POISSON RACE MODELS: THEORY AND APPLICATION IN CONJOINT CHOICE ANALYSIS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Shiling Ruan, M.S.

The Ohio State University

2007

Dissertation Committee:

Steven N. MacEachern, Co-Adviser
Angela M. Dean, Co-Adviser
Thomas Otter

Approved by

Co-Adviser

Co-Adviser

Graduate Program in Statistics
Conjoint choice experiments are widely used to study consumer preference among a set of product alternatives. Traditional models (such as the multinomial logit model) for conjoint choice analysis imply some unrealistic consumer behaviors which are often not observed in real conjoint choice data. Additionally, they are at lack of connection with underlying psychological decision process. In this dissertation, a stochastic model, the Poisson race model, is studied and applied to the context of conjoint choice analysis. It assumes that decision making is a process of accumulating information in favor of each alternative. When the information accumulated reaches certain threshold, a choice is made. The accumulation of information for each alternative follows a Poisson process and is independent of those of other alternatives.

A set of theoretical results are derived for the Poisson race model, including expressions for choice probabilities, monotonicity, effect of thresholds and the behavior implications. The behavior implications includes results on whether the family of Poisson race models has the properties of Independence of Irrelevant Alternatives (IIA) and transitivity and under what conditions such properties fall apart. Theoretically, the Poisson race model not only captures the traditional multinomial logit model as a special case, but also describes a much broader range of decision making behavior.
A new class of Poisson race model is proposed to model dependence in conjoint choice data. It incorporates a dependence structure which captures the relationship between the attributes of the choice alternatives and which appropriately moderates the randomness inherent in the race. The proposed dependent model assumes that there is a shared process tracking the information shared by the alternatives in a choice set. The formulae to calculate the choice probabilities are derived and the tie-breaking mechanism are discussed. The new model is also extended to the conjoint choice data with more than two alternatives in a choice set. The new model is applied to real conjoint choice data on consumer preference of credit cards with binary and trinary choice sets and is shown to have markedly superior performance to independent Poisson race models and to the multinomial logit model.
Dedicated to my parents and grandparents
ACKNOWLEDGMENTS

I would like to express my greatest gratitude to both of my advisors, Steven N. MacEachern and Angela M. Dean for their intellectual guidance, consistent encouragement, and genuine enthusiasm that have made this thesis possible. Their mentorship and expertise have guided me through many obstacles during this research. They always find time to listen to my questions and have taught me innumerable lessons and insights on how to conduct research in general. Their editorial advice (including correcting my writing errors) was essential to the completion of this dissertation. I can not thank them enough for their understanding and patience. After all, they have made my years in graduate school more enjoyable and memorable.

I am greatly indebted to my committee member, Thomas Otter, for his active presence and fruitful collaboration during this research, for his numerous inspiring discussions and suggestions about the dependence race models and for his invaluable help in collecting the conjoint choice data.

I would like to thank Trish Van Zandt for serving in my candidacy committee and for her insightful comments on the dissertation work. Also, I wish to thank Greg Allenby, Mario Peruggia for their stimulating discussion on this topic during the marketing and psychology seminars.
I wish to thank a wonderful friend, Qing Liu, for allowing me to use the conjoint rating data, for proving me the credit card templates and for her assistance in going through the graduation process.

My sincere thanks also go to many faculty and staff in the Department of Statistics who have helped me in numerous ways: Doug Wolfe, for granting me a Teaching Assistantship and later a Research Assistantship that enabled me to enter the graduate program in Statistics; Prem Goel, for supporting me with a Research Assistantship; Tom Bishop, for giving me opportunities to participate in consulting projects. I thank many faculty members who had taught me in various subject matters in statistics. I also wish to thank the department staff, especially the computing support staff, for providing me great help in completing this research work.

Last, but not least, my family receive my deepest gratitude and love: My parents, Deqin Chen and Yongquan Ruan, for raising me during the difficult years, for always supporting me to pursue my dreams, for their unconditional love no matter what; My husband Deng Huang, for going through the ups and downs of life with me with his unwavering love and constant support; My parents-in-law, Shunle Huang and Meiqing Liu, for their great help during these special months; Yaxi and Gesi, my twin daughters, for bringing me so much joy.

The dissertation work was partially funded by the NSF under grant numbers SES-0437251 and DMS-0605041 and by the NSA under award number H98230-05-1-0065.
VITA

May 15, 1974 .......................... Born - Luchuan, China

1997 ................................. B.S. Materials Science and Engineering, Tsinghua University, China

2000 ................................. M.S. Materials Science and Engineering, The Ohio State University

1998-2006 ......................... Graduate Teaching and Research Associate, The Ohio State University

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Statistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td></td>
<td>List of Figures</td>
<td>xvi</td>
</tr>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>History and development of Poisson race models</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Conjoint choice experiments</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Research goals</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Organization of this dissertation</td>
<td>9</td>
</tr>
<tr>
<td>2.</td>
<td>Choice Probabilities of Independent Poisson Race Models and Related Properties</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Expressions of choice probabilities</td>
<td>12</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Notation and definition</td>
<td>12</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Using the gamma distribution</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Using the beta distribution</td>
<td>16</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Using negative binomial and multinomial distributions</td>
<td>19</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Using binomial formulation for $m = 2$</td>
<td>21</td>
</tr>
</tbody>
</table>
4.2.1 Formulation of Poisson rates ........................................ 166
4.2.2 Computation of choice probabilities .............................. 169
4.2.3 Tie-breaking rule .................................................... 172
4.3 Modeling conjoint choice data ......................................... 175
4.3.1 Dependent Poisson race model ................................. 175
4.3.2 Alternative models ................................................. 177
4.4 Estimation .............................................................. 178
4.5 A conjoint choice study with three alternatives .................... 179
4.5.1 Design and data description ...................................... 179
4.5.2 Measures of holdout performance .............................. 184
4.5.3 Results and discussion ........................................... 186
4.6 Conclusion ............................................................... 210
5. Contributions and Future Work .......................................... 212
5.1 Contributions .......................................................... 212
5.2 Future directions ...................................................... 213
Appendices:
A. Proof of Lemma 2.3.5 .................................................. 215
B. A Conjoint Choice Experiment with Two Credit Cards ............ 218
C. A Conjoint Choice Experiment with Three Credit Cards ........... 230
Bibliography ............................................................... 239
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>An example of conjoint choice set with credit cards</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Percentage of games in which the $i$th team defeats the $j$th team.</td>
<td>37</td>
</tr>
<tr>
<td>2.2</td>
<td>Estimated parameters and negative log-likelihood for Gamma(1) and Gamma($K_1, \ldots, K_{12}$) models.</td>
<td>40</td>
</tr>
<tr>
<td>2.3</td>
<td>Predicted percentage of games that the $i$th team (row) defeats the $j$th team (column).</td>
<td>41</td>
</tr>
<tr>
<td>2.4</td>
<td>The estimated $K_i$ and $\lambda_i$ for Gamma($K_1, \ldots, K_{12}$) model with different starting points.</td>
<td>42</td>
</tr>
<tr>
<td>2.5</td>
<td>The test statistic for the Gamma(1) and Gamma($K_1, \ldots, K_{12}$) model.</td>
<td>49</td>
</tr>
<tr>
<td>2.6</td>
<td>Violation of weak stochastic transitivity.</td>
<td>74</td>
</tr>
<tr>
<td>2.7</td>
<td>Another example with violation of strong stochastic transitivity.</td>
<td>75</td>
</tr>
<tr>
<td>2.8</td>
<td>The ranking of teams from the strongest to the weakest according to the percent of games that the $i$th team defeats other teams.</td>
<td>78</td>
</tr>
<tr>
<td>3.1</td>
<td>An example of dominance with credit cards.</td>
<td>82</td>
</tr>
<tr>
<td>3.2</td>
<td>Two alternatives $a_1$ and $a_2$.</td>
<td>95</td>
</tr>
<tr>
<td>3.3</td>
<td>Design matrix for the simulation study</td>
<td>109</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulated Data.</td>
<td>110</td>
</tr>
</tbody>
</table>
3.5 Mean squared errors for the Dep($\epsilon = 0.0001, 0.001, 0.01, 0.1, 1, \infty$) model, Indep model, Indep($K = 1$) model and the MNL model 112
3.6 Estimated $\mu_\gamma$ for all models. 115
3.7 Design matrix for the credit card data. 116
3.8 MSE statistics for the first holdout sample. 127
3.9 Correct choice probability and likelihood of the first holdout sample. 128
3.10 Likelihood ratio of the first holdout sample for all models vs. Indep model. 129
3.11 $MSE$ statistics for the second holdout sample. 132
3.12 Correct choice probabilities and likelihood of the second holdout sample. 133
3.13 Likelihood ratio of the second holdout sample for all models vs. Indep model. 134
3.14 Design matrix of the conjoint choice data experiment. 137
3.15 Summary of the conjoint choice data. 139
3.16 $MSE_1$, $MSE_2$, $P_{cor}$ and $R_{cor}$ for the first holdout sample. 143
3.17 The logarithm of quasi-likelihood($logQL_{holdout}$) and quasi-likelihood ratio($QLR_{holdout}$) for the first holdout sample. 145
3.18 Posterior means of $\mu_\gamma$ in all models for the first holdout sample. 146
3.19 Two mean squared errors, correct choice probabilities and correct choice rates for the second holdout sample. 149
3.20 The logarithm of the quasi-likelihood, ($log(QL_{holdout})$) and quasi-likelihood ratio, ($QLR_{holdout}$) for the second holdout sample. 150
3.21 Averages of the posterior means of $\mu_\gamma$ for all models for the second calibration sample. 151
3.22 M SE1 , M SE2 , Pcor and Rcor for the third holdout sample. . . . . . . 154
3.23 The logarithm of quasi-likelihood, log(QLholdout , and quasi-likelihood
ratio QLRholdout for the third hold out sample. . . . . . . . . . . . . . 155
3.24 Posterior means of µγ for all models for the third calibration sample.

156

3.25 M SE1 , M SE2 , Pcor and Rcor for the fourth holdout sample. . . . . . 158
3.26 The quasi-loglikelihood, log(QLholdout , and quasi-likelihood ratio QLRholdout
for the fourth holdout sample. . . . . . . . . . . . . . . . . . . . . . . 159
3.27 Posterior means of µγ for all models for the fourth holdout sample. . 160
4.1

An example of dominance with three credit cards. . . . . . . . . . . . 166

4.2

Number of rows in Ñ for given K. . . . . . . . . . . . . . . . . . . . . 171

4.3

Design matrix of the conjoint choice experiment and summary of the
raw data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 181

4.4

Summary of the demographic questions . . . . . . . . . . . . . . . . . 182

4.5

Statistics for the first holdout data set. . . . . . . . . . . . . . . . . . 189

4.6

M SE1 , M SE2 , Pcorr and Rcorr for the second holdout sample. . . . . 192

4.7

log(QL), log(L) and likelihood ratios for the second holdout sample. . 193

4.8

Statistics for the second holdout sample. . . . . . . . . . . . . . . . . 195

4.9

log(QL), log(L) and likelihood ratio for the third holdout sample. . . 196

4.10 Holdout statistics for the fourth holdout sample with two trinary sets
and one binary set. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 199
4.11 The quasi-likelihood (QL), joint likelihood (L) and the likelihood ratios
for the fourth holdout sample. . . . . . . . . . . . . . . . . . . . . . . 200
4.12 Averages of the posterior means of µγ for all models for the first and
the fourth calibration sets. . . . . . . . . . . . . . . . . . . . . . . . . 203
xiv


4.13 Holdout statistics based on case-deleted estimates. . . . . . . . . . . 208

4.14 Minimum mean effective sample sizes (average over 10 MCMC chains)
for all subjects for each choice set. . . . . . . . . . . . . . . . . . . . . . 209
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The probability of choosing alternative $a_1$ for different $K_1$ and $K_2$ values.</td>
</tr>
<tr>
<td>2.2</td>
<td>The choice probability of alternative $a_1$ ($p(1</td>
</tr>
<tr>
<td>2.3</td>
<td>The observed probability that team $i$ defeats team $j$.</td>
</tr>
<tr>
<td>2.4</td>
<td>The predicted probability ($\hat{p}_{ij}$) that team $i$ defeats team $j$ under the Gamma(1) model.</td>
</tr>
<tr>
<td>2.5</td>
<td>The predicted probability ($\hat{p}<em>{ij}$) that team $i$ defeats team $j$ under the Gamma($K_1, \ldots, K</em>{12}$) model.</td>
</tr>
<tr>
<td>3.1</td>
<td>Ratio of Choice Probabilities plotted against values of $\lambda_{u2}$.</td>
</tr>
<tr>
<td>3.2</td>
<td>A suspicious consumer.</td>
</tr>
<tr>
<td>3.3</td>
<td>Evidence of dominance.</td>
</tr>
<tr>
<td>4.1</td>
<td>Scatterplots of the estimated mean choice probabilities of the chosen cards in the eight holdout binary sets for all respondents for MNL model vs. Dep($\epsilon = 0.001$) model.</td>
</tr>
<tr>
<td>4.2</td>
<td>Scatterplots of posterior means of choice probabilities for the MNL model vs those of the Dep($\epsilon$ differ), Indep, Dep($\epsilon \to \infty$) model.</td>
</tr>
<tr>
<td>4.3</td>
<td>Histogram of posterior means of $\epsilon_i$ for the Dep($\epsilon_i$ differ) model.</td>
</tr>
<tr>
<td>4.4</td>
<td>The estimated $K_i$ from the first calibration set vs. $K_i$ from the second calibration set. The first calibration set consists of all trinary sets and the second calibration set consists of all binary sets.</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 History and development of Poisson race models

A basic problem in cognitive psychology is the modeling of response latency, also called response time (RT), which is the amount of time taken by an observer to make a decision (or response) based on attributes of the presented stimuli in a task (Townsend & Ashby, 1983). For example, in a simple perceptual matching experiment described by Van Zandt, Proctor and Colonius (2000), respondents are presented with two letters in each task and asked to indicate whether the stimuli are the same or different. The stimulus responses are “same” and “different”. The response time is recorded as the interval from the onset of the stimuli on the screen to the time that the respondents actually indicate their decision.

