient experimental design method was used. This survey has 57 effective interviews for car commuters, resulting in 912 observations.

4.2.7 Australia, 2004 (Aust04b)

Like Aust04a, the data are also obtained in 2004 from another study undertaken in Australia. The sample of 243 effective interviews, with each respondent answering 16 choice sets, resulted in 3,888 observations for model estimation. To ensure that a large number of travel circumstances were captured, the sample consisted of individuals who had recently undertaken trips of various travel times, in locations where toll roads currently exist. To ensure some variety in trip length, three segments were investigated: no more than 30 minutes, 31 to 60 minutes, and more than 61 minutes (capped at two hours). A telephone call was used to establish eligible participants from households stratified geographically. A statistically efficient design that is pivoted around the knowledge base of travellers is used to establish the attribute packages in each choice scenario. The trip attributes associated with each route are free flow time, congestion time, trip time variability, running cost and toll cost.

4.2.8 New Zealand, 2007 (NZ07)

The main field survey was undertaken in New Zealand in 2007. Like Aust04a and Aust04b, the trip attributes associated with each route are free flow time, congestion time, trip time variability, running cost and toll cost. For all attributes except the toll cost, values in the hypothetical alternatives were variations around the values for the current trip. Given the lack of exposure to tolls for many NZ travellers in the study catchment area, the toll levels were fixed over a range, varying from no toll to $4, with the upper limit determined by the trip length of the sampled trip. 115 car commuters were sampled, resulting in 1,864 choice observations for this survey, and a D-efficient design was used to structure the SC experiment.

4.3 SUMMARY STATISTICS

This section provides a description of the datasets from a statistical perspective. To reiterate, trip time variability is excluded in all model estimation given previous evidence that the variability attribute was poorly specified and often not statistically significant. Hence in Aust08, NZ99 and Aust05, the attributes of the alternatives that are modelled in the utility function are: free flow time (FF), slowed down time (SDT), stop/start/crawling time (SST), running cost (RC) and toll cost (TC). In Aust00, Aust04a, Aust04b and NZ07, the attributes modelled in the utility function are: free flow time (FF), congestion time (CT), running cost (RC) and toll cost (TC).

4.3.1 Summary of Key Attributes in the Reference and Hypothetical Alternatives

Table 4.2 presents the summary statistics of each of the datasets in terms of the reference trip and some selected socio-economic characteristics of the respondents. On average, respondents in Aust04a reported the highest average trip travel time across all the seven datasets. Commensurate with this, reported running costs are also highest in Aust04a. The lowest average reported trip time belonged to respondents in NZ99. Similarly, the time spent in congested conditions (i.e., either a sum of slowed down and stop start time or just simply congestion time) is found to be lowest in the New Zealand datasets (NZ99 and NZ07). Recalling that respondents in the New Zealand datasets all did not have any tolling experience in their reference trip, the average toll cost for the reference trip in these datasets is zero.

Some summary statistics of the total trip time and the total trip cost by alternative type and by dataset are presented in Table 4.3. The total trip time may be defined by the summation of free flow, slowed down and stop start time in Aust08, NZ99 and Aust05, and by the summation of free flow time and congestion time in Aust00, Aust04a, Aust04b and NZ07. In all datasets, the total trip cost is defined to be the sum of running costs and toll costs.

As a result of the experimental designs in all datasets, the mean of total travel costs is higher in the hypothetical alternatives than in the reference trip. This feature largely arises from the experimental increases in the toll costs compared to the reported toll costs in the reference trip. On the other hand, in Aust08, NZ99 and Aust05, the means of the total travel time are higher for the hypothetical alternatives than for the reference alternative, whereas in Aust00, Aust04a, Aust04b and NZ07, the means of the total travel time are lower for the hypothetical alternatives than for the reference. The situation for Aust08, NZ99 and Aust05 is rather surprising if it is believed that a higher travel time and travel cost induces disutility, since the “average” hypothetical alternative is now dominated by the “average” reference alternative. Of course, modelling is conducted at the level of individual choice sets, using disaggregated time and cost attributes, and so, this may not be of significant concern in the end.

Table 4.2: A Summary of the Seven Tollroad Studies

	Number of sampled car commuters*	Mean Attribute Levels of Recent Trip (standard deviation in parentheses)						Means of Selected Socioeconomic Variables (standard deviation in parentheses)			
		Free flow minutes (mins)	Slowed down time (mins)	Stop/start/ crawling (mins)	Congestion time (mins)	Running cost ($)	Toll Cost ($)	Age	Gender (proportion male)	Annual Personal Income (thousands)	Hours worked per week
Aust08	280	13.30 (11.42)	11.53 (9.82)	13.59 (14.59)	-	2.95 (2.31)	5.80**	39.44 (13.01)	0.575 (0.494)	53.32*** (31.59)	37.79 (13.65)
Aust00	147	23.35 (15.52)	-	-	16.78 (15.95)	2.72 (2.15)	1.02 (1.53)	42.48 (9.84)	0.694 (0.461)	82.82 (38.27)	39.93 (15.60)
NZ99	152	13.76 (19.49)	5.84 (7.45)	3.96 (5.05)	-	1.61 (3.15)	0	40.74 (12.23)	Not available	16.45 (12.11)	34.30 (15.65)
Aust05	304	12.19 (9.73)	12.66 (9.74)	10.92 (11.09)	-	1.76 (1.33)	2.23**** (0.34)	42.30 (11.86)	0.661 (0.473)	53.59 (30.90)	41.04 (13.77)
Aust04a	57	38.86 (20.83)	-	-	34.93 (19.55)	7.03 (3.75)	1.43 (2.33)	42.74 (11.70)	0.895 (0.307)	82.55*** (34.32)	42.56 (12.81)
Aust04b	243	22.53 (12.37)	-	-	31.80 (19.28)	3.44 (1.66)	2.05 (1.67)	41.70 (11.26)	0.642 (0.479)	92.07*** (35.17)	41.91 (11.60)
NZ07	115	29.03 (15.66)	-	-	9.87 (9.93)	3.99 (2.01)	0	48.02 (12.26)	0.374 (0.484)	42.20*** (21.56)	41.92 (13.83)

Notes: * Model estimation has a multiple of 16 times the number of sampled car commuters.

** Only one sampled respondent paid a toll. This is the reported toll amount paid.

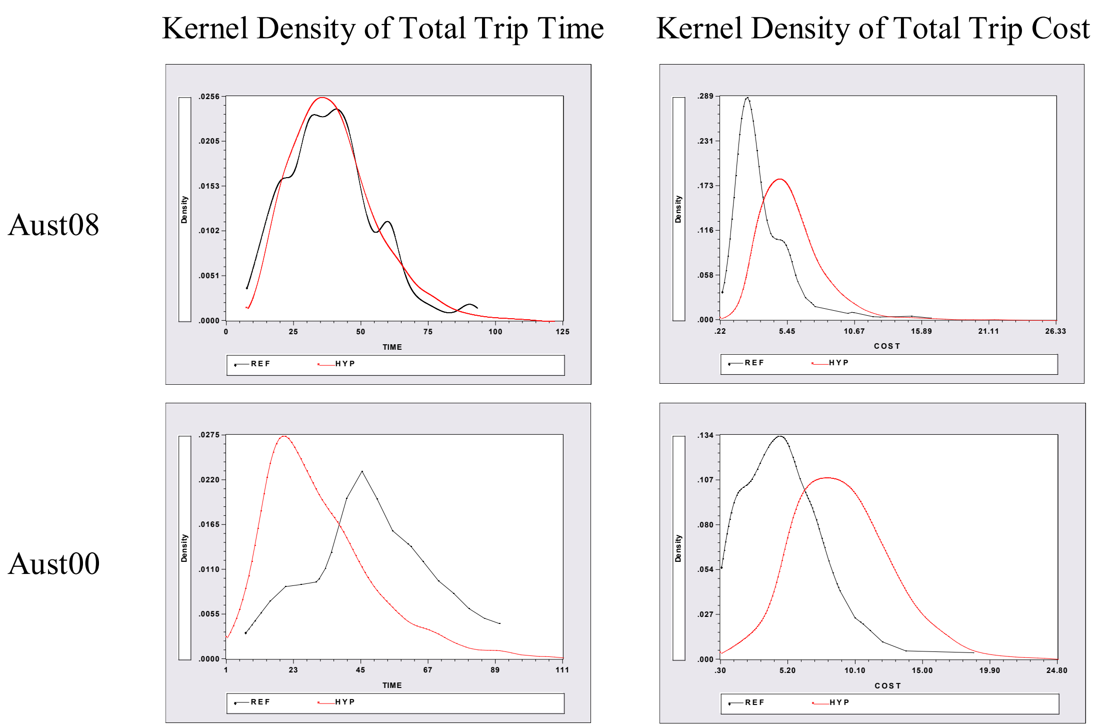
*** Some respondents did not report an income variable in these datasets.

**** Averaged over the 24 respondents who reported paying a toll.

Table 4.3: Summary Statistics of Total Trip Time and Total Trip Cost

	Mean Attribute Levels of Recent Trip (standard deviation in parentheses)		Mean Attribute Levels of Hypothetical Alternatives (standard deviation in parentheses)	
	Total travel time (mins)	Total travel costs ($)	Total travel time (mins)	Total travel costs ($)
Aust08	38.42 (16.67)	2.97 (2.37)	39.72 (16.51)	5.35 (3.00)
Aust00	40.13 (20.81)	3.74 (3.05)	26.93 (18.85)	7.91 (4.04)
NZ99	23.56 (23.72)	1.61 (3.15)	24.21 (24.63)	2.82 (3.46)
Aust05	35.77 (15.51)	1.94 (1.59)	43.24 (18.03)	4.55 (3.01)
Aust04a	73.79 (26.27)	8.45 (4.68)	68.36 (29.95)	10.98 (5.36)
Ausralia04b	54.33 (20.98)	5.49 (2.77)	50.27 (23.50)	6.45 (3.04)
NZ07	38.90 (19.84)	3.99 (2.01)	37.33 (20.46)	5.55 (2.46)

Figure 4.3 illustrates the kernel density plots of the total trip time and total trip cost in each of the datasets. The kernel density plots of these two attributes are partitioned by the type of alternative, either the reference alternative or the two hypothetical alternatives.



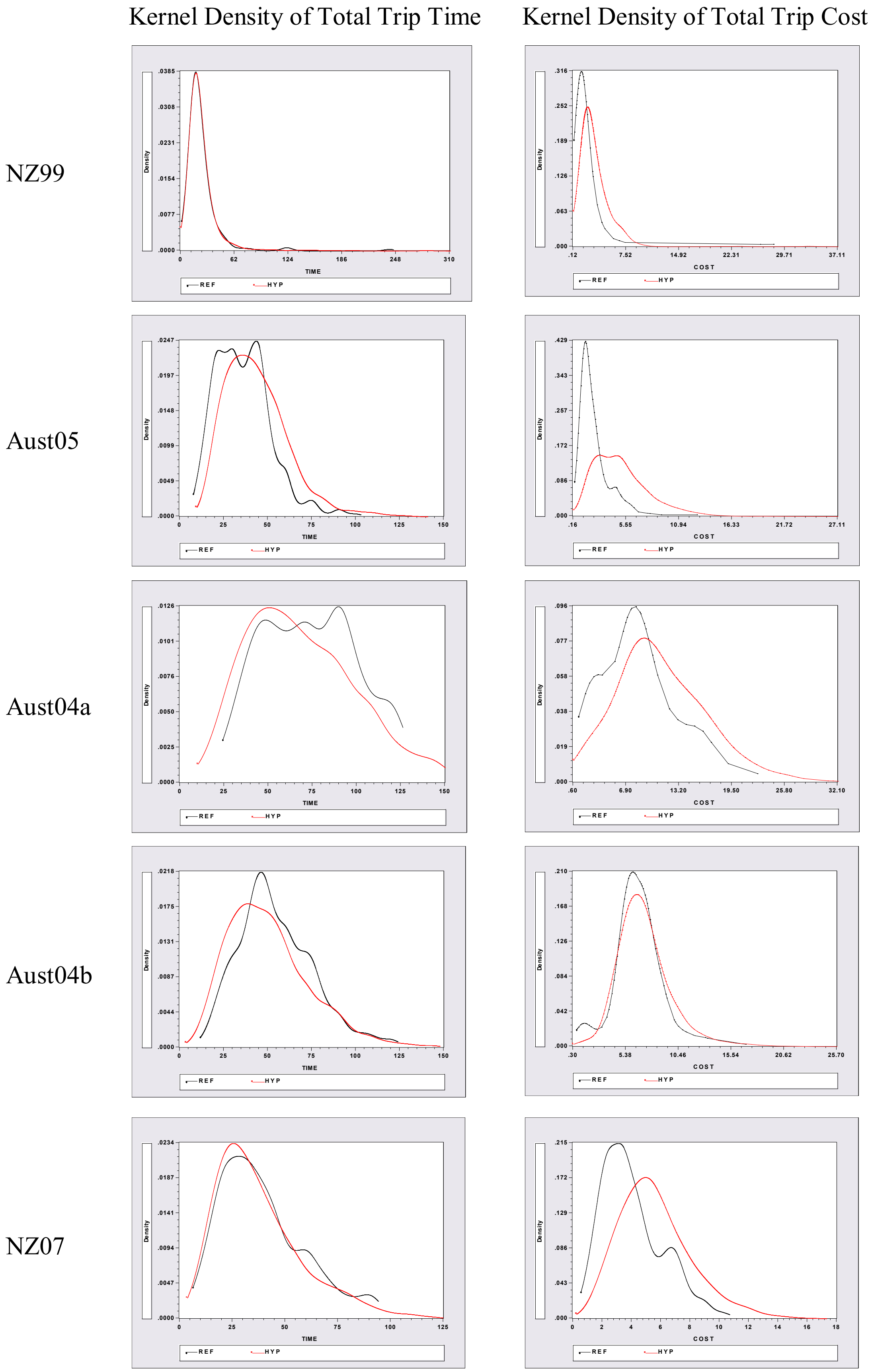


Figure 4.3: Kernel Densities of Total Trip Time and Total Trip Cost, by Dataset

Across all seven datasets, it may be observed that the distributions of total trip time and total trip cost in the reference alternative (indicated in black) are likely to be multi-modal, in the sense that several local maxima exist in the distribution, whereas the same distributions in the hypothetical alternative (indicated in red) are closer to having just one clearly defined mode.

The key attribute levels in the two hypothetical alternatives of each of the seven datasets are summarised in Table 4.4. Compared with the summary statistics of the reference trip in Table 4.2, the attribute levels in the hypothetical alternatives are of roughly the same magnitude as those of the reference trip. This feature of the datasets is more clearly seen in Table 4.5 which offers a slightly different perspective to Table 4.4 by summarising the percentage difference of the time and running cost attributes of the constructed hypothetical alternatives to the reference alternative. As for the toll costs attribute, the differences reported in Table 4.5 are absolute (level) differences, rather than percentage differences, in the hypothetical alternatives relative to the pivot alternative The reason for comparing the toll cost attribute on absolute rather than percentage differences is that many respondents reported not having to pay a toll for their reference trip, even in the case of the Australian datasets.

Table 4.4: Summary Statistics of the Key Attributes in the Hypothetical Alternatives

	Mean Attribute Levels of Hypothetical Alternatives (standard deviation in parentheses)					
	Free flow time (mins)	Slowed down time (mins)	Stop/start/ crawling time (mins)	Congestion time (mins)	Running cost ($)	Toll Cost ($)
Aust08	13.48 (10.77)	11.88 (9.09)	14.36 (13.45)	-	3.25 (2.68)	2.10 (1.37)
Aust00	15.58 (12.51)	-	-	11.35 (12.56)	2.28 (1.90)	5.63 (3.58)
NZ99	13.85 (19.98)	6.06 (8.10)	4.29 (5.30)	-	1.61 (3.22)	1.21 (1.42)
Aust05	14.48 (10.97)	14.95 (11.12)	13.81 (12.00)	-	1.85 (1.58)	2.70 (2.14)
Aust04a	35.93 (23.24)	-	-	32.43 (21.34)	6.68 (4.45)	4.30 (2.93)
Aust04b	21.01 (13.50)	-	-	29.26 (20.65)	3.27 (2.03)	3.19 (2.47)
NZ07	27.42 (16.60)	-	-	9.91 (9.63)	3.78 (2.19)	1.77 (1.19)

Table 4.5: Summary of Variation of Attribute Levels around the Pivot Alternative

		Attribute Levels of Hypothetical Alternatives, Relative to Pivot Alternative					
		Free flow time (mins)	Slowed down time (mins)	Stop/start/ crawling time (mins)	Congestion time (mins)	Running cost ($)	Toll Cost ($)
Aust08	Mean Variation (standard deviation)	– 4.99% (22.97)	– 5.01% (23.08)	– 4.98% (22.95)	-	+10% (22.91)	+ $2.08 (1.42)
	Largest Negative Difference/ Largest Positive Difference	– 50%/ +50%	– 50%/ +50%	– 50%/ +50%	-	– 25%/ +45%	– $5.80/+$4.20
Aust00	Mean Variation (standard deviation)	– 34.25% (24.60)	-	-	– 33.96% (25.16)	– 17.59% (13.21)	+$4.62 (3.78)
	Largest Negative Difference/ Largest Positive Difference	– 80%/ +35%	-	-	– 100%/+33%	– 50%/ +21%	– $6/+$15
NZ99	Mean Variation (standard deviation)	0% (19.00)	– 0.6% (42.03)	– 0.4% (43.2)	-	+ 0.1% (19.49)	+$1.21 (1.42)
	Largest Negative Difference/ Largest Positive Difference	– 33%/ +33%	– 100%/ +100%	– 100%/ +100%	-	– 28%/ +33%	– $0/ +$6
Aust05	Mean Variation (standard deviation)	+13.78% (22.19)	+12.83% (27.89)	+12.97% (28.07)	-	+5% (33.54)	+$2.53 (2.20)
	Largest Negative Difference/ Largest Positive Difference	– 33%/ +50%	– 33%/ +60%	– 33%/ +60%	-	– 40%/ +50%	– $3.70/ +$8
Aust04a	Mean Variation (standard deviation)	– 7.53% (29.54)	-	-	– 7.52% (29.52)	– 4.91% (33.56)	+$2.87 (3.25)
	Largest Negative Difference/ Largest Positive Difference	– 60%/ +20%	-	-	– 60%/ +20%	– 52%/ +42%	– $7/ +$6.40
Aust04b	Mean Variation (standard deviation)	– 7.54% (29.49)	-	-	– 7.87% (29.51)	– 4.86% (33.47)	+$1.13 (2.71)
	Largest Negative Difference/ Largest Positive Difference	– 60%/ +25%	-	-	– 60%/ +23%	– 54%/ +43%	– $10/ +6.40
NZ07	Mean Variation (standard deviation)	– 5.64% (23.15)	-	-	– 5.42% (24.00)	– 5.31% (23.14)	+$1.77 (1.19)
	Largest Negative Difference/ Largest Positive Difference	– 33%/ +33%	-	-	– 50%/ +50%	– 30.9%/ +30.4%	– $0/ +$3.51

Computed at the sample means, the running cost attribute and the time components of the hypothetical alternatives are within 10 percent of their respective counterparts in the reference (pivot) alternative in most datasets, with Aust00 being the only exception to this rule. The similarity of the hypothetical alternatives to the reference pivot afforded by the experimental design at the means generally suggests that there is a high degree of realism and balance in the construction of the hypothetical alternatives. At the same time, there is also quite a substantial variation in the attribute levels of the hypothetical alternatives, evidenced by both the standard deviation and also by the largest negative and largest positive change from the reference. As expected, with the datasets being toll road stated choice studies, the average toll cost attribute of the hypothetical alternatives differs substantially from the reference alternative.

4.3.2 Summary of the "Worst" Attribute Levels in Choice Sets

The maximum attribute level for a given attribute *k* over all alternatives *j* in choice set *s* is denoted by $\max_{j \in s}(X_{jks})$. Arising from the discussion in Chapter 2 on the contextual concavity model, $\max_{j \in s}(X_{jks})$ would be of particular empirical interest, since $\max_{j \in s}(X_{jks})$ may be considered to be a reference attribute level in the choice set, in addition to any referencing effects that may arise from the presence of the pivot alternative. As higher levels of time and cost attributes in a route choice alternative are generally associated with increasing disutility; $\max_{j \in s}(X_{jks})$ might therefore be regarded as the worst level that a given attribute *k* can take in a given choice set *s*. Table 4.6 provides some summary statistics of $\max_{j \in s}(X_{jks})$ in each of the datasets for each attribute of interest.

Table 4.7 shows, for each attribute of interest, how $\max_{j \in s}(X_{jks})$ is distributed among the reference and hypothetical alternatives in the choice sets of each dataset. To illustrate this table, in the case of Aust08, the maximum level of the free flow time attribute (across the three alternatives) occurs in the reference alternative in 1,413 choice situations. Similarly, the maximum level of the free flow time attribute occurs in either of the hypothetical alternatives or both in 3,615 choice situations.

Table 4.6: Summary Statistics of $\max_{j \in s}(X_{jks})$

	Mean of $\max_{j \in s}(X_{jks})$ (standard deviation in parentheses)					
	k = Free flow time (mins)	*k* =Slowed down time (mins)	*k* =Stop/start/ crawling time (mins)	*k* =Congestion time (mins)	*k* =Running cost ($)	*k* =Toll Cost ($)
Aust08	15.99 (12.09)	14.11 (10.16)	17.08 (15.31)	-	3.71 (2.96)	2.92 (1.08)
Aust00	23.68 (15.81)	-	-	17.04 (16.31)	2.75 (2.19)	7.68 (3.03)
NZ99	15.85 (22.45)	7.96 (9.86)	5.54 (6.23)	-	1.85 (3.63)	1.72 (1.56)
Aust05	16.43 (11.89)	17.26 (12.27)	16.39 (13.63)	-	2.23 (1.76)	3.73 (1.92)
Aust04a	44.69 (24.25)	-	-	40.30 (22.53)	8.79 (4.84)	5.94 (1.84)
Aust04b	26.10 (14.09)	-	-	36.43 (22.10)	4.30 (2.15)	4.40 (1.97)
NZ07	33.41 (18.47)	-	-	11.94 (11.06)	4.52 (2.36)	2.47 (1.02)

Table 4.7: Incidence of $\max_{j \in s}(X_{jks})$ Across Alternatives

		Incidence of $\max_{j \in s}(X_{jks})$ across alternatives					
	Incidence of $\max_{j \in s}(X_{jks})$	*k* = Free flow time (mins)	*k* = Slowed down time (mins)	*k*= Stop/start/ crawling time (mins)	*k*= Congestion time (mins)	*k*= Running cost ($)	*k*= Toll Cost ($)
Aust08* (4480)	In reference alt	1413	1405	1107	-	365	150
	In hypothetical alts	3615	3674	3904	-	4115	4464
Aust00 (2352)	In reference alt	2100	-	-	2136	2071	75
	In hypothetical alts	477	-	-	738	597	2277
NZ99 (2432)	In reference alt	661	564	668	-	613	152
	In hypothetical alts	1875	2052	2085	-	1827	2432
Aust05 (4864)	In reference alt	540	1025	579	-	1520	85
	In hypothetical alts	4864	4699	4596	-	3344	4781
Aust04a (912)	In reference alt	228	-	-	224	0	0
	In hypothetical alts	684	-	-	688	912	912
Aust04b (3888)	In reference alt	952	-	-	980	0	0
	In hypothetical alts	2952	-	-	2924	3888	3888
NZ07 (1840)	In reference alt	684	-	-	619	745	0
	In hypothetical alts	1156	-	-	1357	1095	1840

*the number of choice sets in each dataset is indicated in parentheses below the dataset identifier

These choice situations are not mutually exclusive, in the sense that if there is a tie between the reference alternative and at least one of the hypothetical alternatives on the maximum attribute level, that choice situation adds to the count of incidence of both the reference alternative and the hypothetical alternatives. However, if the tie for the maximum only occurs between the two hypothetical alternatives, then that choice situation is counted only once. Since this count of the choice situations across the reference and hypothetical alternatives is not mutually exclusive, the sum can exceed the total number of choice sets in the datasets.

Table 4.7 shows that across most datasets, the maximum values of the time and cost components are more likely to be found in either of the hypothetical alternatives. This is probably not unexpected given the pivot nature of the designs. The incidence of either of the two hypothetical alternatives possessing the maximum of a certain time or cost attribute in the choice set generally exceeds 60 percent of the number of choice sets in each dataset, with the major exception being Aust00. Aust00 contrasts with the other datasets in that the reference alternative has a much larger incidence of the maximum values of the free flow, congestion time and running cost attributes. Only in the toll cost attribute do the hypothetical alternatives possess a higher incidence of its maximum attribute level. In Aust04a and Aust04b, the experiment was constructed such that the maxima of the running cost and the toll cost attributes in the choice set are always to be found in at least one of the hypothetical alternatives. Finally, in NZ99 and NZ07, since the reference trip always had zero toll costs across all respondents, the experimental design required the hypothetical alternatives to be always the ones to contain the maximum of the toll cost attribute. In NZ99 particularly, as the reference alternative is found to contain the maximum toll cost attribute level in 152 choice situations and since the toll cost attribute in the reference is equal to zero, the hypothetical alternatives were constructed with zero toll costs in these 152 choice situations.

4.3.3 Summary of Pair Wise Comparisons in Choice Sets

A study into the extremeness aversion heuristic will also include the estimation of the RRM, Hybrid RRM-RUM and the RAM models. A critical behavioural assumption common to these models is the comparison by a respondent of an attribute *k* across all possible pairs of alternatives in the choice set. Therefore, some summary statistics of these pair wise comparisons, reported in Table 4.8a for Aust08, NZ99 and Aust05, and Table 4.8b for Aust00, Aust04a, Aust04b and NZ07, are useful at this point.

Table 4.8a: Summary of Pairwise Comparisons in Aust08, NZ99 and Aust05

		Aust08 (% of 4480)		NZ99 (% of 2432)		Aust05 (% of 4864)	
		Alt A	*Alt B*	*Alt A*	*Alt B*	*Alt A*	*Alt B*
$k = FF$	*Ref* <	48.13	48.13	49.84	49.84	60.59	60.59
	Ref =	11.38	11.38	4.28	4.28	23.01	23.01
	Ref >	40.49	40.49	45.89	45.89	16.40	16.40
	Alt A <		40.38		34.29		44.12
	Alt A =		20.78		31.41		16.39
	Alt A >		38.84		34.29		39.49
$k = SDT$	*Ref* <	48.97	48.97	54.93	54.93	56.00	56.00
	Ref =	11.70	11.70	8.22	8.22	22.12	22.12
	Ref >	39.33	39.33	36.84	36.84	21.88	21.88
	Alt A <		42.99		33.39		39.64
	Alt A =		15.78		33.22		28.08
	Alt A >		41.23		33.39		32.28
$k = SST$	*Ref* <	56.21	56.21	53.62	53.62	62.99	62.99
	Ref =	9.24	9.24	14.97	14.97	18.75	18.75
	Ref >	34.55	34.55	31.41	31.41	18.26	18.26
	Alt A <		43.44		31.87		43.54
	Alt A =		13.75		36.27		17.52
	Alt A >		42.81		31.87		38.94
$k = RC$	*Ref* <	62.5	62.5	49.84	49.84	50	50
	Ref =	0	0	0.33	0.33	0	0
	Ref >	37.5	37.5	49.84	49.84	50	50
	Alt A <		48.88		37.34		37.50
	Alt A =		10.96		25.33		29.81
	Alt A >		40.16		37.34		32.69
$k = TC$	*Ref* <	87.19	87.19	75	75	78.04	80.51
	Ref =	12.46	12.46	25	25	18.22	16.41
	Ref >	0.36	0.36	0	0	3.74	3.08
	Alt A <		40.40		37.5		44.72
	Alt A =		16.38		25		11.06
	Alt A >		43.21		37.5		44.22

Table 4.8b: Summary of Pair Wise Comparisons in Aust00, Aust04a, Aust04b and NZ07

		Aust00 (% of 2352)		Australia 04a (% of 912)		Aust04b (% of 3888)		NZ07 (% of 1840)	
		Alt A	*Alt B*	*Alt A*	*Alt B*	*Alt A*	*Alt B*	*Alt A*	*Alt B*
$k = FF$	*Ref* <	6.04	5.82	50	50	51.44	51.44	38.48	37.83
	Ref =	5.99	6.25	0	0	0.31	0.31	0	0
	Ref >	87.97	87.93	50	50	48.25	48.25	61.52	62.17
	Alt A <		37.07		50		49.59		36.74
	Alt A =		5.36		25		25.41		22.99
	Alt A >		57.57		25		25		40.27
$k = CT$	*Ref* <	5.27	5.31	50.88	50.88	50	50	46.79	47.17
	Ref =	17.56	17.60	0	0	0.31	0.31	5.98	5.98
	Ref >	77.17	77.08	49.12	49.12	49.69	49.69	47.23	46.85
	Alt A <		39.97		25		25		37.17
	Alt A =		20.41		25		25.10		28.91
	Alt A >		39.63		50		49.90		33.91
$k = RC$	*Ref* <	7.10	6.38	50	50	50	50	38.32	37.5
	Ref =	7.53	7.31	0	0	0	0	0	0
	Ref >	85.37	86.31	50	50	50	50	61.68	62.5
	Alt A <		45.75		50		50		40.22
	Alt A =		17.52		0		0		28.53
	Alt A >		36.73		50		50		31.25
$k = TC$	*Ref* <	87.42	85.63	75	75	75	75	83.97	90.76
	Ref =	3.91	5.14	16.67	16.67	6.17	6.17	16.03	9.24
	Ref >	8.67	9.23	8.33	8.33	18.83	18.83	0	0
	Alt A <		49.87		75		75		56.79
	Alt A =		6.25		0		0		5.98
	Alt A >		43.88		25		25		37.23

Since each choice set contains three alternatives, three possible pairs of comparison, that is, *Ref/Alt A*, *Ref/Alt B* and *Alt A/Alt B*, are possible. To understand Table 4.8a and Table 4.8b, note that each cell in the main body of the table represents a pair wise comparison. Therefore, the cell that is at the intersection of the row labelled "*Ref* <" and the column labelled "*Alt A*" indicates the pair wise comparison of the reference alternative and Alternative A in the given attribute *k*. The number reported in the cell is the percentage of choice sets in the dataset where $X_{ref,k} < X_{AltA,k}$. Likewise, the cell at the intersection of the row labelled "*Ref* =" and

the column labelled "*Alt A*" indicates the percentage of choice sets where $X_{ref,k} = X_{AltA,k}$ and
the column labelled "*Alt A*" indicates the percentage of choice sets where $X_{ref,k} = X_{AltA,k}$ and
the cell at the intersection of the row labelled "*Ref >*" and the column labelled "*Alt A*" indicates the percentage of choice sets where $X_{ref,k} > X_{AltA,k}$. The table is similarly interpreted for the alternative pairs *Ref/Alt B* and *Alt A/Alt B*.

As an illustration, taking the example of Aust08 again, the free flow time attribute in the reference alternative is less than its counterpart in alternative A in 48 percent of all choice sets. The free flow time attribute in the reference is equal to its counterpart in Alternative A in 11 percent of the datasets and it is less than its Alternative A counterpart in 40 percent of all Aust08 choice sets.

A summary of the statistics across all datasets shows that on pair wise comparisons alone, the toll cost attribute in the reference alternative is less than the toll cost attribute value in either of the competing (hypothetical) alternatives in at least 75 percent of choice sets. Again, this is not surprising given that choice responses to proposed increases/introduction to toll charges are one of the key research questions addressed by these seven studies. However, pair wise comparisons between Alternatives A and B on the toll cost attribute show a more balanced picture where in many cases, such as in Aust08, Aust00, NZ99 and Aust05, the incidence of $TC_A < TC_B$ is somewhat closer to the incidence of $TC_A > TC_B$. This feature reflects the structure of the experimental design with its randomised allocation to alternatives A and B. With the exception of Aust00, on the other attributes FF, SDT, SST (or CT, where appropriate) and RC, the experimental design has also resulted in more balanced choice sets across all pair wise comparisons across all datasets, in the sense that the incidence of the attribute being higher than its pair wise competition is not too dissimilar from the incidence of the attribute being lower than its pair wise competition. On the other hand, the design of Aust00 resulted in the vast majority of choice sets in which the reference alternative possesses higher levels of free flow time, congestion time and running cost attributes compared with the hypothetical alternatives.