Among the various models proposed for RT in the past, the sequential-sampling models are the most widely used (Luce, 1986; Townsend & Ashby, 1983; Vikers, 1979). These models assume that the process of making a decision by a respondent is a process of sequentially sampling from the stimuli and gradually accumulating information toward some response. A response is generated when the accumulated
information exceeds some threshold for that response. The sequential-sampling models describe a speed-accuracy tradeoff. If the thresholds for the responses are large, more information is required to reach the thresholds and thus more time is required to respond. The models presume that information accumulates more quickly in favor of the better response. Consequently, with large thresholds, it is less likely for the system to produce an erroneous response. If the thresholds for the responses are small, less information is required and the response time is faster. It is more likely that the system will produce an erroneous response.

An important class of sequential-sampling models is the class of race models (or counting models), first proposed by LaBerge (LaBerge, 1962). The race models assume the existence of a separate internal “counter” associated with each potential response. The information favoring each potential response accumulates in parallel on the associated counters. A potential response is signaled when some preset threshold is exceeded by the accumulated information on its counter. The information stored in all counters is non-negative, which is different from the random walk model. The latter is another class of the sequential-sampling models. It assumes that a single counter stores information and that the information can be either positive or negative (Ratcliff, 1978).

Various assumptions have been proposed in the literature on how the information favoring each response accumulates on the counters. One natural interpretation is that neural spikes are emitted when a stimulus is presented. There are internal detectors (“counters”) associated with spikes, and the appropriate response is made when a preset number of spikes is recorded (Green & Luce, 1972; Siebert, 1970; Townsend & Ashby, 1983). Following this view, Audley and Pike (Audley & Pike,
1965; Pike, 1973) proposed the “Poisson race model”. In this model, the waiting time between successive increments to the counter (neural spikes) is exponentially distributed and the accumulation process can be represented as a Poisson process. Alternatively, Vikers (1970) assumed that information is accumulated in discrete time but the increments to the information total are continuous, which leads to the accumulator model.

Townsend and Ashby (1983) investigated the Poisson race models extensively. In a two-alternative forced choice experiment where respondents are asked to choose from two alternatives, the alternatives are assumed to have separate counters and separate thresholds associated with the independent Poisson processes. They derived the expressions for choice probabilities, distributions of RTs and mean RTs conditioning on the responses. Since then, the Poisson race models have attracted attention in modeling RT in experimental psychology. VanZandt, Colonius and Proctor (2000) applied Poisson race models to perceptual matching experiments and suggested that the race models may be a more useful representation for the choice making process than the random walk models. In these studies, the rates of the Poisson processes are assumed to be constant over time. Smith and Van Zandt (2000) further generalized the Poisson race model by allowing the Poisson rates to vary with time.

An interesting application of the Poisson race model is in the ranking problem. Stern (1987, 1990) considered ranking \( k \) objects according to the result of a competition, for example, the percent (number) of games won by a baseball team over a season. He assumed that each object, i.e., the \( i \)th team, scores points according to a Poisson process with rate \( \lambda_i \) and that all \( k \) processes are independent. The time taken for the \( i \)th team to score \( r \) points has a gamma distribution with shape parameter
and scale parameter $\lambda_i$. The probability that the teams are ranked according to a certain permutation is determined by the probability that the $k$ independent gamma random variables are ranked according to that permutation. The resulting model, refered to as a gamma permutation model by Stern, belongs to a special family of Poisson race models with a single fixed threshold for all teams. Stern studied such models in detail within the context of paired comparison models.

### 1.2 Conjoint choice experiments

Conjoint choice experiments (Louviere & Woodworth, 1983) are used widely in marketing to study consumer preferences amongst a set of alternative products. An alternative is a description of values of the attributes (“levels”) that define the product. For example, the two alternative credit cards presented in Table 1.1 each have four attributes in their profiles; the level of the third attribute (travel points) is the same for each alternative, but the levels of the other three attributes differ. These two credit cards form a choice set. In a conjoint choice task, a respondent is presented with a few choice sets of two or more product alternatives and is asked to choose a single preferred alternative in each set. Thus, the response data from conjoint choice experiments are of 0/1 type.

The conjoint choice data are used to estimate the importance of the levels of an attribute and to predict the choice probabilities for choice sets that may involve new combinations of attribute levels. As seen in the choice set in Table 1.1, the first credit card has a higher interest rate and no annual fee, while the second card has an annual fee but a lower interest rate and cash reward. Thus, these alternatives pose a trade-off relationship and a respondent must resort to some decision rule to make a choice.
Different assumptions about the decision rule lead to different families of models.
The most popular conjoint choice models are those based on a utility maximization assumption, which assumes that a respondent subconsciously maps the alternatives in the choice set into a utility space and then chooses whichever alternative has maximum utility (Gustafsson, Herrmann & Huber, 2000). These models include Luce’s choice models (Luce, 1959), and random utility models (Bierlaire, 1998; Manski, 1977; Thurstone, 1927). Luce’s choice models are based on Luce’s Choice Axiom (Luce, 1959) which implies a utility for every product alternative and assumes that choice probabilities are proportional to these utility values. In the random utility models, the utility has a systematic component and a random component. The systematic component is related to a set of fixed regression coefficients and the random component is represented by the alternative-specific error terms. The multinomial logit (MNL) model (McFadden, 1974) is derived with the assumption that the error terms of the utility function are independently and identically distributed as Gumbel random variables. The choice probabilities derived from the MNL model are the same as those derived from Luce’s Choice Axiom. The probit model (Thurstone,

<table>
<thead>
<tr>
<th></th>
<th>Credit card a₁</th>
<th>Credit card a₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest rate</td>
<td>17.99</td>
<td>14.99</td>
</tr>
<tr>
<td>Annual fee</td>
<td>$0</td>
<td>$30</td>
</tr>
<tr>
<td>Travel points</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cash reward</td>
<td>No</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 1.1: An example of conjoint choice set with credit cards
1927) assumes the errors terms of the utility function are independently and identically distributed normal random variables. Other random utility models include the generalized extreme value (GEV) model (McFadden, 1978) and the mixed logit model (Train, 2002), which place different distributional assumptions on the error terms.

These two types of models differ slightly in their assumptions on how the uncertainty of choice decisions is captured. Luce’s Choice Axiom assumes that the decision rule is stochastic in nature and even a complete knowledge of the problem would not overcome the uncertainty. Thus, each alternative is assigned a probability of being selected. The random utility models, on the other hand, assume that the decision-maker has a perfect discrimination capability. However, the researcher is supposed to have incomplete information and, therefore, uncertainty must be taken into account. Both models have their drawbacks. Luce’s choice model, or the MNL model, leads to a set of often unrealistic implications, such as Independence of Irrelevant Alternatives (IIA) and transitivity. Experimental choice data (Gigerenzer, Todd & The ABC Research Group, 1999; Huber, Payne & Puto, 1982; Payne, Bettman & Johnson, 1992; Tversky & Russo, 1969) often show violations of these implications. The random utility models, in order to avoid such restrictive assumptions, resort to amending the distributional assumptions of the error terms. For example, in the probit model and the nested logit model (Williams, 1977), correlated error terms are used to model dependence among alternatives. In the GEV model, the generalized extreme value distribution is placed on the error terms to avoid IIA. However, such formulation does not account for the characteristics of the alternatives and the decision maker. In most applications, the error terms are merely random residual effects and the random utility models do not lend themselves to a process interpretation of choice.
1.3 Research goals

Apart from the utility based models, there are psychological models attempting to represent the underlying human decision making process. Although the actual decision making process is still unknown, some of these models have empirical support. For example, Van Zandt, Colonius and Proctor (2000) evaluated the performance of Poisson race models empirically and showed that the Poisson race model can predict the RT distribution, mean RTs and accuracy of choices. These psychological models acknowledge that the human decision making process is intrinsically stochastic, which makes them more appealing than the random utility models which assume deterministic decision rules (Allenby, Otter & Van Zandt, 2006; Liu, 2006).

Compared with many other psychological models, the class of Poisson race models not only has empirical support, but also has greater tractability and ease of interpretation. However, current understanding of the Poisson race models is still very limited. There is a lack of systematic development of the theories for this family of models. Also, there very limited understanding of the underlying behavioral implications of the models. It is important to provide theoretical grounds for the practical application of this class of models.

Behavioral decision theory (e.g., Bouyssou & Pirlot, 2004; French, 1998) distinguishes between attribute-based and alternative-based processing. In the latter, preferences for attribute levels are integrated to overall preferences, independently for each alternative. The overall preferences for alternatives are the basis for choice. In attribute-based processing, the chosen alternative is identified by comparing levels only within attributes; the integration of preferences for attribute levels across attributes is not required for choice. Attribute-based processing implies extreme choice
probabilities in the case of dominance, even if alternative-based processing does not result in an overwhelming advantage of one of the alternatives. Strictly attribute-based processing, by definition, is incapable of capturing compensatory trade-offs between attribute levels. Strictly alternative-based processing contradicts the familiar observation that decision makers repeatedly shift their attention back and forth between alternatives before a choice is observed. In particular, alternative-based processing is not suited to handling dominance relationships among alternatives.

Traditional Poisson race models assume independence among the Poisson processes. This assumption is analog to the alternative-based processing and ignores the dependence of alternatives through the attributes. To handle dominance relationship, a new class of models based on attribute-based processing is proposed. However, a model that can based on strictly attribute-based processing lacks of real use since most choice sets are trade-off sets. The proposed model has attribute (level)-based stochastic components and a mechanism to integrate the interpretation and evaluation of attribute levels within alternatives. As such, the model is capable of reproducing important stylized facts associated with strictly attribute-based processing. Specifically, the model implies extreme choice probabilities in the case of dominance and dependence among alternatives that share the same or similar levels of attributes. At the same time, compensatory trade-off relationships are well represented.

In summary, the research described in this dissertation is an attempt to develop models with good predictive performance for conjoint choice data based on psychological models of human choice behavior. The objectives of this dissertation are two-fold. First, to study the Poisson race model from a theoretical point of view and to provide a good summary of the properties of the independent Poisson race models, especially
those properties that have not been discussed in the literature. Second, to develop a class of new models within the framework of the Poisson race models that allow us to incorporate dependence among alternatives. The predictive performance of such “dependent models” is assessed via comparison with other models. Additionally, extension of the new models to the case of choice sets consisting of more than two alternatives is discussed.

1.4 Organization of this dissertation

The rest of this dissertation is organized as follows. Chapter 2 derives expressions for choice probabilities under the Poisson race model and investigates their properties, including monotonicity, violations of the assumption of Independence from Irrelevant Alternatives (IIA), and transitivity. The effect of threshold values in Poisson race models is also discussed. In Chapter 3, a dependent Poisson race model is proposed for binary choice sets. The expression of choice probabilities and the formulation of rates are discussed. Then, the performance of the model is evaluated via a simulation study and through analysis of a conjoint choice experiment. In chapter 4, the dependent Poisson race model is extended to choice sets with more than two alternatives. Chapter 5 presents the conclusion of this dissertation and suggests directions for future work.
CHAPTER 2

CHOICE PROBABILITIES OF INDEPENDENT POISSON RACE MODELS AND RELATED PROPERTIES

2.1 Introduction

In the conjoint choice context, marketing researchers are interested in estimating the preferences of consumers. Thus, the choice probability of an alternative is of particular interest. Behavior psychologists are more interested in RT distributions. Yet, under the Poisson race model, the choice probabilities are closely tied to the RT distribution as they are both functions of the Poisson rate parameters and threshold values. This research work mainly discusses choice probabilities, but the response time distributions are also discussed in detail whenever deemed necessary to understand the behaviors of the choice probabilities.

Townsend and Ashby (Townsend & Ashby, 1983) considered a simple judgment task with two alternative responses. Each of the two alternative responses is assumed to be associated with a Poisson process with some rate and a fixed threshold. The rates of both Poisson processes are assumed to be constant over time and the thresholds are integers. They derived expressions for choice probabilities, conditioning density of RTs and mean RTs. Additionally, Townsend and Ashby (1983) proved
that a stochastic ordering of the two RT distributions implies an ordering of the choice probabilities. Smith and Van Zandt (2000) further generalized the Poisson race model in which the Poisson rate varies with time, that is, $\lambda_i = \lambda_i(t)$, where $\lambda_i$ is the Poisson rate associated with the $i$th alternative response and $t$ is time. They derived general expressions for the RT distribution for both binary and multiple alternative cases. They also obtained closed form expressions for RT distributions and choice probabilities under proportional-rates assumptions and integer threshold values. Under proportional-rates assumptions, $\lambda_i = c_i U(t)$, where $c_i$ are positive, real constants and $U(t)$ is a real function. In the time-homogeneous Poisson race model considered by Townsend and Ashby, $U(t) = t$. Stern (1987, 1980) discussed a family of generalized Poisson race models for paired comparisons, that is, the Poisson race model with a single fixed threshold value for the two alternatives considered. The models are “generalized” in the sense that the threshold values are not necessarily integers, although the information is accumulated in discrete chunks. With these assumptions, he derived expressions of choice probabilities and discussed stochastic transitivity of the choice probabilities. He also made a number of conjectures on the monotonicity of the choice probabilities.

The previous research on the theoretical aspects of the models has provided useful but still limited understanding of the family of Poisson race models. It has considered mainly the case with two alternatives. There are no unified expressions of choice probabilities for a choice set with two or more alternatives and for non-integer threshold values. Stern made some conjectures on the monotonicity of the choice probabilities without theoretical proof, partly due to the lack of closed form expressions of
the choice probabilities. In most of this work, little attention has been paid to the behavior implications of the family of Poisson race models.

In this chapter, an attempt is made to address these issues, mainly from theoretical point of view. First, a set of expressions for choice probabilities of the Poisson race model is derived under various assumptions. These expressions provide a means to calculate the choice probabilities easily. Second, a set of monotonicity results for the choice probabilities is proved. Additionally, the effect of thresholds on the choice probabilities and subsequently on empirical model fitting results is discussed. Lastly, the behavior implications of the models are investigated; specifically, whether the family of Poisson race models has the properties of IIA and transitivity and under what conditions such properties fall apart.

2.2 Expressions of choice probabilities

2.2.1 Notation and definition

In this dissertation, the Poisson race model with \( m \) parallel and independent Poisson processes is defined as the independent Poisson race model.

Under the independent Poisson race model, for a choice set \( A_m \) with \( m \) alternatives \( a_1, \ldots, a_m \), each alternative is associated with a Poisson process. These Poisson processes are represented by

\[
X_i(t) \sim \text{Poisson}(\lambda_i), \quad i = 1, \ldots, m,
\]

and \( X_1(t), \ldots, X_m(t) \) are independent. Also, a separate counter is assumed to be associated with each alternative to record the hits generated. Once the hits generated by \( X_i(t) \) on any counter exceeds its threshold value \( K_i \), the \( i \)th \((i = 1, \ldots, m)\) alternative is chosen as the response. Equivalently, define \( T_i = \text{inf}(X_i(t) \geq K_i) \) to be the
minimum time that the information in favor of alternative $a_i$ reaches its threshold, which is the response time (RT) for alternative $a_i$. The alternative with the minimum response time $T = \min(T_1, T_2, \ldots, T_m)$ is the chosen as the response. Since $T_i$ is the sum of $K_i$ exponentially distributed time intervals, it is known that (Karlin & Taylor, 1975)

$$T_i \sim \text{Gamma}(K_i, \lambda_i), \quad i = 1, \ldots, m$$

and that $T_1, \ldots, T_m$ are independent. Thus, the choice-making process becomes a comparison between a set of gamma random variables.

In this section, the various expressions of choice probabilities are derived within the context of a choice task in which a respondent is asked to select an alternative from a set of alternatives $A_m = \{a_1, a_2, \ldots, a_m\}$, where $m$ is the size of the choice set. Let $p(i|A_m)$ denote the choice probability of alternative $a_i$ in choice set $A_m$. The first interpretation motivates the derivation in terms of discrete distributions such as negative binomial, multinomial and binomial distributions when the thresholds $K_i, \ i = 1, \ldots, m$ are integers. In the second interpretation, $p(i|A_m)$ are derived in terms of gamma and beta distributions. These expressions for $p(i|A_m)$ are more general in the sense that $K_i, i = 1, \ldots, m$ are not necessarily integers. Without loss of generality, the probability of choosing the first alternative $p(1|A_m)$ is discussed.

### 2.2.2 Using the gamma distribution

**Proposition 2.2.1.** For any binary choice set $A_2$ with two alternatives $a_1$ and $a_2$, the probability of choosing alternative $a_1$ is

$$p(1|A_2) = E_{T_2}[F_1(t_2)],$$

where $F_1(t_2)$ is the cdf of gamma random variable $T_1$ at cut-off point $t_2$. 

13
**Proof.** From the above discussion, for any choice set \( A_2 = \{a_1, a_2\} \), the probability of choosing alternative \( a_1 \) from \( A_2 \) is the probability that the response time \( T_1 < T_2 \).

Let \( f_i(t) \) denote the pdf of \( T_i \) and

\[
f_i(t) = \frac{1}{\Gamma(K_i)} t^{K_i-1} \lambda_i^K_i \exp(-\lambda_i t), \quad t > 0.
\]

Then, the probability of choosing alternative \( a_1 \) is:

\[
p(1|A_2) = P(T_1 < T_2) = \int_0^\infty \int_0^{t_2} f(t_1, t_2) dt_1 dt_2
\]

\[
= \int_0^\infty f(t_2) \int_0^{t_2} f(t_1) dt_1 dt_2, \quad \text{by independence of } T_1 \text{ and } T_2
\]

\[
= \int_0^\infty f(t_2) F_1(t_2) dt_2, \quad F_1(t_2) \text{ is the cdf of } T_1 \sim \text{gamma}(K_1, \lambda_1)
\]

\[
= E_{T_2}[F_1(t_2)]. \quad \Box
\]

Similar to the binary case, the choice probabilities for a choice set with \( m > 2 \) alternatives is given in the following proposition.

**Proposition 2.2.2.** For a choice set \( A_m \) with \( m \) alternatives, the probability of choosing alternative \( a_1 \) is

\[
p(1|A_m) = E_Y[F_1(y)],
\]

where \( Y \) is the minimum order statistic of gamma random variables \( T_i, i = 2, \ldots, m \).

**Proof.** Similar to the binary case, for any choice set \( A_m = \{a_1, a_2, \ldots, a_m\} \), the probability of choosing alternative \( a_1 \) from the set \( A_m \) is the probability that the response
time $T_1$ is the minimum among all $T_i$, $i = 1, \ldots, m$.

$$p(1|A_m) = P(T_1 < T_2, T_1 < T_3, \ldots, T_1 < T_m)$$

$$= P[T_1 < \min(T_2, T_3, \ldots, T_m)]$$

$$= \int_0^\infty \int_0^y f(t_1, y) dt_1 dy, \quad Y = \min(T_2, T_3, \ldots, T_k) \text{ with pdf } f_Y(y)$$

$$= \int_0^\infty f_Y(y) \int_0^y f(t_1) dt_1 dy, \quad \text{by independence of } T_1 \text{ and } Y$$

$$= \int_0^\infty f_Y(y) F_1(y) dy, \quad F_1(y) \text{ is the cdf of } T_1$$

$$= E_Y[F_1(y)].$$

The distribution of the minimum order statistic is derived as below:

$$P(Y > y) = P(T_2 > y, T_3 > y, \ldots, T_m > y)$$

$$= \prod_{i=2}^{m} [1 - F_i(y)] \quad \text{and}$$

$$f_Y(y) = \frac{d}{dy} [1 - P(Y > y)] = \sum_{i=2}^{m} f_i(y) \prod_{j=2,j\neq i}^{m} [1 - F_j(y)], \quad y > 0. \quad \square$$

There is usually no closed form expression for the cdf and the minimum order statistic of the gamma distributions. Therefore, $p(1|A_m)$ is usually evaluated by simulation. One exception is the case that $K_1 = K_2 = \cdots = K_m = 1$. In this case, $T_i \sim \exp(\lambda_i), \ i = 1, \ldots, m$.

**Proposition 2.2.3.** For any choice set $A_m = \{a_1, a_2, \ldots, a_m\}$, when $K_1 = K_2 = \cdots = K_m = 1$, the probability of choosing $a_1$ is

$$p(1|A_m) = \frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i}.$$

**Proof.** When $K_1 = K_2 = \cdots = K_m = 1$, the pdf and cdf of the exponential distributions are:

$$f_i(t) = \lambda_i e^{-\lambda_i t}, \quad t > 0,$$
\[ F_i(t) = \int_0^t \lambda_i e^{-\lambda_i t} dt = 1 - e^{-\lambda_i t}, \quad t > 0. \]

The distribution of the minimum order statistic \( Y = \min(T_2, ..., T_m) \) is given by

\[
f_Y(y) = \sum_{i=2}^m f_i(y) \prod_{j=2, j \neq i}^m [1 - F_j(y)]
= \lambda_i e^{-\lambda_i y} \prod_{j=2, j \neq i}^m e^{-\lambda_j y} \\
= \sum_{i=2}^m \lambda_i e^{-\sum_{j=2}^m \lambda_j y}, \quad y > 0.
\]

The probability of choosing the alternative \( a_1 \) from the choice set \( A_m \) is then

\[
p(1|A_m) = E_Y[F_1(y)] \\
= \int_0^\infty (1 - e^{-\lambda_1 y}) \sum_{i=2}^m \lambda_i e^{-\sum_{j=2}^m \lambda_j y} dy \\
= 1 - \sum_{i=2}^m \frac{\lambda_i}{\sum_{i=1}^m \lambda_i} \\
= \frac{\lambda_1}{\sum_{i=1}^m \lambda_i}. \quad \Box
\]

The choice probability is the same as that derived based on the multinomial logit model. It implies that the Poisson race model includes the multinomial logit model as a special case.

2.2.3 Using the beta distribution

Another derivation of the choice probabilities is based on the relationship between beta and gamma random variables, which is given by the following two lemmas.

Lemma 2.2.4. For any random variable \( T_i \sim \text{gamma}(K_i, \lambda_i) \), the transformation \( \lambda_i T_i \) is distributed as \( \text{gamma}(K_i, 1) \).
Proof. Let \( U_i = \lambda_i T_i \). The Jacobian is \( \lambda_i^{-1} \). Then, the distribution of \( U_i \) is

\[
f_{U_i}(u) = \frac{\lambda_i^{K_i}(\frac{u}{\lambda_i})^{K_i-1}e^{-u/\lambda_i}}{\Gamma(K_i)} \frac{1}{\lambda_i} = \frac{u^{K_i-1}e^{-u}}{\Gamma(K_i)}, \quad u > 0. \quad \Box
\]

**Lemma 2.2.5.** For any two independent gamma random variables \( U_1 \) and \( U_2 \) with the same rate parameter \( \lambda_1 = \lambda_2 = 1 \), the transformation \( \frac{U_1}{U_1+U_2} \) is distributed as \( beta(K_1, K_2) \).

Proof. Let \( U = U_1 + U_2, V = \frac{U_1}{U_1+U_2} \). Then, \( U_1 = UV \) and \( U_2 = U - UV \). The Jacobian is given by \( J = U \). Then, the joint distribution of \( U \) and \( V \) is

\[
f_{U,V}(u, v) = \frac{(uv)^{K_1-1}e^{-uv} (u-uv)^{K_2-1}e^{-(u-uv)}}{\Gamma(K_1) \Gamma(K_2)} u
\]

where \( u > 0 \) and \( 0 < v < 1 \). The marginal distribution of \( V = \frac{U_1}{U_1+U_2} \) is given by

\[
f_V(v) = \int_0^\infty f_{U,V}(u,v) du
\]

\[
= \int_0^\infty \frac{\Gamma(K_1 + K_2)u^{K_1-1}(1-v)^{K_2-1} u^{K_1+K_2-1}e^{-u}}{\Gamma(K_1) \Gamma(K_2)} du
\]

\[
= \frac{\Gamma(K_1 + K_2)v^{K_1-1}(1-v)^{K_2-1}}{\Gamma(K_1) \Gamma(K_2)}, \quad 0 < v < 1
\]

\[
\sim beta(K_1, K_2). \quad \Box
\]

Based on these two lemmas, we can express the choice probabilities in terms of beta distributions.

**Proposition 2.2.6.** For any binary choice set \( A_2 = \{a_1, a_2\} \), under the independent Poisson race model, with rate parameter \( \lambda_1 \) and \( \lambda_2 \), and threshold values \( K_1 \) and \( K_2 \), the probability of choosing alternative \( a_1 \) is

\[
P(1|A_2) = \text{betacdf}(\frac{\lambda_1}{\lambda_1 + \lambda_2}, K_1, K_2),
\]

17
where \( \frac{\lambda_1}{\lambda_1 + \lambda_2} \) is the cut-off point for the cdf of beta\((K_1, K_2)\).

**Proof.** With Lemma 2.2.4 and Lemma 2.2.5, the following is true for any two independent gamma random variables \( T_1 \) and \( T_2 \):

\[
p(1|A_2) = P(T_1 < T_2) = P\left( \frac{\lambda_1 T_1}{\lambda_1 T_1 + \lambda_2 T_2} < \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) = P(V < \frac{\lambda_1}{\lambda_1 + \lambda_2}) = \text{betacdf}\left( \frac{\lambda_1}{\lambda_1 + \lambda_2}, K_1, K_2 \right).
\]

Notice that the choice probability depends on the ratio of the rates only. Similarly, we can express the choice probability for the case \( m > 2 \).