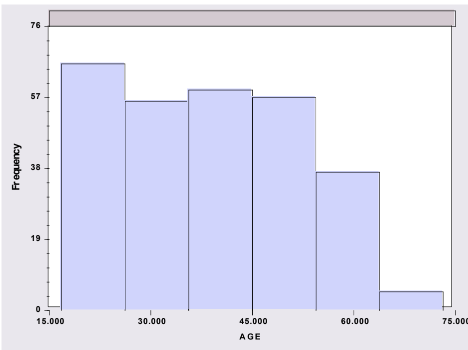
4.3.4 Socio-economic Characteristics

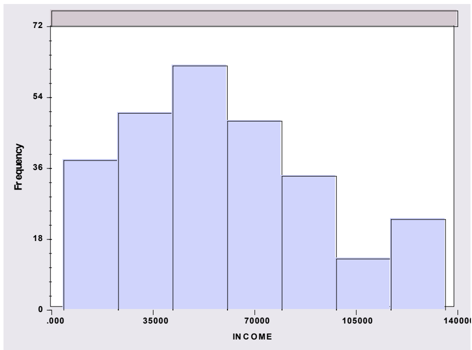
In order to mitigate the intrusive nature of questions asking respondents to report their age and income, the surveys did not ask for exact age and income levels, but merely required respondents to indicate the appropriate age and income bracket to which they belonged. For the age variable, the interval classifications for all datasets are: $age \leq 24$, $24 < age \leq 30$, $30 < age \leq 40$, $40 < age \leq 50$ and $50 < age \leq 60$. For datasets other than Aust00, another category $60 < age \leq 70$ is also available. For modelling purposes, the age of the respondent is coded at the top of the range of the indicated category; hence a discrete distribution for age is obtained. Likewise, the income variable is also coded at the top of the range. These are not unreasonable assumptions if it is believed that people tend to underreport their age and income. Figure 4.4 plots the histogram of the age and income variables for all datasets. In the majority of datasets, the most commonly reported age is between 30 and 40 years old, with modal income between $40,000 and $60,000. Compared to other datasets, it appears that younger people are more heavily represented in Aust08.

Histogram of age variable

Histogram of income variable

Aust08

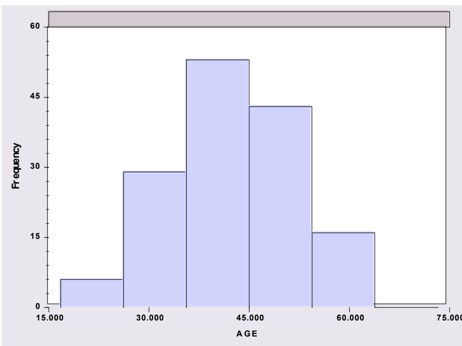


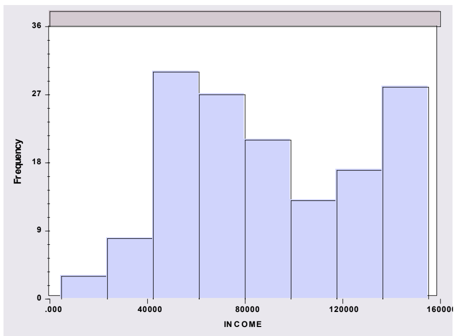


Mode: < 24 years

Mode: $40,000 - $60,000

Aust00



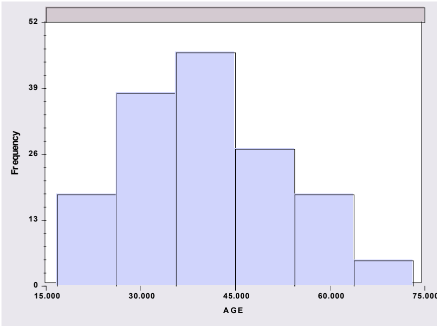


Mode: 30 – 40 years

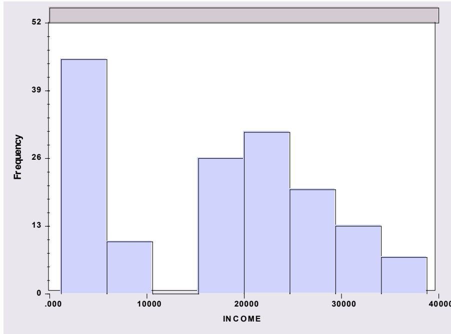
Mode: $40,000 - $60,000

Histogram of age variable | Histogram of income variable

NZ99

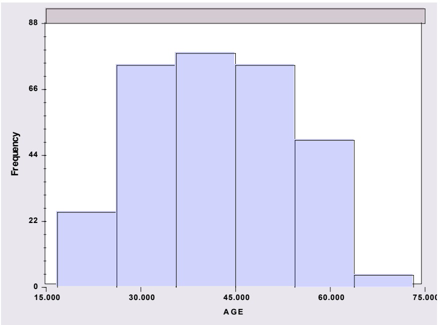


Mode: 30 – 40 years

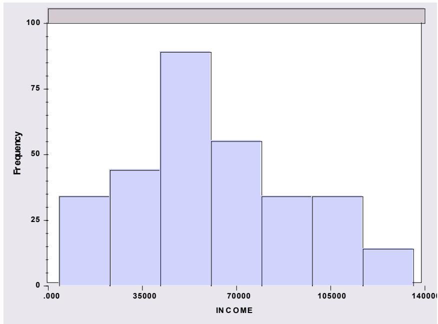


Mode: $0 - $5,000

Aust05

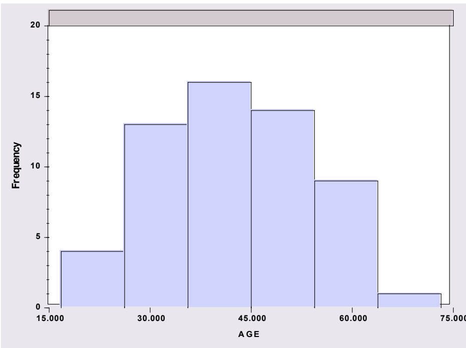


Mode: 30 – 40 years

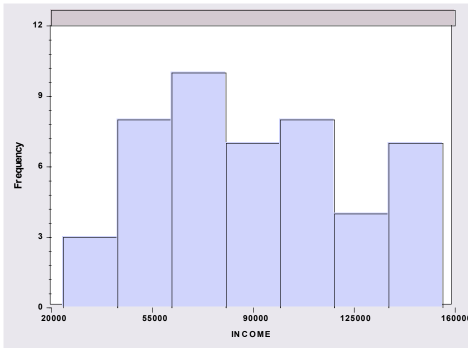


Mode: $40,000 - $60,000

Aust04a

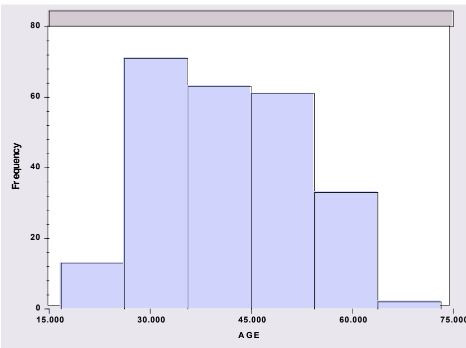


Mode: 30 – 40 years

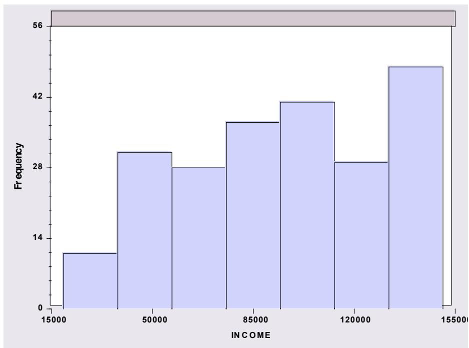


Mode: $60,000 - $80,000

Aust04b



Mode: 24 – 30 years

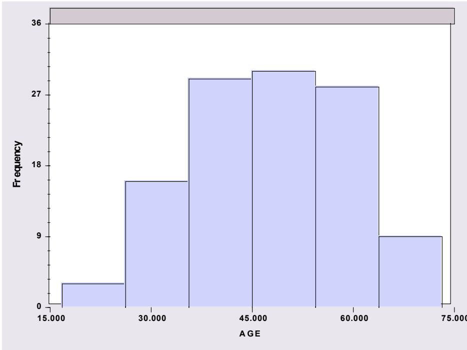


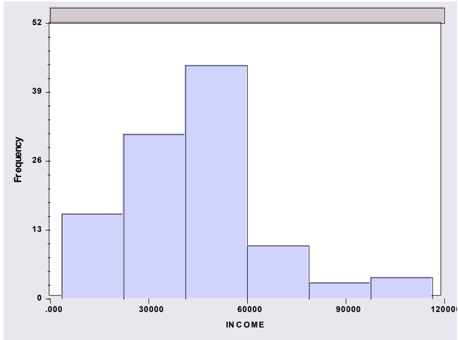
Mode: $120,000 - $140,000

Histogram of age variable

Histogram of income variable

NZ07





Mode: 40 – 50 years

Mode: $40,000 - $60,000

Figure 4.4: Histograms of Age and Income Variables

The income variable is reported in the currency of the country where the experiment was conducted. Some significant differences in the income distribution across datasets may be observed, for example, the mean income in NZ99 is considerably lower than in the other datasets, even when compared to the other New Zealand dataset, NZ07. In Aust04b, and to a lesser extent, in Aust00 and Aust04b, the highest income bracket also happens to be heavily represented.

4.4 SOME CONCLUDING REMARKS

This chapter has provided a description of the datasets which will be used in the empirical analysis in Chapter 5. Respondents in all datasets have been presented with a very similar stated, unlabelled, route choice problem involving pivot designs and three (two hypothetical plus one status quo) alternatives in each choice set. There are some differences across the datasets in terms of how each of the choice experiments has been designed, but to a large extent, the similarity of these seven datasets controls for the data context. Therefore, using these seven datasets will present a unique opportunity to test if the heuristics discussed in Chapter 2 are independent of respondents, that is, whether heuristics are portable across datasets, given that the differences in the data context have been largely controlled for. At the same time, it is worth noting that the fact that very similar datasets are used can limit the degree to which evidence in the form of model fit differences is in fact convincing and generalisable.

CHAPTER 5 MODEL RESULTS AND ANALYSIS

5.1 INTRODUCTION

This chapter reports the results of the data analysis using a subset of the heuristics that have been discussed in Chapter 2. The heuristics that are discussed in this chapter are the majority of confirming dimensions (MCD), the extremeness aversion heuristic and the reference point revision heuristic. As previously discussed, the work on the MCD rule and the reference point revision rule is an extension of Hensher and Collins (2011) while the discussion on the extremeness aversion heuristic extends the work of Tversky and Simonson (1993), Kivetz *et al.* (2004) and Chorus (2010).

The outline of this chapter is as follows. Firstly, Section 5.2 presents the results of the standard linear additive MNL model with no heuristics. The standard model might be seen as a benchmark to which more behaviourally realistic models of heuristics may be compared. Section 5.3 which follows discusses the majority of confirming dimensions (MCD) rule. Section 5.4 then moves on to a specific application of the extremeness aversion heuristic using a non-linear logit specification that makes reference to the worst attributes level in the choice set. The model presented is an extension of the contextual concavity model. After this, Section 5.5 discusses the random regret minimisation (RRM) model and its cousin, the hybrid RRM-RUM model. Section 5.6 reports an extensive analysis into the relative advantage maximisation (RAM) model. This is important since the RAM model is a competitor model to the RRM model on several dimensions. Section 5.7 then compares the models of the extremeness aversion heuristic. Following this section, Section 5.8 discusses the reference revision heuristic. Next, Section 5.9 discusses the results of using the heuristic weighting function approach to embedding multiple heuristics into the choice model. Section 5.10 presents the results of some mixed logit estimations and Section 5.11 discusses calculations related to the value of travel time savings. Section 5.12 provides a brief summary of this chapter.

5.2 THE STANDARD CONTEXT INDEPENDENT MNL MODEL

In this section, the results of the standard, linear additive context independent MNL model are reported and discussed. To make the nomenclature more convenient, this model will be referred to as the 'standard RUM model'. The analysis focuses on the seven toll road datasets described in Chapter 4. The model is written as shown in Equation (5.1):

$$U_j = \beta_{0,j} + \sum_k \beta_k X_k + \varepsilon_j \tag{5.1}$$

As discussed previously, the utility specification written in Equation (5.1) is an example of a context independent specification because the indirect utility for alternative *j* is assumed to be a function of only its own attributes, and not of the attributes of any other competing alternative. In other words, U_j is assumed to be the same regardless of how good or how bad a competing alternative is.

The ε_j term represents an error or may be thought of as a representation of all other unobserved influences on utility. In the MNL model, ε_j is assumed to be independent and identically distributed (*i.i.d*), with an extreme value type I distribution. The MNL model does not account for repeated observations made by the same individual, i.e., the panel nature of the dataset, which may induce correlations across choice observations. Since the toll road datasets consist of unlabelled experiments, it will be preferable to estimate the taste parameters β_k as generic across all alternatives, as recommended by Hensher *et al.* (2006). The reason is that the labels (or lack thereof) associated with each alternative do not confer any additional meaning to the respondent on what the alternatives represent in real life and there is no information that can be used to systematically distinguish one alternative from another. All models in this chapter are estimated using a post release version of Nlogit 5.

As each choice set contains three alternatives, the alternative specific constant (ASC) $\beta_{0,j}$ can appear in at most two alternatives. For the purposes of estimation, $\beta_{0,j}$ is included in the utility expression for the current alternative and the hypothetical alternative A. One justification for the inclusion of ASCs in the modelling for choice experiments with some unlabelled alternatives is to control for potential left-to-right effects that may or may not be present in the choice response. In pivot designs where the attribute levels of the hypothetical alternatives are pivoted off the current or reference alternative, another justification for the use of an alternative specific constant, at least for the reference alternative which is labelled in all datasets, is to allow the practitioner to control for potential status quo biases.

The results from the MNL estimation are shown in Table 5.1. Only parameters that are statistically significant at the five per cent level are reported. Where parameter estimates are statistically significant, they are negative, as expected. A negative sign for the taste parameters of the time and cost attributes reflects the reduced desirability of an alternative as its time and cost attribute values increase. This is consistent with expectations and also a large body of evidence in the literature. The normalised AIC in Table 5.1 and in subsequent tables is defined as equal to (– 2(log likelihood at convergence)) + 2K)/n, where K is the number of estimated parameters and n is the number of observations in the dataset.

It may be observed that $\hat{\beta}_{RC}$ in Aust00 and $\hat{\beta}_{FF}$ in NZ99 are found to be statistically indistinguishable from zero. This result may mean one of a few things. Firstly, respondents may on average be genuinely unconcerned about the levels of the RC attribute in Aust00 and the levels of the FF attribute in NZ99. This may be linked to attribute non-attendance in the sense that respondents may be ignoring the attribute at the levels presented in the choice task, but higher levels of that attribute would elicit some aversive response. For example, among all datasets, the average of the $\frac{\text{RC}}{\text{Total Travel Cost}}$ ratio across all alternatives is the lowest in Aust00, which may have led to a larger proportion of respondents ignoring the RC attribute in Aust00. However, without any further information on say the incidence of self-reported attribute non-attendance, it is difficult to make any more definitive statements on the apparent non-attendance of the RC attribute in Aust00 and the FF attribute in NZ99.

Another reason may be linked to experimental design. Aust00 and NZ99 are the only two datasets out of the seven that are based on an orthogonal design. The analysis by Rose *et al.* (2008) shows that orthogonal designs in general produce much larger asymptotic standard errors and are therefore not as statistically efficient as D-efficient designs. In the specific example highlighted by Rose *et al.* (2008), some of the t-ratios associated with the estimated parameters in the orthogonal design were too low to permit a rejection of the null hypothesis

Table 5.1: Results from Standard MNL Estimation

Standard RUM Model	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio)			
Free flow time (FF) (min)	– 0.0516*** (– 7.36)	– 0.1256*** (–14.97)	n.s.	– 0.0761*** (– 7.94)	– 0.0351*** (– 8.37)	– 0.0683*** (–17.71)	– 0.0994*** (–14.67)
Congestion time (CT) (min)		– 0.1192*** (–14.04)			– 0.0351*** (– 7.40)	– 0.0904*** (–28.53)	– 0.1271*** (–10.06)
Slowed down time (SDT) (min)	– 0.0723*** (– 9.92)		– 0.0788*** (– 6.45)	– 0.1167*** (–13.60)			
Stop start time (SST) (min)	– 0.0805*** (–14.28)		– 0.1701*** (– 9.76)	– 0.1684*** (–20.29)			
Running Cost (RC) ($)	– 0.3425*** (– 9.19)	n.s.	– 0.2597*** (– 4.05)	– 0.5629*** (– 10.16)	– 0.1166*** (– 6.54)	– 0.3159*** (–14.40)	– 0.4671*** (–10.27)
Toll Cost (TC) ($)	– 0.2770*** (–12.46)	– 0.5568*** (–21.42)	– 0.8152*** (–13.39)	– 0.5041*** (–22.85)	– 0.1575*** (– 9.34)	– 0.3633*** (–28.74)	– 0.6488*** (–21.95)
Alternative Specific Constants							
-*current* alternative	0.9116*** (17.62)	0.5984*** (8.55)	1.0937*** (14.72)	0.3621*** (6.49)	– 0.5491*** (– 5.17)	0.0920** (2.17)	n.s.
-Alternative *A*	n.s.	0.1434** (2.02)	0.2295*** (2.84)	n.s.	0.2396*** (2.79)	n.s.	n.s.
No. of observations	4480	2352	2432	4864	912	3,888	1840
Log-Likelihood at convergence	– 3434.58	–1862.23	– 1694.93	– 2670.14	– 847.75	– 3031.58	– 1631.79
Normalised AIC	1.536	1.588	1.399	1.100	1.872	1.562	1.778
LL(0)	– 4921.78	–2583.94	– 2671.83	– 5343.65	– 1001.93	– 4271.40	– 2021.45

** denotes significance at the five percent level.
*** denotes significance at the one percent level.
n.s.: not significant

that the parameter was equal to zero, even when non-zero priors were assumed in the "true" model. Likewise, in the case of Aust00 and NZ99, the use of an orthogonal experimental design may not have allowed for some of the parameters to be recovered with statistical significance, even if the true values of these parameters are negative.

Regarding the ASCs, it might be noted in passing that with the exception of Aust04a and NZ07, all the ASCs associated with the current alternative are positive at the five percent significance level. In addition, it can also be observed that $\hat{\beta}_{0,A} < \hat{\beta}_{0,current}$. Taken together, these are consistent with the hypotheses of left-to-right bias and the status quo effect. While such biases do not seem to be present in NZ07, it is interesting that the modelling results from Aust04a show that the status quo alternative is relatively less preferred to the hypothetical alternatives, *ceteris paribus*. This may have something to do with the trip time variability attributes that have been omitted from the model specification.

5.3 MAJORITY OF CONFIRMING DIMENSIONS HEURISTIC AS A CONTEXT DEPENDENT HEURISTIC

5.3.1 Discussion of the *mcd* Variable

The analysis of context dependent, or more precisely, choice set dependent heuristics begins with the majority of confirming dimensions (MCD) heuristic. As with several of the heuristics such as lexicography and EBA discussed earlier in Chapter 2, an appeal may be made to the effort-accuracy framework in order to justify the use of the MCD heuristic in certain choice contexts; that is, respondents are searching for a mental shortcut to use that will deliver a relatively high level of accuracy but with lower cognitive effort compared to the standard RUM decision rule. The MCD is a heuristic that is easy to use, since the respondent is not required to compute the magnitude of the differences in the attribute levels across alternatives (Payne *et al.*, 1993). Instead, respondents simply need to ascertain the relative ranking of the attributes, that is, identify the best attribute level across all alternatives in the choice set. The alternative with a larger number of best ranked attributes will be more preferred to another alternative with a smaller number of best ranked attributes. While most, if not all heuristics, have in common the idea of context dependency embedded in them, another advantage of using the MCD heuristic in modelling is its convenient representation for context dependency, since it can simply be specified as another linear additive term to the standard RUM model.

As suggested by Hensher and Collins (2011), one way of testing for the MCD heuristic involves generating a total count of the number of attributes in each alternative *j* that have the best levels across all alternatives in choice set *s*. This generated variable is labelled *mcd*. To contribute to the count, the attribute has to have an attribute level strictly better than all the levels of the same attribute in the other alternatives. Given the results of the MNL estimation in Section 5.1, where all statistically significant taste parameters of the attributes are found to be negative, the definition of a 'strictly better' attribute level is one where the attribute level under consideration is strictly less than the counterpart attribute level of a competing alternative in the same choice set. Each alternative is associated with an *mcd* value and Table 5.2 describes how each of the possible discrete values of the *mcd* variable is distributed across each of the three alternatives in each of the seven datasets.

Table 5.2: Frequency Distribution of the *mcd* Variable

	Frequency of Occurrence						
mcd_j	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
0	2373 (17.7%)	2530 (35.9%)	2286 (31.3%)	4815 (33.0%)	551 (20.1%)	2246 (19.3%)	2002 (36.3%)
1	5225 (38.9%)	2879 (40.8%)	3211 (44.0%)	4867 (33.4%)	1253 (45.8%)	5038 (43.2%)	2732 (49.5%)
2	3899 (29.0%)	1645 (23.3%)	1371 (18.8%)	3058 (20.1%)	705 (25.8%)	3122 (26.8%)	716 (13.0%)
3	1585 (11.8%)	2 (0.03%)	394 (5.4%)	1361 (9.3%)	207 (8.3%)	1258 (10.8%)	70 (1.3%)
4	334 (2.5%)	N.A.	34 (0.5%)	432 (3.0%)	0	0	0
5	24 (0.2%)	N.A.	N.A.	59 (0.4%)	N.A.	N.A.	N.A.
TOTAL	13440	7056	7296	14592	2736	11664	5520

The assumption used in creating the *mcd* variable for these datasets is that an attribute is only considered if its associated taste parameter was found to be statistically significant in the MNL estimation. Recall that in Aust00 and NZ99 respectively, $\hat{\beta}_{RC}$ and $\hat{\beta}_{FF}$ are not statistically significant at the five percent level. Hence, in Aust00, the *mcd* variable was constructed using only the FF, CT and TC attributes, while in NZ99, the *mcd* variable was constructed on the SDT, SST, RC and TC attributes. With regards only to the time and cost attributes that have been included in the model, which are not the full set of attributes that describe the alternatives in the choice experiment, it might be observed that in Aust08, NZ99 and Aust05, there is a handful of alternatives which dominate all other competing alternatives in the choice set; that is, all the attribute levels of these alternatives are strictly better than those in all other competing alternatives in the choice set.

5.3.2 Generic versus Alternative Specific Parameters for the *mcd* Variable

The approach of Hensher and Collins (2011) of including *mcd* as a linear additive term in the utility function of an alternative seems to be a good starting point on how the *mcd* variable might be modelled. An interesting question arises as to whether the generated *mcd* variable should enter into the utility expressions of all three alternatives, or only into the utility of just the hypothetical ones. In support of having a generic β_{mcd} parameter across all alternatives, it may be argued that there is consistency in the decision process when applying a particular heuristic. Hensher and Collins (2011) choose this approach and restrict the parameter for the *mcd* variable to be the same across all alternatives, but a more nuanced understanding of *mcd* can also be reasonably advocated. If the Payne *et al.* (1993) view of decision making is adopted, then in pivot designs, where the attribute levels of the current alternative are fixed throughout the experiment, the cognitive burden of evaluating the current alternative using the weighted linear additive function in a fully compensatory sense is not very onerous: once respondents calculate the value for the current alternative in the first choice set or if they are able to refer to a stored utility value for the current alternative from memory, the same value may be applied to the current alternative in all other choice sets. On the other hand, using the MCD heuristic to evaluate the current alternative may actually require a larger cognitive effort as the relative rankings of each of the attributes in the current alternative will change from choice set to choice set which means that the *mcd* variable for the current alternative has to be recalculated for every choice set in the experiment. Therefore, for the reference alternative in pivot designs, the fully compensatory linear additive function may be adequate as a decision rule and there is no necessity to appeal to the MCD heuristic.

As a first step in the analysis of the MCD rule, adapting the approach of Hensher and Collins (2011), it might be appropriate to consider models involving just the standard RUM specification and the *mcd* variable in order to gain a better understanding of how *mcd* influences utility. The following empirical questions are of interest at this point:

1. Whether the inclusion of the *mcd* variable into the utility function improves model fit.
2. Whether $\beta_{mcd}^{curr} = \beta_{mcd}^{A,B}$, that is, whether the *mcd* parameter is generic across all three alternatives or whether $\beta_{mcd}^{curr} \neq \beta_{mcd}^{A,B}$, i.e., whether the *mcd* parameter is generic only across the hypothetical alternatives and assumed to be alternative specific to the current alternative. In particular, β_{mcd}^{curr} might be expected to less than $\beta_{mcd}^{A,B}$ if the MCD heuristic is not used by some segment of respondents to evaluate the current status quo alternative.
3. Whether $\beta_{mcd}^{curr} = 0$ and $\beta_{mcd}^{A,B} > 0$ which implies a stronger condition that all respondents are rejecting the use of the MCD heuristic when evaluating the current alternative, i.e., the standard specification suffices.

A simple model to address the questions raised above is proposed in Equation (5.2).

$$\begin{aligned} U_{curr} &= \beta_0^{curr} + \sum_k \beta_k X_k + \beta_{mcd}^{curr} mcd + \varepsilon_j \\ U_A &= \beta_0^A + \sum_k \beta_k X_k + \beta_{mcd}^{A,B} mcd + \varepsilon_j \\ U_B &= \sum_k \beta_k X_k + \beta_{mcd}^{A,B} mcd + \varepsilon_j \end{aligned} \tag{5.2}$$

In all models, $\beta_{mcd}^{A,B}$ is specified to be generic across the hypothetical alternatives.

Table 5.3 summarises the results of the estimation. The parameter estimates associated with the attributes of the alternatives are not reported in order to facilitate comparison across the datasets as to the various hypotheses about the *mcd* parameter.

Indeed, just by considering the model where β_{mcd} is assumed generic across all three alternatives (results of Column 1), the relationship between *mcd* and utility is positive and statistically significant at the five percent level in all datasets except Aust04a. Generally speaking therefore, it is possible to conclude that the inclusion of the *mcd* variable improves model fit with respect to the standard RUM model. However, without further consideration of

Table 5.3: Comparison of *mcd* Parameter Assumptions Using MNL

		Column 1	Column 2	
		$\beta_{mcd}^{curr} = \beta_{mcd}^{A,B}$	$\beta_{mcd}^{curr} \neq \beta_{mcd}^{A,B}$	$\beta_{mcd}^{curr} = \beta_{mcd}^{A,B}$ rejected at 5% level?
		$\hat{\beta}$ (z-ratio)	$\hat{\beta}$ (z-ratio)	
Aust08	β_{mcd}^{curr}	0.0673** (2.51)	– 0.0686* (– 1.77)	Yes
	$\beta_{mcd}^{A,B}$		0.2374*** (5.35)	
	LL at convergence	– 3431.44	– 3419.71	
	No. of observations	4,480		
Aust00	β_{mcd}^{curr}	0.2949*** (6.81)	1.3387*** (10.44)	Yes
	$\beta_{mcd}^{A,B}$		0.0029 (0.05)	
	LL at convergence	– 1838.71	– 1798.72	
	No. of observations	2,352		
NZ99	β_{mcd}^{curr}	0.4698*** (11.63)	0.7900*** (11.85)	Yes
	$\beta_{mcd}^{A,B}$		0.1564** (2.46)	
	LL at convergence	– 1624.57	– 1604.09	
	No. of observations	2,432		
Aust05	β_{mcd}^{curr}	0.1496*** (4.78)	0.1202*** (2.69)	No
	$\beta_{mcd}^{A,B}$		0.1826*** (3.84)	
	LL at convergence	– 2658.71	– 2658.29	
	No. of observations	4,864		
Aust04a	β_{mcd}^{curr}	0.1008 (1.47)	0.1328 (1.17)	No
	$\beta_{mcd}^{A,B}$		0.0813 (0.93)	
	LL at convergence	– 846.67	– 846.61	
	No. of observations	912		
Aust04b	β_{mcd}^{curr}	0.0812** (2.40)	0.1057* (1.88)	No
	$\beta_{mcd}^{A,B}$		0.0631 (1.34)	
	LL at convergence	– 3028.71	– 3028.56	
	No. of observations	3,888		
NZ07	β_{mcd}^{curr}	0.0902** (2.03)	0.0433 (0.86)	Yes
	$\beta_{mcd}^{A,B}$		0.1573*** (2.79)	
	LL at convergence	– 1629.73	– 1627.86	
	No. of observations	1,840		

*denotes significance at the ten percent level.
** denotes significance at the five percent level.
*** denotes significance at the one percent level.

the possibility of an alternative specific *mcd* parameter, the dangers of assuming a generic *mcd* parameter are also evident. Notably in Aust08, Aust04b and NZ07, the point estimate of the generic β_{mcd} parameter is a fraction of that in Aust00 and NZ99. Moreover, the z-ratios of the parameter in these datasets are close to the value of two, suggesting that *mcd* is not a strong determinant of choice. However, if the true values of β_{mcd}^{curr} and $\beta_{mcd}^{A,B}$ are such that β_{mcd}^{curr} is zero and $\beta_{mcd}^{A,B}$ is positive, then assuming a generic *mcd* parameter would result in a downward bias for the estimated β_{mcd}.

Column 2 of Table 5.3 reports the results of estimating alternative specific β_{mcd}^{curr} and $\beta_{mcd}^{A,B}$. These results directly address Hypothesis 2 and Hypothesis 3. It can be seen that in Aust08, Aust00, NZ99 and NZ07, the hypothesis $\beta_{mcd}^{curr} = \beta_{mcd}^{A,B}$ may be rejected at the five percent level, that is, alternative specific *mcd* parameters may be assumed. Among all the datasets, the results from Aust08 and NZ07 are most consistent with the hypotheses that $\beta_{mcd}^{curr} = 0$ and $\beta_{mcd}^{A,B} > 0$. In particular, inclusion of an alternative specific *mcd* parameter in the model for Aust08 results in a substantial improvement in the log-likelihood statistic. On the other hand, $\beta_{mcd}^{curr} = \beta_{mcd}^{A,B}$ is not rejected at the five percent level in Aust05, Aust04a and Aust04b. For Aust04a in particular, neither the β_{mcd}^{curr} parameter nor the $\beta_{mcd}^{A,B}$ parameter is statistically significant, while in the other two datasets Aust05 and Aust04b, the assumption of a generic *mcd* parameter across all alternatives, that is, $\beta_{mcd}^{curr} = \beta_{mcd}^{A,B}$ and $\beta_{mcd}^{curr}, \beta_{mcd}^{A,B} > 0$, is supported by the data. Unexpectedly, results from Aust00 and NZ99 seem to suggest that respondents place a higher important weight on using the MCD heuristic when the current alternative is assessed vis-à-vis the hypothetical alternatives; in Aust00 for example, β_{mcd}^{curr} is found to be greater than zero, while $\beta_{mcd}^{A,B} = 0$ is not rejected at the five percent level.

The MCD effect hypothesised so far can be simply stated as the higher the *mcd* count, the more desirable the alternative. It is also of interest to note that the extremeness aversion heuristic expressed through the compromise effect may in theory oppose the MCD effect. Since the extreme alternatives are defined by both extremely good and bad attribute levels, it is quite possible for an extreme alternative to have a higher *mcd* count than a compromise alternative, by virtue of the extreme alternative having some attributes that are the 'best' in the choice set. On the other hand, the compromise alternative may have a low *mcd* count, possibly even having a *mcd* count of zero, if none of its attributes are the 'best' in the choice set. Therefore, through the compromise effect, it might be the case that the lower the *mcd* count, the more desirable the alternative becomes. At even higher *mcd* counts, where an alternative is close to becoming, or is in effect, the dominating alternative, the MCD effect may run up against the believability of an alternative. To the extent that such an alternative lacks believability, its utility may be negatively impacted. These observations may explain some of the empirical anomalies described above.

To summarise, there does not appear to be a generalisable rule governing how the MCD heuristic can be modelled in different datasets. Nevertheless, these results suggest that some empirical testing needs to be undertaken if MCD is to be used as a representation of context dependence. Hensher and Collins (2011) also conclude that the *mcd* variable on its own is not sufficient to explain the observed choice made by the respondent. Attributes of the alternatives do matter.

5.3.3 A Probabilistic Decision Process Model

One possible avenue of further research into the role of the MCD heuristic may be undertaken through a probabilistic decision process (PDP) model such as that suggested in McNair *et al.* (2012). A PDP model may be used to capture the possibility of heterogeneity in decision rules among the sample of respondents. In the PDP model considered here, assume that there are two latent decision rules. The first decision rule corresponds to the MCD rule and the second decision rule corresponds to the standard RUM linear additive rule. Therefore, in the first decision rule, all taste parameters associated with the attributes of the alternatives are set to zero, and in the second decision rule, the coefficient of the MCD variable is set to zero. The model accounts for the panel nature of the dataset. The results are reported in Table 5.4.