**Proposition 2.2.7.** For any choice set \( A_m = \{a_1, a_2, \ldots, a_m\} \), the probability of choosing alternative \( a_1 \) is given by

\[
p(1|A_m) = \int_0^{\frac{\lambda_1}{\lambda_1 + \lambda_2}} \cdots \int_0^{\frac{\lambda_1}{\lambda_1 + \lambda_m}} f_{V_2, \ldots, V_m}(v_1, \ldots, v_m) dv_2 \cdots dv_m,
\]

where \( f_{V_2, \ldots, V_m}(v_2, \ldots, v_m) \) is the joint distribution of beta random variables \( V_1, \ldots, V_m \) and \( V_i \sim \text{beta}(K_i, K_i) \) for \( i = 2, \ldots, m \).

**Proof.**

\[
p(1|A_m) = P(T_1 < T_2, T_1 < T_3, \ldots, T_1 < T_m) = P(V_2 < \frac{\lambda_1}{\lambda_1 + \lambda_2}, \ldots, V_m < \frac{\lambda_1}{\lambda_1 + \lambda_m}) = \int_0^{\frac{\lambda_1}{\lambda_1 + \lambda_2}} \cdots \int_0^{\frac{\lambda_1}{\lambda_1 + \lambda_m}} f_{V_2, \ldots, V_m}(v_1, \ldots, v_m) dv_2 \cdots dv_m.
\]

Similarly, the choice probability depends on the ratio of the rates only. There is usually no closed form expression. However, for the special case that \( T_i, i = 2, \ldots, m \)
are all exponential random variables (so that $K_1 = \cdots = K_m = 1$), then, $Y = \min(T_2, \ldots, T_m)$ is distributed as gamma($1, \sum_{i=2}^{m} \lambda_i$). Then, $p(1|A_m)$ is given by

$$p(1|A_m) = \text{betacdf}(\frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i}, K_1, 1); \quad (2.2.1)$$

### 2.2.4 Using negative binomial and multinomial distributions

**Proposition 2.2.8.** For any binary set $A_2 = \{a_1, a_2\}$, under the independent Poisson race model defined in Section 2.2.1,

1. **When $K_1$ is an integer,**

$$p(1|A_2) = 1 - \sum_{i=0}^{K_1-1} \frac{\Gamma(i+K_2)}{\Gamma(i+1)\Gamma(K_2)} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{K_2} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{i}.$$  \quad (2.2.2)

2. **When both $K_1$ and $K_2$ are integers,**

$$p(1|A_2) = \sum_{i=0}^{K_2-1} \binom{K_1 + i - 1}{i} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{i} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{K_1}.$$  \quad (2.2.3)

**Proof.** Under the independent Poisson race model, $T_1 \sim \text{gamma}(K_1, \lambda_1)$ and $T_2 \sim \text{gamma}(K_2, \lambda_2)$. From Proposition 2.2.1,

$$p(1|A_2) = E_{T_2}[F_1(t_2)].$$

When $K_1$ is an integer, using the relationship between the gamma cdf and the Poisson distribution, we have

$$F_1(t_2) = 1 - \sum_{i=0}^{K_1-1} \frac{(t_2\lambda_1)^i e^{(-t_2\lambda_1)}}{i!},$$

where $F_1(t_2)$ is the cdf of $T_1$ at cut-off point $t_2$. Then, take expectation with respect to the distribution of $T_2$, the first part is proved.
When both $K_1$ and $K_2$ are integers, the gamma functions can be written out as factorials. Then, from (2.2.2),

$$p(1|A_2) = 1 - \sum_{i=0}^{K_1-1} \binom{K_2 + i - 1}{i} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{K_2} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^i$$

$$= 1 - p(2|A_2)$$

because $p(1|A_2) + p(2|A_2) = 1$. By symmetry of the expressions for $p(1|A_2)$ and $p(2|A_2)$, we have

$$p(1|A_2) = \sum_{i=0}^{K_2-1} \binom{K_1 + i - 1}{i} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^i \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{K_1} . \quad \Box$$

The expression in (2.2.3) when both $K_1$ and $K_2$ are integers is also derived in Townsend and Ashby (1983) with a different approach. To interpret the expression, let $(K_2 + K_1 - 1)$ be the number of successes (information chunks) resulting from both Poisson processes, where the two rates are $(\lambda_1)$ and $(\lambda_2)$. Suppose the $(K_2 + K_1 - 1)$th success is from $\text{Poisson}(\lambda_1)$. Let $i$ be the number of successes resulting from $\text{Poisson}(\lambda_2)$. Then, $i$ has a negative binomial distribution $(i+K_1-1, \frac{\lambda_2}{\lambda_1+\lambda_2})$. $i$ has to be less than $K_2$ when process 1 reaches its threshold $K_1$. Therefore, the probability of choosing $a_1$ is equivalent to the probability that there are at most $K_2 - 1$ successes from $\text{Poisson}(\lambda_2)$.

The expression when $K_2$ is not an integer can be interpreted in similar fashion. In this case, the information need not be in discrete chunks, but the process moves one step when certain amount of information has accumulated. For $a_2$ to be chosen, $K_2$ is the amount of information required on the counter for $a_2$ while at most $K_1 - 1$ successes can accumulate on the counter for $a_1$. 
When \( m > 2 \), alternative \( a_1 \) is chosen if its counter has accumulated \( K_1 \) counts while counter \( i, i = 2, \ldots, m \), have accumulated at most \( K_i - 1 \) counts. By a straightforward extension of the expression in the binary case, we can write \( p(1|A_m) \) for the case of \( m > 2 \), and all \( K_i, i = 1, \ldots, m \) are integers.

**Proposition 2.2.9.** For any choice set, \( A_m \), with \( m > 2 \) alternatives, assume that all \( K_i, i = 1, \ldots, m \) are integers. The probability of choosing alternative \( a_1 \) is given by

\[
p(1|A_m) = \sum_{k_2}^{K_2-1} \sum_{k_3}^{K_3-1} \cdots \sum_{k_m}^{K_m-1} \frac{(K_1 + k_2 + \cdots + k_m - 1)!}{(K_1 - 1)! k_2 \cdots k_m!} \frac{\lambda_1^{K_1} \lambda_2^{k_2} \cdots \lambda_m^{k_m}}{(\sum_{i=1}^{m} \lambda_i)^{K_1 + k_2 + \cdots + k_m}}.
\]

This expression is also given in Smith and Van Zandt (2000), which is based on the so called negative multinomial distribution.

### 2.2.5 Using binomial formulation for \( m = 2 \)

**Proposition 2.2.10.** For any binary choice set \( A_2 = \{a_1, a_2\} \), assume \( K_1 \) and \( K_2 \) are integers, the probability of choosing alternative \( a_1 \) is given by

\[
p(1|A_2) = 1 - \sum_{i=0}^{K_1-1} \binom{K_2 + K_1 - 1}{i} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^i \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{K_2 + K_1 - 1 - i}.
\]

**Proof.** Let \( K_2 + K_1 - 1 \) be the total number of successes resulting from Poisson(\( \lambda_1 \)) or Poisson(\( \lambda_2 \)). Let \( X \) be the number of successes resulting from Poisson(\( \lambda_1 \)). Then, \( X \) has a binomial distribution \( (K_2 + K_1 - 1, \frac{\lambda_1}{\lambda_1 + \lambda_2}) \). The probability of choosing alternative \( a_1 \) is equivalent to at least \( K_1 \) successes from Poisson(\( \lambda_1 \)). Therefore,

\[
p(1|A_2) = P(X \geq K_1)
\]

\[
= 1 - P(X \leq K_1 - 1)
\]

\[
= 1 - \sum_{i=0}^{K_1-1} \binom{K_2 + K_1 - 1}{i} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^i \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{K_2 + K_1 - 1 - i}. \quad \square
\]
2.2.6 Summary

These expressions can serve different purposes. The derivation based on the gamma distribution has a nice interpretation based on the response time and does not restrict the threshold values to be integers. However, the computation of the choice probabilities requires the simulation of the density of the minimum order statistic. The expression based on the beta distribution can also be applied to general cases, and the computation for the binary case is fast as the cdf of the beta distribution is a built-in function in much statistical software. But simulation is also required for the case $m > 2$. The expressions based on Poisson and Binomial distributions are restricted as some or all of the $K_i$, $(i = 1, \ldots, m)$ values have to be integers, but the choice probabilities can be computed accurately and easily for $m > 2$.

2.3 Properties of choice probabilities

In the following section, without loss of generality, the behavior of the probability of choosing alternative $a_1$ from a choice set $A_m$ is discussed. First, a set of monotonicity results are discussed, mainly for the binary choice sets. Then, the limiting behaviors of the choice probabilities are also discussed. Again, alternative $a_i$ is associated with a Poisson process with rate $\lambda_i$ and a threshold value $K_i$.

2.3.1 Monotonicity

**Proposition 2.3.1.** For fixed $\lambda_1$ and $\lambda_2$, the probability of choosing alternative $a_1$, denoted by $p(1|A_2)$, is an increasing function of $K_2$ and a decreasing function of $K_1$. 

22
Proof. To prove the proposition, recall the expression of \( p(1|A_m) \) in Proposition 2.2.6, that is,

\[
p(1|A_2) = betacdf\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}, K_1, K_2\right) = P(V < \frac{\lambda_1}{\lambda_1 + \lambda_2}|K_1, K_2)
\]

where \( V \sim beta(K_1, K_2) \). When \( K_1 \) is fixed, the beta family \( f_V(v|K_1, K_2) \) has monotone likelihood ratio (MLR) because

\[
\frac{f_V(v|K_1', K_2')}{f_V(v|K_1, K_2)} \propto (1 - v)^{K_2' - K_2} \quad 0 < v < 1
\]

is a decreasing function in \( v \) for \( K_2' > K_2 \). As a result, the family of cdf is stochastically decreasing in \( K_2 \) (Casella & Berger, 2002). By the definition of stochastic ordering, \( P(V \leq v|K_1, K_2) \) is an increasing function of \( K_2 \) for any fixed value of \( v \). Since the cut-off point \( v = \frac{\lambda_1}{\lambda_1 + \lambda_2} \) is fixed and \( p(1|A_2) = P(V \leq v|K_1, K_2) \), \( p(1|A_2) \) is an increasing function of \( K_2 \).

Note \( p(1|A_2) = 1 - p(2|A_2) \). By symmetry, \( p(2|A_2) \) is decreasing in \( K_1 \) and \( p(1|A_2) \) is decreasing in \( K_1 \).

An alternative proof that is less involved is through the interpretation of the Poisson race models. For fixed \( \lambda_1 \) and \( \lambda_2 \), if \( K_1 \) is fixed and \( K_2 \) becomes larger, it will take longer for the second counter to reach the threshold and its corresponding alternative has less chance being chosen and the probability of choosing alternative \( a_1 \) increases. Similarly, if \( K_2 \) is fixed but \( K_1 \) gets larger, \( p(1|A_2) \) will decrease. \( \square \)

Figure 2.1 gives an example of \( p(1|A_2) \) as a function of \( K_1 \) and \( K_2 \).

The result in Proposition 2.3.1 can be easily extended to the case \( m > 2 \). \( p(1|A_m) \) is decreasing in \( K_1 \) and increasing in \( K_i, \ i = 1, \ldots, m \) when \( \lambda_i, \ i = 1, \ldots, m \) is fixed.
Figure 2.1: The probability of choosing alternative $a_1 \ (p(1|A_2))$ for different $K_1$ and $K_2$ values. $\lambda_1$ and $\lambda_2$ are fixed while $K_1$ and $K_2$ can vary freely. The vertical axis is $p(1|A_2)$ and the horizontal axis is $K_2$. The curves correspond to different $K_1$ values. It is an increasing function of $K_2$ and a decreasing function of $K_1$ for fixed $\lambda_1$ and $\lambda_2$. 
Proposition 2.3.2. For fixed $K_1$ and $K_2$, $p(1|A_2)$ is increasing in $\lambda_1$ and decreasing in $\lambda_2$.

Proof. Since the threshold values $K_1$ and $K_2$ are fixed, the greater the rate of generating hits in favor of one alternative, the larger the probabilities that this process will reach the threshold value first. Alternatively, when $K_1$ and $K_2$ are fixed, $p(1|A_2)$ depends only on the cut off point $v = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. When $\lambda_1$ increases, $v$ increases. Thus, $p(1|A_2)$ increases. Similarly, the increase in $\lambda_2$ decreases $v$ and $p(1|A_2)$ decreases. □

Notice that if both $\lambda_1$ and $\lambda_2$ are multiplied by the same constant, $p(1|A_2)$ will not change. The same result can be extended to the case $m > 2$ as well. From Proposition 2.2.1, if the rate for one alternative increases, the choice probability of this alternative also increases.

Proposition 2.3.3. For $K_1 = K_2 = K$, $p(1|A_2)$ is increasing in $K$ for $\lambda_1 > \lambda_2$ and decreasing in $K$ for $\lambda_1 < \lambda_2$. When $\lambda_1 = \lambda_2$, $p(1|A_2) = 0.5$.