Across all seven datasets, as expected, the coefficient associated with the *mcd* variable is positive and statistically significant at the five percent level. In contrast to the RUM model results reported earlier, all coefficients associated with the time and cost attributes in Class 2 for Aust00 and NZ99 are now negative and statistically significant. Likewise in the other

Table 5.4: Estimation of Probabilistic Decision Process Models

	$\hat{\beta}$ (z-ratio)					
	Aust08		NZ99		Aust05	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Free flow time (FF) (min)		– 0.0460*** (– 3.13)		– 0.1649*** (– 12.49)		– 0.1160*** (– 8.63)
Slowed down time (SDT) (min)		– 0.1201*** (– 8.38)		– 0.2336*** (– 6.46)		– 0.1596*** (– 12.35)
Stop start time (SST) (min)		– 0.0847*** (– 7.44)		– 0.3477*** (– 6.46)		– 0.2052*** (– 17.27)
Running Cost (RC) ($)		– 0.8154*** (– 8.54)		– 2.5528*** (– 17.41)		– 0.8090*** (– 10.50)
Toll Cost (TC) ($)		– 0.6678*** (– 8.56)		– 6.0570*** (– 11.86)		– 0.8327*** (– 18.47)
MCD	0.3715*** (9.19)		0.3991*** (5.64)		0.4972*** (9.36)	
Alternative Specific Constants						
-*current* alternative	– 0.6516*** (– 4.20)	1.4913*** (16.53)	1.3102*** (8.34)	n.s.	– 0.2414** (– 2.08)	0.4316*** (6.25)
- Alternative *A*	n.s.	n.s.	0.3148** (2.14)	n.s.	n.s.	n.s.
Probability	0.2928*** (8.65)	0.7072*** (20.90)	0.3573*** (12.27)	0.6427*** (22.06)	0.1433*** (5.98)	0.8567*** (35.78)
Log-likelihood at convergence	– 2783.53		– 1494.29		– 2420.12	
Normalised AIC	1.247		1.236		0.999	

*** denotes significance at the one percent level.
** denotes significance at the five percent level.
n.s.: not significant.
Models account for panel nature of dataset

	$\hat{\beta}$ (z-ratio)							
	Aust00		Aust04a		Aust04b		NZ07	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Free flow time (FF) (min)		– 0.2744*** (– 7.20)		– 0.0864*** (– 9.77)		– 0.1054*** (– 16.63)		– 0.1627*** (– 13.87)
Congestion time (CT) (min)		– 0.2697*** (– 7.14)		– 0.0174* (– 1.81)		– 0.1240*** (– 24.16)		– 0.1878*** (– 9.33)
Running Cost (RC) ($)		– 0.9195*** (– 2.53)		– 0.1335*** (– 3.08)		– 0.4309*** (– 14.19)		– 0.7851*** (– 10.79)
Toll Cost (TC) ($)		– 1.4968*** (– 8.01)		– 0.3090*** (– 9.03)		– 0.5270*** (– 23.73)		– 1.1265*** (– 15.54)
MCD	0.3654*** (3.82)		0.5739*** (8.77)		0.3915*** (7.12)		0.3483*** (7.12)	
Alternative Specific Constants								
-*current* alternative	1.0930*** (7.53)	n.s.	– 2.3377** (– 10.80)	0.9366*** (5.10)	0.3287*** (3.16)	n.s.	– 2.5647*** (– 9.12)	0.6094*** (5.90)
-Alternative *A*	n.s.	n.s.	n.s	0.5789*** (2.99)	n.s.	n.s.	0.3727*** (4.09)	n.s.
Probability	0.3501*** (11.57)	0.6499*** (21.48)	0.5962*** (9.16)	0.4038*** (6.20)	0.2062*** (6.53)	0.7938*** (25.15)	0.2970*** (6.88)	0.7030*** (16.29)
Log-likelihood at convergence	– 1812.38		– 714.61		– 2889.23		– 1271.94	
Normalised AIC	1.547		1.587		1.490		1.392	

*** denotes significance at the one percent level.
** denotes significance at the five percent level.
n.s.: not significant
Models account for panel nature of dataset

datasets, the time and cost parameters are negative and statistically significant at the five percent level, with the sole exception of the CT parameter in Aust04a. The log-likelihood statistics of the PDP model indicate a vast improvement in model fit, which is also to be expected since the panel nature of the data is now accounted for. All class probabilities are statistically significant, with the probability of a respondent belonging to class 1 (use of the MCD heuristic) ranging from around 15 percent in Aust05 to almost 60 percent in Aust04a. These results show that the incidence of respondents not fully trading amongst the attributes can be very high.

As an aside, it may be observed that the MCD heuristic is a decision rule that captures semi-compensatory decision making, that is, both compensatory and non-compensatory decisions can be captured. It is compensatory in the sense that the loss engendered by a previously 'best' attribute moving to 'non-best' can be compensated by another attribute moving from 'non-best' to 'best'. However, it can also be non-compensatory at the same time in the sense that once an attribute is 'best' or 'non-best', any improvements/worsening of that attribute will not affect the MCD count.

The consistency across datasets that is revealed in the (non-zero) probability of using the MCD heuristic as opposed to the standard linear additive RUM rule suggests the need for more careful consideration on how to embed the MCD heuristic in choice models. On its own, the MCD heuristic provides little information to the practitioner on the outputs of interest – typically marginal willingness to pay measures, welfare estimates or elasticities – so it is envisaged that some other kind of decision rule will have to be assumed for the model as well. The extent to which the use of the MCD heuristic is itself a function of the experimental design is another interesting research question. For example, explanations for why there might be perceived non-attendance to certain attributes (Hensher *et al.*, 2012b) might also be used to explain the perceived presence or absence of the MCD heuristic across various choice contexts.

A criticism that can be levelled against the MCD approach is that the researcher would have to *a priori* identify the 'best' level for the attribute based on whether, say, more is better or less is better. This identification poses a significant challenge particularly in mixed logit models where in some cases, parameter distributions may be assumed to be unconstrained and such models may reveal substantial heterogeneity in the sign of the parameters[6].

5.4 THE EXTREMENESS AVERSION HEURISTIC

As discussed in Chapter 2, a major class of heuristics explored in this thesis relates to the extremeness aversion heuristic. With extremeness aversion, the assumption of context independence is immediately relaxed, since a reference point is now introduced into the utility function. The modelling approach discussed in this section is adopted from one of Kivetz *et al.*'s (2004) specifications for extremeness aversion, which is the so-called contextual concavity model written out in Equation (5.3). In this model, the context dependence stems from the part-utilities of each attribute in the utility function being expressed as gains relative to the minimum part-utility of the same attribute in that choice set. The nomenclature of concavity arises from the assumption that the gains in part-utility are concave functions, although the empirical results that are presented later in this section will question this assumption. Hence, for the moment, this model will be referred to as the non-linear logit model. Equation (5.3) expresses the utility for alternative *j* in each choice set *s* as:

$$U_{js} = \beta_{0,j} + \sum_{k} (\beta_k (X_{jks} - \max_{j \in s}(X_{jks})))^{\varphi_k} + \varepsilon_j \tag{5.3}$$

$\max_{j \in s}(X_{jks})$ is the maximum value attained by attribute *k* among all *j* alternatives in choice set *s*, to recognise that the minimum part-utility of an attribute is associated with the worst attribute level it takes in choice set *s*. φ_k is the power parameter associated with attribute *k*. It is at least thought to be greater than zero so that U_{js} is an increasing function in the gains.

---

[6] For this matter, this is a criticism against many models which make use of a reference point. For example, in lexicography or EBA, 'best' or 'worst' levels of attributes typically need to be known/assumed beforehand.

Like the MCD heuristic, the reference point in the utility function in the case of the extremeness aversion heuristic is choice set dependent, meaning that preferences are context specific. For now, the model is estimated assuming *i.i.d.* Extreme Value I error terms.

It is also important to note that written in the form above, the utility function U_{js} is estimable only if β_k is non-positive for each attribute *k*. Otherwise, if β_k was positive, an estimation error would occur since φ_k is not generally an integer value and a negative quantity raised to a non-integer power is not a real number. While there is overwhelming evidence in the transportation literature regarding the sign of β_k on the time and cost attributes of an alternative, in practice, where the sign of β_k is not known *a priori*, it might be necessary to estimate, say the standard RUM model first, in order to determine the sign of β_k. That information can then be used to determine which of $\max(X_{jks})$ or $\min(X_{jks})$ is the appropriate reference, i.e., “worst” attribute level in the choice set, to be used in the model.

If prospect theory applies, in other words, if there are diminishing returns to the gains in part utilities, as suggested by Kivetz *et al.* (2004), the prior expectation is for φ_k to satisfy the inequality $0 < \varphi_k < 1$. From the econometric perspective however, it is not necessary for such a constraint to be imposed on φ_k, hence φ_k is not bound by this restriction and may take on values greater than one. To highlight one application of this non-linear logit approach, Table 5.5 reports the results of the estimation for Aust08, where the error terms are assumed to be *i.i.d.* EV type I. In the case of $\hat{\varphi}_k$, 95 percent confidence intervals are also reported in order to facilitate the comparison of the estimated value with the value of one, which is the linear form of the utility function.

The results obtained from the non-linear logit model show that all the $\hat{\beta}_k$ estimates are statistically significant at the one percent level and all have the expected negative sign. It is also reassuring to note that all $\hat{\varphi}_k$ are statistically significant as well, indicating that all the attributes that have been modelled in the utility function do matter. Notice that when $\varphi_k = 1$ for all *k*, the model collapses to the standard context independent MNL model since within each choice set, the same term $\beta_k \max(X_{jks})$ is added to all the utility functions and the

Table 5.5: Results from Applying a Non-Linear Multinomial Logit Model to Aust08

	$\hat{\beta}$ (z-ratios)	$\hat{\varphi}$ (z-ratios) [95% confidence interval]
Free flow time (FF) (minutes)	– 0.0637*** (– 15.62)	1.842*** (8.75) [1.430 – 2.255]
Slowed down time (SDT) (minutes)	– 0.0756*** (– 11.79)	1.192*** (8.35) [0.912 – 1.472]
Stop start time (SST) (minutes)	– 0.0815*** (– 13.97)	0.934*** (11.43) [0.774 – 1.094]
Running Cost (RC) ($)	– 0.3155*** (– 5.45)	0.571*** (6.07) [0.386 – 0.755]
Toll Cost (TC) ($)	– 0.2672*** (– 9.48)	0.751*** (6.09) [0.509 – 0.993]
Alternative Specific Constants		
-*current* alternative	0.9168*** (14.98)	
-Alternative *A*	not significant	
No. of observations	4,480	
Log-likelihood at convergence	– 3414.41	

*** denotes significance at the one percent level.

addition of the same term across all utility functions will not affect the estimation results. Hence, the standard model is nested in this non-linear logit model, and to test if the non-linear model is providing a better statistical fit for the data, the likelihood ratio test will be appropriate. The result shows that embedding a contextual heuristic into the model provides a better statistical fit at the one percent level [Non-linear model *vs.* Standard RUM model: $\text{Prob}\,(\chi^2_{(5)} > 40.34) < 0.01$].

The estimated $\hat{\beta}_k$ parameters in the non-linear model have a natural counterpart in the $\hat{\beta}_k$ parameters of the standard representation, in the sense that these parameters can be interpreted as weights for the attribute levels in the utility function. However, while the $\hat{\beta}_k$ parameters in the RUM specification have another interpretation as the marginal utility/disutility associated with a unit increase in the attribute, the marginal utilities in the non-linear contextual specification will depend on the values of $\hat{\beta}_k$, $\hat{\varphi}_k$, X_{jks} and $\max(X_{jks})$.

Diminishing returns to utility gains or the concavity of the functional form can be tested by comparing the null $H_0 : \hat{\varphi}_k = 1$ (no concavity, i.e., linear in the attributes) against the alternative $H_1 : \hat{\varphi}_k < 1$ (concavity). In Aust08, the null cannot be rejected at the five percent level for the power parameters φ_{SDT} and φ_{SST} associated with the SDT and SST attributes. Hence, in the SDT and SST time components, respondents are behaving as if they are standard linear additive utility maximisers. On the other hand, the data allow the rejection of the null hypotheses $\hat{\varphi}_{RC} = 1$ and $\hat{\varphi}_{TC} = 1$, in favour of the alternative hypotheses $\hat{\varphi}_{RC} < 1$ and $\hat{\varphi}_{TC} < 1$, at the five percent significance level, although the 95 percent confidence interval of $\hat{\varphi}_{TC}$ just barely excludes the value of one. This implies that the gains in the *RC* and *TC* part-utilities relative to the reference point are concave and therefore consistent with prior expectations.

For this non-linear logit model, a more interesting result relates to the rejection of the null hypothesis $\hat{\varphi}_{FF} = 1$ in favour of the alternative hypothesis $\hat{\varphi}_{FF} > 1$, leading to the conclusion that respondents are treating the part-utility gains in the free flow time attribute as a convex function, rather than a concave function. This is a surprising observation as a convex function implies that there are increasing, rather than diminishing, returns to gains in the part utility. Previous studies (Kivetz *et al.*, 2004; Chorus and Bierlaire, 2013) do not find non-concavity for the power parameters and so the theoretical reason for this result is not entirely clear.

The non-linear logit model is also estimated for the other datasets, with the results for the β_k estimates reported in Table 5.6. Since the power parameters $\hat{\varphi}_k$ are of empirical interest, Table 5.7 makes a comparison of these parameters across all the seven datasets. In addition to the point estimates, 95 percent confidence intervals are again reported to facilitate the comparison of these estimates to critical values such as zero or one.

The first observation relates to model fit. Applying the likelihood ratio test, the non-linear logit model outperforms the standard RUM model in all datasets, with the exception of Aust00. In other words, there seems to be a rather consistent pattern that accounting for some form of non-linearity in the utility functions is an important element of describing choice behaviour. It might also be observed that all the estimated β_k are negative and statistically

Table 5.6: Estimates from a Non-Linear Multinomial Logit Estimation

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				Estimates of β_k (z-ratios)			
Free flow time (FF) (min)	– 0.0637*** (– 15.62)	– 0.1156*** (– 7.71)	n.s.	– 0.0629*** (– 3.56)	– 0.0333*** (– 12.02)	– 0.0726*** (– 15.34)	– 0.0958*** (– 14.64)
Congestion time (CT) (min)		– 0.1468*** (– 6.88)			– 0.0364*** (– 6.61)	– 0.0819*** (– 21.71)	– 0.1233*** (– 12.43)
Slowed down time (SDT) (min)	– 0.0756*** (– 11.79)		– 0.0717*** (4.82)	– 0.1294*** (– 11.66)			
Stop start time (SST) (min)	– 0.0815*** (– 13.97)		– 0.1771*** (– 8.27)	– 0.1602*** (– 15.98)			
Running Cost (RC) ($)	– 0.3155*** (– 5.45)	n.s.	– 0.3073*** (– 7.46)	– 0.7320*** (– 9.45)	– 0.1206*** (– 6.45)	– 0.3227*** (– 11.77)	– 0.4797*** (– 10.38)
Toll Cost (TC) ($)	– 0.2672*** (– 9.48)	– 0.5884*** (– 13.51)	– 10.322** (– 2.10)	– 1.2691*** (– 6.05)	– 0.2250*** (– 3.86)	– 0.4279*** (– 14.30)	– 0.7961*** (– 9.51)
Alternative Specific Constants							
- *current* alternative	0.9168*** (17.59)	0.6188*** (7.57)	0.8558*** (11.25)	0.2408*** (3.81)	– 0.7312*** (– 6.85)	n.s.	n.s.
- Alternative *A*	n.s.	0.1407** (1.98)	0.2756*** (3.15)	n.s.	n.s.	n.s.	n.s.
LL at convergence	– 3414.41	– 1859.55	– 1579.82	– 2617.33	– 837.29	– 3019.63	– 1625.99
Normalised AIC	1.529	1.588	1.307	1.081	1.856	1.557	1.776

*** denotes significance at the one percent level.
n.s.: not significant

Table 5.7: A Comparison of the φ_k Estimates across the Seven Toll Road Datasets (MNL Estimation)

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				Estimates of φ_k (z-ratios) [95% confidence interval]			
$\hat{\varphi}_{FF}$	1.842*** (8.75) [1.430 – 2.255]	1.053*** (13.72) [0.903 – 1.204]	not estimated; $\beta_{FF} = 0$	0.505*** (3.99) [0.257 – 0.753]	1.587*** (6.40) [1.101 – 2.073]	0.848*** (11.77) [0.707 – 0.989]	1.128*** (14.76) [0.978 – 1.278]
$\hat{\varphi}_{CT}$		0.874*** (12.36) [0.735 – 1.013]			0.796*** (4.74) [0.467 – 1.125]	1.148*** (23.69) [1.053 – 1.243]	1.264*** (10.16) [1.020 – 1.5108]
$\hat{\varphi}_{SDT}$	1.192*** (8.35) [0.912 – 1.472]		0.853*** (5.31) [0.538 – 1.167]	0.784*** (9.76) [0.626 – 0.941]			
$\hat{\varphi}_{SST}$	0.934*** (11.43) [0.774 – 1.094]		0.901*** (8.37) [0.690 – 1.112]	1.047*** (15.98) [0.918 – 1.175]			
$\hat{\varphi}_{RC}$	0.571*** (6.07) [0.386 – 0.755]	not estimated; $\beta_{RC} = 0$	Restricted to $\varphi_{RC} = 1$	0.720*** (8.75) [0.558 –0.881]	0.896*** (3.75) [0.427 – 1.365]	0.780*** (7.87) [0.585 – 0.974]	0.914*** (7.56) [0.677 – 1.151]
$\hat{\varphi}_{TC}$	0.751*** (6.09) [0.509 – 0.993]	0.975*** (28.01) [0.906 – 1.043]	0.301*** (6.83) [0.215 – 0.387]	0.476*** (10.99) [0.391 – 0.560]	0.382*** (3.93) [0.191 – 0.573]	0.835*** (15.74) [0.731 – 0.939]	0.812*** (10.36) [0.659– 0.966]
R	Yes	No	Yes	Yes	Yes	Yes	Yes

*** denotes significance at the one percent level.
R = reject standard RUM utility specification at the 5% level?

significant at the five percent level. An interesting observation is the greater difficulty in fitting the non-linear logit model to NZ99 (φ_{RC} had to be constrained to one in order for sensible results to be obtained) and the emergence of an apparently strong aversion to toll costs (say, relative to the running cost attribute). At the same time, estimating the non-linear logit model for NZ99 allows a vast improvement in the log-likelihood statistic, from – 1694.9 in the RUM model to – 1579.8 in the non-linear logit model.

Upon closer inspection of the data, it is observed that there is a much stronger consistent pattern of choice for the reference alternative in NZ99, compared to say the other New Zealand dataset which is NZ07. In NZ99, it was found that the reference alternative was chosen in 72 percent of the choice sets, whereas in NZ07, the reference alternative was chosen only 49 per cent of the time. In the New Zealand datasets, there was no prior experience with tolling and hence, the reference alternatives all had zero toll costs. In order for the non-linear logit model to provide a better fit to the choice patterns observed, it seems that the TC parameter has to be allowed to be highly negative, indicating a strong aversion to toll costs and to the hypothetical alternatives, which makes sense since the reference alternative had no toll costs.

Turning to the φ_k parameters, it is observed that while the null hypothesis that $\varphi_{FF} = 0$ can always be rejected in favour of $\varphi_{FF} > 0$ across all the six datasets where φ_{FF} is estimated, φ_{FF} itself can take on values between zero and one (Aust05 and Aust04b), equal to one (Aust00 and NZ07), or greater than one (Aust08 and Aust04a).

Across the three datasets Aust08, NZ99 and Aust05 where the SDT and SST attributes are available for modelling, estimates of φ_{SDT} reveal that φ_{SDT} is never statistically greater than one; hence gains in the part utilities in the SDT attribute can be represented as either a linear or concave function, but not a convex function. Respondents in Aust08 and NZ99 may be thought of taking the SDT attribute to be linear in the gains, while in Aust05, it is preferable to treat the gains in SDT as a concave function. Across these three datasets, the null of $\varphi_{SST} = 1$ is not rejected at the five percent level, that is, the SST attribute may be modelled as entering the utility function as a linear function.

Across the four datasets where CT is an attribute of the alternatives, it is found that $\hat{\varphi}_{CT} = 1$ is not rejected at the five percent level in Aust00 and Aust04a. However, in Aust04b and NZ07, $\hat{\varphi}_{CT}$ is found to take on a value greater than one, which is again somewhat of a surprising result in light of evidence that gains should be concave. Therefore, along the time attributes of the alternatives, it is possible for respondents to exhibit extremeness seeking, extremeness aversive or extremeness neutrality (linear-in-attributes and linear-in-parameters) behaviour, and it is very much an empirical question as to what kind of behaviour is on display.

Along the RC and TC cost attributes, the estimated φ_{RC} and φ_{TC} parameters of the model are, in the vast majority of cases, statistically less than one and in all cases, do not statistically exceed one. It therefore appears that on the monetary dimension at least, behaviour of respondents is generally consistent with prospect theory and extremeness aversion. There are six instances of the φ_{TC} parameter and four instances of the φ_{RC} parameter being less than one. This might suggest that respondents are generally more extremeness averse in the TC attribute than in the RC attribute, which sounds intuitively correct since not only are toll costs more salient than running costs, but the emphasis of the choice experiments was on tolled roads as well.

The overall conclusion from examining the data through the lens of a non-linear logit model such as Equation (5.3) suggests that while some consistent behavioural patterns across datasets may be observed, there is also quite a substantial amount of heterogeneity in the way the same heuristic is applied across groups of respondents. With the exception of Aust00, the estimation results from Model 2 show that accounting for some form of referencing and accounting for non-linearity in the utility function are important, at the very least, for some attributes of the alternative. Whether the demonstrated behaviour is extremeness aversive or extremeness seeking is ultimately an empirical question that depends on the magnitude of the power parameters.

In view of the evidence presented, instead of what has hitherto been known as a contextual concavity model, which makes the prior assumption that differences in the part-utilities are concave in the gains, it may be more appropriate to label such a functional specification as a "non-linear worst level referencing" (NLWLR) model. In cases such as in Aust04a where some of the power parameters are greater than one, and other power parameters less than one, the NLWLR model can display not the compromise effect, but the polarisation effect where there is an increased tendency to pick one of the extreme alternatives.

Table 5.8 illustrates the polarisation effect using data from an actual choice set in Aust04a. In this example, alternative *A* is the compromise alternative since the part utilities of its *FF* and *TC* attributes lie in-between the other two extreme alternatives of *current* and *B*. As a matter of notation, elements within the round brackets indicate the choice set under consideration. For example, $V_{curr_{(curr,A,B)}}$ denotes the observed utility of the current alternative in the choice set comprising of the current alternative, alternative A and alternative B. As calculated, the introduction of an extreme alternative *B* into the binary choice set of {*curr*, *A*} causes the choice share of *A* relative to the choice share of *curr* to fall from 3.20 to 1.82. This is extremeness aversion away from *curr*. On the other hand, the introduction of *curr* into the binary choice set of {*A*, *B*} causes the choice share of *B* relative to *A* to increase. This represents extremeness seeking behaviour towards *B*, which has the highest part-utility in the FF attribute. Together, these observations constitute an example of polarisation since extremeness seeking/extremeness aversion operates only in one direction in this case, that is, away from *curr* and towards *B*.

The initial motivation behind the development of the contextual concavity model was to find a way to explain the compromise effect, but as has been shown, the NLWLR model, which is a more appropriate nomenclature for the contextual concavity model, is capable of generating not only compromise effects, but also polarisation effects. In cases where all of the φ_k parameters are greater than one, the NLWLR model might even account for extremeness seeking behaviour, where all the extreme alternatives are preferred to the compromise.

As a matter of empirical estimation, the NLWLR model can be tricky to estimate. It is possible for unconstrained estimation to lead to errors since the β_k parameters need to be non- positive. In the case of an alternative whose attribute X_{jks} takes on the maximum value in the choice set, the gains in the part-utility of that attribute relative to the reference is zero

Table 5.8: An Illustration of the Polarisation Effect Using Parameters Obtained from Estimation on Aust04a

	current	*Alternative A*	*Alternative B*
Part-utility of FF attribute, $V_{j,FF}$	$V_{curr,FF} = -0.54$	$V_{A,FF} = -0.45$	$V_{B,FF} = -0.24$
Part-utility of TC attribute, $V_{j,TC}$	$V_{curr,TC} = 0$	$V_{A,TC} = -1.56$	$V_{B,TC} = -1.89$
Assume NLWLR model for V_j; $V_j = (V_{j,FF} - \min(V_{j,FF}))^{\varphi_{FF}} + (V_{j,TC} - \min(V_{j,TC}))^{\varphi_{TC}}$ Assumptions: $\varphi_{FF} = 1.59$, $\varphi_{TC} = 0.38$			
Observed utility in binary choice situation $\{curr, A\}$	$V_{curr_{(curr,A)}} = 1.19$	$V_{A_{(curr,A)}} = 0.022$	
Observed utility in binary choice situation $\{A, B\}$		$V_{A_{(A,B)}} = 0.65$	$V_{B_{(A,B)}} = 0.084$
Observed utility in choice situation $\{curr, A, B\}$	$V_{curr_{(curr,A,B)}} = 1.27$	$V_{A_{(curr,A,B)}} = 0.67$	$V_{B_{(curr,A,B)}} = 0.15$
Relative probabilities (choice shares)			
In binary choice situation $\{curr,A\}$	$\frac{\Pr(curr_{(curr,A)})}{\Pr(A_{(curr,A)})} = 3.20$		
In binary choice situation $\{A,B\}$	$\frac{\Pr(B_{(A,B)})}{\Pr(A_{(A,B)})} = 0.57$		
In choice situation $\{curr,A,B\}$	$\frac{\Pr(curr_{(curr,A,B)})}{\Pr(A_{(curr,A,B)})} = 1.82$ $\frac{\Pr(B_{(curr,A,B)})}{\Pr(A_{(curr,A,B)})} = 0.59$		

Note: $V_{j,k} = \beta_k X_{jk}$.

and an error will occur if φ_k happens to be non-positive. It is possible to avoid these errors by a sensible choice of starting values for the parameters, for example, by not starting with $\varphi_k \leq 0$ and $\beta_k \geq 0$. Nevertheless, a significant amount of trial and error may be required before an appropriate set of starting values is found which allows the model to converge.

To aid in the estimation, an algebraic trick may be employed to avoid the problem of division by zero. The utility function may be re-written as in Equation (5.4):

$$U_{js} = \beta_{0,j} + \sum_{k} \left[\beta_k (X_{jks} - \max_{j \in s}(X_{jks})) + d_k \right]^{\varphi_k} - \sum_{k} d_k + \varepsilon_j, \tag{5.4}$$

where $d_k = 1$ if $X_{jks} = \max_{j \in s}(X_{jks})$ and 0 otherwise. Hence, if it is true that β_k is always less than zero, then the term in the square brackets will always be positive and the utility function will be defined for all values of φ_k.

5.5 MODELS WITH REGRET MINIMISATION

5.5.1 The Random Regret Minimisation (RRM) Model

Regret based theories and models are built on the premise that people aim to minimise anticipated regret when making a choice. The RRM model was primarily developed to analyse riskless choice involving multi-attribute alternatives, and in this context regret is said to occur when the attributes of a non-chosen alternative perform better than the attributes of the chosen alternative (Chorus, 2012).

The Chorus (2010) version of the RRM model was described in Chapter 2, Section 2.5.2, but for the reader's convenience, salient points of the model are outlined and reproduced below. Define, in Equation (5.5), the binary or pair-wise regret associated with considering alternative j as opposed to alternative j':

$$reg(j, j') = \sum_{k} \ln[1 + \exp(\beta_{j'k} X_{j'k} - \beta_{jk} X_{jk})] \tag{5.5}$$

The total regret associated with alternative j is the sum of binary regrets over all alternatives j' in choice set s, that is, Equation (5.6):

$$reg(j) = \sum_{\substack{j' \in s, \\ j' \neq j}} reg(j, j') \tag{5.6}$$

The RRM model can be estimated by observing that minimising the regret function is equivalent to maximising the negative of regret. Estimation of the RRM model is easily accomplished in Nlogit 5.0 as it already comes with a built-in procedure for estimating the model.

In the RRM model, preferences for each alternative depend not only on the attribute values for that alternative, but also on the relative performance of these attributes against their counterpart attribute levels in all the other alternatives in the choice set. In other words, preferences for an alternative are context dependent or choice set specific. The RRM allows preferences to change even if an alternative's attribute levels remain constant from choice set to choice set, as long as there is a change in the attribute levels of any of the other alternatives to which an alternative is being compared.

As explained in Chapter 2, the compromise effect due to the RRM model is also closely related to semi-compensatory behaviour. Since extreme alternatives contain both best and worst performing attributes in the choice set, the higher amount of regret engendered by their worst performing attributes, described in Equation 5.5, is not fully offset by their best performing attributes, and this can lead to the extreme alternative being relatively less preferred and the compromise alternative being relatively more preferred. Hence, the discussion of the RRM model is closely related to the extremeness aversion heuristic, and therefore linked to the NLWLR model elaborated in Section 5.4 as well as the relative advantage model to be discussed in Section 5.6.

Table 5.9 summarises the results of estimating the RRM model on each of the seven toll road datasets. With the exception of the free flow time taste parameter in NZ99 which is not statistically significant, all other parameters are statistically significant at the five percent level and of the correct sign. In particular, estimating the RRM model on Aust00 has allowed the taste parameter for the RC attribute to become significant, although the log-likelihood statistic for the model is lower than for the standard RUM model.

Implications on the value of travel time savings arising from fitting the RRM model to the data are discussed later in Section 5.9, but for now, it suffices to note that the results obtained from the estimation of the RRM model on the seven toll road datasets are consistent with earlier findings that the empirical performance of the RRM model is somewhat mixed when compared with the standard linear additive RUM model. As reported in previous studies (see for example, Chorus, 2012), the full RRM model, compared to the standard linear additive

Table 5.9: Estimation Results from the RRM (MNL) Model

RRM Model	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio in parentheses)			
Free flow time (FF) (min)	– 0.0332*** (– 7.05)	– 0.0308*** (– 5.21)	n.s.	– 0.0504*** (– 7.86)	– 0.0238*** (– 7.96)	– 0.0465*** (– 16.92)	– 0.0622*** (– 14.31)
Congestion time (CT) (min)		– 0.0358*** (– 7.23)			– 0.0242*** (– 7.17)	– 0.0631*** (– 26.50)	– 0.0837*** (– 9.51)
Slowed down time (SDT) (min)	– 0.0472*** (– 9.36)		– 0.0549*** (– 6.02)	– 0.0754*** (– 12.92)			
Stop start time (SST) (min)	– 0.0549*** (– 13.50)		– 0.1160*** (– 9.29)	– 0.1114*** (– 19.03)			
Running Cost (RC) ($)	– 0.2171*** (– 8.54)	– 0.2727*** (– 3.05)	– 0.1653*** (– 3.45)	– 0.4142*** (– 10.77)	– 0.0770*** (– 6.36)	– 0.2091*** (– 14.15)	– 0.2872*** (– 9.51)
Toll Cost (TC) ($)	– 0.1813*** (– 12.46)	– 0.2328*** (– 17.96)	– 0.5757*** (– 13.08)	– 0.3390*** (– 21.84)	– 0.1083*** (– 9.49)	– 0.2578*** (– 27.77)	– 0.4175*** (– 21.28)
Alternative Specific Constants							
-*current* alternative	0.9156*** (18.32)	0.6599*** (11.46)	1.0784*** (14.62)	0.3578*** (6.34)	– 0.5837*** (– 5.59)	n.s.	n.s.
-Alternative *A*	n.s.	n.s.	0.2319*** (2.86)	n.s.	0.2394*** (2.79)	n.s.	n.s.
No. of observations	4480	2352	2432	4,864	912	3,888	1840
Log-Likelihood at convergence	– 3439.32	– 1951.54	– 1691.55	– 2683.70	– 847.55	– 3044.11	– 1639.22
Normalised AIC	1.538	1.664	1.396	1.106	1.872	1.568	1.786
Log-Likelihood (standard RUM)	– 3434.02	–1861.64	– 1694.73	– 2669.73	– 847.75	– 3031.58	– 1630.62
Reject standard RUM model in favour of RRM?	No	No	Yes	No	Yes	No	No

***denotes significance at the one percent level.
n.s.: not significant

RUM, offers improvement in model fit in some, but not all, of the datasets considered. Out of the seven datasets studied, the RRM model only fits NZ99 and Aust04a better. In the latter case, only a very marginal improvement is observed over the RUM model.