Proof. When $K_1 = K_2 = K$, the beta distribution $f(x|K, K)$ is symmetric around the point $x = 0.5$. It is easy to show that such a beta family has MLR for $x > \frac{1}{2}$ and $x < \frac{1}{2}$. When $x > \frac{1}{2}$, the family of cdf is stochastically decreasing in $K$ and thus $p(1|A_2)$ is increasing in $K$. On the other hand, when $x < \frac{1}{2}$, the family of cdf is stochastically increasing in $K$ and thus $p(1|A_2)$ is decreasing in $K$. Notice that, let $x = \frac{\lambda_1}{\lambda_1 + \lambda_2}$, then $x > \frac{1}{2}$ implies that $\lambda_1 > \lambda_2$ and $x < \frac{1}{2}$ implies that $\lambda_1 < \lambda_2$. When $\lambda_1 = \lambda_2$, $p(1|A_2) = 0.5$ since $a_1$ and $a_2$ have the same thresholds and rates. □

This result is similar to a conjecture made by Stern(1983). Stern did not prove the result theoretically. Instead, with the recurrence formula for $p(1|A_2)$ under the
independent Poisson race model with \( K_1 = K_2 = \cdots = K_m = K \), Stern reaches the same result with numerical confirmation.

Proposition 2.3.3 can be interpreted in this way: when \( K \) increases, the process associated with the larger rate is more likely to reach the threshold first and its choice probabilities increases. In fact, as \( K \) increases, the response time \( T_i, i = 1, 2 \) approaches normal distribution as \( T_i \) is the sum of \( K \) exponential random variables. When \( K \to \infty \), the distribution of \( T_i \) is degenerate to its mean \( K/\lambda_i \). In this case, \( T_1 < T_2 \) if \( \lambda_1 > \lambda_2 \) and \( a_1 \) will be chosen with certainty. In general, if \( K_1 = K_2 = \cdots = K_m = K \), when \( K \) increases, the process with the largest rate is most likely to reach the threshold first and so the choice probability for the corresponding alternative is also the largest.

Another special case is to assume \( T_1 \) and \( T_2 \) have the same mean, that is, \( K_1/\lambda_1 = K_2/\lambda_2 \). The choice probabilities in this special case have not been studied before and it has shown to have some interesting properties. Proposition 2.3.4 is a result for the monotonicity of \( p(1|A_2) \) in this special case.

**Proposition 2.3.4.** When \( \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} \) and \( K_1 \) is a fixed integer \((< 90)\), \( p(1|A_2) \) is an increasing function of \( K_2 \).

This proposition is proved for \( K_1 \) up to 90. For \( K_1 > 90 \), the same proof can be applied but it was not checked due to the limitation of computation power we have.

For the proof of Proposition 2.3.4, the following lemma is needed.

**Lemma 2.3.5.** For any integers \( K_1 \) and \( K_2 \) and \( i \geq 0 \), the expression of \( \frac{(K_2+K_1-1)!}{(K_2+K_2-1-i)!} \) can be written as

\[
\frac{(K_2 + K_1 - 1)!}{(K_2 + K_2 - 1 - i)!} = \sum_{j=0}^{i} c_{j}^{(i)} b K_2^j
\]
where \( b = K_1 - i \) and \( c_{j,b}^{(i)} \) is the coefficient of \( K_2^j \) and does not depend on \( K_2 \). The coefficient is defined through a recursive relationship:

\[
c_{j,b}^{(q)} = c_{j-1,b}^{(q-1)} + (b + q - 1)c_{j,b}^{(q-1)} \quad \text{for any integer } q \geq 0
\]

\[ (2.3.1) \]

with

\[
c_{0,b}^{(1)} = b,
\]

\[
c_{j,b}^{(q)} = 1 \quad \text{if } j = q,
\]

\[
c_{j,b}^{(q)} = 0 \quad \text{if } j < 0 \text{ or } j > q.
\]

This lemma is proved by induction method and the details are given in Appendix A.

Proof of Proposition 2.3.4. Case 1: When \( K_1 = 1 \), \( T_1 \sim \text{exponential}(\lambda_1) \). Also, \( T_2 \sim \text{gamma}(K_2, \lambda_2) \). From Proposition 2.2.1,

\[
p(1|A_2) = E_{T_2}[F_1(t_2)]
\]

\[
= \int_0^\infty (1 - e^{-t_2/\lambda_1}) f_{T_2}(t_2) dt_2
\]

\[
= 1 - \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{K_2}
\]

\[
= 1 - \left( \frac{K_2}{1 + K_2} \right)^{K_2} \quad \text{since } \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} \text{ and } K_1 = 1.
\]

(2.3.2)

Then,

\[
p(2|A_2) = 1 - p(1|A) = \left( \frac{K_2}{1 + K_2} \right)^{K_2}.
\]

Take the first and second derivative of \( p(2|A_2) \) with respect to \( K_2 \), we have

\[
\frac{dp(2|A_2)}{dK_2} = ln \frac{K_2}{K_2 + 1} + \frac{1}{K_2 + 1}
\]
and
\[
\frac{d^2 p(1|A_2)}{dK_2^2} = -\frac{1}{K_2 + 1} + \frac{1}{K_2} \left( \frac{1}{K_2 + 1} - \frac{1}{K_2} \right) = \frac{1}{K_2(K_2 + 1)^2}
\]
(2.3.3)

Since \( \frac{d^2 p(1|A_2)}{dK_2^2} > 0 \), \( \frac{dp(2|A_2)}{dK_2} \) is increasing in \( K_2 \). Also, \( \lim_{K_2 \to \infty} \frac{dp(2|A_2)}{dK_2} = 0 \) so that \( \frac{dp(2|A_2)}{dK_2} \) must be smaller than 0. Therefore, \( p(2|A) \) is decreasing in \( K_2 \) and \( p(1|A_2) = 1 - p(2|A_2) \) is increasing in \( K_2 \).

Case 2: When \( K_1 > 1 \) and both \( K_1 \) and \( K_2 \) are integers.

When both \( K_1 \) and \( K_2 \) are integers and \( \frac{K_1}{A_1} = \frac{K_2}{A_2} \), \( p(1|A) \) can be written in term of the binomial distribution by Proposition 2.2.10.

\[
p(1|A_2) = 1 - \sum_{i=0}^{K_1-1} \left( \begin{array}{c} K_2 + K_1 - 1 \\ i \end{array} \right) \left( \frac{K_1}{K_2 + K_1} \right)^i \left( \frac{K_2}{K_1 + K_2} \right)^{K_2 + K_1 - 1 - i}
\]

\[
= 1 - \left( \frac{K_2}{K_1 + K_2} \right)^{K_2 + K_1 - 1} \sum_{i=0}^{K_1-1} \left( \begin{array}{c} K_2 + K_1 - 1 \\ i \end{array} \right) \left( \frac{K_1}{K_2} \right)^i
\]

\[
= 1 - \left( \frac{K_2}{K_1 + K_2} \right)^{K_2 + K_1 - 1} \frac{1}{K_2^{K_1-1}} \sum_{i=0}^{K_1-1} \frac{(K_2 + K_1 - 1)!}{(K_2 + K_1 - 1 - i)!} K_1^{K_2 - 1 - i} K_i^{K_1}
\]

\[
= 1 - \frac{K_2^{K_2}}{(K_1 + K_2)^{K_2 + K_1 - 1}} B
\]
(2.3.4)

where
\[
B = \sum_{i=0}^{K_1-1} \frac{(K_2 + K_1 - 1)!}{(K_2 + K_1 - 1 - i)!} K_2^{K_1-1 - i} K_i^i
\]
is a polynomial of \( K_2 \).

From (2.3.4), it is easy to see that
\[
p(2|A_2) = \frac{K_2^{K_2}}{(K_1 + K_2)^{K_2 + K_1 - 1}} B.
\]

Take the logarithm of \( p(2|A_2) \), we have
\[
ln[p(2|A_2)] = K_2 lnK_2 - (K_2 + K_1 - 1)ln(K_2 + K_1) + lnB.
\]
Now, the first derivative of $\ln p(2|A_2)$ with respect to $K_2$ is
\[
\frac{d\ln[p(2|A_2)]}{dK_2} = \ln \frac{K_2}{K_2 + K_1} + \frac{1}{K_1 + K_2} + 1 \frac{dB}{B dK_2}.
\]

When $K_2 \to \infty$, the first derivative goes to 0 since the limit of the three terms in the first derivative all go to 0. The limits of the first two terms are obvious. For the third term, since $B$ is a polynomial of $K_2$, $\frac{dB}{dK_2}$ is also a polynomial of $K_2$ of lower order. Therefore, as $K_2 \to \infty$, $B \to \infty$, the third term $\frac{1}{B} \frac{dB}{dK_2} \to 0$.

The second derivative of $\ln p(2|A_2)$ is given by
\[
\frac{d^2\ln[p(2|A_2)]}{dK_2^2} = \frac{K_2^2 + K_2(K_1 - 1)}{K_2(K_1 + K_2)^2} + \frac{B \frac{d^2B}{dK_2^2} - \left( \frac{dB}{dK_2} \right)^2}{B^2}
\]
\[
= \frac{B^2(K_1^2 + K_2(K_1 - 1)) + (B \frac{d^2B}{dK_2^2} - \left( \frac{dB}{dK_2} \right)^2)K_2(K_1 + K_2)^2}{B^2(K_1 + K_2)^2}
\]
\[
= \frac{D}{B^2(K_1 + K_2)^2},
\]
where $D$ represents the numerator in the above expression.

The next step is to show that $\frac{d\ln[p(2|A_2)]}{dK_2}$ is an increasing function of $K_2$, which is equivalent to show $\frac{d^2\ln[p(2|A_2)]}{dK_2^2} > 0$ and thus $D > 0$. From (2.3.5),
\[
D = B^2(K_1^2 + K_2(K_1 - 1)) + (B \frac{d^2B}{dK_2^2} - \left( \frac{dB}{dK_2} \right)^2)K_2(K_1 + K_2)^2.
\]

$D$ consists of two terms which are polynomial in $K_2$ since $B$ is polynomial of $K_2$. $B$ can in terms of $K_2$ using Lemma 2.3.5.

With Lemma 2.3.5, $B$ can be written as
\[
B = \sum_{i=0}^{K_1-1} \frac{(K_2 + K_1 - 1)!}{(K_2 + K_1 - 1 - i)!i!} K_2^{K_1-1-i} K_1^i
\]
\[
= \sum_{i=0}^{K_1-1} \sum_{j=0}^{i} c_j(i) b^j K_2^j K_2^{K_1-1-i} K_1^i
\]

29
Now collect all the terms with $K_i^j$, $i = 0, \ldots, K_1 - 1$, 

$$B = \sum_{i=0}^{K_1-1} K_2^i \sum_{j=K_1-1-i}^{K_1-1} \frac{K_1^j c_j^{(j)}}{j!} c_j^{(j-K_1-1-i)},$$

which gives

$$B = \sum_{i=0}^{K_1-1} K_2^i d_i,$$

where

$$d_i = \sum_{j=K_1-1-i}^{K_1-1} \frac{K_1^j c_j^{(j)}}{j!} c_j^{(j-K_1-1-i)},$$

are the coefficients for the polynomial in $K_2$.

With (2.3.6), take the first and second derivative of $B$ in terms of $K_2$, we have

$$\frac{dB}{dK_2} = \sum_{i=1}^{K_1-1} i d_i K_2^{i-1} = \sum_{i=1}^{K_1-1} e_i K_2^{i-1},$$

and

$$\frac{d^2 B}{dK_2^2} = \sum_{i=2}^{K_1-1} i(i-1) d_i K_2^{i-2} = \sum_{i=2}^{K_1-1} g_i K_2^{i-2},$$

where

$$e_i = id_i \text{ and } g_i = i(i-1)d_i.$$
Also,

\[
B^2 = \left( \sum_{i=0}^{K_1-1} d_i K_2^i \right)^2
= \sum_{i=1}^{2K_1-1} f_i K_2^{i-1},
\]

where

\[
f_i = \begin{cases} 
\sum_{j=1}^{i} d_{j-1} d_{i-j} & \text{for } i = 1, \ldots, K_1 \\
\sum_{j=i-K_1}^{i-1} d_j d_{i-j-1} & \text{for } i = K_1 + 1, \ldots, 2K_1 - 1
\end{cases}
\]

With these expressions, the two polynomial terms in \( D \) can be expanded. For the first term,

\[
B^2(K_1^2 + K_2(K_1 - 1)) = \sum_{i=1}^{2K_1-1} f_i K_2^{i-1}(K_1^2 + K_2(K_1 - 1))
= \sum_{i=0}^{2K_1-1} w_i K_2^i.
\]

where

\[
w_i = \begin{cases} 
f_1 K_1^2 & \text{for } i = 0 \\
f_{i+1} K_1^2 + f_i (K_1 - 1) & \text{for } i = 1, \ldots, 2K_1 - 2 \\
f_{2K_1-1}(K_1 - 1) & \text{for } i = 2K_1 - 1
\end{cases}
\]

For the second term,

\[
(B \frac{d^2 B}{dK_2^2} - (\frac{dB}{dK_2})^2)K_2(K_1 + K_2)^2 = \sum_{i=1}^{2K_1-3} v_i K_2^{i-1}(K_2^3 + 2K_2^2 K_1 + K_2 K_1^2)
= \sum_{i=0}^{2K_1-1} q_i K_2^i,
\]

where

\[
q_i = \begin{cases} 
0 & \text{for } i = 0 \\
v_1 K_1^2 & \text{for } i = 1 \\
2v_1 K_1 + v_2 K_1^2 & \text{for } i = 2 \\
v_{i-2} + 2v_{i-1} K_1 + v_i K_1^2 & \text{for } i = 3, \ldots, 2K_1 - 3 \\
v_{2K_1-4} + 2v_{2K_1-3} K_1 & \text{for } i = 2K_1 - 2 \\
v_{2K_1-3} & \text{for } i = 2K_1 - 1
\end{cases}
\]

The coefficients of \( D \) are then given by \( q_i + w_i \) for \( i = 0, \ldots, 2K_1 - 1 \). For a fixed \( K_1 \), both \( q_i \) and \( w_i \) can be computed. If \( q_i + w_i > 0 \) for all \( i \), \( D > 0 \) and by the
argument mentioned earlier, $p(1|A_2)$ is an increasing function of $K_2$. It is conjectured that $q_i + w_i > 0$ for all $K_1$. However, due to the limitation of computing power, only the cases for $K_1 < 90$ are verified via computation. For $K_1 \geq 90$, the magnitude of $q_i$ and $w_i$ increase drastically and cannot be obtained accurately. In this case, $p(1|A_2)$ can be shown as a increasing function of $K_2$ by computing the probability directly using the expression given in Proposition 2.2.3.

Case 3: $K_1$ is an integer but $K_2$ is not an integer.

From (2.2.2), we have

$$p(1|A_2) = 1 - \sum_{i=0}^{K_1-1} \frac{\Gamma(i + K_2)}{\Gamma(K_2)\Gamma(i + 1)} \left( \frac{K_2}{K_1 + K_2} \right)^{K_2} \left( \frac{K_1}{K_1 + K_2} \right)^i ,$$

$$= 1 - \frac{K_2^{K_2}}{(K_1 + K_2)^{K_1 + K_2 - 1}} \sum_{i=0}^{K_1-1} \frac{\Gamma(i + K_2)}{\Gamma(K_2)\Gamma(i + 1)} (K_1 + K_2)^{K_1-i-1} \frac{K_1^i}{i!} ,$$

$$= 1 - \frac{K_2^{K_2}}{(K_1 + K_2)^{K_1 + K_2 - 1}} B^* ,$$

where

$$B^* = \sum_{i=0}^{K_1-1} \frac{\Gamma(i + K_2)}{\Gamma(K_2)\Gamma(i + 1)} (K_1 + K_2)^{K_1-i-1} \frac{K_1^i}{i!} .$$

From the recursion formula of gamma function

$$\Gamma(x + 1) = x\Gamma(x) ,$$

we have

$$\frac{\Gamma(i + K_2)}{\Gamma(K_2)} = (K_2 + i - 1) \ldots (K_2 + 1)K_2$$

$$= \prod_{j=1}^{i} (K_2 + j - 1)$$

$$= \sum_{j=0}^{i} c_{j, b=0}^{(i)} K_2^j ,$$
where \( c^{(i)}_{j, b=0} \) is as defined in (2.3.1). Also,

\[
(K_1 + K_2)^{K_1-1-i} = \sum_{k=0}^{K_1-i-1} \binom{K_1-i-1}{k} K_1^{K_1-i-1-k} K_2^k = \sum_{k=0}^{K_1-i-1} h_k K_2^k ,
\]

where \( h_k = \binom{K_1-i-1}{k} K_1^{K_1-i-1-k} \). Then,

\[
B^* = \sum_{i=0}^{K_1-1} \frac{\Gamma(i + K_2)}{\Gamma(K_2) \Gamma(i + 1)} (K_1 + K_2)^{K_1-i-1} \frac{K_i}{i!} = \sum_{i=0}^{K_1-1} \sum_{j=0}^{K_2} (\sum_{j=0}^{K_2} c^{(i)}_{j, b=0} K_2^j)(\sum_{k=0}^{K_1-i-1} h_k K_2^k) \frac{K_i}{i!} = \sum_{i=0}^{K_1-1} d_i^* K_2^i .
\]

The coefficients \( d_i^* \) are obtained by collecting all the coefficients for each \( K_2^i, i = 0, \ldots, K_1 - 1 \). These coefficients depend on \( K_1 \) only. We can show that \( d_j^* = d_j \) in (2.3.6). Therefore, the same approach for integer values of \( K_2 \) could be extended to the case that \( K_2 \) is not integer.

Figure 2.2 depicts \( p(1|A_2) \) as a function of \( K_1 \) and \( K_2 \) for the case that \( K_1/\lambda_1 = K_2/\lambda_2 \). \( p(1|A_2) \) is an increasing function of \( K_2 \) and a decreasing function of \( K_1 \).
Figure 2.2: The choice probability $p(1|A_2)$ as a function of $K_1$ and $K_2$ when $K_1/\lambda_1 = K_2/\lambda_2$. The vertical axis is $p(1|A_2)$ and the horizontal axis is $K_2$. Different curves correspond to different $K_1$ values. $p(1|A_2)$ is an increasing function of $K_2$ and a decreasing function of $K_1$. 
Proposition 2.3.6. \( p(1|A_m) \) is a decreasing function of \( m \).

Proof. Under the Poisson race model, the probability of choosing alternative \( a_1 \) is the probability that its response time \( T_1 \) is smaller than those for other alternatives. That is, \( p(1|A_m) = p(T_1 < \tau|A_m) \), where \( \tau = \min(T_2, \ldots, T_m) \) is the minimum order statistic of \( m - 1 \) gamma random variables. \( \tau \) is independent of \( T_1 \). As \( m \) increases, \( \tau \) is a decreasing function. Therefore, \( p(T_1 < \tau|A_m) \) is decreasing in \( m \). \( \square \)

2.3.2 Limiting behaviors

Next, the limits of the choice probabilities for choice set \( A_2 \) when the threshold values go to zero or infinity are discussed.

Proposition 2.3.7. When \( \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} = \mu \), for fixed \( K_1 \),

(i) \( \lim_{K_2 \to \infty} p(1|A_2) = F_1(\mu) \), where \( F_1 \) is the cdf of \( T_1 \sim \text{gamma}(K_1, \lambda_1) \).

(ii) \( \lim_{K_2 \to 0} p(1|A_2) = 0 \).

Proof. When \( K_2 \to \infty \), the distribution of \( T_2 \sim \text{gamma}(K_2, \lambda_2) \) is degenerated to a point mass at \( \mu \). From Proposition 2.2.1, and let \( K_2 \to \infty \), then, \( \lim_{K_2 \to \infty} p(1|A_2) = \lim_{K_2 \to \infty} E_T[F_1(t)] = F_1(\mu) \). If \( p(1|A_2) = 0.5 \), \( \mu \) is the median of the distribution of \( T_1 \).

When \( K_2 \to 0 \), the distribution of \( T_2 \) degenerates to a point mass at 0.

\[
\lim_{K_2 \to 0} p(1|A_2) = \lim_{K_2 \to 0} E_T[F_1(t)] = F_1(0) = 0 .
\]

Proposition 2.3.8. When \( K_1 = K_2 = K \),

\[
\lim_{K \to 0} p(1|A_2) = \lim_{K \to \infty} p(1|A_2) = 0, \text{ if } \lambda_1 < \lambda_2 ,
\]
\[
\lim_{K \to 0} p(1|A_2) = \lim_{K \to \infty} p(1|A_2) = 1, \text{ if } \lambda_1 > \lambda_2.
\]

**Proof.** For \(K_1 = K_2 = K\), if \(\lambda_1 < \lambda_2\), \(T_1\) is stochastically greater than \(T_2\). When \(K \to 0\), the distribution of \(T_1\) and \(T_2\) both degenerate to point mass 0 and \(T_2\) approaches 0 faster than \(T_1\). Therefore, \(\lim_{K \to 0} F_2(t_1) = 1\) and by Proposition 2.2.1, \(\lim_{K \to 0} p(1|A_2) = \lim_{K \to 0} (1 - E_{T_1}[F_2(t_1)]) = 0\)

When \(K \to \infty\), both \(T_1\) and \(T_2\) are both degenerate normal located at their means. Since \(T_1\) is stochastically greater than \(T_2\), \(\lim_{K \to \infty} F_2(t_1) = 1\) and \(\lim_{K \to \infty} p(1|A_2) = 0\).

This coincides with the results of Stern(1987, 1990).

### 2.4 Comparisons of the general form of Poisson race model with the Gamma(\(K\)) model

Stern(1987, 1990) investigated the gamma permutation model which is the Poisson race model with the restriction that all of the shape parameters of the gamma distributions are equal, that is, \(K_1 = \ldots = K_m = K\), referred to as the Gamma(\(K\)) model hereafter. Stern used the Gamma(\(K\)) model to analyze sports data involving a set of 12 baseball teams. The teams appeared in sets of size 2, yielding, for a single piece of the experiment, a binary set \(A = \{i, j\}\), with \(i, j \in \{1, \ldots, 12\}\). Table 2.1 shows the percentage of games that team \(i\) (in the rows) defeats team \(j\) (in the column) over a season. The Gamma(\(K\)) model can be interpreted as follows: the scoring process of a team (winning over another team through the season) is assumed to follow a Poisson process. The team that first scores \(K\) points wins the competition. The time taken for the \(i\)th team to score \(K\) points follows a gamma distribution with
Table 2.1: Percentage of games in which the $i$th team (row) defeats the $j$th team (column).

<table>
<thead>
<tr>
<th>Team</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.44</td>
<td>0.67</td>
<td>0.56</td>
<td>0.67</td>
<td>0.94</td>
<td>0.58</td>
<td>0.67</td>
<td>0.58</td>
<td>0.83</td>
<td>0.75</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.56</td>
<td>0.33</td>
<td>0.56</td>
<td>0.47</td>
<td>0.61</td>
<td>0.50</td>
<td>0.42</td>
<td>0.75</td>
<td>0.50</td>
<td>0.58</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>0.67</td>
<td>0.50</td>
<td>0.41</td>
<td>0.61</td>
<td>0.42</td>
<td>0.42</td>
<td>0.58</td>
<td>0.58</td>
<td>0.33</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.44</td>
<td>0.44</td>
<td>0.50</td>
<td>0.56</td>
<td>0.61</td>
<td>0.33</td>
<td>0.42</td>
<td>0.42</td>
<td>0.33</td>
<td>0.58</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>0.53</td>
<td>0.59</td>
<td>0.44</td>
<td>0.39</td>
<td>0.33</td>
<td>0.42</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.06</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.61</td>
<td>0.39</td>
<td>0.33</td>
<td>0.67</td>
<td>0.33</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.42</td>
<td>0.50</td>
<td>0.58</td>
<td>0.67</td>
<td>0.67</td>
<td>0.50</td>
<td>0.78</td>
<td>0.50</td>
<td>0.56</td>
<td>0.56</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.33</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.83</td>
<td>0.22</td>
<td>0.50</td>
<td>0.50</td>
<td>0.56</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.42</td>
<td>0.25</td>
<td>0.42</td>
<td>0.58</td>
<td>0.50</td>
<td>0.67</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.56</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.17</td>
<td>0.50</td>
<td>0.42</td>
<td>0.67</td>
<td>0.50</td>
<td>0.33</td>
<td>0.44</td>
<td>0.50</td>
<td>0.44</td>
<td>0.44</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td>11</td>
<td>0.25</td>
<td>0.42</td>
<td>0.67</td>
<td>0.42</td>
<td>0.50</td>
<td>0.67</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.33</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.33</td>
<td>0.33</td>
<td>0.50</td>
<td>0.36</td>
<td>0.75</td>
<td>0.42</td>
<td>0.28</td>
<td>0.33</td>
<td>0.39</td>
<td>0.67</td>
<td>0.56</td>
<td></td>
</tr>
</tbody>
</table>

The shape parameter $K$ and scale parameter $\lambda_i$. Then, the comparison of any two teams becomes a comparison of their corresponding gamma random variables.

Starting from $K_i = 1$, $i = 1, \ldots, 12$, Stern computed the predicted winning percentage of each team and the goodness of fit test statistics for different value of $K_i$. It seems that for a wide range of $K$ (from 0.1 to 20), the statistics of Gamma($K$) model are not statistically different according to the likelihood ratio test defined later in Section 2.4.2. A simulation example indicates that a large sample size is required to differentiate between the fits of different Gamma($K$) models. Overall, the main conclusion is that the Gamma($K$) model is insensitive to different values of $K$.

The Gamma($K$) model investigated by Stern is only a special case of the Poisson race model. A natural question is whether the fit of the data would improve if the threshold values $K_i$, $i = 1, \ldots, m$ are not restricted to be the same values. The
answer to this question might help us understand the effect of the threshold values on model fitting. In the following discussion, we investigate this more general version of the gamma permutation model, denoted by $\text{Gamma}(K_1, \ldots, K_m)$ model with the same data set in Table 2.1.