5.5.2 The Hybrid RRM-RUM Model

One attempt to refine the RRM model is to introduce heterogeneity in decision making along the dimension of how attributes are processed. For example, in a hybrid RRM/linear additive RUM model, respondents are assumed to process a subset of attributes according to RRM, and the remaining attributes of the alternatives according to a linear additive processing rule (Chorus *et al.*, 2013). If it is assumed that attributes *1,..., m* of alternative *j* are processed according to linear additive RUM, and attributes *m+1, ..., K* are processed according to RRM, then the observed component of utility can be described by Equation (5.7):

$$V_j^{RRM-RUM} = \beta_{0,j} + \sum_{k=1,\ldots,m} \beta_{jk} X_{jk} - \sum_{\substack{j' \in S,\\ j' \neq j}} \sum_{k=m+1,\ldots,K} \ln(1 + \exp(\beta_{j'k} X_{j'k} - \beta_{jk} X_{jk})) \tag{5.7}$$

As there is no theory to *a priori* determine which attributes are RUM-processed and which are RRM-processed, one approach that can be adopted is to search over all possible RUM/RRM combinations of attributes to find the one combination that results in the best model fit. The results obtained by estimating the best performing hybrid RRM-RUM model, together with an indication of whether an attribute is RRM or RUM processed, are reported in Table 5.10. The estimated taste parameters in the hybrid model are for the most part statistically significant at the five percent level and of the correct sign. The two exceptions, which were also found in the case of the standard RUM estimation of Section 5.2, are for the RC attribute in Aust00 and the FF attribute in NZ99. As far as goodness of fit is concerned, a comparison of the hybrid RRM-RUM model with the standard RUM utility estimation shows that the empirical performance of the hybrid model is also somewhat mixed, like the RRM model. The hybrid model might be preferred in NZ99, Aust04a, Aust04b and NZ07, whereas the RUM specification might still be retained for Aust08, Aust00 and Aust05. However, the hybrid model can be said to be a better performer than the RRM model for all datasets, suggesting that applying the RRM assumption to all attributes might be overly constrictive. In a sense, the hybrid model may be viewed as a counterpart to the NLWLR model, from which

Table 5.10: Results from Hybrid RRM-RUM (MNL) Estimation

Hybrid RRM-RUM Model	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio in parentheses)			
Free flow time (FF) (min)	– 0.0515*** (– 7.35) (U)	– 0.1214*** (– 14.56) (U)	n.s.	– 0.0507*** (–7.89) (R)	– 0.0351*** (– 8.36) (U)	– 0.0472*** (– 17.04) (R)	– 0.0998*** (– 14.74) (U)
Congestion time (CT) (min)		– 0.0686*** (– 13.22) (R)			– 0.0243*** (– 7.19) (R)	– 0.0898*** (– 28.70) (U)	– 0.1280*** (– 10.12) (U)
Slowed down time (SDT) (min)	– 0.0725*** (– 9.94) (U)		– 0.0543*** (– 6.01) (R)	– 0.1163*** (– 13.58) (U)			
Stop start time (SST) (min)	– 0.0804*** (– 14.26) (U)		– 0.1701*** (– 9.68) (U)	– 0.1682*** (– 20.28) (U)			
Running Cost (RC) ($)	– 0.3417*** (– 9.17) (U)	n.s.	– 0.2605*** (– 4.05) (U)	– 0.5649*** (– 10.19) (U)	– 0.1172*** (– 6.55) (U)	– 0.2131*** (– 14.34) (R)	– 0.4475*** (– 9.95) (U)
Toll Cost (TC) ($)	– 0.1823*** (– 12.50) (R)	– 0.5281*** (– 21.14) (U)	– 0.5796*** (– 13.11) (R)	– 0.5055*** (– 22.84) (U)	– 0.1083*** (– 9.48) (R)	– 0.3693*** (– 29.69) (U)	–0. 4385*** (– 21.27) (R)
Alternative Specific Constants							
-*current* alternative	0.9294*** (18.64)	0.6105*** (8.74)	1.0851*** (14.69)	0.3571*** (6.33)	– 0.5597*** (– 5.32)	n.s.	n.s.
-Alternative *A*	n.s.	0.1432** (2.02)	0.2324*** (2.86)	n.s.	0.2409*** (2.80)	n.s.	n.s.
No. of observations	4480	2352	2,432	4864	912	3,888	1840
Log-Likelihood	– 3434.79	– 1872.09	– 1690.49	– 2670.20	– 846.53	– 3030.78	– 1631.14
Normalised AIC	1.536	1.596	1.395	1.100	1.870	1.561	1.777
Reject standard RUM model in favour of hybrid model?	No	No	Yes	No	Yes	Yes	Yes
Reject RRM model in favour of hybrid model?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*** denotes significance at the one percent level.
(R): Attribute is RRM-processed
(U): Attribute is RUM-processed
n.s.: not significant

one might recall that some attributes are processed according to linear additive utility (if the power parameters are not statistically different from one) while others are processed through a non-linear rule.

Another empirical observation relates to the frequency of occurrence of each attribute in the RRM component of the model. While there does not appear to be a general rule as to the allocation of each attribute either to the RRM or RUM part of the utility, the TC attribute appears in the RRM part of the utility function in four of the seven datasets, which is the highest occurrence in the RRM for any of the attributes considered. The TC attribute has the highest likelihood of being processed according to a regret minimisation heuristic which would be consistent with prospect theory and loss aversion, considering that it is a very salient attribute of the alternative and one which also invokes a substantial amount of negative emotion.

5.6 THE RELATIVE ADVANTAGE MAXIMISATION MODEL AS A DESCRIPTION OF EXTREMENESS AVERSION

5.6.1 Setting up the RAM Model

As discussed in Chapter 2, the componential contextual model first proposed by Tversky and Simonson (1993) and later relabelled a relative advantage model by Kivetz *et al.* (2004) is another candidate representation for the extremeness aversion heuristic. In the context of discrete choice modelling, it might be useful to term such a model as a "relative advantage maximisation" model, since the relative advantage component of the model enters directly into the utility specification for alternative *j* as a term to be maximised.

Like the RRM model, the RAM model assumes that each alternative is assessed against all other alternatives in the choice set. However, one key difference between RAM and RRM is that the RAM model explicitly considers the disadvantages and advantages of an alternative, with the advantages of an alternative expressed as a ratio to the sum of advantage and disadvantage.

Since the RAM model, due to Kivetz *et al.* (2004), has been described extensively in Chapter 2, again, some salient points regarding the model are simply noted here by way of recall.

$A_k(j, j')$, which is the advantage of *j* over *j'* with respect to attribute *k*, might be defined as the increase in the part utility of alternative *j* over alternative *j'* with respect to attribute *k*. Following the standard representation of part utility used in the standard RUM model, i.e., the multiplication of the attribute value by its parameter weight, $A_k(j, j')$ can be specified in Equation (5.8) as follows:

$$A_k(j, j') = \begin{cases} \beta_{jk} X_{jk} - \beta_{j'k} X_{j'k} \text{ if } \beta_{jk} X_{jk} - \beta_{j'k} X_{j'k} \geq \tau_k^{j \to j'} \\ 0 \text{ otherwise.} \end{cases} \tag{5.8}$$

Equation (5.8) indicates that the minimum value that $A_k(j, j')$ can take is zero, that is, $A_k(j, j')$ is always non-negative. Given the values of $X_{j'k}$, β_{jk} and $\beta_{j'k}$, $A_k(j, j')$ is a piecewise linear function in X_{jk}. $\tau_k^{j \to j'}$ can be interpreted as a lower bound threshold level specific to attribute *k* for which $A_k(j, j')$ is recognised by respondents as having a non-zero value. In other words, the advantage $A_k(j, j')$ is perceived only when there are "large enough" differences in the part utilities of *j* over *j'*. For simplicity, $\tau_k^{j \to j'}$ is assumed to be zero for all respondents over all attributes *k*, ignoring this notion of "large enough" differences, so that any increase in the part utility $\beta_{jk} X_{jk}$ over $\beta_{j'k} X_{j'k}$ will imply a non-zero advantage of *j* over *j'*. Future work might explore the possibility of incorporating non-zero $\tau_k^{j \to j'}$ thresholds into the RAM model.

The specification of the disadvantage variable $D_k(j, j')$ is dependent on which assumptions are used in the model. In the simplest case, if symmetry between advantage and disadvantage is assumed, that is, if the disadvantage of alternative *j* over alternative *j'* with respect to an attribute *k* is the corresponding advantage of *j'* over *j* with respect to the same attribute, then Equation (5.9) follows:

$$D_k(j, j') = A_k(j', j) \tag{5.9}$$

To maintain a parsimonious representation for the RAM model, this assumption of symmetry in the advantage and disadvantage functions will be used throughout the analysis. This means that the RAM, RUM and RRM models use the same number of taste parameters, in contrast to the NLWLR model which requires additional power parameters to be estimated. Empirically, as Kivetz *et al.* (2004) demonstrated, the loss aversion parameters associated with $D_k(j, j')$ are not statistically significant, and so the symmetry assumption is not necessarily overly restrictive. Nevertheless, further work may consider extensions to allow the $D_k(j, j')$ functions to be increasing and convex in $A_k(j', j)$.

For notational convenience, represent $A_k(j, j')$ and $D_k(j, j')$ by Equation (5.10):

$$A_k(j, j') = \max\left(\left[\beta_{jk} X_{jk} - \beta_{j'k} X_{j'k}\right], 0\right) \text{ and}$$
$$D_k(j, j') = \max\left(\left[-\beta_{jk} X_{jk} + \beta_{j'k} X_{j'k}\right], 0\right). \tag{5.10}$$

Then, Equation (5.11) follows from the definition of the relative advantage of alternative j over alternative j', $RA(j, j')$:

$$RA(j, j') = \frac{A(j, j')}{A(j, j') + D(j, j')} = \frac{\sum_k A_k(j, j')}{\sum_k A_k(j, j') + \sum_k D_k(j, j')} \tag{5.11}$$

In cases where the denominator $A(j, j') + D(j, j')$ is equal to zero, such that alternative j confers neither an advantage nor a disadvantage over alternative j', $RA(j, j')$ is mathematically undefined. In this case, it would be convenient to assume that alternative j confers no relative advantage over alternative j' and so $RA(j, j')$ is assumed to be zero. Such a scenario may arise when all attribute values of alternatives j and j' that enter into the $R(j, j')$ function are equal, or when thresholds are assumed for $A(j, j')$, such that small enough differences in the attribute values between the alternatives are either not perceived or disregarded by the respondent. The first scenario might occur especially in a stated choice experiment where due to the experimental design, a small number of choice situations exist in which all attribute values of alternatives j and j' that enter into the $RA(j, j')$ function are equal. In the toll road datasets, a handful of such cases were encountered. The denominator $A(j, j') + D(j, j')$ may also be equal to zero when all the taste parameters β_{jk} are equal to zero, and this is a potential drawback of the model if during the estimation process, all the β_{jk} have to be set to zero.

Having defined $RA(j, j')$, the modelled component of utility for alternative j is then written as a linear combination of the linear additive RUM model and the relative advantage component $RA(j, j')$, as shown in Equation (5.12):

$$V_j^{RAM} = \beta_{0,j} + \sum_k \beta_{jk} X_{jk} + \sum_{\substack{j' \in s, \\ j' \neq j}} RA(j, j') \tag{5.12}$$

Since this particular functional specification assumes $RA(j, j')$ is composed of piecewise linear functions of $A_k(j, j')$ and $D_k(j, j')$, it is useful to call such a model the "piecewise RAM" model, to be contrasted with another version of the RAM model discussed in Section 5.6.6. Unlike the RRM model, the RAM model allows for a combination of context independent ('inherent') preferences and context dependent preferences, which is consistent with Kivetz *et al.*'s (2008) hypothesis that preferences may not be entirely context dependent. The RAM model acknowledges that preferences are to a certain extent shaped by the choice context, but also allows each person a set of context free, innate preferences which are brought to bear on each choice situation.

A question that may arise is whether the RAM model and in particular, whether the $RA(j, j')$ component, imposes a prior sign constraint on β_{jk}. A model that requires a prior sign constraint to be imposed on the parameters is not desirable since it goes against the scientific method of allowing the data to impact the conclusions. By definition, while $RA(j, j')$ is restricted to only non-negative values, the same is not true for β_{jk}. In other words, barring the special circumstances discussed earlier, $RA(j, j')$ is defined for all values of β_{jk}. If higher values of an attribute are preferred to lower values, one expects β_{jk} to be positive and the combination of a positive β_{jk} and a positive $X_{jk} - X_{j'k}$ leads to a positive $A_k(j, j')$. Similarly, if lower values of an attribute are preferred to higher values, one expects β_{jk} to be negative and the combination of a negative β_{jk} and a negative $X_{jk} - X_{j'k}$ again leads to a positive

$A_k(j, j')$. Figure 5.1 illustrates the graph of $A_k(j, j')$ under two sign assumptions of β_k, where β_k has been assumed to be generic for ease of illustration.

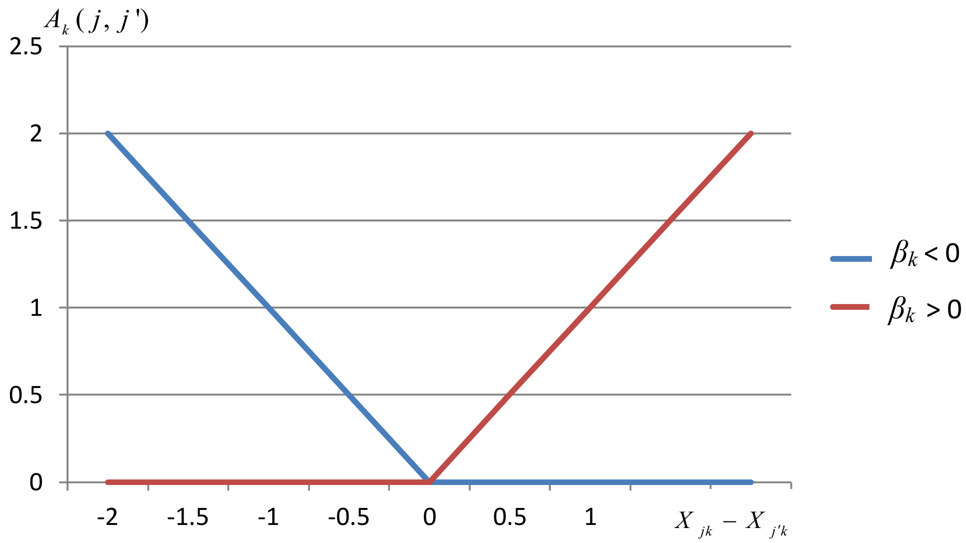


Figure 5.1: Graph of $A_k(j, j')$ under Two Different Sign Assumptions on β

The role of $RA(j, j')$ is to allow the overall utility for alternative j to be augmented or modified depending on how well this alternative performs compared to the reference alternative, the reference being all other competing alternatives j' in the choice set. Under the RAM model, an attribute of an alternative j that does 'better' than its counterpart attribute in alternative j' contributes to the utility of alternative j beyond what is specified by the context independent rule. An attribute that does 'worse' than its counterpart in another alternative adds nothing to the utility of alternative j.

In most empirical work in this chapter, it will be assumed that all attributes which appear in the context independent component RUM component will also be included in the relative advantage component of V_j^{RAM}. The two exceptions to this assumption will be discussed later in sub-section 5.6.4, when the empirical results of the RAM model are presented. It might also be observed that both the context independent component and the relative advantage components, as written in Equation (5.12) are given equal weights in V_j^{RAM}. In Section 5.7, the weight of the context independent component and/or the relative advantage component of utility is allowed to be heterogeneous across respondents, by conditioning on certain socio-economic characteristics, using a multiple heuristics approach in the utility function.

Using the additive error term structure, the total utility for an alternative in the piecewise RAM model is simply the sum of the observed component and an unobserved error component ε_j, i.e., $U_j^{RAM} = V_j^{RAM} + \varepsilon_j$.

5.6.2 Some Theoretical Properties of the RAM Model

Some theoretical properties of the RAM model are briefly discussed in this section. Like the RRM model, the RAM model does not exhibit the IIA property, even though *i.i.d.* error terms are assumed[7]. This should not be entirely surprising considering that the utility function for an alternative j in the RAM model makes explicit reference and comparison to all competitor alternatives j' in the choice set. The non-IIA feature of the RAM model is easily observed if we take the ratio of choice probabilities of alternatives p and q from a choice set consisting of three alternatives p, q and r in Equation (5.13).

$$\frac{\text{Prob}(p)}{\text{Prob}(q)} = \frac{\exp(V_p^{RAM})}{\exp(V_q^{RAM})} = \frac{\exp\left(\beta_{0,p} + \sum_k \beta_{pk} X_{pk} + \dfrac{\sum_k A_k(p,q)}{\sum_k A_k(p,q) + \sum_k D_k(p,q)} + \dfrac{\sum_k A_k(p,r)}{\sum_k A_k(p,r) + \sum_k D_k(p,r)}\right)}{\exp\left(\beta_{0,q} + \sum_k \beta_{qk} X_{qk} + \dfrac{\sum_k A_k(q,p)}{\sum_k A_k(q,p) + \sum_k D_k(q,p)} + \dfrac{\sum_k A_k(q,r)}{\sum_k A_k(q,r) + \sum_k D_k(q,r)}\right)} \tag{5.13}$$

Clearly, $\dfrac{\text{Prob}(p)}{\text{Prob}(q)}$ is dependent on the attribute levels of alternative r.

It has been argued that that the $D(j, j')$ function needs to be an increasing and convex function of $A(j, j')$ in order for the RAM model to display the compromise effect (Simonson and Tversky, 1993; Kivetz et al., 2004). It turns out that this is not a necessary condition for the piecewise RAM model to display the compromise effect. To demonstrate this, consider

[7] Although see an example at the end of this section for a caveat.

the simplest case of two attributes $\{1,2\}$ and three alternatives $\{p, q, r\}$. Assume that q is the in-between alternative and p and r are the extreme alternatives. Let the context independent component of preferences be denoted by v*(.)*. In the standard RUM model, $\text{v}(p) = \text{v}_{p1} + \text{v}_{p2}$, $\text{v}(q) = \text{v}_{q1} + \text{v}_{q2}$ and $\text{v}(r) = \text{v}_{r1} + \text{v}_{r2}$.

Let the probability ratio $\frac{\text{Prob}(q)}{\text{Prob}(p)}$ in the binary choice problem $\{p,q\}$ be denoted by $\text{Pr}(q / p)_{(p,q)}$. As before, the alternatives present in the choice set are explicitly listed in the subscript. Hence, the same probability ratio $\frac{\text{Prob}(q)}{\text{Prob}(p)}$ in the three alternative problem $\{p, q, r\}$ is denoted by $\text{Pr}(q / p)_{(p,q,r)}$. Figure 5.2, which is modified from Figure 4 and Figure 5 in Tversky and Simonson (1993), provides a graphical illustration of the RAM model for two attributes, with the axes v_1 and v_2 denoting the part utility of attributes 1 and 2 respectively.

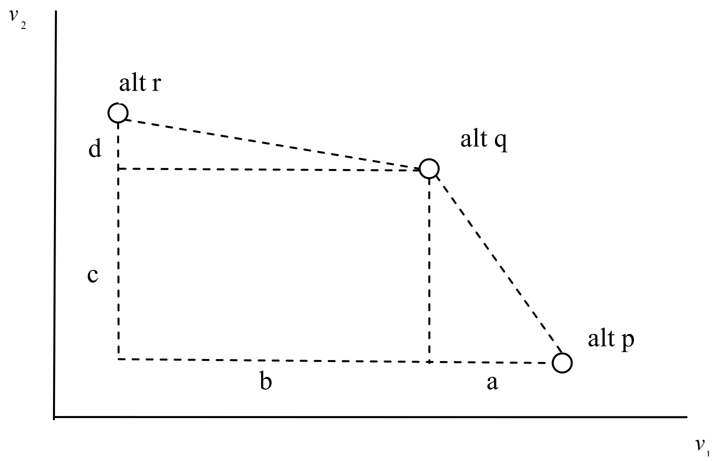


Figure 5.2: Piecewise RAM Model with Compromise Effect

In the two alternatives $\{p,q\}$ case, where alternative r is unavailable, the values of the various advantage and disadvantage functions are as follows:

$A_1(p,q) = D_1(q,p) = a$
$A_2(p,q) = D_2(q,p) = 0$
$D_1(p,q) = A_1(q,p) = 0$
$D_2(p,q) = A_2(q,p) = c$

Therefore, with two alternatives $\{p,q\}$ in the choice set, the piecewise RAM model is defined as follows in Equation (5.14):

$$\begin{aligned} V^{RAM}_{p_{(p,q)}} &= \mathrm{v}(p) + RA(p,q) = \mathrm{v}(p) + \frac{a}{a+c} \\ V^{RAM}_{q_{(p,q)}} &= \mathrm{v}(q) + RA(q,p) = \mathrm{v}(q) + \frac{c}{c+a} \end{aligned} \tag{5.14}$$

In the three alternatives $\{p,q,r\}$ case, the piecewise RAM model is given by Equation (5.15):

$$\begin{aligned} V^{RAM}_{p_{(p,q,r)}} &= \mathrm{v}(p) + RA(p,q) + RA(p,r) = \mathrm{v}(p) + \frac{a}{a+c} + \frac{a+b}{a+b+c+d} \\ V^{RAM}_{q_{(p,q,r)}} &= \mathrm{v}(q) + RA(q,p) + RA(q,r) = \mathrm{v}(q) + \frac{c}{c+a} + \frac{b}{b+d} \\ V^{RAM}_{r_{(p,q,r)}} &= \mathrm{v}(r) + RA(r,p) + RA(r,q) = \mathrm{v}(r) + \frac{c+d}{c+d+a+b} + \frac{d}{d+b} \end{aligned} \tag{5.15}$$

In the classic definition of the compromise effect, the compromise effect is said to occur when $\Pr(q/p)_{(p,q,r)} > \Pr(q/p)_{(p,q)}$. In the piecewise RAM model with *i.i.d.* EV type I error terms, $\Pr(q/p)_{(p,q,r)} = \exp(V^{RAM}_{q_{(p,q,r)}} - V^{RAM}_{p_{(p,q,r)}})$ and $\Pr(q/p)_{(p,q)} = \exp(V^{RAM}_{q_{(p,q)}} - V^{RAM}_{p_{(p,q)}})$. Hence, showing $\Pr(q/p)_{(p,q,r)} > \Pr(q/p)_{(p,q)}$ is equivalent to showing whether the inequality $(V^{RAM}_{q_{(p,q,r)}} - V^{RAM}_{p_{(p,q,r)}}) - (V^{RAM}_{q_{(p,q)}} - V^{RAM}_{p_{(p,q)}}) > 0$ is satisfied, since the exponential function is a one-to-one increasing function.

After some algebraic simplification, Equation (5.16) is obtained:

$$(V^{RAM}_{q_{(p,q,r)}} - V^{RAM}_{p_{(p,q,r)}}) - (V^{RAM}_{q_{(p,q)}} - V^{RAM}_{p_{(p,q)}}) = \frac{bc - ad}{(b+d)(a+b+c+d)} \tag{5.16}$$

For the term on the right hand side of the equation to be strictly greater than zero, the condition $bc > ad$ must hold, since the denominator is always positive, by definition. This occurs unambiguously when $c \geq a$ and $d \leq b$, with at least one of these inequalities holding strictly. It can be seen from Figure 5.2 that $\mathrm{v}_{p1} = \mathrm{v}_{q1} + a$ and $\mathrm{v}_{p2} = \mathrm{v}_{q2} - c$, and the addition of these two equations results in $\mathrm{v}(p) = \mathrm{v}(q) + a - c$. Therefore, $c \geq a$ implies $\mathrm{v}(p) \leq \mathrm{v}(q)$.

Similarly, for the condition $d \le b$ to hold, the inequality $\mathrm{v}(q) \ge \mathrm{v}(r)$ must be satisfied. Hence, the compromise effect can be obtained in the piecewise RAM model if $\mathrm{v}(q) \ge \mathrm{v}(p)$ and $\mathrm{v}(q) \ge \mathrm{v}(r)$, with one of these inequalities holding strictly. A similar argument can be used to show that $\Pr(q/r)_{(p,q,r)} > \Pr(q/r)_{(q,r)}$ when $\mathrm{v}(q) \ge \mathrm{v}(p)$ and $\mathrm{v}(q) \ge \mathrm{v}(r)$, with at least one strict inequality, which completes the demonstration of the compromise effect from the perspective of adding an extreme alternative p to the binary set of $\{q, r\}$.

As an aside, it is well known that the standard RUM MNL model will not display the compromise effect because of the IIA property of the model. Observe also that if the signs of the inequalities are reversed, that is, $\mathrm{v}(q) \le \mathrm{v}(p)$ and $\mathrm{v}(q) \le \mathrm{v}(r)$, and again, with at least one of these inequalities holding strictly, the piecewise RAM model leads to extremeness seeking, in which there is an increase in the probability share of the extreme alternatives p and r relative to q in the three alternative case compared with the binary case, in other words, $\Pr(q/p)_{(p,q,r)} < \Pr(q/p)_{(p,q)}$ and $\Pr(q/r)_{(p,q,r)} < \Pr(q/r)_{(q,r)}$. Extremeness seeking seems to be less common than extremeness aversion, but it has been documented in a few cases (Gourville and Soman, 2007). It is also possible for the piecewise RAM model to exhibit the polarisation effect. For example, if there is a polarisation that favours alternative p away from alternative r, then $\Pr(q/r)_{(p,q,r)} > \Pr(q/r)_{(q,r)}$ and $\Pr(p/r)_{(p,q,r)} > \Pr(p/r)_{(p,r)}$ which would happen when $bc > ad$ and $ba > cd$.

One instance where the piecewise RAM model gives rise to an IIA-like property happens when $\mathrm{v}(p) = \mathrm{v}(q) = \mathrm{v}(r)$. Under this scenario, illustrated in Figure 5.3, observe that $a = c$ and $b = d$. Therefore, each of the binary relative advantage terms in Equation (5.15) is equal to 0.5 and the pairwise differences in the utilities of each of the alternatives is equal to zero, whether in the standard RUM model or in the piecewise RAM model. Hence, the inclusion of a third alternative into a binary choice set has no impact on the ratio of choice probabilities. Having said this however, an occurrence of the condition $\mathrm{v}(p) = \mathrm{v}(q) = \mathrm{v}(r)$ is unlikely to ever happen in practice, and it would be reasonable to conclude that in general, the RAM model does not display the IIA property.

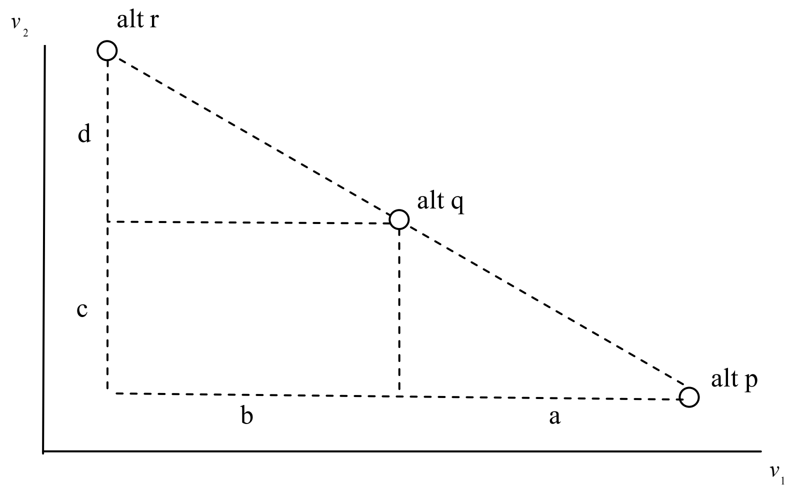


Figure 5.3: Piecewise RAM Model with IIA-like property

5.6.3 A Relative Disadvantage Model and its Equivalence to the RAM Model

This sub-section considers a counterpart model to the RAM model, where instead of maximising relative advantage, decision makers might be thought to choose an alternative on the basis of relative disadvantage minimisation. The motivation for considering such a model may be derived analogously to the motivation behind the RRM model, where negative emotions associated with the choice of an alternative are to be avoided as far as possible. Like relative advantage, the relative disadvantage of alternative j over alternative j', $RD(j,j')$, may be defined in Equation (5.17) as:

$$RD(j,j') = \frac{D(j,j')}{A(j,j') + D(j,j')} \tag{5.17}$$

Like the standard regret function, the relative disadvantage function of Equation (5.17) also preserves the notion of context dependency and the key idea that disutility, in the form of disadvantage, occurs whenever a competing alternative fares better than the considered alternative. However, unlike the regret function, one key difference is that in relative disadvantage, the disadvantage of an alternative is defined relative to the sum of advantage and disadvantage. If one accepts that disadvantage and regret are synonymous concepts, then the notion of 'relative regret' is on view in $RD(j,j')$. A relative disadvantage minimising model (RDM) might then be specified in Equation (5.18) as:

$$V_j^{RDM} = \beta_{0,j} + \sum_k \beta_{jk} X_{jk} - \sum_{\substack{j' \in s, \\ j' \neq j}} RD(j, j') \tag{5.18}$$

As with the conventional understanding of regret, a negative sign is inserted before the relative disadvantage component to indicate that the relative disadvantage of an alternative is to be minimised, or that the negative of relative disadvantage is to be maximised.

It turns out that such a relative disadvantage model as defined in Equation (5.18) is econometrically equivalent to the RAM model defined earlier in Equation (5.12). This is easily seen by noting that $RA(j, j') + RD(j, j') = 1$ and so $RA(j, j') = 1 - RD(j, j')$. Taking the summation over all alternatives $j' \in s, j' \neq j$ results in $\sum_{\substack{j' \neq j \\ j' \in s}} RA(j, j') = \sum_{\substack{j' \neq j \\ j' \in s}} 1 - \sum_{\substack{j' \neq j \\ j' \in s}} RD(j, j')$.

Therefore, V_j^{RAM} differs from V_j^{RDM} by a constant term and since it is only differences in the utility that matter, the constant term washes out and the RAM model specified in Equation (5.12) will produce the same results and output as the RDM model of Equation (5.18).

5.6.4 Empirical Estimation of the Piecewise RAM model with Three Alternatives in Choice Set

For the moment, assume that in the piecewise RAM model, ε_j is *i.i.d.* EV type I distributed, so that all models discussed in this section are of the fixed parameter, MNL form.

Recall from Section 5.4 that the empirical performance of the RRM and the hybrid RRM-RUM models for the seven datasets was mixed when compared to the standard RUM model. Turning to the estimation of the piecewise RAM model using a non-linear logit command, a substantial improvement in model fit compared to the standard RUM, the RRM and the hybrid RRM-RUM models was obtained in all the datasets studied. Results are reported in Table 5.11. In some cases such as Aust00, NZ99 and Aust05, this improvement is close to, or even exceeds, 100 log-likelihood units. Considering that all models contain the same number of estimated parameters, this is a very encouraging finding. Unlike the NLWLR model, estimation of the symmetric RAM model was not particularly sensitive to the choice of starting values, despite the highly non-linear nature of the utility function. Table 5.11

summarises the log-likelihood statistics for the standard RUM, RRM, hybrid RUM-RRM and RAM models. The common feature across all these four models is that they share the same number of parameters, that is, the models are equally parsimonious.

Table 5.11: Comparison of Log-likelihood Statistics across the Four Models

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
	Log-likelihood at convergence						
Standard RUM	– 3434.58	– 1862.23	– 1694.93	– 2670.14	– 847.75	– 3031.58	– 1631.79
RRM	– 3439.32	– 1951.54	– 1691.55	– 2683.70	– 847.55	– 3044.11	– 1639.22
Hybrid RRM-RUM	– 3434.79	– 1872.09	– 1690.49	– 2670.20	– 846.53	– 3030.78	– 1631.14
Piecewise RAM	– 3383.95	– 1783.53	– 1550.81	– 2580.28	– 826.23	– 2990.51	– 1605.15

An examination of the parameter estimates for Aust08, Aust00 and Aust05 through NZ07 (Table 5.12 below) shows that all parameter estimates are of the expected sign and are statistically significant at the five percent level. In Aust00, the RC taste parameter, which was not statistically significant in the RUM and hybrid models, has now turned statistically significant. The RRM is the only other model examined so far that allows the RC taste parameter to be estimated with statistical significance. However, the clear advantage of the piecewise RAM model is that it exhibits a big decrease in the LL statistic compared to the RRM model. This result for the piecewise RAM model was obtained by assuming that the RC attribute appeared only in the context independent RUM component and not in the relative advantage component as either an advantage or a disadvantage.