2.4.1 Model estimation

For paired comparison data with $m = 12$ teams, with Stern’s notation, we assume that $n_{ij}$ is the number of times that team $i$ defeats team $j$ and $n_{ji}$ is the number of times that team $j$ defeats team $i$. Then, the total number of comparisons between $i, j$ pair is $N_{ij} = n_{ij} + n_{ji}$. Define vectors $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $K = (K_1, \ldots, K_m)$. Let $p_{ij}$ be the probability that team $i$ defeats team $j$, also called “the preference probability”. This is the choice probability that team $i$ is “chosen” as the winner from the binary set $A_2 = \{i, j\}$. Assuming no ties and independence of the results of each paired comparison, the likelihood function is the product of $m(m-1)/2$ binomial likelihoods:

$$L(n_{ij}, i \neq j; \lambda, K) = \prod_{i=1}^{m} \prod_{j>i}^{n_{ij}} \left( \frac{N_{ij}}{n_{ij}} \right)^{n_{ij}} (1 - p_{ij})^{n_{ji}},$$

and $p_{ij} = p(T_i < T_j)$, where $T_i \sim \text{gamma}(K_i, \lambda_i)$ and $T_j \sim \text{gamma}(K_j, \lambda_j)$. Then,

$$p_{ij} = \int_0^\infty \int_0^{t_j} \frac{\lambda_i^{K_i-1}t_i^{K_i-1}e^{-\lambda_i t_i} \lambda_j^{K_j-1}t_j^{K_j-1}e^{-\lambda_j t_j}}{\Gamma(K_i) \Gamma(K_j)} dt_1 dt_2.$$

Thus, the maximum likelihood estimate (MLE) of $K$ and $\lambda$ can be expressed as

$$(\hat{K}, \hat{\lambda}) = \text{argmax} \{ L(n_{ij}, i \neq j; \lambda, K) \}.$$

Since the likelihood function can not be differentiated analytically, it is difficult to find a closed form solution for $\lambda$ and $K$. Numerical methods can be used to find the MLEs. The maximum of likelihood function is the same as the minimum of
the negative of the log-likelihood function. Also, as seen in Proposition(2.2.6), $p_{ij}$ depends on the ratio of $\lambda_i/\lambda_j$ and will not change if both $\lambda_i$ and $\lambda_j$ are multiplied by the same constant. We add a constraint $\sum_1^m \lambda_i = 1$ to normalize $\Lambda$. A constrained nonlinear optimization problem is thus formulated:

$$
\begin{align*}
\min & -\ln L(n_{ij}, i \neq j; \lambda, K) \\
\text{s. t.} & \sum_1^m \lambda_i = 1, \\
& \lambda_1, \ldots, \lambda_m > \epsilon \\
& K_1, \ldots, K_m > \epsilon.
\end{align*}
$$

Ideally, $\epsilon$ is equal to 0 since the constraint $\lambda_i > 0$ and $K_i > 0$, $i = 1, \ldots, m$. However, for computational simplification, $\epsilon$ is pre-specified as a small positive constant. The optimization is carried out with Matlab. The program gives the same values of the MLEs for $\Lambda$ as Stern obtained in his example when $K = 1$ for all teams. For the Gamma($K_1, \ldots, K_{12}$) model, several starting points are used to test the stability of the MLEs. The results of the optimization are compared with the Gamma(1) model.

Table 2.2 shows the optimization solution produced by the Gamma(1) model and the Gamma($K_1, \ldots, K_{12}$) model. The mean and variance of the associated gamma random variables are given by $K_i/\lambda_i$ and $K_i/\lambda_i^2$, respectively. The predicted percentage of games won by team $i$ over team $j$ is given in Table 2.3 based on the estimates in Table 2.2. Table 2.4 listed the optimization solution under Gamma ($K_1, \ldots, K_{12}$) for ($\Lambda, K$) from two different starting points. It seems that different starting points produce different estimates for ($\Lambda, K$), but the difference in negative log-likelihood is very small ($< 0.1$). It is likely that the likelihood function is relatively flat near the maximum. Furthermore, although the solutions for ($\Lambda, K$) are not the same, the
<table>
<thead>
<tr>
<th>Team</th>
<th>( \lambda_i )</th>
<th>( K_i )</th>
<th>Mean</th>
<th>Variance</th>
<th>( \lambda_i )</th>
<th>( K_i )</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.151</td>
<td>1</td>
<td>6.623</td>
<td>43.858</td>
<td>0.119</td>
<td>3.814</td>
<td>32.130</td>
<td>270.680</td>
</tr>
<tr>
<td>2</td>
<td>0.091</td>
<td>1</td>
<td>10.989</td>
<td>120.758</td>
<td>0.000</td>
<td>0.126</td>
<td>1255.000</td>
<td>12550000</td>
</tr>
<tr>
<td>3</td>
<td>0.078</td>
<td>1</td>
<td>12.821</td>
<td>164.366</td>
<td>0.002</td>
<td>0.319</td>
<td>199.563</td>
<td>124726.563</td>
</tr>
<tr>
<td>4</td>
<td>0.076</td>
<td>1</td>
<td>13.158</td>
<td>173.130</td>
<td>0.004</td>
<td>0.453</td>
<td>125.694</td>
<td>34915.123</td>
</tr>
<tr>
<td>5</td>
<td>0.064</td>
<td>1</td>
<td>15.625</td>
<td>244.141</td>
<td>0.003</td>
<td>0.484</td>
<td>166.966</td>
<td>57574.316</td>
</tr>
<tr>
<td>6</td>
<td>0.054</td>
<td>1</td>
<td>18.519</td>
<td>342.936</td>
<td>0.659</td>
<td>41.505</td>
<td>62.944</td>
<td>95.456</td>
</tr>
<tr>
<td>7</td>
<td>0.113</td>
<td>1</td>
<td>8.850</td>
<td>78.315</td>
<td>0.002</td>
<td>0.260</td>
<td>123.857</td>
<td>58979.591</td>
</tr>
<tr>
<td>8</td>
<td>0.091</td>
<td>1</td>
<td>11.111</td>
<td>123.457</td>
<td>0.151</td>
<td>6.943</td>
<td>45.982</td>
<td>304.517</td>
</tr>
<tr>
<td>9</td>
<td>0.084</td>
<td>1</td>
<td>11.905</td>
<td>141.723</td>
<td>0.016</td>
<td>1.019</td>
<td>62.914</td>
<td>3883.554</td>
</tr>
<tr>
<td>10</td>
<td>0.068</td>
<td>1</td>
<td>14.706</td>
<td>216.263</td>
<td>0.014</td>
<td>1.167</td>
<td>83.386</td>
<td>5956.122</td>
</tr>
<tr>
<td>11</td>
<td>0.067</td>
<td>1</td>
<td>14.925</td>
<td>222.767</td>
<td>0.011</td>
<td>0.966</td>
<td>88.615</td>
<td>8129.787</td>
</tr>
<tr>
<td>12</td>
<td>0.066</td>
<td>1</td>
<td>15.152</td>
<td>229.568</td>
<td>0.020</td>
<td>1.458</td>
<td>74.749</td>
<td>3833.268</td>
</tr>
</tbody>
</table>

\[-\log L\] 653.5835 646.7709

Table 2.2: Estimated \( K_i, \lambda_i \) and mean and variance of the associated gamma random variables \( T_i \), for \( i = 1, \ldots, 12 \). The negative log-likelihood \(-\log L\) is also listed. The Gamma(\( K_1, \ldots, K_{12} \)) has slightly smaller \(-\log L\).

The estimated preference probability of each \((i, j)\) pair differs by less than 0.01 from those shown in Table 2.3 under the Gamma(\( K_1, \ldots, K_{12} \)) model. On the other hand, the variation in \((\hat{K}_i, \hat{\lambda}_i)\) across teams is quite large. Especially, some \( \hat{\lambda}_i \) lie on the boundary specified by \( \epsilon \). Overall, the solutions obtained by the maximum likelihood do not seem to be very stable in terms of the \((\hat{K}, \hat{\lambda})\) pairs but do seem to be stable in terms of the log-likelihood function and \( \hat{p}_{ij} \).

In the observed data (Table 2.1), some instance of intransitivity are observed. For example, \( p_{21} > 0.5, p_{32} > 0.5 \), but \( p_{31} < 0.5 \). However, the predicted winning probabilities under the Gamma(1) model exhibit strong stochastic transitivity, that is, if \( p_{ij} > 0.5 \) and \( p_{jk} > 0.5 \), then \( p_{ik} > max(p_{ij}, p_{jk}) \). The prediction under that the
Table 2.3: Predicted percentage of games that the $i$th team (row) defeats the $j$th team (column).

```plaintext
<table>
<thead>
<tr>
<th>Team</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>0.62</td>
<td>0.66</td>
<td>0.67</td>
<td>0.70</td>
<td>0.74</td>
<td>0.57</td>
<td>0.63</td>
<td>0.64</td>
<td>0.69</td>
<td>0.69</td>
<td>0.70</td>
</tr>
<tr>
<td>2</td>
<td>0.38</td>
<td>-</td>
<td>0.54</td>
<td>0.54</td>
<td>0.59</td>
<td>0.63</td>
<td>0.45</td>
<td>0.50</td>
<td>0.52</td>
<td>0.57</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
<td>0.46</td>
<td>-</td>
<td>0.51</td>
<td>0.55</td>
<td>0.59</td>
<td>0.41</td>
<td>0.46</td>
<td>0.48</td>
<td>0.53</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>4</td>
<td>0.33</td>
<td>0.46</td>
<td>0.49</td>
<td>-</td>
<td>0.54</td>
<td>0.58</td>
<td>0.40</td>
<td>0.46</td>
<td>0.48</td>
<td>0.53</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
<td>0.41</td>
<td>0.45</td>
<td>0.46</td>
<td>-</td>
<td>0.54</td>
<td>0.36</td>
<td>0.42</td>
<td>0.43</td>
<td>0.48</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>6</td>
<td>0.26</td>
<td>0.37</td>
<td>0.41</td>
<td>0.42</td>
<td>0.46</td>
<td>-</td>
<td>0.32</td>
<td>0.38</td>
<td>0.39</td>
<td>0.44</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>7</td>
<td>0.43</td>
<td>0.55</td>
<td>0.59</td>
<td>0.60</td>
<td>0.64</td>
<td>0.68</td>
<td>-</td>
<td>0.56</td>
<td>0.57</td>
<td>0.62</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>8</td>
<td>0.37</td>
<td>0.50</td>
<td>0.54</td>
<td>0.54</td>
<td>0.58</td>
<td>0.63</td>
<td>0.44</td>
<td>-</td>
<td>0.52</td>
<td>0.57</td>
<td>0.57</td>
<td>0.58</td>
</tr>
<tr>
<td>9</td>
<td>0.36</td>
<td>0.48</td>
<td>0.52</td>
<td>0.53</td>
<td>0.57</td>
<td>0.61</td>
<td>0.43</td>
<td>0.48</td>
<td>-</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>10</td>
<td>0.31</td>
<td>0.43</td>
<td>0.47</td>
<td>0.47</td>
<td>0.52</td>
<td>0.56</td>
<td>0.38</td>
<td>0.43</td>
<td>0.45</td>
<td>-</td>
<td>0.50</td>
<td>0.51</td>
</tr>
<tr>
<td>11</td>
<td>0.31</td>
<td>0.42</td>
<td>0.46</td>
<td>0.47</td>
<td>0.51</td>
<td>0.55</td>
<td>0.37</td>
<td>0.43</td>
<td>0.44</td>
<td>0.50</td>
<td>-</td>
<td>0.50</td>
</tr>
<tr>
<td>12</td>
<td>0.30</td>
<td>0.42</td>
<td>0.46</td>
<td>0.46</td>
<td>0.51</td>
<td>0.55</td>
<td>0.37</td>
<td>0.42</td>
<td>0.44</td>
<td>0.49</td>
<td>0.50</td>
<td>-</td>
</tr>
</tbody>
</table>
```