NZ99 provides an interesting case study on the importance of embedding context dependence effects into the utility function. On analysis, only the TC attribute enters into the relative advantage component of the model; the SDT, SST and RC attributes are excluded from the relative advantage function and only appear in the context independent part of the model (note that the parameter weight for FF is insignificant across all models). Since there is only one attribute in the relative advantage component, $RA(j, j')$ essentially reduces to a [0,1] variable, where “1” indicates that the TC attribute of alternative j confers an advantage over

Table 5.12: Estimates from the Symmetric Piecewise RAM (MNL) Model

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio)			
Attribute							
Free flow time (FF) (min)	– 0.0057*** (– 2.68)	– 0.0487*** (–5.94)	n.s.	– 0.0416*** (– 8.27)	– 0.0060*** (– 2.69)	– 0.0325*** (–13.72)	– 0.0386*** (– 8.31)
Congestion time (CT) (min)		– 0.0532*** (–6.70)			– 0.0066*** (– 2.68)	– 0.0422*** (–17.50)	– 0.0508*** (– 7.13)
Slowed down time (SDT) (min)	– 0.0074*** (– 2.81)		– 0.0726*** (– 6.09)	– 0.0588*** (–11.86)			
Stop start time (SST) (min)	– 0.0103*** (– 3.06)		– 0.1624*** (– 9.01)	–0.0787*** (–14.48)			
Running Cost (RC) ($)	– 0.0564*** (– 3.00)	– 0.2543** (–2.02)	– 0.2958*** (– 7.16)	– 0.2844*** (– 9.59)	– 0.0244*** (– 2.69)	– 0.1473** (–11.56)	– 0.1978*** (– 7.25)
Toll Cost (TC) ($)	– 0.0401*** (– 2.95)	– 0.2697*** (–10.51)	– 0.2375*** (– 6.05)	– 0.2895*** (–17.33)	– 0.0350*** (– 3.00)	– 0.1819*** (–18.23)	– 0.2693*** (–10.19)
Alternative Specific Constants							
-*current* alternative	0.8004*** (15.34)	0.7693*** (9.78)	0.6739*** (9.23)	0.2239*** (3.83)	– 0.7277*** (– 6.69)	n.s.	n.s.
-Alternative *A*	n.s.	0.1775** (2.47)	0.2438*** (2.81)	n.s.	0.2171** (2.44)	n.s.	n.s.
Number of observations	4480	2,352	2,432	4,864	912	3,888	1,840
LL at convergence	– 3383.95	–1783.53	– 1550.81	– 2580.28	– 826.23	– 2990.51	– 1605.15

** denotes significance at the five percent level.
*** denotes significance at the one percent level.
n.s. : not significant

the TC attribute of alternative j', and "0" otherwise[8]. What is surprising is the magnitude of improvement in model performance when a simple indicator function for context dependence is entered into what is essentially the standard RUM specification.

In summary, the empirical estimates from the RAM model suggest that there is a case to be made for using the piecewise RAM model as a superior representation of context dependency compared to say the RRM or the hybrid RRM-RUM models. Secondly, if the results are viewed from the perspective of relative disadvantage minimisation, then, the notion of regret, which has so far been based on some function of absolute differences of attribute values, might benefit from being recast into relative terms, into relative disadvantage, for example.

5.6.5 Estimation of the Piecewise RAM with Two Alternatives in Choice Set

Since their primary motivation for introducing the RAM/componential context model was to explain the compromise effect, Tversky and Simonson (1993) posited that $RA(j, j')$ would be equal to zero in cases where only two alternatives are presented in the choice set. As a matter of empirical estimation, however, it is not necessary for the RAM to be restricted to cases with three or more alternatives in the choice sets, and so the RAM may be estimated even where there are only two alternatives in the choice set. This sub-section therefore illustrates an innovative use of the RAM model by applying the RAM model to binary choice data.

With binary choice data, the RRM model faces one key limitation in that it collapses to the standard RUM model. Therefore, the RRM model is meaningless as a model of context dependency and regret when it comes to binary choice data. Proof that the RRM model reduces to the standard RUM model is shown below and this is similar to the proof presented in Chorus (2010).

---

[8] $RA(j, j')$ was also set to zero when $TC_j = TC_{j'}$

Consider the two alternatives i and j in the choice set, where the regret of alternative i over alternative j on an attribute k is given by the usual regret formula:

$$reg_k(i, j) = \ln[1 + \exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})]$$

Recalling that the minimisation of regret is equivalent to maximising the negative of regret, the "part-utility" of alternative i on attribute k is given by

$$V_{ik}^{RRM} = -\ln[1 + \exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})]$$

The differences in the part utility on attribute k between alternatives i and j is given by

$$\begin{aligned}
V_{ik}^{RRM} - V_{jk}^{RRM} &= -\ln[1 + \exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})] - (-\ln[1 + \exp(\beta_{ik} X_{ik} - \beta_{jk} X_{jk})] \\
&= \ln \frac{1 + \exp(\beta_{ik} X_{ik} - \beta_{jk} X_{jk})}{1 + \exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})} \\
&= \ln \left[\frac{1 + \exp(\beta_{ik} X_{ik} - \beta_{jk} X_{jk})}{\exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})[\frac{1}{\exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})} + 1]} \right] \\
&= \ln \left[\frac{1 + \exp(\beta_{ik} X_{ik} - \beta_{jk} X_{jk})}{\exp(\beta_{jk} X_{jk} - \beta_{ik} X_{ik})[\exp(\beta_{ik} X_{ik} - \beta_{jk} X_{jk}) + 1]} \right] \\
&= \ln[\exp(\beta_{ik} X_{ik} - \beta_{jk} X_{jk})] \\
&= \beta_{ik} X_{ik} - \beta_{jk} X_{jk}
\end{aligned}$$

Denoting $V_j^{RUM} = \sum_k \beta_{jk} X_{jk}$, then

$$\begin{aligned}
V_i^{RRM} - V_j^{RRM} &= \sum_k V_{ik}^{RRM} - \sum_k V_{jk}^{RRM} \\
&= \sum_k \beta_{ik} X_{ik} - \sum_k \beta_{jk} X_{jk} \\
&= V_i^{RUM} - V_j^{RUM}
\end{aligned}$$

Since only differences in utility matter, the RRM model is equivalent to the RUM model in the case of two alternatives.

Fortunately, the RAM model (and equivalently, the RDM model) does not exhibit this property. In other words, where binary choice data are concerned, the RAM/RDM model can still meaningfully capture the notion of context dependency and in the case of the RDM model, the notion of regret, which is not possible with the RRM model. This feature of the RAM/RDM model seems important considering that there is no reason to suspect that context dependency and regret become non-existent whenever there are two alternatives in the choice set. Since the RRM cannot be estimated with binary choice data, the comparison is therefore restricted to between the RAM model and the RUM model.

Also quite fortuitously, the data in the toll road datasets allow for the possibility of comparing the performance of the piecewise RAM model with the standard RUM model in the experimental situation with only two alternatives in the choice set. Such a possibility arises because in all datasets except for NZ99, each respondent was required to answer two choice questions per choice scenario, where the first question required respondents to choose one alternative among the current alternative, and two hypothetical alternatives A and B, and the second choice question asking respondents to choose among the hypothetical alternatives only. Modelling on the responses to the second choice question alone, the results are presented in Table 5.13.

Again, it may be observed that the piecewise RAM model outperforms the RUM model in all six of the datasets analysed. Again, all parameter estimates of the RAM model are statistically significant at the five percent level and of the correct sign. In the case of Aust00, one key difference (or improvement) in the RAM model over the standard RUM model is that the taste parameter for the *RC* attribute has turned significant. Moreover, unlike the three alternative case where the *RC* attribute was excluded from the $RA(j, j')$ component, the RC attribute is now a determinant of preferences in both the context independent component and the relative advantage component of the model.

Overall, the empirical findings suggest that context dependency, as represented by relative advantage maximisation, is still relevant even in the case of binary choice data.

Table 5.13: Comparison of Standard RUM and Piecewise RAM (MNL) Models With Two Alternatives in Choice Set

	$\hat{\beta}$ (z-ratio in parentheses)											
	Aust08		Aust00		Aust05		Aust04a		Aust04b		NZ07	
	RUM	RAM	RUM	RAM	RUM	RAM	RUM	RAM	RUM	RAM	RUM	RAM
Free flow time (FF) (min)	– 0.0509*** (– 8.27)	– 0.0181*** (– 5.85)	– 0.1172*** (– 10.55)	– 0.0357*** (– 4.63)	– 0.1197*** (– 15.33)	– 0.0682*** (– 12.18)	– 0.0301*** (– 6.76)	– 0.0153*** (– 4.89)	– 0.0664*** (– 15.76)	– 0.0423*** (– 13.01)	– 0.0977*** (– 13.89)	– 0.0577*** (– 11.28)
Congestion time (CT) (min)			– 0.1297*** (– 10.95)	– 0.0543*** (– 5.57)			– 0.0373*** (– 6.90)	– 0.0186*** (– 4.98)	– 0.0825*** (– 21.45)	– 0.0523*** (– 15.78)	– 0.0922*** (– 6.99)	– 0.0589*** (– 6.87)
Slowed down time (SDT) (min)	– 0.0489*** (– 7.17)	– 0.0174*** (– 5.19)			– 0.0856*** (– 13.58)	– 0.0508*** (– 11.40)						
Stop start time (SST) (min)	– 0.0699*** (– 13.15)	– 0.0293*** (– 8.71)			– 0.1382*** (– 20.10)	– 0.0891*** (– 17.32)						
Running Cost (RC) ($)	– 0.3695*** (– 12.46)	– 0.1807*** (– 9.80)	n.s.	– 0.3343*** (– 3.46)	– 0.4842*** (– 10.24)	– 0.3264*** (– 10.28)	– 0.1037*** (– 6.09)	– 0.0560*** (– 5.06)	– 0.2845*** (– 14.15)	– 0.1798*** (– 11.99)	– 0.5533*** (– 10.28)	– 0.3643*** (– 10.37)
Toll Cost (TC) ($)	– 0.4869*** (– 25.34)	– 0.2282*** (– 13.63)	– 0.6568*** (– 19.87)	– 0.3685*** (– 11.67)	– 0.6231*** (– 27.18)	– 0.4189*** (– 21.85)	– 0.2006*** (– 10.43)	– 0.1107*** (– 7.30)	– 0.3590*** (– 23.57)	– 0.2423*** (– 18.88)	– 0.8179*** (– 18.19)	– 0.5057*** (– 13.54)
ASC – Alt *A*	n.s.	n.s.	0.1657*** (3.20)	0.1132** (2.04)	n.s.	n.s.	0.2749*** (3.52)	0.2435*** (2.94)	0.2084*** (4.95)	0.1912*** (4.30)	0.1179** (2.04)	n.s.
LL at convergence	– 2515.95	– 2454.37	– 1137.09	– 1080.82	– 2365.11	– 2289.52	– 499.85	– 487.85	– 1820.94	– 1788.64	– 928.45	– 886.62
Log-likelihood (0)	– 3104.44		– 1625.87		– 3371.23		– 627.50		– 2691.24		– 1267.53	
No. of observations	4480		2352		4864		912		3888		1840	
Reject standard RUM model in favour of RAM model?	Yes		Yes		Yes		Yes		Yes		Yes	

*** denotes significance at the one percent level.
** denotes significance at the five percent level.
n.s.: not significant

5.6.6 An Alternative Specification for the RAM Model Using the Regret Function

From an empirical point of view, one concern that might be raised with estimating the piecewise RAM model as originally proposed by Kivetz *et al.* (2004) is the non-differentiability of the utility function at the kink. There may also be stability issues in the model arising from the possibility that the denominator in the $RA(j, j')$ component of the piecewise RAM model may be zero. The z-ratios of the parameter estimates are, in several cases such as Aust08 and Aust04a, a fraction of the counterpart z-ratios in the standard RUM model, arguably making the RAM model estimates less useful for practitioners because less precise parameter estimates are obtained. To address these potential concerns, this section proposes an alternative specification of the RAM model.

A simple starting point for such an alternative specification might be to note that since regret and disadvantage are practically synonymous, the regret function may be used to represent $D_k(j, j')$ such as in Equation (5.20):

$$D_k(j, j') = \ln[1 + \exp(\beta_{j'k} X_{j'k} - \beta_{jk} X_{jk})] \tag{5.20}$$

Again, assuming symmetry between advantage and disadvantage, Equation (5.21) follows:

$$A_k(j, j') = D_k(j', j) = \ln[1 + \exp(\beta_{jk} X_{jk} - \beta_{j'k} X_{j'k})] \tag{5.21}$$

In this regret form of the RAM model, the definitions of $A(j, j') = \sum_k A_k(j, j')$, $D(j, j') = \sum_k D_k(j, j')$ and $RA(j, j') = \dfrac{A(j, j')}{A(j, j') + D(j, j')}$ are now based on Equation (5.20) and Equation (5.21). In this case, $RA(j, j')$ only takes values in the open interval between zero and one.

A graph of the function $f(z) = \dfrac{\ln(1 + \exp(z))}{\ln(1 + \exp(z)) + \ln(1 + \exp(-z))}$ is plotted in Figure 5.4 as a means of visualising $RA(j, j')$ in the case of a single attribute. Where the alternatives contain multiple attributes, changing the value of one attribute holding all else constant is akin to adding a constant to both the numerator and the denominator and this does not fundamentally alter the shape of the function.

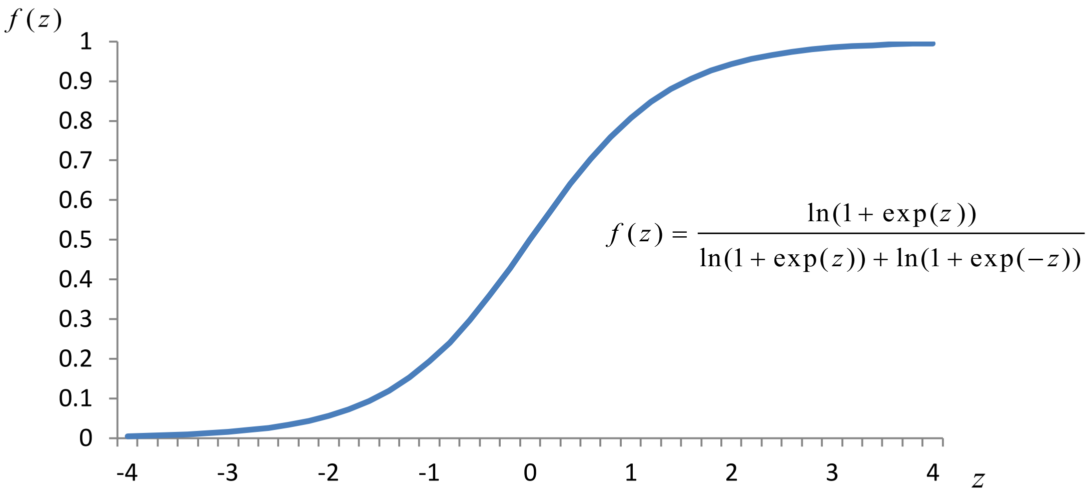


Figure 5.4: Graph of the $RA(j, j')$ function

It is immediately obvious from Figure 5.4 that $RA(j, j')$ follows an S-shaped curve reminiscent of the value function from prospect theory (Kahneman and Tversky, 1979) and that it captures the notion of concavity in gains and convexity in the losses. This is a rather nice result which suggests that the use of the symmetry assumption $A_k(j, j') = D_k(j', j)$ is not entirely inappropriate. Importantly, the purpose of introducing an alternative specification for $RA(j, j')$ has been met, which is that $RA(j, j')$ is now a smooth function, unlike the kinked functions used by Kivetz *et al.* (2004). This model might be referred to as the regret-RAM model. Estimating the regret-RAM model on the toll road datasets (and dropping all insignificant parameters) yields the results reported in Table 5.14. Table 5.15 provides a comparison of the log-likelihood statistics across the standard RUM, RRM, hybrid RRM-RUM and regret-RAM models.

As with the piecewise RAM model, estimating the regret-RAM model results in a better fit of the data, compared to the standard RUM, the RRM, and the hybrid RRM-RUM models, in almost all of the datasets studied, with the single exception that the hybrid model outperforms the regret-RAM model in Aust04a. However, the overall improvement in log-likelihood is now much more modest compared to the piecewise form of the regret model estimated in Section 5.5.4. Nevertheless, this improvement can still be quite substantial, as seen in Aust00,

Table 5.14: Estimates from the Regret Form of RAM (MNL) Model (Three Alternatives in Choice Set)

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio)			
Attribute							
Free flow time (FF) (min)	– 0.0427*** (– 7.84)	– 0.1030*** (–14.39)	n.s.	– 0.0649*** (– 8.87)	– 0.0280*** (– 8.46)	– 0.0553*** (–19.05)	– 0.0804*** (– 16.07)
Congestion time (CT) (min)		– 0.0983*** (–13.32)			– 0.0282*** (– 8.24)	– 0.0733*** (–29.68)	– 0.1028*** (– 11.00)
Slowed down time (SDT) (min)	– 0.0599*** (– 10.12)		– 0.0650*** (– 6.62)	– 0.0985*** (–15.26)			
Stop start time (SST) (min)	– 0.0669*** (– 14.98)		– 0.1389*** (– 9.42)	– 0.1418*** (–21.37)			
Running Cost (RC) ($)	– 0.2866*** (– 10.88)	n.s.	– 0.2214*** (– 6.35)	– 0.4732*** (– 11.04)	– 0.0937*** (– 7.29)	– 0.2554*** (–14.81)	– 0.3806*** (– 11.21)
Toll Cost (TC) ($)	– 0.2302*** (– 13.19)	– 0.4578*** (–21.62)	– 0.7037*** (– 20.83)	– 0.4307*** (–25.69)	– 0.1261*** (– 9.72)	–0.2948*** (–29.69)	– 0.5240*** (–22.17)
Alternative Specific Constants							
-*current* alternative	0.9076*** (18.62)	0.5991*** (8.13)	1.0561*** (14.77)	0.3405*** (6.23)	– 0.5521*** (– 5.30)	0.0897** (2.08)	n.s.
-Alternative *A*	n.s.	0.1447** (2.04)	0.2321*** (2.74)	n.s.	0.2397*** (2.77)	n.s.	n.s.
Number of observations	4,480	2,352	2,432	4,864	912	3,888	1,840
LL at convergence	– 3433.77	–1854.09	– 1688.83	– 2664.50	– 847.29	– 3027.75	– 1630.32

*** denotes significance at the one percent level.
n.s.: not significant

Table 5.15: Comparison of Standard RUM, RRM, Hybrid and Regret-RAM Models (all MNL) with Three Alternatives in Choice Set

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
	Log-likelihood at convergence						
Standard RUM	– 3434.58	– 1862.23	– 1694.93	– 2670.14	– 847.75	– 3031.58	– 1631.79
RRM	– 3439.32	– 1951.54	– 1691.55	– 2683.70	– 847.55	– 3044.06	– 1639.22
Hybrid RRM-RUM	– 3434.79	– 1872.09	– 1690.49	– 2670.20	– 846.53	– 3030.78	– 1631.14
Regret-RAM	– 3433.77	– 1854.09	– 1688.83	– 2664.50	– 847.29	– 3027.75	– 1630.32

Note: For each dataset, the number of free parameters used across all model types is the same.

NZ99 and Aust05. Overall, this is still a remarkable finding and points to the robustness of the RAM model in explaining choice behaviour. The z-ratios of the parameter estimates are also comparable to those obtained from the standard RUM model. The regret-RAM model also has an advantage over the piecewise-RAM model in that the denominator is never zero and this provides for more stability in the estimation especially if for some reason, all the β_k parameters happen to be zero. A worked example of the regret-RAM model with illustrative values for $A(j, j')$, $D(j, j')$ and $RA(j, j')$ is shown in Table 5.16.

The regret-RAM model may also be estimated on binary choice data. In terms of model fit and parameter estimates, the results are qualitatively similar to those reported for the piecewise RAM model in Section 5.6.5 but with the caveat that like the regret-RAM models in the three alternative case, the magnitude of improvement in fit is much reduced.

5.6.7 Relative versus Absolute Regret

Since the RAM model is comprised of both context independent and context dependent components, it is worth exploring if the addition of a context independent component to the standard RRM model results in the same pattern of improvement to the model fit as the piecewise and regret-RAM models discussed earlier. Following the spirit of the RAM model, a (modified) RRM_1 model may be defined according to Equation (5.22):

$$V_j = \beta_{0,j} + \sum_k \beta_k X_{jk} - \sum_{\substack{j' \in s, \\ j' \neq j}} \sum_k \ln(1 + \exp(\beta_k (X_{j'k} - X_{jk}))) \tag{5.22}$$

Table 5.16: Regret-RAM Model with Example Choice Set from Aust00

	current	Alternative A	Alternative B
FF (mins) CT (mins) RC ($) TC ($)	8 67 3.40 0	4 29 2.30 7.40	6 46 2.70 3.70
	$A(j, j')$		
	j' = current	j' = Alternative A	j' = Alternative B
j = Current		2.02	1.42
j = Alternative A	2.35		1.52
j = Alternative B	1.67	1.44	
	$D(j, j')$		
	j' = current	j' = Alternative A	j' = Alternative B
j = Current		2.35	1.67
j = Alternative A	2.02		1.44
j = Alternative B	1.42	1.52	
	$RA(j, j')$		
	j' = current	j' = Alternative A	j' = Alternative B
j = Current		0.46	0.46
j = Alternative A	0.54		0.51
j = Alternative B	0.54	0.49	
RUM Probabilities	0.271	0.429	0.299
RAM Probabilities	0.276	0.425	0.299

If the improvement in model fit is as hypothesised, that is, it comes about as a result of adding the standard RUM component to the RRM model, then the RRM_1 model, together with the RAM model, should outperform both the standard RRM and the standard RUM models in all seven datasets. All models within the same dataset are estimated with the same number of parameters. A comparison of the model fit of this RRM_1 model with the standard RUM and RRM models is reported in Table 5.17.

Comparing the standard RRM model to the RRM_1 model, it may be observed that in five out of the seven datasets, the RRM_1 model performs better than the standard RRM model.

Table 5.17: Comparison of Log-likelihoods of RRM_1 Model (MNL) with Standard RUM (MNL) and Standard RRM (MNL) Models

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
	Log-likelihood at convergence						
Modified RRM_1	– 3435.99	– 1901.01	– 1692.51	– 2672.97	– 847.60	– 3035.04	– 1633.75
Standard RRM	– 3439.32	– 1951.54	– 1691.55	– 2683.70	– 847.55	– 3044.11	– 1639.22
Standard RUM	– 3434.58	– 1862.23	– 1694.93	– 2670.14	– 847.75	– 3033.93	– 1631.79
Reject standard RRM model in favour of RRM_1?	Yes	Yes	No	Yes	No	Yes	Yes
Reject standard RUM model in favour of RRM_1?	No	No	Yes	No	Yes	No	No

Note: For each dataset, the number of free parameters used across all model types is the same.

In the other two datasets (NZ99 and Aust04a), the performance of the RRM model is only slightly better than the RRM_1 model. This finding supports the hypothesis of Kivetz *et al.* (2008) that preferences are not purely context dependent and therefore suggests that introducing a context independent component to the RRM model might be preferable in many cases to simply representing preferences by the regret function alone.

However, a comparison of the RRM_1 model to the standard RUM specification shows that like the standard RRM and the hybrid RRM-RUM models, the RRM_1 model does not unambiguously exhibit a clear improvement in model fit across all datasets. Much like the standard RRM, the RRM_1 model outperforms the standard RUM model in NZ99 and Aust04a even though in these datasets, the RRM_1 model is not as good as the conventional RRM. In the other datasets, the standard RUM can be said to still be the preferred model. Overall, the empirical results demonstrate that it is not simply a matter of adding a context independent specification to the standard RRM model that will improve the model fit. These suggest that the specification of the regret function context does matter significantly.

As an aside, on these datasets, the RRM_1 model is well defined. However, the model runs into potential problems for larger choice sets, as the RRM-parameters become smaller than their RUM-counterparts due to the additive regret function (Chorus, 2012). As a consequence, RRM_1's intrinsic assumption that both RUM and RRM parameter sets are equal is likely to become restrictive in the context of such larger choice sets.

The standard RRM and RRM_1 models also differ from the two versions of the RAM model explored up to this point in how regret is entered into the utility functions. Fundamentally, the standard RRM and RRM_1 models represent regret in an absolute sense, whereas the RAM models use regret in a relative sense, albeit in two different functional specifications. It is worth considering if it is the relative form of regret that matters, rather than the absolute level of regret.

To explore this hypothesis further, a hybrid version of the two forms of relative regret might also be considered, where $A_k(j, j')$ is still given by a piecewise linear function, such as in Equation (5.23):

$$A(j, j') = \sum_k A_k(j, j') = \sum_k \max\left(\beta_k\left[X_{jk} - X_{j'k}\right], 0\right) \tag{5.23}$$

However, let $D(j, j')$ be represented by the regret function as stated in Equation (5.24):

$$D(j, j') = \sum_k D_k(j, j') = \sum_k \ln[1 + \exp(\beta_k(X_{j'k} - X_{jk}))] \tag{5.24}$$

Using Equations (5.23) and (5.24) means that the model is no longer symmetric in $A(j, j')$ and $D(j, j')$. In this case, the RDM model is estimated, where $RD(j, j')$ may be recalled from Equation (5.17) to be:

$$RD(j, j') = \frac{D(j, j')}{A(j, j') + D(j, j')}$$

The log-likelihood statistics of such a modified RRM_2 model and a comparison with the standard RUM model are presented in Table 5.18. Again, the number of parameters estimated for each dataset is the same across models.

The results show that like the piecewise RAM and regret-RAM models, this RRM_2 model outperforms both the standard RUM model and the standard RRM model in all the seven datasets. The consistency of the evidence from the three forms of relative regret that have

Table 5.18: Summary and Comparison of Log-likelihood Statistics of RRM_2 (MNL) with RUM (MNL) Model

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
	Log-likelihood at convergence						
Modified RRM_2	– 3429.78	– 1833.75	– 1676.36	– 2652.14	– 846.63	– 3018.87	– 1626.95
Standard RUM	– 3434.58	– 1862.23	– 1694.93	– 2670.14	– 847.75	– 3031.58	– 1631.79
Standard RRM	– 3439.32	– 1951.54	– 1691.55	– 2683.70	– 847.55	– 3044.11	– 1639.22
Reject standard RUM in favour of modified RRM_2?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reject standard RRM in favour of modified RRM_2?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

been studied indicates that respondents are concerned not just with the absolute level of regret, but rather with the relative magnitude of regret that encompasses both the advantages and disadvantages embodied in an alternative.

The result that relative regret seems to matter is reminiscent of what Tversky and Kahneman (1981) discovered when they posed the following problem, in two versions, to two group of respondents:

> *Imagine that you are about to purchase a jacket for ($125)[$15], and a calculator for ($15)[$125]. The calculator salesman informs you that the calculator you wish to buy is on sale for ($10)[$120] at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?* (Tversky and Kahneman, 1981, p. 457)

The first group of respondents received the problem with the numerical values in parentheses ( ), while the second group of respondents received the same problem, but with the numerical values in square brackets [ ]. Tversky and Kahneman (1981) found that the responses to the two versions of the problem were markedly different. In the first version, 68 percent of respondents were willing to travel 20 minutes to save an extra $5 off the purchase price of $15, while in the second version, only 29 percent of the respondents were willing to make the extra trip. It appears that the discount of $5 has a greater impact when the price of the calculator is low than when the price of the calculator is high. Tversky and Kahneman (1981) suggest that respondents are evaluating the potential saving in a more inclusive mental account, which includes the price of the calculator, rather than a more minimal account which focuses only on the incremental savings.

The conclusion drawn by Tversky and Kahneman (1981) may explain why using a relative form of regret in modelling is likely to be preferable to using an absolute form of regret. Using an absolute level of regret is akin to assuming a minimal account which focuses only on the incrementals. On the other hand, if a more inclusive mental account is being used by respondents to evaluate alternatives, then it would be more appropriate to consider both advantages and disadvantages of an alternative using a relative advantage component as a representation of context dependency.

5.7 COMPARISONS OF MODEL FIT

Having examined the NLWLR, RRM and the RAM models in some detail, this section provides a comparison of model fit across these models, with a primary focus on the comparison between the NLWLR and the RAM models, since this comparison has not been dealt with directly in previous discussions. The focus of the discussion is now back to the multinomial choice data of three alternatives in the choice set. Where models are not nested within one another, for example the NLWLR and the RRM models, they are compared using the Ben Akiva and Swait (1986) test, The Ben Akiva and Swait (1986) test has been described in Chapter 3. Results are reported in Table 5.19.

Table 5.19: Ranking of Models based on Model Fit

	NLWLR	RUM	RRM	Piecewise-RAM	Regret-RAM
Aust08	2	4	5	1	3
Aust00	3	4	5	1	2
NZ99	2	5	4	1	3
Aust05	2	4	5	1	3
Aust04a	2	5	4	1	3
Aust04b	2	4	5	1	3
NZ07	2	4	5	1	3

If the comparison includes the piecewise-RAM model, then the piecewise-RAM is the best performing model in all datasets. Between the NLWLR model and the regret-RAM model, the differences in model fit are significant at the five percent level. Excluding the piecewise-RAM model from the comparison, then the best performing model is the NLWLR in all the datasets, except in Aust00, where the regret-RAM model is better performing. In all cases, the NLWLR performs better than the RRM. This result reinforces the conclusion obtained by Chorus and Bierlaire (2013) who find that for the one stated choice dataset that was analysed, the contextual concavity model, which might be seen as the forerunner to the NLWLR model, is the better performing model compared to the RRM model.

5.8 THE REFERENCE POINT REVISION HEURISTIC

This section explores the role of a reference point revision heuristic in the determination of choices. As mentioned in Chapter 2, in the context of stated choice experiments, respondents may apply a reference point revision heuristic to the current choice task whenever a non status quo alternative was chosen in some previous choice task, typically assumed to be the choice task immediately preceding the current one (Hensher and Collins, 2011). Essentially, this hypothesis states that the respondents' utility for an experimentally constructed alternative shifts (upwards) whenever a hypothetical alternative was chosen in the previous choice set. Recall that this version of reference revision may be distinguished from a broader concept of value learning. In the latter, underlying preferences in the form of the taste parameters may initially be poorly formed or unknown to the respondent and are discovered as respondents work through the sequence of choice tasks. For example, McNair *et al.* (2011) show that preferences are sensitive to the sequence and attribute levels shown to respondents.

In reference revision, the taste parameters are assumed to be stable but preferences can be affected by previous choices. Reference revision can also be interpreted from the perspective of the status quo bias (Samuelson and Zeckhauser, 1988). Here, the respondent exhibits an increased willingness to consider hypothetical alternatives which have non-zero toll costs as the preceding choice of a hypothetical alternative acts as a force of habit or inertia on the decision to be made in the current choice task. Econometrically, the specification used to model reference point revision is a more explicit and appropriate way of treating choice set interdependence, compared to using a correlated error variance structure (Hensher and Collins, 2011).

Specifically, a dummy variable (*refrev*) was created that equals one whenever a hypothetical experimental alternative (i.e., Alternative A or Alternative B) was chosen in the previous choice set. This dummy variable was then inserted into the Alternative A/Alternatie B utility functions of the current choice set. *refrev* was set to zero for the first choice set encountered by the respondent. As with the *mcd* variable, the *refrev* variable can be specified as entering the utility function in a linear additive manner, as in Equation (5.25). Again, such a model may be thought of as a mixture of two decision rules. This is a simple model of reference revision used by Hensher and Collins (2011) which does not account for potential endogeneity issues.

$$U_j = \beta_{0,j} + \sum_k \beta_k X_{jk} + \beta_r refrev + \varepsilon_j \tag{5.25}$$

In this simple model of reference revision, the log-likelihood statistics estimated in all datasets show a remarkable improvement over the standard RUM model with the inclusion of just one additional variable, *refrev*. As hypothesised, the parameter associated with the *refrev* variable is positive and statistically significant across all seven datasets, with the z-ratios of the estimated parameter found to be in excess of 10. Table 5.20 reports these results.

Inclusion of the reference revision heuristic into the utility expressions for the hypothetical alternatives introduces a dummy variable indicating the type of alternative (current or hypothetical) chosen in the previous choice scenario. This reference revision variable is linked to the unobserved effects of the previous choice set and potentially induces endogeneity and correlation across choice sets for specific alternatives. Extending the work of Hensher and Collins (2011), error components logit models are next estimated to address the issue. At the same time, these models are allowed to account for the panel nature of the data (recall that each respondent was asked to answer 16 choice sets each). The error components model is written in Equation (5.26).

$$\begin{aligned}
U_{curr} &= \beta_{0,curr} + \sum_k \beta_k X_{curr,k} + \varepsilon_{curr} \\
U_{Route\ A} &= \beta_{0,A} + \sum_k \beta_k X_{A,k} + \beta_r refrev + \theta_{AB} E_{AB} + \varepsilon_A \\
U_{Route\ B} &= \sum_k \beta_k X_{B,k} + \beta_r refrev + \theta_{AB} E_{AB} + \varepsilon_B \\
E_{AB} &\sim i.i.d.N(0,1)
\end{aligned} \tag{5.26}$$

Table 5.20: Estimation Results Embedding Reference Revision Heuristic (MNL Model)

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z– ratios)			
Free flow time (FF) (min)	– 0.0583*** (– 7.89)	– 0.1156*** (– 13.09)	n.s.	– 0.0870*** (– 8.80)	– 0.0358*** (– 8.32)	– 0.0711*** (– 18.10)	– 0.1089*** (– 14.84)
Congestion Time (CT) (min)		– 0.1096*** (– 12.17)			– 0.0348*** (– 7.15)	– 0.0939*** (– 28.60)	– 0.1240*** (– 9.39)
Slowed down time (SDT) (min)	– 0.0770*** (– 9.82)		– 0.0814*** (– 6.54)	– 0.1153*** (– 13.19)			
Stop start time (SST) (min)	– 0.0846*** (– 14.16)		– 0.1745*** (– 9.80)	– 0.1686*** (– 19.87)			
Running Cost (RC) ($)	– 0.3643*** (– 9.62)	n.s.	– 0.2342*** (– 3.74)	– 0.6244*** (– 10.99)	– 0.1155*** (– 6.50)	– 0.3249*** (– 14.75)	– 0.4685*** (– 9.67)
Toll Cost (TC) ($)	– 0.3180*** (– 13.53)	– 0.5341*** (– 19.61)	– 0.8709*** (– 13.99)	– 0.5283*** (– 23.18)	– 0.1643*** (– 9.51)	– 0.3530*** (– 27.34)	– 0.7824*** (– 17.16)
refrev	2.2078*** (28.47)	1.6620*** (16.37)	1.1835*** (11.11)	1.5375*** (17.58)	1.8736*** (10.55)	1.2995*** (15.52)	2.3756*** (20.25)
Alternative Specific Constants							
– *current* alternative	1.6120*** (26.28)	1.2232*** (14.79)	1.4033*** (17.26)	0.7416*** (12.11)	0.6224*** (4.21)	0.8810*** (13.36)	0.9146*** (9.86)
– Alternative *A*	n.s.	0.1508** (2.10)	0.2578*** (3.15)	n.s.	0.2485*** (2.88)	n.s.	n.s.
No. of choice observations	4480	2,352	2,432	4,864	912	3,888	1,840
LL at convergence	– 2987.40	– 1718.97	– 1633.49	– 2511.44	– 789.60	– 2906.72	– 1390.34

***, ** denote significance at the one percent and five percent level respectively.
n.s.: not significant

This specification of the error components logit model assumes a correlation between the hypothetical alternatives (through the unobserved effect) but no correlation between the hypothetical alternative and the current alternative. Results of the estimation are shown in Table 5.21.

With the error components logit specification, the parameter on the *refrev* variable has become insignificant at the five percent level in Aust00, NZ99 and Aust05. Additionally, in Aust08 and Aust04b, this parameter is barely significant at the five percent level. β_r is only significant in Aust05 and NZ07, but even so, these parameters are now estimated at a much lower value of around 0.6 compared to 1.5 to 1.9 in the MNL model. This is a reduction of a factor of about three to four. Likewise, the z-ratios of the β_r estimates are reduced by a factor of about four. Another observation worth nothing is that in the error components model, the z-ratios associated with the estimates of the taste parameters of the attributes are in general much higher than in the MNL model.

The results relating to the lack of significance of the reference revision heuristic might be expected as some of the choice set interdependence embodied in reference revision is now picked up by the correlated error structure, so the reference revision effect identified earlier in is due in large part to what was then un-modelled similarity in preferences. However, in some of the datasets, there is still a significant effect to *refrev* even after accounting for the panel nature of the data and for a possible correlation in error structure. Overall, a comparison ofthe results across the seven datasets suggests that there is a lack of consistency in the role of the reference revision heuristic across different choice contexts.

5.9 HEURISTIC WEIGHTING FUNCTIONS

Recall from Chapter 3 that a suggested alternative to the probabilistic decision process model is to weight each heuristic directly in the utility function. In the models such as the various RAM models that have been estimated so far in this chapter, the implicit assumption is that the context independent RUM component and the context dependent relative advantage component are assigned equal weights in the utility function. In a utility function where a total of M heuristics are embedded and weighted, the weights of each heuristic, denoted by W_m, $m=1,2,\ldots,M$ can be given by means of an exponential function shown in Equation (5.27):

Table 5.21: **Estimation Results Embedding Reference Revision Heuristic (Error Components Logit Model)**

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z– ratios)			
Free flow time (FF) (min)	– 0.0752*** (– 11.06)	– 0.1406*** (– 24.72)	n.s.	– 0.1231*** (– 17.22)	– 0.0383*** (– 16.43)	– 0.0865*** (– 35.59)	– 0.1266*** (– 30.41)
Congestion Time (CT) (min)		– 0.1366*** (– 24.59)			– 0.0418*** (– 12.32)	– 0.1099*** (– 46.49)	– 0.1532*** (– 12.74)
Slowed down time (SDT) (min)	– 0.1121*** (– 16.24)		– 0.0801*** (– 5.83)	– 0.1283*** (– 14.62)			
Stop start time (SST) (min)	– 0.1236*** (– 23.48)		– 0.1782*** (– 9.68)	– 0.1988*** (– 26.09)			
Running Cost (RC) ($)	– 0.4297*** (– 15.68)	n.s.	– 0.2232*** (– 6.38)	– 0.7228*** (– 12.32)	– 0.1218*** (– 9.59)	– 0.3876*** (– 20.24)	– 0.5407*** (– 12.38)
Toll Cost (TC) ($)	– 0.4234*** (– 18.58)	– 0.6362*** (– 37.36)	– 1.0717*** (– 40.37)	– 0.6214*** (– 34.14)	– 0.1861*** (– 14.95)	– 0.4212*** (– 45.14)	– 0.8934*** (– 25.21)
refrev	0.2194** (2.13)	0.2590* (1.69)	0.0319 (0.23)	0.5643*** (5.52)	0.2474 (1.27)	0.1580* (1.75)	0.6270*** (4.32)
θ_{AB}	2.6507*** (15.88)	1.8381*** (10.60)	1.5125*** (12.22)	1.6561*** (18.58)	2.6127*** (3.90)	1.8604*** (19.63)	2.5906*** (10.15)
Alternative Specific Constants							
– *current* alternative	1.4149*** (7.34)	0.7671*** (3.55)	1.1388*** (6.30)	0.5381*** (4.14)	– 1.6271*** (– 3.61)	n.s.	n.s.
– Alternative *A*	n.s.	0.1427** (2.42)	0.2533*** (2.72)	n.s.	0.2595*** (4.17)	0.1165*** (2.63)	n.s.
No. of choice observations	4,480	2,352	2,432	4,864	912	3,888	1,840
LL at convergence	– 2539.60	– 1559.13	– 1486.72	– 2304.48	– 709.68	– 2640.36	– 1209.87

Simulation based on 250 Halton draws
***, **, * denote significance at the one percent, five percent and ten percent levels respectively.
n.s.: not significant.
Models account for panel nature of datasets

$$W_m = \exp(\sum_l \delta_{lm} Z_l) \tag{5.27}$$

Assuming each of the components in the utility function contributes positively to the utility of the alternative, the exponential function is appropriate as a functional representation of the weight since W_m , by construction, will be positive. The parameters δ_{lm} may be unconstrained in sign. The signs of the parameters are informative as the partial derivatives of W_m with respect to each of its l arguments are functions that take the same sign as δ_{lm} . In other words, a positive δ_{lm} means that an increase in the level of socio economic characteristic l of a respondent is associated with an increased reliance on the use of heuristic m in the decision process.

5.9.1 Application of the Multiple Heuristic approach to the RUM and NLWLR Mixture

As an illustration of the multiple heuristic approach, a 'mixture' of the standard context independent decision rule and the NLWLR heuristic is explored in this section. Subsequently, a 'mixture' approach using the standard RUM and the relative advantage component is analysed.

For this model, define the standard context independent and NLWLR specifications as illustrated in Equation (5.28):

$$\begin{aligned} H_1 &= \sum_k \beta_k X_k \text{ and} \\ H_2 &= \sum_k (\beta_k (X_{jks} - \max(X_{jks})))^{\varphi_k} \end{aligned} \tag{5.28}$$

If the weighting function W_1 is specified by $W_1 = \exp(\sum_l \delta_l Z_l)$, then a normalisation of W_2 by $W_2 = \exp(-\sum_l \delta_l Z_l)$ is used, that is, W_2 is the reciprocal of W_1. Such a specification of the weighting function is appealing for at least two reasons. Firstly, the use of the exponential function ensures that W_1 and W_2 are positive, which is to be expected since a higher value obtained from each of the decision rules is positively associated with an increase in utility. Secondly, any change to W_1 is associated with a change in the opposite direction for W_2, which is not a behaviourally implausible assumption. Finally, as a matter of empirics, it was observed that other specifications of W_2, such as $W_2 = 1 - W_1$ led to problems of identification. Therefore, the observable component of the utility functions for each of the three alternatives in the choice set can be written in the form of Equation (5.29):

$$
\begin{aligned}
V_{curr} &= \beta_{0,curr} + W_1 H_1 + W_2 H_2 \\
V_A &= \beta_{0,A} + W_1 H_1 + W_2 H_2 \\
V_B &= W_1 H_1 + W_2 H_2
\end{aligned}
\tag{5.29}
$$

where $W_2 = \dfrac{1}{W_1}$.

By writing V_j in this manner of Equation (5.29), the relative weight of H_2 to H_1 may be given by $RW = \dfrac{W_2}{W_1 + W_2}$ in all three utility expressions. Additionally, using the interpretation offered by Tversky and Simonson (1993), RW may be considered to be an indication of the strength of the context dependent component relative to the context independent RUM component.

By way of illustration, the heuristic weighting function W_1 is conditioned on the *age* and *gender* characteristics of the respondents, such that Equation (5.30) is obtained:

$$
W_1 = \exp(\delta_o + \delta_1 * gender + \delta_2 * age) \tag{5.30}
$$

In principle, other socio-economic characteristics or choice task characteristics such as the choice task number, or the length of time taken to answer a particular choice question, might be used as conditioning variables as well. It might also be appropriate to consider using income as a conditioning variable; however, as the income variable was not captured for some respondents in Aust04a, Aust04b and NZ07, the use of this variable to condition the heuristic weighting function is not considered further.

To simplify the estimation of the RUM/NLWLR mixture model, φ_k is set equal to one if, in the simple NLWLR version of the model, the null hypothesis $\varphi_k = 1$ is not rejected at the five percent level. For each attribute *k*, the taste parameters β_k are assumed to be equal across the RUM and the NLWLR decision rules. Table 5.22 reports the results of the estimation for selected datasets.

As the NLWLR heuristic is not found to be supported in Aust00, the multiple heuristics approach combining the RUM and the NLWLR is not applied to this dataset. Another point to note is that unlike the other datasets, there appears to be some issue with identification and so the utility for the current alternative in Aust05 is written as:

$$V_{curr} = \beta_{0,curr} + H_1 + \frac{1}{W_1} H_2$$

However, V_A and V_B in Aust05 are unchanged from Equation (5.29).

The parameter estimates $\hat{\beta}_k$ and $\hat{\varphi}_k$ have already been discussed extensively, and these will not be discussed here, except perhaps to note that the magnitude of $\hat{\varphi}_{FF}$ has increased substantially from the simple NLWLR model in Aust08 and Aust04a. The signs and the statistical significance of the parameters conform to expectations. Using the likelihood ratio test, the multiple heuristics model outperforms either the single RUM or the single NLWLR heuristic in all datasets considered, indicating that this multiple heuristic approach of using a heuristic weighting function has its merits. Focussing on the parameters of interest in this section, which are the $\hat{\delta}_l$, in all datasets, at least some, if not all, of the $\hat{\delta}_l$ are statistically significant. Moreover, the socio-economic characteristics of the respondent do not all enter the heuristic weighting function in exactly the same way. Some variability is observed. For example, $\hat{\delta}_{age}$ is negative in Aust04a, Aust04b and NZ07, and statistically indistinguishable from zero in Aust08 and Aust05. This implies that an older respondent in Aust04a, Aust04b and NZ07 is more reliant on the NLWLR heuristic than a younger person, but in Aust08 and Aust05, no such conclusion can be drawn. Similarly, $\hat{\delta}_{gender}$ can take on a range of values, be

Table 5.22: Empirical Estimation of a RUM and NLWLR Mixture (MNL Model)

	Aust08	Aust05	Aust04a	Aust04b	NZ07
	$\hat{\beta}$ (z-ratio)				
Free flow time (FF) (min)	– 0.0426*** (– 17.60)	– 0.0316*** (– 4.67)	– 0.0199*** (– 13.33)	– 0.0256*** (– 12.16)	– 0.0388*** (– 13.06)
Congestion time (CT) (min)			– 0.0131*** (– 7.14)	– 0.0389*** (– 25.37)	– 0.0552*** (– 11.09)
Slowed down time (SDT) (min)	– 0.0349*** (– 9.74)	– 0.0552*** (– 10.84)			
Stop start time (SST) (min)	– 0.0388*** (– 13.41)	– 0.0747*** (– 18.29)			
Running Cost (RC) ($)	– 0.1105*** (– 4.49)	– 0.2300*** (– 5.97)	– 0.0439*** (– 6.37)	– 0.1144*** (– 9.70)	– 0.1871*** (– 10.97)
Toll Cost (TC) ($)	– 0.1080*** (– 6.16)	– 0.2654*** (– 14.92)	– 0.0494*** (– 4.90)	– 0.1525*** (– 20.30)	– 0.2436*** (– 12.91)
ASC -*current* alternative	0.9042*** (17.93)	0.7387*** (6.56)	– 0.6392*** (– 5.70)	n.s.	n.s.
-Alternative *A*	n.s	n.s.	0.2099** (2.18)	n.s.	n.s.
	$\hat{\varphi}$ (z-ratio) [95% confidence interval]				
$\hat{\varphi}_{FF}$	3.2167*** (4.27) [1.740 – 4.693]	0.6078*** (4.16) [0.321 – 0.895]	3.078*** (3.96) [1.556 – 4.600]	0.5896*** (5.73) [0.388 – 0.791]	$\varphi_{FF} = 1$
$\hat{\varphi}_{CT}$			$\varphi_{CT} = 1$	1.2932*** (9.37) [1.023 – 1.564]	1.3656*** (6.17) [0.931 – 1.800]
$\hat{\varphi}_{SDT}$	$\varphi_{SDT} = 1$	0.7281*** (6.68) [0.514 – 0.942]			
$\hat{\varphi}_{SST}$	$\varphi_{SST} = 1$	$\varphi_{SST} = 1$			
$\hat{\varphi}_{RC}$	0.5261*** (5.35) [0.333 – 0.719]	0.5743*** (6.35) [0.397 – 0.751]	$\varphi_{RC} = 1$	0.5898*** (4.93) [0.356 – 0.824]	$\varphi_{RC} = 1$
$\hat{\varphi}_{TC}$	0.6985*** (4.59) [0.400 – 0.997]	0.3421*** (4.41) [0.190 – 0.494]	0.6437*** (4.12) [0.337 – 0.950]	0.7009*** (5.75) [0.462 – 0.940]	0.6455*** (6.86) [0.461 – 0.830]
	$\hat{\delta}$ (z-ratio)				
constant	n.s.	– 0.1428*** (– 2.66)	2.0654*** (3.96)	2.1280*** (10.48)	2.7409*** (9.17)
age	n.s.	n.s.	– 0.0709*** (– 7.48)	– 0.0422*** (– 8.71)	– 0.0526*** (– 8.49)
gender	– 0.4109*** (– 2.80)	– 0.1308*** (– 3.08)	0.8282*** (3.59)	n.s.	– 0.7719*** (– 4.53)
Number of obs	4,480	4,864	912	3,888	1840
LL at convergence	– 3412.90	–2598.12	– 820.61	– 3003.68	– 1609.61
Normalised AIC	1.528	1.073	1.824	1.550	1.759

*** denotes significance at the one percent level.
n.s.: not significant

they positive (Aust04a), zero (Aust04b) or negative (Aust08, Aust05 and NZ07)[9].

The estimated coefficients in the heuristic weighting function may be used to construct W_1 and W_2. Differences in the socio-economic characteristics across respondents will provide some heterogeneity in the values of W_1 and W_2 within the same dataset. Summary statistics for W_1, W_2 and RW are provided in Table 5.23 and the associated kernel density plots are shown in Figure 5.5. For Aust05, RW as defined is applicable only to the hypothetical alternatives, since the weights in the utility function for the current alternative are different to those in the hypothetical alternatives.

Table 5.23: Summary Statistics of W_1, W_2 and RW

	Aust08	Aust05	Aust04a	Aust04b	NZ07
W_1					
Mean	0.806	0.797	1.078	1.605	1.238
Std Deviation	0.167	0.050	0.780	0.706	0.974
Minimum	0.663	0.761	0.112	0.436	0.181
Maximum	1.0	0.867	3.29	3.047	4.392
W_2					
Mean	1.292	1.260	1.83	0.778	1.415
Std Deviation	0.251	0.076	1.91	0.391	1.092
Minimum	1.0	1.153	0.303	0.328	0.228
Maximum	1.508	1.315	8.91	2.291	5.524
$RW = \frac{W_2}{W_1 + W_2}$					
Mean	0.612	0.571	0.559	0.350	0.527
Std Deviation	0.096	0.032	0.283	0.197	0.283
Minimum	0.5	0.536	0.084	0.097	0.049
Maximum	0.695	0.633	0.988	0.840	0.968

[9] Recall that the dummy variable gender takes the value of one if the respondent is male.

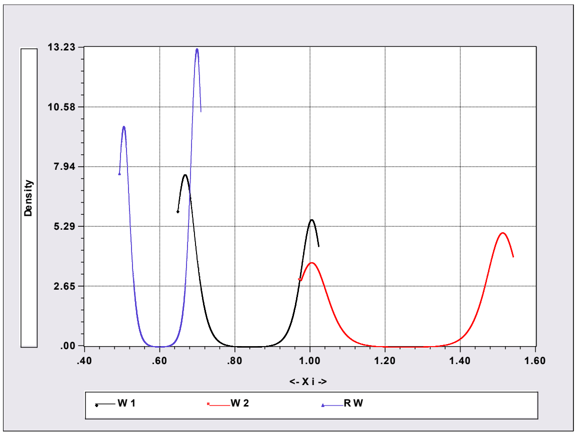


Figure 5.5a: Kernel Density Plots from Aust08

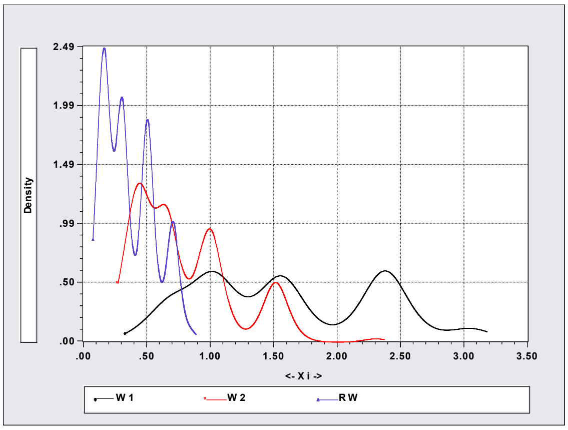


Figure 5.5d: Kernel Density Plots from Aust04b

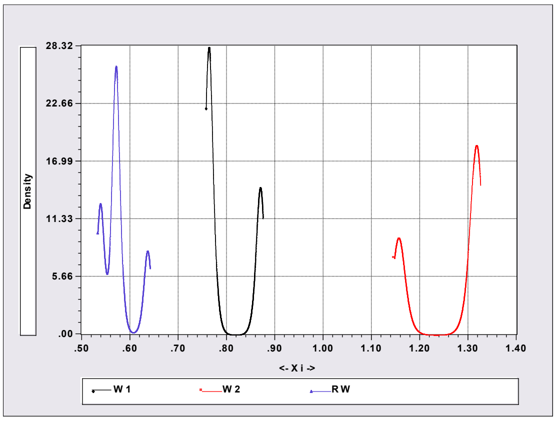


Figure 5.5b: Kernel Density Plots from Aust05

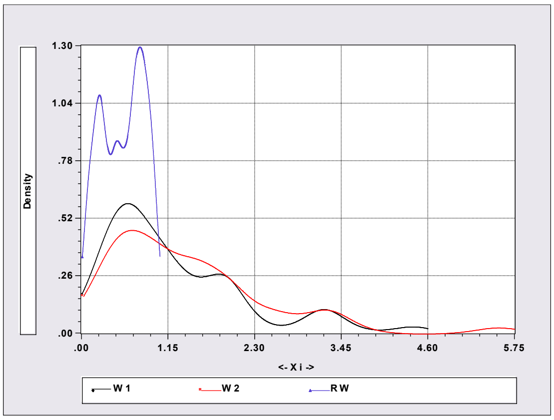


Figure 5.5e: Kernel Density Plots from NZ07

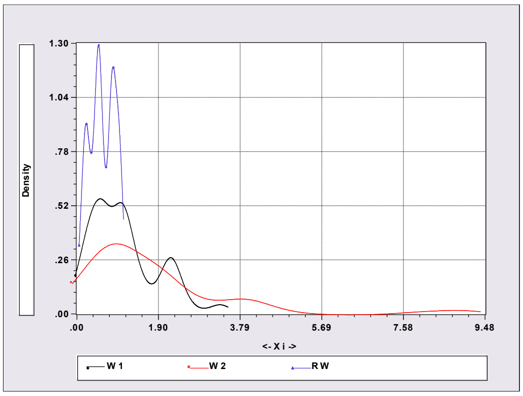


Figure 5.5c: Kernel Density Plots from Aust04a

Figure 5.5: Kernel Density Plots of W_1, W_2 and RW from Selected Datasets

As discussed, the distributions of W_1, W_2 and RW provide an indication of the strength of the context independent effect against the context dependent NLWLR heuristic. In a mixture model where all $\delta_l = 0$, $W_1 = W_2 = 1$ and $RW = 0.5$. Hence a RW value greater than 0.5 might be taken to indicate that the context dependent heuristic is stronger than the context independent RUM decision rule. In all datasets except Aust04b, the average value of RW exceeds 0.5. The NLWLR heuristic can therefore be considered to feature, on average, quite strongly in decision making in such datasets. In datasets where the $\hat{\delta}_{age}$ parameter was found to be significant, the variability in RW is high, due to the variability in the age variable in the dataset. The kernel density plots show a bimodal distribution for RW in Aust08 and Aust05, which should be the case since the only variable that enters into the heuristic weighting function is a dummy variable for gender.

5.9.2 Application of the Multiple Heuristic Approach to the RAM Model

The multiple heuristic approach may also be applied to the piecewise RAM model. In this case, let

$$H_1 = \sum_k \beta_k X_k \text{ and}$$

$$H_2 = \sum_{j' \in s} RA(j, j')$$

The heuristic weighting functions are specified in the same way as in the RUM and NLWLR mixture. The results obtained from the multiple heuristic estimation of the piecewise RAM model are presented in Table 5.24. The gender variable is not available in NZ99 and gender is therefore not included in the W_1 specification for NZ99. In the RUM/piecewise RAM mixture, the utility functions of all three alternatives in Aust05 may now follow Equation (5.29).

Other than Aust08, the multiple heuristic approach to the piecewise RAM model leads to an improvement in model fit in all the other datasets, compared to the equal weights piecewise RAM model. Estimation of the multiple heuristics RUM/piecewise RAM model for Aust08 was somewhat problematic and the final model was estimated assuming that the FF attribute did not appear in the $RA(j, j')$ component, that is, respondents were assumed to disregard the advantages and disadvantages in the FF attribute. This resulted in a poorer model fit. Putting aside Aust08, there again appears to be no consistent pattern as to how the socio economic characteristics of the respondents enter into the heuristic weighting function, as was observed

Table 5.24: Estimation Results from a Multiple Heuristics Approach to the Piecewise RAM (MNL) Model

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio)			
Free flow time (FF) (min)	– 0.0242*** (– 3.97)	– 0.0244*** (– 4.90)	n.s.	– 0.0410*** (– 6.97)	– 0.0070*** (– 4.77)	– 0.0264*** (– 11.13)	– 0.0284*** (– 7.46)
Congestion time (CT) (min)		– 0.0322*** (– 5.35)			– 0.0075*** (– 4.76)	– 0.0342*** (– 12.89)	– 0.0367*** (– 6.68)
Slowed down time (SDT) (min)	– 0.0120*** (– 4.91)		– 0.0544*** (– 4.55)	– 0.0488*** (– 7.23)			
Stop start time (SST) (min)	– 0.0159*** (– 5.64)		– 0.1488*** (– 8.71)	– 0.0614*** (– 7.49)			
Running Cost (RC) ($)	– 0.0811*** (– 5.31)	– 0.8304*** (– 5.53)	– 0.3568*** (– 7.47)	– 0.2376*** (– 7.00)	– 0.0278*** (– 4.66)	– 0.1199*** (– 9.99)	– 0.1457*** (– 6.99)
Toll Cost (TC) ($)	– 0.0595*** (– 5.52)	– 0.1879*** (– 7.05)	– 0.1528*** (– 3.30)	– 0.2645*** (– 8.74)	– 0.0421*** (– 5.40)	– 0.1500*** (– 13.12)	– 0.2002*** (– 8.54)
Alternative Specific Constants							
-*current* alternative	0.8150*** (15.73)	0.8428*** (9.81)	0.8497*** (10.62)	n.s.	– 0.7692*** (– 6.94)	n.s.	n.s.
-Alternative *A*	n.s.	0.1870*** (2.58)	0.2435*** (2.80)	n.s.	0.2213** (2.51)	n.s.	n.s.
				$\hat{\delta}$ (z-ratio)			
constant	1.2048*** (4.69)	1.0091*** (3.44)	– 0.5658*** (– 4.90)	– 0.6765*** (– 11.30)	1.5297*** (2.78)	– 1.3980*** (– 6.44)	– 2.1374*** (– 8.15)
age	n.s.	– 0.0208*** (– 4.26)	0.0094** (3.31)	n.s.	– 0.0457*** (– 4.88)	0.0286*** (4.88)	0.0463*** (7.14)
gender	– 1.093*** (– 3.80)	– 0.6075*** (– 3.67)	Not available	0.2097*** (2.72)	0.8178*** (3.36)	n.s.	0.6955*** (3.67)
Number of obs	4,480	2,352	2,432	4,864	912	3,888	1,840
LL at convergence	– 3385.13	– 1766.07	– 1528.83	– 2562.28	– 818.00	– 2984.15	– 1591.51

***, ** denotes significance at the one percent and five percent level respectively.
n.s.: not significant

with the multiple heuristic approach to the RUM and NLWLR mixture. As before, the parameters $\hat{\delta}_{age}$ and $\hat{\delta}_{gender}$ may or may not be statistically significant, and even when statistically significant, may be positive or negative. For example, in Aust00 and Aust04a, an older respondent is associated with the tendency to use less of the RUM decision rule and more of the relative advantage rule, while the opposite conclusion is obtained in Aust04b and NZ07. The signs of $\hat{\delta}_{age}$ and $\hat{\delta}_{gender}$ may even vary from the RUM/NLWLR mixture to the RUM/ piecewise RAM mixture. For example, in Aust04b, $\hat{\delta}_{age}$ is negative in the RUM/NLWLR mixture but positive in the RAM mixture. Likewise in NZ07, $\hat{\delta}_{age}$ and $\hat{\delta}_{gender}$ are both negative in the RUM/NLWLR mixture model but become positive in the RAM model. The multiple heuristics approach may also be estimated on the regret form of the RAM model. The results are qualitatively similar to what has already been discussed.

5.10 MIXED LOGIT ESTIMATIONS

5.10.1 Mixed Logit Estimation of the NLWLR Model

Besides the fixed (non-random) parameters model, the NLWLR model may be estimated using a mixed logit model which is an innovation by itself. In particular, the fixed parameters NLWLR model imposes the assumption of homogenous power parameters across all respondents. It would be of empirical interest to test if there is heterogeneity in the power parameters, since the magnitude of the power parameters gives an indication of whether mixtures of extremeness seeking ($\varphi_k > 1$), extremeness aversive ($\varphi_k < 1$) or extremeness neutral ($\varphi_k = 1$) behaviour is being exhibited. The result would be especially interesting in cases where the null hypothesis of $\hat{\varphi}_k = 1$ could not be rejected, since heterogeneity around the value of one may now be revealed, that is, there may be heterogeneity in the extremeness seeking/extremeness aversive behaviour in the respondent sample. Hence, in the mixed logit NLWLR model, the power parameters are now assumed to be random.

Since the utility function for an alternative is expressed as a function of the gains in part-utility relative to the worst level in the choice set, it is logical for this gain to be an increasing function. Hence it would make sense to constrain the power parameters to be non-negative, which can be achieved through the commonly used constrained triangular distribution, where the mean of the distribution is set to be the same as the spread of the distribution. The estimation accounts for the panel nature of the dataset. Parameters are assumed to be uncorrelated. Estimation results are presented in Table 5.25 for selected datasets. 250 Halton draws are used to estimate the mixed logit models in this section. Consideration was made as to whether a higher number of draws was necessary; however, simulations involving 500 or 1000 Halton draws did not substantially change the estimation results for one dataset (Aust04a) and the decision was therefore made to use 250 draws for estimation.

Overall, it might be observed that the lack of a consistent pattern across datasets as to the magnitude of the φ_k parameters highlights the point that the same NLWLR heuristic rule may lead to different behavioural implications of extremeness aversion depending on the context. For example, looking at just the FF attribute, the mean of $\hat{\varphi}_{FF}$ is greater than one in Aust08 and less than one in Aust00 and NZ07. There are also differences arising from the mixed logit model of the NLWLR heuristic compared to the MNL model. In Aust04a and NZ07, the TC parameter is much more strongly negative. In terms of the power parameters, estimation results from Aust04a and Aust04b show that the null hypothesis of the mean of $\hat{\varphi}_{FF}$ equal to the value of one is not rejected at the five percent significance level. On the other hand, in NZ07, the null hypotheses that the mean of $\hat{\varphi}_{FF}$ and $\hat{\varphi}_{RC}$ are equal to one can be rejected. Opposite conclusions were obtained in the MNL NLWLR model.

For Aust08, all the power parameters associated with the time components have mean values greater than one, indicating an extremeness seeking behaviour in the time components of the alternative. This is another point of difference from the MNL version of the NLWLR model, where only $\hat{\varphi}_{FF}$ was found to be greater than one. As for $\hat{\varphi}_{TC}$, in the MNL form of the NLWLR, this was estimated at less than one, and in the mixed logit model, although its mean is no longer statistically different from one, the spread of the parameter indicates a certain amount of heterogeneity with a fairly significant probability that $\hat{\varphi}_{TC}$ is indeed below one.

The calculated probability that all φ_k are less than one is also of interest since this corresponds to the conventional understanding of extremeness aversion (compromise effect). For a constrained symmetric triangular distribution bounded below at zero, centred at c and

Table 5.25: Mixed Logit Estimation of the NLWLR Model

		Aust08	Aust05	Aust04a	Aust04b	NZ07
Non random parameters		$\hat{\beta}$ (z-ratio)				
Free flow time (FF) (min)		– 0.0642*** (– 19.58)	– 0.0989*** (– 17.29)	– 0.0421*** (– 15.95)	– 0.1057*** (– 23.51)	– 0.1701*** (– 13.03)
Congestion time (CT) (min)				– 0.0728*** (– 10.52)	– 0.1078*** (– 28.27)	– 0.1257*** (– 12.91)
Slowed down time (SDT) (min)		– 0.0755*** (– 19.16)	– 0.1250*** (– 15.56)			
Stop start time (SST) (min)		– 0.0960*** (– 39.68)	– 0.1853*** (– 24.42)			
Running Cost (RC) ($)		– 0.2802*** (– 5.98)	– 0.7829*** (– 13.96)	– 0.2258*** (– 8.40)	– 0.5772*** (– 15.32)	– 0.8496*** (– 14.66)
Toll Cost (TC) ($)		– 0.1858*** (– 6.40)	– 1.7505*** (– 7.83)	– 2.7591*** (– 2.71)	– 0.6451*** (– 17.16)	– 2.1580*** (– 7.61)
ASC -*current* alternative		1.06550*** (34.53)	n.s.	– 1.2815*** (– 15.67)	n.s.	– 0.7852*** (– 17.32)
-Alternative *A*		n.s.	n.s.	– 0.1624** (– 2.32)	n.s.	n.s.
Random Parameters		$\hat{\varphi}$ (z-ratio) [95% confidence interval]				
$\hat{\varphi}_{FF}$	Mean = Spread	1.8107*** (9.70) [1.445 – 2.177]	0.7731*** (8.14) [0.587 – 0.959]	1.1654*** (7.93) [0.878 – 1.453]	1.0636*** (16.25) [0.935 – 1.192]	0.7402*** (16.90) [0.654 – 0.826]
$\hat{\varphi}_{CT}$	Mean = Spread			0.7623*** (9.05) [0.597 – 0.928]	1.0730*** (25.95) [0.992 – 1.154]	1.1543*** (8.25) [0.880 – 1.428]
$\hat{\varphi}_{SDT}$	Mean = Spread	1.2888*** (9.93) [1.034 – 1.543]	0.9384*** (9.88) [0.752 – 1.124]			
$\hat{\varphi}_{SST}$	Mean = Spread	1.250*** (17.45) [1.109 – 1.390]	1.1123*** (19.53) [1.001 – 1.224]			
$\hat{\varphi}_{RC}$	Mean = Spread	0.5592*** (5.33) [0.354 – 0.765]	1.2065*** (9.95) [0.969 – 1.444]	0.6889*** (4.04) [0.355 – 1.022]	0.8429*** (13.61) [0.721 – 0.964]	0.7875*** (8.47) [0.605 – 0.970]
$\hat{\varphi}_{TC}$	Mean = Spread	0.7912*** (3.47) [0.344 – 1.238]	0.6711*** (14.93) [0.583 – 0.759]	0.3902*** (7.09) [0.282 – 0.498]	0.8111*** (21.91) [0.739 – 0.884]	0.8355*** (12.94) [0.709 – 0.962]
Number of obs		4,480	4,864	912	3,888	1,840
LL at convergence		– 3359.72	– 2406.46	– 800.80	– 2772.14	– 1361.69

Note: Simulations based on 250 Halton draws
*** denotes significance at the one percent percent level respectively.
n.s.: not significant
Estimations account for panel nature of datasets

bounded above at *2c*, the cumulative distribution function (CDF) is given by Equation (5.31):

$$CDF = \begin{cases} 0 \text{ for } x \le 0 \\ \dfrac{x^2}{2c^2} \text{ for } 0 \le x \le c \\ 1 - \dfrac{(2c-x)^2}{2c^2} \text{ for } c \le x \le 2c \\ 1 \text{ for } 2c < x \end{cases} \tag{5.31}$$

Given the estimated parameter c of the triangular distribution, the CDF can be used to calculate the probability that the power parameter $\hat{\varphi}_k$ is less than one. This is done by setting $x = 1$. Then, the CDF is a function of the parameter c alone, and it can be simulated by the delta method. Results are presented in Table 5.26.

Table 5.26: Simulations of the CDFs for $\Pr(\varphi_k < 1)$

	Mean value of $\Pr(\varphi_k < 1)$ (standard deviation in parentheses)				
	Aust08	Aust05	Aust04a	Aust04b	NZ07
$\Pr(\varphi_{FF} < 1)$	0.153 (0.031)	0.750 (0.112)	0.368 (0.093)	0.442 (0.054)	0.789 (0.052)
$\Pr(\varphi_{CT} < 1)$			0.763 (0.100)	0.434 (0.033)	0.375 (0.091)
$\Pr(\varphi_{SDT} < 1)$	0.301 (0.061)	0.564 (0.101)			
$\Pr(\varphi_{SST} < 1)$	0.320 (0.037)	0.404 (0.041)			
$\Pr(\varphi_{RC} < 1)$	0.978 (0.071)	0.343 (0.069)	0.850 (0.197)	0.669 (0.071)	0.733 (0.109)
$\Pr(\varphi_{TC} < 1)$	0.729 (0.268)	0.870 (0.051)	1.0	0.706 (0.043)	0.678 (0.074)
$\prod_k \Pr(\varphi_k < 1)$	0.010 (0.005)	0.051 (0.017)	0.239 (0.076)	0.091 (0.018)	0.147 (0.046)

Indications from Table 5.26 are that the incidence of full extremeness aversion, that is, the extremeness aversion across all attributes, is relatively small in the respondent sample. This is particularly true for Aust08 where the estimated means of the power parameters for the time components FF, SDT and SST are all statistically greater than one. Looking at the individual marginal probabilities, it can be generally said that for the cost components, the probability that their associated power parameters is less than one is a lot higher than their counterparts for the time components. This suggests that respondents tend to be extremeness averse in costs rather than in time. Perhaps this is a reflection that losses in monetary terms are a lot more salient than losses in time.

5.10.2 Mixed Logit Estimation of the Regret-RAM Model

The regret form of the RAM model can also be estimated using the mixed logit model. There is no particular theory to determine whether a parameter is random or not; however, there is reason to suspect heterogeneity in the congestion time (CT) parameter, or in the slowed down time (SDT) and stop-start time (SST) parameters where available, and in the toll cost (TC) parameters. Heterogeneity in the CT/SDT/SST parameters is possible since these could be related to travel time variability which has not been explicitly modelled. The TC parameter may also be considered random since the aim of the toll road studies was to observe choice behaviour under various attribute combinations of tolled/non-tolled routes and also given that in many instances, no prior experience with tolling was reported. In particular, the mixing distribution for these parameters is assumed to be constrained triangular, as a way of imposing consistency with the behavioural implication that increasing levels of the CT and TC attributes provide disutility rather than utility. Other parameters of the model remain non-random. The model accounts for the panel nature of dataset. Estimation results are presented in Table 5.27.

In the mixed logit estimation, all parameters continue to possess the appropriate sign, as expected. More so than the travel time parameters, there appears to be a substantial amount of heterogeneity in the TC parameter, as can be observed from the large spread of the distribution in some of the datasets.

5.11 CALCULATIONS OF VALUE OF TRAVEL TIME SAVINGS (VTTS)

5.11.1 Some Preliminaries

The marginal willingness-to-pay measure is a typical parameter of interest derived from choice models. This section explores some of the implications on the values of travel time savings (VTTS) of embedding the previously identified heuristics into the model. The focus

Table 5.27: Estimation Results of the Mixed Logit Regret RAM Model

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
				$\hat{\beta}$ (z-ratio)			
Non random parameters							
Free flow time (FF) (min)	– 0.0578*** (– 13.37)	– 0.1237*** (– 11.75)	n.s.	– 0.0862*** (– 20.42)	– 0.0310*** (– 17.49)	– 0.0618*** (– 34.45)	– 0.1018*** (– 31.61)
Running Cost (RC) ($)	– 0.3103*** (– 15.01)	n.s.	– 0.2294*** (– 8.14)	– 0.4705*** (– 13.44)	– 0.1048*** (– 10.37)	– 0.2940*** (– 19.38)	– 0.5111*** (– 17.68)
ASC							
-*current* alternative	0.1091*** (4.08)	0.3585*** (6.16)	0.3760*** (5.75)	n.s.	– 0.7835*** (– 11.56)	0.1482*** (4.50)	– 0.6581*** (– 14.12)
-Alternative *A*	n.s.	0.2141*** (3.28)	0.3055*** (3.23)	n.s.	0.2714*** (4.22)	0.1228*** (2.65)	n.s.
Random parameters							
Congestion time (CT) (min) [constrained triangular] |Mean| = Spread		– 0.1980*** (– 17.28)			– 0.0290*** (– 13.17)	– 0.0785*** (– 31.81)	– 0.0769*** (– 12.78)
Slowed down time (SDT) (min) [constrained triangular] |Mean| = Spread	– 0.0611*** (– 17.00)		– 0.0811*** (– 6.38)	– 0.1016*** (– 15.98)			
Stop-start time (SST) (min) [constrained triangular] |Mean| = Spread	– 0.0595*** (– 33.26)		– 0.1321*** (– 10.25)	– 0.1489*** (– 26.98)			
Toll Cost (TC) ($) [constrained triangular] |Mean| = Spread	– 1.1377*** (– 24.54)	– 1.1137*** (– 26.92)	– 3.5492*** (– 15.77)	– 0.7663*** (– 33.79)	– 0.2651*** (– 14.48)	– 0.3893*** (– 31.66)	– 1.3813*** (– 17.04)
Number of observations	4,480	2,352	2,432	4,864	912	3888	1,840
LL at convergence	– 3043.11	–1314.47	–1415.95	– 2468.03	– 799.20	– 2844.45	– 1405.85

Note: Simulations based on 250 Halton draws
***, ** denotes significance at the one percent and five percent level respectively.
n.s.: not significant
Estimations account for panel nature of datasets

of the discussion will be on the models that are related to the extremeness aversion heuristic. The VTTS for these models are more interesting since they depend not only on the alternative being chosen, but also on the attribute levels of other competing alternatives in the choice set.

In general, the value of travel time savings (VTTS) or the marginal willingness to pay for a one unit reduction in travel time is given in Equation (5.32) by:

$$VTTS_j = \frac{\partial V_j / \partial(time)}{\partial V_j / \partial(cost)} \tag{5.32}$$

Equation (5.32) is the ratio of the marginal utility with respect to time to the marginal utility with respect to cost. In general, since there are two cost components modelled in the utility function, $\frac{\partial V_j}{\partial(cost)}$ can be expressed as a weighted average of $\frac{\partial V_j}{\partial RC_j}$ and $\frac{\partial V_j}{\partial TC_j}$ as shown in Equation (5.33):

$$\frac{\partial V_j}{\partial(cost)} = \frac{RC_j}{RC_j + TC_j} \times \frac{\partial V_j}{\partial RC_j} + \frac{TC_j}{RC_j + TC_j} \times \frac{\partial V_j}{\partial TC_j} \tag{5.33}$$

In Aust00, Aust04a, Aust04b and NZ07, where the only time components modelled are FF and CT, the VTTS measure for an alternative *j* (in \$/person-hour) is obtained as a weighted average of $\frac{\partial V_j / \partial(FF_j)}{\partial V_j / \partial(cost)}$ and $\frac{\partial V_j / \partial(CT_j)}{\partial V_j / \partial(cost)}$ as in Equation (5.34)[10]:

$$VTTS_j = 60 \times \left[\frac{FF_j}{FF_j + CT_j} \times \frac{\partial V_j / \partial(FF_j)}{\partial V_j / \partial(cost)} + \frac{CT_j}{FF_j + CT_j} \times \frac{\partial V_j / \partial(CT_j)}{\partial V_j / \partial(cost)} \right] \tag{5.34}$$

[10] A multiplication by 60 is appropriate since the time attributes are presented in minutes.

In Aust08, NZ99 and Aust05, where the FF, SDT and SST time attributes are modelled in the utility function, the VTTS expression is simply an extension of Equation (5.34), i.e., Equation (5.35):

$$WTP_j = 60 \times \left[\frac{FF_j}{TT_j} \times \frac{\partial V_j / \partial (FF_j)}{\partial V_j / \partial (cost)} + \frac{SDT_j}{TT_j} \times \frac{\partial V_j / \partial (SDT_j)}{\partial V_j / \partial (cost)} + \frac{SST_j}{TT_j} \times \frac{\partial V_j / \partial (SST_j)}{\partial V_j / \partial (cost)} \right] \tag{5.35}$$

$$TT_j = FF_j + SDT_j + SST_j$$

For the standard RUM model, $\dfrac{\partial V_j}{\partial X_{jk}} = \beta_k$.

5.11.2 The NLWLR model

With the NLWLR model, the willingness to pay function will be non-linear in the attribute levels. The marginal utility with respect to any single attribute X_{jks} can be written in Equation (5.36) as:

$$\frac{\partial V}{\partial X_{jks}} = \beta_k \varphi_k \left(\beta_k (X_{jks} - \max(X_{jks})) \right)^{\varphi_k - 1} \tag{5.36}$$

If $\varphi_k < 1$, $\dfrac{\partial V}{\partial X_{jks}}$ is not defined when $X_{jks} = \max(X_{jks})$, that is, when the level of attribute *k* is the maximum or worst level it can take in the particular choice set. In this case, there is a division by zero. One may either disregard the calculation of $\dfrac{\partial V}{\partial X_{jks}}$ in such cases or else the function may be evaluated at the sample mean of $\left[X_{jks} - \max(X_{jks}) \right]$, which will typically have a non-zero value. This latter approach is used in the calculations presented in Section 5.11.4.

A similar issue arises for the RC and TC attributes since they appear in the denominator of the VTTS equation. Even where $\hat{\varphi}_{RC} > 1$ and $\hat{\varphi}_{TC} > 1$, a situation in which both the RC and TC attribute levels are simultaneously the maximum values in the choice set leads to the denominator of the VTTS equation taking on the value of zero, since $\frac{\partial V}{\partial RC_{js}} = 0$ and $\frac{\partial V}{\partial TC_{js}} = 0$. Evaluating the function at the sample means also avoids this complication.

5.11.3 The RRM Model and the Hybrid RRM-RUM Model

Recall that in the RRM model, $V_j^{RRM} = -\sum_{\substack{j' \in s \\ j' \neq j}} \sum_k \ln(1+\exp(\beta_k (X_{j'k} - X_{jk})))$.

Where an attribute is processed according to random regret minimisation, the marginal utility $\frac{\partial V_j}{\partial X_{jk}}$ with respect to X_{jk} will be specific to the alternative and also specific to the choice set as the attribute value X_{jk} of the alternative and its counterpart values in all other competitor alternatives enter into this expression. Therefore, Equation (5.37) follows:

$$\frac{\partial V_j}{\partial X_{jk}} = \sum_{\substack{j' \in s \\ j' \neq j}} \frac{\beta_k \exp[\beta_k (X_{j'k} - X_{jk})]}{1+\exp[\beta_k (X_{j'k} - X_{jk})]} \tag{5.37}$$

By appropriate substitution of $\frac{\partial V_j}{\partial X_{jk}}$ in Equation (5.37) into Equations (5.33) to (5.35), the VTTS expressions in the case of the full RRM model follow analogously.

In the hybrid RRM-RUM model, the expressions for the partial derivatives of V_j with respect to X_{jk} will follow either Equation (5.37) if the attribute is processed according to regret minimisation or will simply be β_k if the attribute is processed according to linear additive utility maximisation. Again, by appropriate substitution, the VTTS expressions can be derived from Equations (5.33) to (5.35).

5.11.4 The RAM Model

Piecewise RAM Model

Recall the observed component of utility in the RAM model, as shown in Equation (5.38):

$$V_j^{RAM} = \beta_{0,j} + \sum_k \beta_k X_{jk} + \sum_{\substack{j' \in s, \\ j' \neq j}} RA(j, j') \tag{5.38}$$

As a result of the advantage function $A(j, j') = \sum_k A_k(j, j')$ and disadvantage function $D(j, j') = \sum_k D_k(j, j')$ appearing in $RA(j, j')$, the partial derivative of $RA(j, j')$ with respect to X_{jk} will be a function of all attributes of all alternatives, and not just a function of attribute *k* alone. In the piecewise RAM model, $\frac{\partial RA(j, j')}{\partial X_{jk}}$ can be distinguished according to three cases.

Case One

If alternative *j* confers a non-zero disadvantage over alternative j' in attribute *k*, $A_k(j, j') = 0$ and $A(j, j')$ is independent of X_{jk}. $D(j, j')$ remains a function of X_{jk}, and Equation (5.39) follows:

$$\frac{\partial RA(j, j')}{\partial X_{jk}} = \frac{A(j, j') * \beta_k}{[A(j, j') + D(j, j')]^2} \tag{5.39}$$

Case Two

If alternative *j* confers a non-zero advantage over alternative j' in attribute *k*, $D_k(j, j') = 0$ and $D(j, j')$ is independent of X_k. $A(j, j')$ is a function of X_{jk} and therefore Equation (5.40) follows:

$$\frac{\partial RA(j, j')}{\partial X_{jk}} = \frac{D(j, j') * \beta_k}{[A(j, j') + D(j, j')]^2} \tag{5.40}$$

In **Case One** and **Case Two**, the marginal utility of alternative j with respect to attribute k is given in Equation (5.41) by:

$$\frac{\partial V_j}{\partial X_{jk}} = \beta_k + \sum_{\substack{j' \in s \\ j' \neq j}} \frac{\partial RA(j, j')}{\partial X_{jk}} \tag{5.41}$$

With a negative β_k, as is expected for the travel time and travel cost attributes, then $\frac{\partial V_j}{\partial X_{jk}}$ would be negative as well.

Case Three

If $A_k(j, j') = D_k(j, j') = 0$, for example, when the attribute values X_{jk} and $X_{j'k}$ are equal, $\frac{\partial R(j, j')}{\partial X_k}$ is not defined and in such cases, $\frac{\partial V_j}{\partial X_{jk}}$ is simply equal to β_k.

Regret-RAM Model

In the regret RAM model, Equation (5.42) follows directly from the definitions of $A(j, j')$ and $D(j, j')$.

$$\frac{\partial A(j, j')}{\partial X_{jk}} = \frac{\beta_k}{1 + \exp[-\beta_k (X_{jk} - X_{j'k})]} \quad \text{and} \quad \frac{\partial D(j, j')}{\partial X_{jk}} = \frac{-\beta_k}{1 + \exp[\beta_k (X_{jk} - X_{j'k})]} \tag{5.42}$$

Equation (5.43) follows.

$$\begin{aligned} \frac{\partial V_j^{RAM}}{\partial X_{jk}} &= \beta_k + \sum_{\substack{j' \in S \\ j' \neq j}} \frac{\partial RA(j, j')}{\partial X_{jk}} \\ &= \beta_k + \sum_{\substack{j' \in S \\ j' \neq j}} \frac{D(j, j').\dfrac{\partial A(j, j')}{\partial X_{jk}} - A(j, j').\dfrac{\partial D(j, j')}{\partial X_{jk}}}{\left[A(j, j') + D(j, j')\right]^2} \end{aligned} \tag{5.43}$$

Again, by appropriate substitutions into Equations (5.33) to (5.35), the VTTS for the regret-RAM model may be obtained.

5.11.5 Estimates and Discussion

The summary results of the VTTS calculations for the various models, computed at the sample means, are reported in Table 5.28. Hence, variability in VTTS calculations is due to parameter estimation error. These estimates are intended as a comparison across the model types. For the context dependent models, the VTTS values obtained may be valid only within the limits of the experimental design, since the VTTS equations for the context dependent models depend, to varying degrees, on the attribute values of the other competing alternatives in the choice set, and in a stated choice experiment, these are hypothetical. The real market VTTS values will depend not only on the real market attribute values of competing alternatives, but possibly on the composition of the choice set itself.

The use of context dependent preferences for VTTS calculations is likely to violate the microeconomic axioms that underpin the VTTS obtained from RUM models. In real life, people are probably not as 'well-behaved' as postulated by neoclassical economic theory, but VTTS and consumer surplus measures are well understood from RUM models, even if it almost inevitably means accepting a loss of behavioural realism. On the other hand, behavioural realism may be an important consideration where more accurate forecasts of future demand are needed.

Alternative specific VTTS will certainly be the norm in the models of context dependence (NLWLR, RRM, Hybrid RRM-RUM and RAM models) since attributes of whichever alternative is chosen and those of any relevant competing alternatives enter into the VTTS function. Hence, to bring the VTTS estimates in agreement to the stated preference of respondents, the VTTS calculations are restricted to only those alternatives actually chosen. As can be seen from Equations (5.34) and (5.35), the VTTS calculations are based on some weightings of the incidence of the time and cost components of the alternatives, and so even in the case of the standard RUM model, the VTTS will be alternative specific.

Table 5.28: Summary of VTTS Calculations for the MNL Form of the Standard RUM, NLWLR, RRM, Hybrid RRM-RUM and Regret-RAM Models

	Aust08	Aust00	NZ99	Aust05	Aust04a	Aust04b	NZ07
	Mean VTTS($/person-hour)						
	Standard RUM						
Mean	12.26	13.25	22.39	12.85	16.23	14.51	12.92
Standard Deviation	1.34	0.35	4.62	1.19	2.04	0.69	1.08
95% confidence interval	9.62– 14.89	12.56 – 13.94	13.34– 31.45	10.53 – 15.18	12.23– 20.22	13.16– 15.86	10.80– 15.03
	NLWLR						
Mean	10.89	13.94	17.59	11.08	16.31	14.98	12.57
Standard Deviation	1.37	0.87	2.22	0.98	2.53	0.74	1.08
95% confidence interval	8.20 – 13.57	12.23 – 15.66	13.24 – 21.95	9.15 – 13.01	11.35 – 21.26	13.53 – 16.43	10.46 – 14.69
	RRM						
Mean	13.04	8.67	24.54	11.41	16.34	14.71	13.50
Standard Deviation	1.51	2.88	6.04	1.03	2.10	0.72	1.26
95% confidence interval	10.07 – 16.01	3.03 – 14.32	12.70 – 36.38	9.39 – 13.42	12.21 – 20.46	13.29 – 16.13	11.04 – 15.96
	Hybrid RRM-RUM						
Mean	12.94	11.53	21.68	11.55	15.57	15.63	14.59
Standard Deviation	1.47	0.35	4.89	1.03	2.06	0.67	1.32
95% confidence interval	10.06 – 15.83	10.85 – 12.22	12.09 – 31.27	9.53 – 13.56	11.52 – 19.61	14.32 – 16.94	12.00 – 17.17
	Regret-RAM						
Mean	12.17	13.51	21.51	12.87	16.18	14.52	12.86
Standard Deviation	1.11	0.38	3.11	1.05	1.92	0.64	1.01
95% confidence interval	10.00 – 14.34	12.78 – 14.25	15.42 – 27.60	10.81 – 14.94	12.43 – 19.93	13.27 – 15.77	10.89 – 14.83

Note: VTTS estimates presented in the currency of the country where experiment was conducted. VTTS calculated using the MNL forms of the models. VTTS computed at the sample mean and VTTS variability is due to parameter estimation error.

In the VTTS calculations for the NLWLR model, the non-linear specification for all attributes was retained, even if the null hypothesis of $\varphi_k = 1$ cannot be rejected at the five percent level. For the RAM models, only the VTTS results for the regret-RAM model are reported. Estimates are obtained using the delta method.

The VTTS in NZ99 appears high, but this result can be explained by observing that the FF attribute does not enter the utility expression, and hence, the VTTS in NZ99 may be expected to be higher since it represents the marginal willingness to pay to avoid an extra minute of slowed down or stop start time.

From a statistical perspective, there is little to suggest that the mean VTTS are significantly different across model types for the same dataset, with the exception in Aust00 where the hybrid model produces VTTS estimates which are lower than those in the standard RUM, NLWLR and regret-RAM models. The much lower VTTS estimate from the RRM model in Aust00 might also be treated with more circumspection in light of the significantly poorer model fit of the RRM and the wider 95 percent confidence interval of the estimates. However, in practice, only mean VTTS values are generally used and the differences in means do reveal substantial time benefit differences when applied to specific projects. For example, in Aust05, the NLWLR model predicts a mean VTTS of $11.08 while the regret-RAM predicts a mean VTTS of $12.87.

Another observation to make is that in most datasets, the standard deviation of the VTTS estimates from across the models are comparable to one another, suggesting in this case at least that even with the inclusion of various alternative specific attribute values into the VTTS expression, the precision of the VTTS estimate is not impacted too negatively. In fact, in several datasets, the standard deviation of the VTTS estimates in the regret-RAM model is the lowest across the models.

Since the attribute values of all available alternatives enter the RAM VTTS equation, the RAM VTTS measures will generally change with changes in the attributes of competing alternatives. This is of course fully in line with the notion that the RAM model, together with the RRM, hybrid and NLWLR models, implies choice set-specific preferences. As suggested by Chorus *et al.* (2012), this allows for a richer interpretation of the implied trade-offs that are made as choice set composition is varied.

Allowing preferences to be choice set specific may seemingly imply that models like the NLWLR, RAM, RRM and hybrid RRM-RUM are less suitable for the derivation of VTTS measures. After all, the VTTS measures are now a function of hypothetical attribute values, which are a function of the experiment design. For the policy analyst, the task of deciding an appropriate VTTS estimate appears to be even more challenging than before, since the real market attribute values of competing alternatives in the person's choice set needs to be known. However, a careful assessment of the VTTS equation will reveal that the range of policy options may actually be expanded under the assumption of choice set specific preferences. The policy maker may be able to influence VTTS quite substantially simply by framing and appropriately defining alternatives and choice sets in the public domain. For example, to increase VTTS and perhaps the chance of a transport project being approved, the policy maker could paint an alternative scenario with very poor time attribute values (if nothing is done), so that the transport project, in comparison, will appear to have a very large relative advantage compared with the status quo or no-improvement alternative.

5.12 SUMMARY

Within the confines of the stated route choice unlabelled datasets, some broad conclusions are beginning to emerge. The empirical results highlighted in this chapter suggest that accounting for some form of context dependency into context independent models like the standard RUM model is an important consideration and should not be ignored in future work. In particular, the NLWLR model and the RAM model, even in its restricted form, offer to hold a lot of promise in how context dependency might be modelled. The estimation results of the RAM/RDM model strongly indicates that a relative form of regret should be seriously considered as well. The data analysis also suggests that assuming a purely context dependent model like the RRM model might be overly restrictive. Instead, a combination of context independent decision rules and context dependent heuristics, possibly weighted by some weighing function, appear to hold substantial promise as a better representation of choice behaviour.

More importantly, there is a word of caution as to how context dependence is specified in the models. For example, consistent with previous findings, and now further reinforced by this research, the performance of the RRM model compared to the standard RUM model can be said to be mixed at best. Through the PDP model, the MCD heuristic appears to hold some validity as a simple model of context dependence. On the other hand, the effect of choice set interdependence, by means of the reference revision heuristic, largely disappears once more advanced models are considered.

The next chapter summarises the overall contributions of this thesis and suggests some potential avenues of inquiry for future research.

CHAPTER 6 CONCLUSIONS

6.1 MAIN CONTRIBUTIONS OF THIS THESIS

Using well-established methods from discrete choice modelling and within the context of unlabelled stated route choice experiments as the data setting, this thesis has focused on examining the evidence for the majority of confirming dimensions (MCD), the extremeness aversion and the reference revision heuristics in choice behaviour. On the MCD and the reference revision heuristics, this study might be seen as an extension of the work by Hensher and Collins (2011).

6.1.1 Reference Revision Heuristic

The reference revision heuristic has been modelled assuming that there is a shift in the reference in the subsequent choice scenario when a non-status quo alternative was chosen previously. From the empirical results, the conclusion is that the impact of this particular heuristic on choice decisions is substantial if simpler models such as the MNL are considered, but its impact is substantially diminished once the correlation between alternatives and the panel nature of the dataset are accounted for. The reference revision heuristic becomes statistically negligible in four of the seven datasets, although it still retains statistical significance at the five percent level in the remaining three datasets, albeit with a much smaller estimated coefficient. This result does echo the kind of findings obtained whenever the RRM model is compared against the RUM model and the lack of a consistent pattern of observable behaviour across datasets suggests that this heuristic might somehow depend on the characteristics of the respondent sample. At the very least, the result suggests that if models with simple error structures have to be used, it would be important to control for the impact of reference revision.

6.1.2 Majority of Confirming Dimensions Heuristic

The MCD heuristic can be interpreted as a simple way of modelling context dependency, since the ‘best’ attribute level in any given choice set will depend on the attribute levels of all competing alternatives. It was hypothesised that in the case of pivot designs, the MCD heuristic might apply to either all alternatives in the choice set, or if the effort-accuracy trade-off framework is correct, only to the hypothetical alternatives, since the value of the current or status quo alternative is well known and might be stored in memory. As it turns out, the results are mixed, which is itself a more nuanced understanding of the role of the MCD compared to what has been previously reported. From an empirical point of view, in a handful of datasets, it is better for the MCD heuristic to appear with a generic coefficient in all three alternatives of the choice set, while in other datasets, it is preferable to restrict the MCD heuristic to only the hypothetical alternatives. Still in other datasets, the MCD heuristic appears only in the reference alternative or its effect may be statistically negligible.

Some further analysis was undertaken by means of a probabilistic decision process model to evaluate the role of this heuristic. The MCD heuristic vis-à-vis the standard linear-in-the parameters RUM model may be thought to be latent in the population of respondents and with the latent class structure, each decision rule may be assigned to one particular class. Under these assumptions, the model output indicates a significant probability that the MCD rule (instead of the standard linear additive rule) is being used to make decisions. This probability is as high as 60 per cent in one of the datasets, implying that future work must take into account the high incidence of semi compensatory behaviour in trading amongst attributes.

6.1.3 Models of the Extremeness Aversion Heuristic

On the extremeness aversion heuristic, this thesis has extended the contextual concavity model of Kivetz *et al.* (2004) by allowing the power parameters to be freed from the constraint that they be less than one. The result is a loss of the concavity property for some attributes, at least from a statistical point of view, but the gain is a deeper understanding that respondents may be extremeness seeking in some attributes, especially for the time attributes of the alternatives, while being extremeness averse in others. This is a finding that has not been reported in the literature before. Therefore, instead of calling such a model a "contextual concavity" model, which by implication suggests that all power parameters are less than one, this thesis has suggested a new nomenclature for this model, the "non-linear worst level referencing" (NLWLR) model. Furthermore, since concavity may be rejected for some attributes of the alternatives, the potential for compromise effects as an explanation for observed choice behaviour is very much reduced. Rather, such combinations of extremeness seeking in some attributes and extremeness aversion in other attributes that is evident across most of the datasets studied suggest that a polarisation effect, rather than a compromise effect, may be more prevalent in decision making. As has been noted in Chapter 2, the polarisation effect may be interpreted as an asymmetric form of the extremeness aversion heuristic.

This thesis has also extensively examined the performance of the random regret model (RRM) model in the context of the seven toll road datasets. The roots of the RRM model lie in regret based theories and regret avoidance behaviour. Although the RRM model is not primarily motivated by extremeness aversive behaviour, it turns out that the convexity of the regret function allows the compromise effect to be modelled. Hence, like the non-linear worst level referencing (NLWLR) model, which is an extension of the contextual concavity model and the relative advantage maximisation (RAM) model, the RRM model may also be thought of as another model of extremeness aversion. Overall, a comparison of the RUM and RRM models produces evidence suggesting that model fit criteria do not consistently favour one model over another. This finding is in line with the results reported by Chorus (2012), meaning that the RRM model remains as a serious candidate in empirical studies.

This thesis has also extensively discussed the RAM model. In addition to the original piecewise specification proposed by Kivetz *et al.* (2004), a new and easily estimable form of the RAM model, based on the regret function of the RRM model, has been introduced. This particular re-specification of the RAM model has been called the regret-RAM model. This is an important innovation especially in light of relatively unsuccessful attempts at estimating a RAM model in the past that resulted in poor model fit (Kivetz *et al.*, 2004). The regret-RAM model has several advantages over the piecewise-RAM model. The regret-RAM model is everywhere differentiable and the denominator of its relative advantage component is always positive. Hence, issues of divisibility by zero do not arise. By assuming that advantages and disadvantages are symmetric, both forms of the RAM model can be just as parsimonious as the RUM, RRM and hybrid RRM-RUM models.

Under certain conditions, the piecewise and regret-RAM models can display the compromise effect, but it is also possible for these models to display polarisation or even extremeness seeking effects. Initial evidence across all seven datasets from this study indicates that there is a lot of potential in the RAM model in terms of providing a better fit for the data, in the retrieval of parameter estimates that match with prior expectations from previous research and from theory, and in obtaining more precise model outputs such as willingness to pay measures, at least from the perspective of smaller standard deviations.

From the evidence gathered, two points of observation may be made. Firstly, from a behavioural perspective, the idea of context dependence, in this case, using competing alternatives as a point of reference is an important one. This is evident from the better statistical fit of the RAM model across all datasets compared to the RUM model. Secondly, the specification of context dependence is also crucial. It has been noted that the RAM model is mathematically equivalent to a relative disadvantage minimisation (RDM) model. If respondents' choice of an alternative is a reflection of an attempt to minimise the negative emotions associated with their choice, then in essence, the terms "relative disadvantage" and "regret" can be treated synonymously. Hence, the notion of regret which, in the RRM model, is based on some function of absolute differences of attribute values, may benefit from being recast into relative terms, for example into a relative form of regret or a relative disadvantage. From the comparisons of the RRM model to the RAM/RDM model, the relative form of regret appears to better capture respondent preferences. For future work, it would be possible to study a RRM-model with a RAM-based relative component which results in a new and interesting hybrid model containing both absolute and relative regret.

Another innovation of this thesis is to apply the RAM model, hitherto used in the context of making a choice from three or more alternatives, to binary choice data. As discussed in Sections 5.6.5 and 5.6.6 of Chapter 5, the results are again promising. In all the datasets studied, the fit of the RAM model is just as good, if not better, than the standard RUM model. This seems especially important given that the standard RRM model simplifies to the linear additive RUM model. There is no reason to suspect that context dependency effects are less important in binary choices than in choices involving more than two alternatives. Since the RRM model is not able to meaningfully account for referencing and context dependency in binary choice data, having an alternative model of context dependence such as the RAM that is generalisable to binary choice data will be particularly useful.

Allowing for heterogeneous weights through the use of heuristic weighting functions on the context independent RUM component and the context dependent component has been shown to be another promising line of research. As a means of embedding multiple heuristics into choice models, this is an alternative to the latent class (probabilistic decision processes) approach. If consumer preferences are thought to be a synthesis of both ‘inherent’ and ‘constructed’ preferences, as more recently advocated by Simonson (2008) and Kivetz *et al.* (2008), then there is an appealing behavioural basis to the heuristic weighting function approach as well. This model may even be interpreted as an early attempt to heed the call by Kivetz *et al.* (2008) to build “synthetised models that capture the relative weight between inherent and constructed preferences.” (Kivetz *et al.*, 2008, p. 185). In fact, Kivetz *et al.* (2008) hypothesised that the weight of constructed preferences may vary according to certain individual differences, which is precisely what the heuristic weighting approach aims to do. Accordingly, the data at hand allow the heuristic weighting functions to be specified as being dependent on socio-economic characteristics such as age and gender. Although the empirical evidence across datasets is mixed as to how each of the socio-economic characteristics influences the weight of the heuristic, this does not detract from the potential merits of this approach in future work.

Finally, another original contribution of the thesis is the estimation of mixed logit models of the NLWLR and the regret-RAM models. Mixed logit models of the NLWLR heuristic are particularly interesting as heterogeneity in the power parameters can be revealed, and these can be used to establish the extent to which the compromise effect occurs in the choice set. The conclusion is that extremeness aversion across all attributes, as indicated by a power parameter of less than one, is not a common occurrence. Rather, a mixture of extremeness seeking and extremeness aversion is more common. Moreover, in the context of choice among various toll road alternatives, extremeness aversion seems to be more likely for the cost attributes than for the time attributes. The mixed logit estimation of the regret-RAM model is a fairly straightforward extension of the fixed parameter, MNL case.

The value of travel time savings (VTTS) expressions for the various extremeness aversion models have also been derived in this thesis. With models such as the RRM, NLWLR and RAM, the VTTS expressions are not only a function of the taste parameters, but also functions of the attributes in the competing alternatives as well, to varying degrees. In the NLWLR model, the reference point is the worst level of any given attribute in the choice set, and so, changes to other attribute levels (as long as they do not become the worst level) will not impact VTTS. This is not the case with the RAM and RRM models, as these models require binary comparisons to be made across all eligible pairs of alternatives in the choice set; thus, a change to any attribute value in the choice set will change the VTTS.

On model outputs, Hensher *et al.* (2011) and Theine *et al.* (2012) claim to find non-negligible differences in the elasticities derived from the linear additive RUM and the RRM models. While it is true that the VTTS formulae are different in the RUM and RRM models, the aforementioned papers only reported mean VTTS values and in the absence of further information, it is difficult to make a judgement on whether the differences are statistically significant or not. In contrast to these findings, on the VTTS measures that have been calculated in Chapter 5, the results generally indicate statistically insignificant differences in the means of the VTTS estimates between the RUM and the RRM models, even though the magnitude of the differences may appear to be quite large. An analysis of the VTTS values obtained from the NLWLR and the regret-RAM models also suggest statistically insignificant differences and in the case of the regret-RAM model, the results are very similar in magnitude to the linear additive RUM model.

Although the empirical results from these datasets do not suggest that there is substantial variation in the VTTS values across the models (valid within the limits of the experimental design), nevertheless, choice set specific preferences such as those encapsulated by the NLWLR, RRM and RAM models produce markedly different behavioural insights which potentially lead to a richer interpretation of the trade-offs as choice set composition is changed. For example, in the RRM and RAM models, the VTTS calculations would be sensitive to changes in the performance of any of the other available alternatives, a feature which may become more apparent in other data contexts. For labelled experiments, it would also be appropriate to generate choice probability forecasts to get a sense of the magnitudes of differences that might be expected across the various behavioural models.

6.2 SOME SUGGESTIONS FOR FUTURE RESEARCH

Attribute non-attendance is one area of research that has been very actively looked into in recent years. However, this issue has by and large been left aside in this thesis, in the sense that all time and cost attributes presented to the respondent are assumed to be attended to. Future work could bring the two research elements of context dependent preferences and attribute non-attendance together. For example, in the study of the MCD heuristic, provision might be made for certain attributes to be ignored in the total MCD count based on say, responses to supplementary questions that may be found at the end of the survey. An extension of this idea might be to integrate the notion of just noticeable differences with MCD. Behind just noticeable differences is a recognition that differences in attributes across alternatives are too small to warrant further consideration and these attributes may therefore be ignored in the editing phase of the decision. For just noticeable differences to work, the idea is that an attribute adds to MCD only if some threshold level is exceeded. To illustrate, a time attribute might contribute to the MCD count only if it exceeds the second best level (in the choice set) by say, five minutes. Another possibility is to consider functional forms for MCD that impose conditions such as lower or upper thresholds in a latent class setting, such as Campbell *et al*. (2013).

Likewise, travel time variability, in particular, the treatment of risk, uncertainty and probabilities applying the heuristics discussed in this thesis, is another area worthy of further research. Models such as the rank dependent utility (RDU) model suggested by Hensher and Li (2012) come to mind, where the decision weight for each event depends on a non-linear probability weighting function. This decision weight itself transforms the cumulative probability distribution of the event, based on its ranking relative to other outcomes. In a modified RDU model for example, there would be ample scope to consider alternative specifications of the utility function, such as those informed by the extremeness aversion heuristic.

Turning to a discussion of the models themselves, it might be observed that while the RAM model has been applied to a limited extent in consumer choice datasets, it is still very early days yet for the RAM model in transportation research. In fact, the RAM model has yet to be more closely evaluated in other disciplines like environmental and health research that also rely significantly on choice modelling tools. Likewise, the NLWLR model or its cousin the contextual concavity model is only just beginning to get attention in the formal choice modelling literature. Beyond this thesis, a number of very fruitful areas of research can be pursued with regards to these models. Very similar datasets have been used which may limit the degree to which evidence in the form of model fit differences is in fact convincing and generalisable. Therefore, testing of the RAM and NLWLR models can be extended to other datasets, for example with revealed preference data where attribute values across alternatives could be quite similar to one another. Labelled datasets are also good candidates for further work. Although choice sets with a smaller number of alternatives such as the ones examined in this thesis might make a good start for exploratory analysis of this kind, it would also be imperative at some point down the road to consider extending these models to larger choice sets, as such choice sets are more likely to be encountered as part of a larger research agenda.

With the RRM and RAM models in particular, there is the intriguing possibility that the respondent may be attending to only a subset of all possible binary pair wise comparisons of an attribute across alternatives. This may be a consequence of just noticeable differences, in which comparisons are not performed by the respondent if the differences are perceived to be negligible. To explore this issue further, models which have been developed to make choice set formation and attribute non-attendance endogenous may be extended to make pair wise comparisons endogenous as well. At a practical level, a reduction in the modelled number of pair wise comparisons (especially for the RAM model which has a more complicated utility expression) is likely to bring greater stability to the estimation and help to make run times more manageable, especially when choice sets become large.

Related to this issue, one working assumption for the RAM model is that all attributes attended to in the context independent RUM component of the model are also attended to in the relative advantage component. This assumption can also be tested in future work by allowing some subset of attributes to appear in either one component only, by assuming a structure of heterogeneity like the hybrid RRM-RUM model. Where data on attribute processing are available, such as whether a respondent reported ignoring an attribute (see Hensher (2010) for details of such models), the link between attribute processing and the relative advantage component can also be studied in greater depth.

Chorus (2012) highlights that since the RRM VTTS is dependent not only on the considered alternative, but also on all alternatives that make up the choice set, the RRM VTTS is likely to violate the microeconomic axioms that underpin the VTTS for the RUM model. The same can be said for the other context dependent models as well. In real life, people are probably not as ‘well-behaved’ as neoclassical economic theory would postulate, but using the conventional RUM model leads to well understood VTTS and consumer surplus measures, even if it almost inevitably means accepting a loss of behavioural realism. On the other side of the trade off, there may be instances where capturing more behavioural realism is preferred which is where context dependent models could come into their own. Such a scenario might come about where forecasting future demand (for a new road, for a new good etc) is concerned. Differences in model fit may offer some guidance, but ultimately, it is up to the analyst to carefully consider the more appropriate model type to use, based on the research question to be addressed. Context dependent models are based on certain reference points, and may be more suited to short term predictions assuming references can be identified and are reasonably stable over the period of evaluation. With longer term predictions, references can change (for example, a new and significant travel mode) and more research is needed to understand the evolution of references over time.

Where the mixture model of heuristics is concerned, this thesis provides some indication as to how various socio economic factors might trigger the pattern of decision making that is associated with context dependent preferences. It would be interesting to see what other factors might tilt the weight either towards or away from such preferences. For example, it has been said (Zeelenberg and Pieters, 2007) that in decisions which matter to the respondent, the anticipation of regret becomes more dominant. Hence, it would be an interesting research question to determine if factors like the importance or the difficulty of the decision matter in the heuristic weighting function or not.

The models described in this thesis are still very much works-in-progress. Until such time when a fuller understanding of the properties of these context dependent models becomes available, a useful first step for the practitioner might be to think of using such alternative behavioural specifications as part of a sensitivity analysis to test the robustness of predictions from our more conventional (and better understood) models. With more research effort, time will tell if, and under what conditions, context dependent models might in fact be preferable to the standard RUM model. More specifically, the validity of the NLWLR and RAM models vis-à-vis the RRM model may be put to the test. These conclusions will have significant implications for practical work.

In the next generation of behavioural models for the utility expression, the results presented suggest that the alternative model forms for context dependence and referencing highlighted in this thesis are prime candidates for further consideration. These are:

(1) The NLWLR heuristic, where the reference is the worst attribute level in the current choice set s, and the power parameter φ_k is greater than zero. In the case of time and cost attributes the worst attribute level would typically be the highest level in the choice set:

$$U_{js}^{NLWLR} = \sum_k (\beta_k (X_{jks} - \max_{j \in s}(X_{jks})))^{\varphi_k} + \varepsilon_j \tag{6.1}$$

(2) The RAM model, where the reference is to counterpart attribute levels of the competing alternatives in the choice set. The RAM model is an example of a model form where preferences are not simply context dependent alone:

$$U_{js}^{RAM} = \sum_k \beta_{jk} X_{jk} + \sum_{\substack{j' \in s, \\ j' \neq j}} RA(j, j') + \varepsilon_j , \tag{6.2}$$

Finally, variants of these models where each heuristic or decision rule is weighted using some form of weighting function, can also be considered.

It is hoped that the work from this thesis will further excite the growing interest in modelling the role of heuristics in choice behaviour. Without overstating the case, there is no doubt that a wealth of new discoveries awaits the intrepid researcher.

REFERENCES

Abou-Zeid, M. and Ben-Akiva, M. 2011, 'The effect of social comparisons on commute well-being', *Transportation Research Part A*, vol. 45, no. 4, pp. 345-361.

Adamowicz, W. and Swait, J. 2013, 'Are food choices really habitual? Integrating habits, variety seeking and compensatory choice in a utility maximising framework', *American Journal of Agricultural Economics*, vol. 95, no. 1, pp. 17-41.

Andersen, S., Harrison, G.W., Lau, M.I. and Rutström, E.E. 2007, 'Dual criteria decisions', *Working Paper 06-11*, Department of Economics, College of Business Administration, University of Central Florida.

Avineri, E. and Prashker, J. N. 2004, 'Violations of expected utility theory in route-choice stated preferences: certainty effect and inflation of small probabilities', *Transportation Research Record*, vol. 1894, pp. 222-229.

Bateman, I.J., Carson, R.T., Day, B., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P. 2008, 'Choice set awareness and ordering effects in discrete choice experiments', *CSERGE Working Paper EDM 08-01*.

Bell, D.E. 1982, 'Regret in decision making under uncertainty', *Operations Research*, vol. 30, no. 5, pp. 961-981.

Ben-Akiva and S. Lerman. 1985, *Discrete choice analysis: theory and application to predict travel demand*, MIT Press, Cambridge, MA.

Ben-Akiva and Swait, J. 1986, 'The Akaike likelihood ratio index', *Transportation Science*, vol. 20, no. 2, pp. 133-136.

Bettman, J. R., Luce, M.F. and Payne, J. W. 1998, 'Constructive consumer choice processes', *Journal of Consumer Research*, vol. 25, no. 3, pp. 187-217.

Boxall, P., Adamowicz, W. L. and Moon, A. 2009, 'Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement', *Australian Journal of Agricultural and Resource Economics*, vol. 53, no. 4, pp. 503–519.

Briesch, R.A., Krishnamurthi, L., Mazumdar, T. and Raj, S.P. 1997, 'A comparative analysis of reference price models', *Journal of Consumer Research*, vol. 24, no. 2, pp. 202-214.

Camerer, C. 1999, 'Behavioral economics: reunifying psychology and economics', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 10575-10577.

Cameron, T.A. and DeShazo, J.R. 2010, 'Differential attention to attributes in utility-theoretic choice models', *Journal of Choice Modelling*, vol. 3, no. 3, pp. 73-115.

Campbell, D., Hensher, D.A. and Scarpa, R. 2013, 'Bounding WTP distributions to reflect the 'actual' consideration set', *3rd International Choice Modelling Conference*, Sydney, Australia, 3-5 July, available online at http://www.icmconference.org.uk/index.php/icmc/ICMC2013/paper/view/713/358

Cantillo, V., Heydecker, B. and Ortuzar, J. de D. 2006, 'A discrete choice model incorporating thresholds for perception in attribute values', *Transportation Research Part B*, vol. 40, no. 9, pp. 807-825.

Cantillo, V. and Ortuzar, J.de D. 2005, 'A semi-compensatory discrete choice model with explicit attribute thresholds of perception', *Transportation Research Part B*, vol. 39, no. 7, pp. 641-657.

Carson, R.T. and Louviere, J.J. 2011, 'A common nomenclature for stated preference elicitation approaches', *Environmental and Resource Economics*, vol. 49, no. 4, pp. 539-559.

Chorus, C.G. 2010, 'A new model of random regret minimization', *European Journal of Transportation and Infrastructure Research*, vol. 10, no. 2, pp. 181-196.

Chorus, C.G. 2012, ‘Random regret minimization: an overview of model properties and empirical evidence’, *Transport Reviews: A Transnational Transdisciplinary Journal*, vol. 32, no. 1, pp 75-92.

Chorus, C.G., Arentze, T.A. and Timmermans, H.J.P. 2008, ‘A random regret minimization model of travel choice’, *Transportation Research Part B*, vol. 42, no. 1, pp. 1-18.

Chorus, C.G. and Bierlaire, M. 2013, ‘An empirical comparison of travel choice models that capture preferences for compromise alternatives’, *Transportation*, vol. 40, pp. 549-562.

Chorus, C.G., Rose, J.M. and Hensher, D.A. 2013, ‘Regret minimization or utility maximization: it depends on the attribute’, *Environment and Planning B: Planning and Design*, vol. 40, pp. 154-169.

Day, B. and Prades, J.P. 2010, ‘Ordering anomalies in choice experiments’, *Journal of Environmental Economics and Management*, vol. 59, no. 3, pp. 271-285.

DeShazo, J.R. 2002, ‘Designing transactions without framing effects in iterative question formats’, *Journal of Environmental Economics and Management*, vol. 43, no. 3, pp. 360-385.

DeShazo, J.R. and Fermo, G. 2002, ‘Designing choice sets for stated preference methods: the effects of complexity on choice consistency’, *Journal of Environmental Economics and Management*, vol. 44, no.1, pp. 123-143.

Fernandez, R. and Rodrik, D. 1991, ‘Resistance to reform: status quo bias in the presence of individual-specific uncertainty’, *American Economic Review*, vol. 81, no. 5, pp. 1146-1155.

Gigerenzer, G., Todd, P.M. and the ABC Research Group. 1999, *Simple heuristics that make us smart*, Oxford University Press, Oxford, U.K.

Gilbride, T.J. and Allenby, G.M. 2004, ‘A choice model with conjunctive, disjunctive and compensatory screening rules’, *Marketing Science*, vol. 23, no. 3, pp. 391-406.

Gilbride, T.J. and Allenby, G.M. 2006, ‘Estimating heterogeneous EBA and economic screening rule choice models’, *Marketing Science*, vol. 25, no. 5, pp. 494-509.

Gilboa, I. and Schmeidler, D. 1995, ‘Case-based decision theory’, *Quarterly Journal of Economics*, vol. 110, no. 3, pp. 605 – 640.

Gourville J.T. and Soman, D. 2007, ‘Extremeness seeking: when and why consumers prefer the extreme’, Harvard Business School Working Paper Number 07-092. Available online at http://hbswk.hbs.edu/item/5700.html.

Heckman, J. 1981, ‘Statistical models for discrete panel data’, in C. F. Manski and D. McFadden (eds.), *Structural analysis of discrete data and econometric applications*, The MIT Press, Cambridge, MA, pp. 114-177.

Hensher, D.A. 2006, ‘How do respondents process stated choice experiments? Attribute consideration under varying information load’, *Journal of Applied Econometrics*, vol. 21, no. 6, pp. 861-878.

Hensher, D.A. 2010, ‘Attribute processing, heuristics and preference construction in choice analysis’, in S. Hess and A. Daly (eds.), *Choice modelling: the state-of-the-art and the state-of-practice – proceedings from the inaugural international choice modelling conference*, Emerald Group Publishing, Bingley, U.K., pp. 35-70.

Hensher, D.A. and Collins, A.T. 2011, ‘Interrogation of responses to stated choice experiments: is there sense in what respondents tell us?’, *Journal of Choice Modelling*, vol. 4, no. 1, pp. 62-89.

Hensher, D.A., Collins, A.T. and Greene, W.H. 2012a, ‘Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding’, *Transportation,* doi: 10.1007/s11116-012-9447-0.

Hensher, D.A. and Greene, W.H. 2010, ‘Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification’, *Empirical Economics*, vol. 39, no. 2, pp. 413-426.

Hensher, D.A., Greene, W.H. and Chorus, C.G. 2011, ‘Random regret minimization or random utility maximization: an exploratory analysis in the context of automobile fuel choice’, *Journal of Advanced Transportation,* doi: 10.1002/atr.188

Hensher, D.A. and Li, Z. 2012, ‘Valuing travel time variability within a rank-dependent utility framework and an investigation of unobserved taste heterogeneity’, *Journal of Transport Economics and Policy*, vol. 46, no. 2, pp. 293-312.

Hensher, D.A., Rose, J.M. and Greene, W.H. 2006, *Applied choice analysis: a primer*, Cambridge University Press, Cambridge, U.K.

Hensher, D.A., Rose, J.M. and Greene, W.H. 2012b, ‘Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design’, *Transportation*, vol. 39, no. 2, pp. 235-245.

Hensher, D.A. and Rose, J.M. 2012, ‘The influence of alternative acceptability, attribute thresholds and choice response certainty on automobile purchase preferences, *Journal of Transport Economics and Policy*, vol. 46, no. 3, pp. 451-468.

Hensher, D.A., Rose, J.M. and Li, Z. 2012c, ‘Does the choice model method and/or the data matter?’ Transportation, vol. 39, no. 2, p. 351-385.

Hensher, D.A., Smith, N.C., Milthorpe, F.W. and Barnard, P.O. 1992, *Dimensions of automobile demand: a longitudinal study of household automobile ownership and use*, Elsevier Science Publishers, Amsterdam.

Hess, S., Rose, J.M. and Hensher, D.A. 2008, ‘Asymmetric preference formation in willingness to pay estimates in discrete choice models’, *Transportation Research Part E*, vol. 44, no. 5, pp. 847-863.

Hess, S., Stathapoulos, A. and Daly, A. 2012, ‘Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies’, *Transportation*, vol. 39, no. 3, pp. 565-591.

Hole, A.R. 2011, ‘A discrete choice model with endogenous attribute attendance’, *Economics Letters*, vol. 110, no. 3, pp. 203-205.

Holmes, T.P. and Boyle, K.J. 2005, ‘Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions’, *Land Economics*, vol. 81, no. 1, pp. 114-126.

Huber, J., Payne, J.W. and Puto, C. 1982, ‘Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis’, *Journal of Consumer Research*, vol. 9, no. 1, pp. 90-98.

Kahneman, D. and Tversky, A. 1979, ‘Prospect theory: an analysis of decision under risk’, *Econometrica*, vol. 47, no. 2, pp. 263-292.

Kivetz, R., Netzer, O. and Srinivasan, V. 2004, ‘Alternative models for capturing the compromise effect’, *Journal of Marketing Research*, vol. 41, no. 3, pp. 237-257.

Kivetz, R., Netzer, O. and Schrift, O. 2008, ‘The synthesis of preference: bridging behavioral decision research and marketing science’, *Journal of Consumer Psychology*, vol. 18, no. 3, pp. 179-186.

Lancaster, K.J. 1966, ‘A new approach to consumer theory’, *Journal of Political Economy*, vol. 74, no. 2, pp. 132-157.

Li, Z. and Hensher, D.A. 2011, ‘Prospect theoretic contributions in understanding traveller behaviour under risk or uncertainty: a review and some comments’, *Transport Reviews,* vol. 31, no. 1, pp. 97-117.

Loomes, G. and Sugden, R. 1982, ‘Regret theory: an alternative theory of rational choice under uncertainty’, *The Economic Journal*, vol. 92, no. 368, pp. 805-824.

Louviere, J.J., Flynn, T.N. and Carson, R.T. 2010, 'Discrete choice experiments are not conjoint analysis', *Journal of Choice Modelling*, vol. 3, no. 3, pp. 57-72.

Louviere, J.J. and Hensher, D.A. 2001, 'Combining sources of preference data' in D. Hensher (ed.), *Travel behaviour research: the leading edge*, Elsevier Science Ltd., Oxford, U.K., pp. 125-144.

Louviere, J.J., Hensher, D.A., and J.D. Swait. 2000, *Stated choice methods: analysis and application*, Cambridge University Press, Cambridge, U.K.

Louviere, J.J. and Meyer, R.J. 2008, 'Formal choice models of informal choices – what choice modelling research can (and can't) learn from behavioural theory', in N. K. Malhotra (ed.), *Review of marketing research volume 4*, M.E. Sharpe, Armonk, NY, pp. 3-32.

Luce, R.D. 1959, *Individual choice behaviour*, John Wiley & Sons, Inc., New York, NY.

Luce, R.D. 1977, 'The choice axiom after twenty years', *Journal of Mathematical Psychology*, vol. 15, no. 3, pp. 215-233.

Manski, C.F. 1977, 'The structure of random utility models, *Theory and Decision*, vol. 8, no. 3, pp. 229-254.

McFadden, D. 1974, 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (ed.), *Frontiers in Econometrics*, Chapter 4, Academic Press, New York, NY and London, U.K., pp. 105-142.

McFadden, D. 1999, 'Rationality for economists?' *Journal of Risk and Uncertainty*, vol. 19, nos. 1-3, pp. 73-105.

McNair, B.J., Bennett, J. and Hensher, D.A. 2011, 'A comparison of responses to single and repeated discrete choice questions', *Resource and Energy Economics*, vol. 33, no. 3, pp. 554-571.

McNair, B.J., Hensher, D.A. and Bennett, J. 2012, ‘Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model’, *Environment and Resource Economics*, vol. 51, no. 4, pp. 599-616.

Payne, J.W. 1976, ‘Task complexity and contingent processing in decision making: an information search and protocol analysis’, *Organizational Behavior and Human Performance*, vol. 16, pp. 366-387.

Payne, J.W. and Bettman, J.R. 1992, ‘Behavioural decision research: a constructive processing perspective’, *Annual Review of Psychology*, vol. 43, pp. 87-131.

Payne, J.W., Bettman, J.R. and E.J. Johnson. 1993, *The Adaptive Decision Maker*, Cambridge University Press, Cambridge, U.K.

Payne, J.W., Bettman, J.R. and Schkade, D.A. 1999, ‘Measuring constructed preferences: towards a building code’, *Journal of Risk and Uncertainty*, vol. 19, nos. 1-3, pp. 243-270.

Rooderkerk, R.P., Van Heerde, H.J. and Bijmolt, T.H.A. 2011, ‘Incorporating context effects into a choice model’, *Journal of Marketing Research*, vol. 48, no. 4, pp. 767-780.

Rose, J.M., Bliemer, M.C.J., Hensher, D.A. and Collins, A.T. 2008, ‘Designing efficient stated choice experiments in the presence of reference alternatives’, *Transportation Research Part B*, vol. 42, no. 4, pp. 395-406.

Russo, J.E. and Dosher, B.A. 1983, ‘Strategies for multiattribute binary choice’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 9, no. 4, pp. 676-696.

Samuelson, W. and Zeckhauser, R. 1988, ‘Status quo bias in decision making’, *Journal of Risk and Uncertainty*, vol. 1, no. 1, pp. 7-59.

Scarpa, R., Gilbride, T.J., Campbell, D. and Hensher D.A. 2009, ‘Modelling attribute non-attendance in choice experiments for rural landscape valuation, *European Review of Agricultural Economics*, vol. 36, no. 2, pp. 151-174.

Scarpa, R., Thiene, M. and Hensher, D.A. 2010, ‘Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: does it matter?’ *Land Economics*, vol. 86, no. 4, pp. 817-839.

Scheufele, G. and Bennett, J. 2010, ‘Ordering effects and response strategies in discrete choice experiments’, *2010 World Congress of Environmental and Resource Economists*, Montreal, Canada, available online at http://purl.umn.edu/107743.

Scheufele, G. and Bennett, J. 2013, ‘Effects of alternative elicitation formats in discrete choice experiments’, *The Australian Journal of Agricultural and Resource Economics*, vol. 57, no. 2, pp. 214-233.

Senbil, M. and Kitamura, R. 2004, ‘Reference points in commuter departure time choice: a prospect theoretic test of alternative decision frames’, *Journal of Intelligent Transportation Systems*, vol. 8, no. 1, pp. 19-31.

Simon, H. A. 1955, ‘A behavioral model of rational choice’, *Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99-118.

Simonson, I. 1989, ‘Choice based on reasons: the case of attraction and compromise effects’, *Journal of Consumer Research*, vol. 16, no. 2, pp. 158-174.

Simonson, I. 2008, ‘Will I like a “medium” pillow? Another look at constructed and inherent preferences’, *Journal of Consumer Psychology*, vol. 18, no. 3, pp. 159-169.

Simonson, I. and Tversky, A. 1992, ‘Choice in context: tradeoff contrast and extremeness aversion’, *Journal of Marketing Research*, vol. 29, no. 3, pp. 281-295.

Stevenson, M.K., Rusemeyer, J.R. and Naylor, J.C. 1990, ‘Judgment and decision-making theory’, in M.D. Dunnette and L.M. Hough (eds.), *Handbook of industrial and organizational psychology*, *vol. I*, Consulting Psychologists Press, Palo Alto, CA, pp. 283-374.

Strazzera, E., Contu, D. and Ferrini, S. 2013, ‘Check it out! A monte carlo analysis of the performance of selection criteria and tests for models fitted to choice experiments data’, Paper presented at the 3rd International Choice Modelling Conference, Sydney, Australia, 3-5 July.

Suzuki, Y. 2007, ‘Modeling and testing the “two-step” decision process of travelers in airport and airline choices’, *Transportation Research Part E*, vol. 43, no. 1, pp. 1-20.

Swait, J. 2001, ‘A non-compensatory choice model incorporating attribute cut-offs’, *Transportation Research Part B*, vol. 35, no. 10, pp. 903-928.

Swait, J. 2009, ‘Choice models based on mixed discrete/continuous pdfs’, *Transportation Research Part B*, vol. 43, no. 7, pp. 766-783.

Swait, J. and Adamowicz, W. 2001, ‘Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice’, *Organizational Behavior and Human Decision Processes*, vol. 86, no. 2, pp. 141-167.

Swait, J., Adamowicz, W., Hanemann, M., Diederich, A., Krosnick, J., Layton,D., Provencher, W., Schkade, D. and Tourangeau, R. 2002, ‘Context dependence and aggregation in disaggregate choice analysis’, *Marketing Letters*, vol. 13, no. 3, pp. 195-205.

Swait, J., Adamowicz, W. and van Buren, M. 2004, ‘Choice and temporal welfare impacts: incorporating history into discrete choice models’, *Journal of Environmental Economics and Management*, vol. 47, no. 1, pp. 94-116.

Swait, J. and Ben-Akiva, M. 1987, ‘Incorporating random constraints in discrete models of choice set generation’, *Transportation Research Part B*, vol. 21, no. 2, pp. 91-102.

Theine, M., Boeri, M. and Chorus, C.G. 2012, ‘Random Regret Minimization: Exploration of a New Choice Model for Environmental and Resource Economics’, *Environmental and Resource Economics*, vol. 51, no. 3, pp. 413-29.

Timmermans, H. 2010, ‘On the (ir)relevance of prospect theory in modelling uncertainty in travel decisions, *European Journal of Transport and Infrastructure Research*, vol. 10, no. 4, pp. 368-384.

Train, K. 1999, ‘Halton sequences for mixed logits’, Department of Economics, University of California at Berkeley, revised August 2, 1999.

Train, K. and Wilson, W.W. 2008, ‘Estimation on stated-preference experiments constructed from revealed-preference choices’, *Transportation Research Part B*, vol. 42, no. 3, pp. 191-203.

Tversky, A. 1969, ‘Intransitivity of preferences’, *Psychological Review*, vol. 76, no. 1, pp. 31-48.

Tversky, A. 1972, ‘Elimination by aspects: a theory of choice’, *Psychological Review*, vol. 79, no. 4, pp. 281-299.

Tversky, A. and Kahneman, D. 1981, ‘The framing of decisions and the psychology of choice’, *Science*, vol. 211, no. 4481, pp. 453-458.

Tversky, A. and Kahneman, D. 1991, ‘Loss aversion in riskless choice: a reference-dependent model’, *Quarterly Journal of Economics*, vol. 106, no. 4, pp. 1039-1061.

Tversky, A. and Kahneman, D. 1992, ‘Advances in prospect theory: cumulative representations of uncertainty’, *Journal of Risk and Uncertainty*, vol. 5, no.4, pp. 297–323.

Tversky, A., Slovic, P. and Kahneman, D. 1990, ‘The causes of preference reversal’, *American Economic Review*, vol. 80, no. 1, pp. 204-217.

Tversky, A. and Simonson, I. 1993, ‘Context-dependent preferences’, *Management Science*, vol. 39, no. 10, pp. 1179-1189.

Van de Kaa, E.J. 2010a, ‘Applicability of an extended prospect theory to travel behaviour research: a meta-analysis’, *Transport Reviews*, vol. 30, no. 6, pp. 771-804.

Van de Kaa, E.J. 2010b, ‘Prospect theory and choice behaviour strategies: review and synthesis of concepts from social and transport sciences’, *European Journal of Transport and Infrastructure Research*, vol. 10, no. 4, pp. 299-329.

Vuong, Q.H. 1989, ‘Likelihood Ratio Test for Model Selection and Non-Nested Hypotheses’, *Econometrica*, vol. 57, no. , pp. 307-333.

Young, W. 1984, ‘A non-tradeoff decision making model of residential location choice’, *Transportation Research Part A*, vol. 18A, no. 1, pp. 1-11.

Zeelenberg, M. and Pieters, R. 2007, ‘A theory of regret regulation 1.0’, *Journal of Consumer Psychology*, vol. 17, no. 1, pp. 3-18.

Zhang, J. and Adamowicz, W.L. 2011, ‘Unraveling the choice format effect: a context-dependent random utility model’, *Land Economics*, vol. 87, no. 4, pp. 730-743.