Prediction by Gamma(1) model.

```plaintext
<table>
<thead>
<tr>
<th>Team</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>0.49</td>
<td>0.59</td>
<td>0.60</td>
<td>0.67</td>
<td>0.94</td>
<td>0.47</td>
<td>0.73</td>
<td>0.62</td>
<td>0.72</td>
<td>0.70</td>
<td>0.73</td>
</tr>
<tr>
<td>2</td>
<td>0.51</td>
<td>-</td>
<td>0.53</td>
<td>0.53</td>
<td>0.56</td>
<td>0.56</td>
<td>0.48</td>
<td>0.54</td>
<td>0.53</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>0.41</td>
<td>0.47</td>
<td>-</td>
<td>0.49</td>
<td>0.53</td>
<td>0.52</td>
<td>0.41</td>
<td>0.47</td>
<td>0.46</td>
<td>0.51</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
<td>0.47</td>
<td>0.51</td>
<td>-</td>
<td>0.55</td>
<td>0.54</td>
<td>0.41</td>
<td>0.47</td>
<td>0.47</td>
<td>0.53</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>0.44</td>
<td>0.47</td>
<td>0.45</td>
<td>-</td>
<td>0.46</td>
<td>0.37</td>
<td>0.40</td>
<td>0.41</td>
<td>0.46</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>6</td>
<td>0.07</td>
<td>0.44</td>
<td>0.48</td>
<td>0.46</td>
<td>0.54</td>
<td>-</td>
<td>0.36</td>
<td>0.19</td>
<td>0.37</td>
<td>0.49</td>
<td>0.49</td>
<td>0.47</td>
</tr>
<tr>
<td>7</td>
<td>0.53</td>
<td>0.52</td>
<td>0.59</td>
<td>0.59</td>
<td>0.63</td>
<td>0.64</td>
<td>-</td>
<td>0.58</td>
<td>0.57</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>8</td>
<td>0.27</td>
<td>0.46</td>
<td>0.53</td>
<td>0.53</td>
<td>0.60</td>
<td>0.81</td>
<td>0.42</td>
<td>-</td>
<td>0.50</td>
<td>0.61</td>
<td>0.60</td>
<td>0.61</td>
</tr>
<tr>
<td>9</td>
<td>0.38</td>
<td>0.47</td>
<td>0.54</td>
<td>0.53</td>
<td>0.59</td>
<td>0.63</td>
<td>0.43</td>
<td>0.50</td>
<td>-</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>10</td>
<td>0.28</td>
<td>0.45</td>
<td>0.49</td>
<td>0.47</td>
<td>0.54</td>
<td>0.51</td>
<td>0.38</td>
<td>0.39</td>
<td>0.42</td>
<td>-</td>
<td>0.50</td>
<td>0.49</td>
</tr>
<tr>
<td>11</td>
<td>0.30</td>
<td>0.45</td>
<td>0.49</td>
<td>0.48</td>
<td>0.54</td>
<td>0.51</td>
<td>0.38</td>
<td>0.40</td>
<td>0.42</td>
<td>0.50</td>
<td>-</td>
<td>0.49</td>
</tr>
<tr>
<td>12</td>
<td>0.27</td>
<td>0.45</td>
<td>0.50</td>
<td>0.48</td>
<td>0.55</td>
<td>0.53</td>
<td>0.38</td>
<td>0.39</td>
<td>0.42</td>
<td>0.51</td>
<td>0.51</td>
<td>-</td>
</tr>
</tbody>
</table>
```

Prediction by Gamma($K_1, \ldots, K_{12}$) model.

Table 2.3: Predicted percentage of games that the $i$th team (row) defeats the $j$th team (column).
Table 2.4: The estimated $K_i$ and $\lambda_i$ Gamma($K_1, \ldots, K_{12}$) models with different starting points. The estimated parameters are quite different in these two solutions but the difference in the negative loglikelihood $-\log L$ is negligible.
Gamma($K_1, \ldots, K_{12}$) model does not always exhibit such transitivity. In fact, it is a weaker form of stochastic transitivity. To see this, compare the percentage of won games for team 5, 6, and 7 in the bottom panel of Table 2.3, $p_{65} > 0.5$, $p_{76} > 0.5$, but $p_{75} < \max\{p_{65}, p_{76}\}$.

Figure 2.3, Figure 2.4 and Figure 2.5 compare the observed $p_{ij}$ with $\hat{p}_{ij}$ under the Gamma(1) and the Gamma($K_1, \ldots, K_{12}$) models. It seems that the Gamma(1) model produces estimates with a decreasing trend from team 1 to team 6 and from team 7 to team 12. Similarly, the estimates of the Gamma($K_1, \ldots, K_{12}$) model shows a decreasing trend from team 1 to team 6, and team 7 to team 12. This is consistent with the observed $p_{ij}$. The estimates of the Gamma($K_1, \ldots, K_{12}$) for some teams (team 1, 6, 8) have large variations, which is close to the observed $p_{ij}$ for these teams.

In the Gamma($K_1, \ldots, K_{12}$) model, when the estimated mean of the gamma random variable is large (which occurs for teams associated with extreme values of $(K_i, \lambda_i)$), the model produce estimates of $p_{ij}$ that are closely clustered (team 2, 3, 4, 5, 7).

From the above results of the estimation, it seems that the predicted choice probabilities of the Gamma model with different thresholds exhibit some interesting differences from those of Gamma model with the same threshold. The impact of these differences is further investigated in the following section.
Figure 2.3: The observed probability that team $i$ defeats team $j$ ($p_{ij}$). The horizontal axis labels the teams and the vertical axis represents the preference probability that a team defeats each of the other 11 teams.
Figure 2.4: The predicted probability ($\hat{p}_{ij}$) that team $i$ defeats team $j$ under the Gamma(1) model. The horizontal axis labels the teams and the vertical axis represents the preference probability that a team defeats each of the other 11 teams.
Figure 2.5: The predicted probability ($\hat{p}_{ij}$) that team $i$ defeats team $j$ under the Gamma($K_1, \ldots, K_{12}$) model. The horizontal axis labels the teams and the vertical axis represents the preference probability that a team defeats each of the other 11 teams. Compared with Gamma(1) model, $\hat{p}_{ij}$ have large variations for team 1, 6, 8 but much smaller for team 2, 3, 4, 5, 7.
2.4.2 Hypothesis tests

To compare the fit of the Gamma($K$) model and Gamma($K_1, \ldots, K_{12}$) model, we test the following two sets of hypotheses. Here The Gamma(1) model is used to represent the Gamma($K$) model since the fit for a wide range of Gamma($K$) models is very similar.

\[ H_0 : p_{ij} = \text{Gamma}(1) \text{ model for all } i \text{ and } j, i \neq j \]
\[ H_1 : p_{ij} \neq \text{Gamma}(1) \text{ model for some } i \text{ and } j , \]

and

\[ H_0 : p_{ij} = \text{Gamma}(K_1, \ldots, K_m) \text{ model for all } i \text{ and } j, i \neq j \]
\[ H_1 : p_{ij} \neq \text{Gamma}(K_1, \ldots, K_m) \text{ for some } i \text{ and } j . \]

where = means that the probability $p_{ij}$ is generated by the testing model.

The likelihood ratio test is constructed for both models. Three statistics $Q_1, Q_2, Q_3$ are proposed by Stern for Gamma($K$) models. These statistics are constructed for the Gamma($K_1, \ldots, K_{12}$) model here. $Q_1$ is the statistic from the likelihood ratio test, which is given below in a general notation to accommodate both models. Thus,

\[ Q_1 = 2 \sum_{i=1}^{k} \sum_{j \neq i} a_{ij} \ln \frac{a_{ij}/n_{ij}}{\hat{p}_{ij}} , \]

where $n_{ij}$ is the total number of games play between team $i$ and team $j$ and $a_{ij}$ is the number of games won by team $i$ over team $j$.

An approximation of $Q_1$ leads to $Q_2$. Define $\frac{a_{ij}/n_{ij}}{\hat{p}_{ij}} = 1 + \epsilon_{ij}$. Then,

\[ Q_1 = 2 \sum_{i=1}^{k} \sum_{j \neq i} n_{ij} \hat{p}_{ij} (1 + \epsilon_{ij}) \ln (1 + \epsilon_{ij}) . \quad (2.4.1) \]
$Q_2$ can be derived from $Q_1$ using the first two terms of the Taylor series expansion of $\ln(1 + \epsilon_{ij})$ when $a_{ij} > 0$.

$$Q_2 = \sum_{i=1}^{k} \sum_{j>i} a_{ij} \frac{(\hat{p}_{ij} - a_{ij}/n_{ij})^2}{(\hat{p}_{ij})(1 - \hat{p}_{ij})/n_{ij}}.$$  

$Q_3$ is the Mosteller’s goodness of fit statistic. It is similar to $Q_2$ but with a variance stabilizing transformation $d_{ij} = \sin^{-1}(2a_{ij}/n_{ij} - 1)$ and $\hat{d}_{ij} = \sin^{-1}(2\hat{p}_{ij} - 1)$. Then, we have

$$Q_3 = \sum_{i=1}^{k} \sum_{j>i} n_{ij}(d_{ij} - \hat{p}_{ij})^2.$$  

These statistics should each approximately follow a chi-square distribution. For a paired comparison of $m$ teams, there are $m(m - 1)/2$ unspecified parameters under $H_1$. Under the Gamma(1) model, the number of unspecified parameters are $\lambda_i$, $i = 1, \ldots, m$ with a constraint that $\sum_{i=1}^{m} \lambda_i = 1$. Therefore, for the first hypothesis test, the degrees of freedom for each of $Q_1$, $Q_2$ and $Q_3$ are the difference $m(m - 1)/2 - (m - 1) = (m - 1)(m - 2)/2$. For the Gamma($K_1, \ldots, K_m$) model, $K_1, \ldots, K_m$ are also parameters. The degrees of freedom are approximately $(m - 1)(m - 2)/2 - m$.

Table 2.5 compares the three statistics under these two different models. It seems that the Gamma($K_1, \ldots, K_{12}$) model produces smaller statistics than the Gamma(1) model and thus seems to fit the data better. However, The difference in $Q_1$ for Gamma(1) and Gamma($K_1, \ldots, K_2$) model is about 13.6 with a degree of freedom about 12. This is not statistically significant (with a P-value $> 0.05$). The difference in $Q_2$ and $Q_3$ are also not statistically different.

The example in this section gives a glimpse of the properties of general independent Poisson race models. This class of models exhibits some special properties, for example, violation of strong stochastic transitivity which the MNL model does not
Table 2.5: The test statistic for the Gamma(1) and Gamma($K_1, \ldots, K_{12}$) model.

<table>
<thead>
<tr>
<th>Model</th>
<th>$Q_1$</th>
<th>$Q_2$</th>
<th>$Q_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma(1)</td>
<td>50.01</td>
<td>47.74</td>
<td>52.28</td>
</tr>
<tr>
<td>General gamma model</td>
<td>36.39</td>
<td>35.63</td>
<td>37.14</td>
</tr>
<tr>
<td>Difference</td>
<td>13.6</td>
<td>12.35</td>
<td>15.14</td>
</tr>
</tbody>
</table>

poccess. To understand the implication of the Poisson race models, the behavioral implications of the independent Poisson race model and its special cases are discussed in the following two sections. Section 2.5 discusses the Independence of Irrelevant Alternative assumption and Section 2.6 discusses the transitive properties.

## 2.5 Independence of irrelevant alternatives assumption and its Violations

Independence from Irrelevant Alternatives (IIA) (Luce, 1959; Ben-Akiva & Lerman, 1985; Bierlaire, 1998) is an important property of the multinomial logit model. Mathematically, assume any two choice sets $A_m$ and $A_{m'}$ such that $A_m \subseteq A_{m'}$, then for any alternatives $a_1$ and $a_2$ in $A_m$, the IIA property implies that

$$\frac{p(1|A_m)}{p(2|A_m)} = \frac{p(1|A_{m'})}{p(2|A_{m'})},$$

where $p(1|A_m)$ and $P(2|A_m)$ are the probabilities of choosing $a_1$ and $a_2$ from set $A_m$, respectively. Similarly we define $p(1|A_{m'})$ and $p(2|A_{m'})$. The implication of the IIA is that the addition of irrelevant alternatives would not change the ratio of the choice probabilities of the other alternatives already in the set. To see this with the MNL
model, recall from Proposition 2.2.3 that for any choice set of size $m$, we have

$$p(1|A_m) = \frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i},$$

and

$$p(2|A_m) = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i}.$$  

Then,

$$\frac{p(1|A_m)}{p(2|A_m)} = \frac{p(1|A_{m'})}{p(2|A_{m'})} = \frac{\lambda_1}{\lambda_2}.$$  

IIA implies that if a new alternative is added to the choice set but is identical to some existing alternative already in the set, the new alternative will reduce the chance of all alternatives. Intuitively, the new alternative should split the probability of the identical one already in the choice set but leave others untouched (Guadagni & Little, 1983). It seems that IIA is usually violated in choice problems with significantly correlated alternatives. The MNL model is not able to capture such violations. Since Poisson race models include the MNL model as a special case, it is of interested to investigate whether this larger class of models can capture the violations of IIA.

In the following discussion, we investigate the conditions in which the Poisson race models might exhibit violation of the IIA. We first set up the necessary notation, then derive the related results needed to prove these theorems. The main results are summarized in two theorems.

To exhibit violation of the IIA, consider the ratio of the choice probabilities of any two alternatives already in $A_m$ ($a_1$ and $a_2$, without loss of generality), and how the ratio changes when more alternatives are included. If this ratio changes, IIA does not hold. That is, the goal is to investigate some conditions under which the ratio
changes, that is,
\[ \frac{p(1|A_m)}{p(2|A_m)} \neq \frac{p(1|A_{m'})}{p(2|A_{m'})}. \]
This expression is equivalent to
\[ \frac{p(1|A_m)}{p(1|A_m) + p(2|A_m)} \neq \frac{p(1|A_{m'})}{p(1|A_{m'}) + p(2|A_{m'})}. \] (2.5.1)

Define \( p(1|C_{1,2}(A_m)) \) to be the conditional probability of choosing \( a_1 \) given that either \( a_1 \) or \( a_2 \) is chosen from the set \( A_m \). Then, (2.5.1) becomes
\[ p(1|C_{1,2}(A_m)) \neq p(1|C_{1,2}(A_{m'})). \]

Note that \( p(1|C_{1,2}(A_2)) = P(1|A_2) \) for binary sets.

To study the behavior of \( p(1|C_{1,2}(A_m)) \), we first consider a Poisson race only between two alternatives \( a_1 \) and \( a_2 \) from a choice set \( A_m \) with \( m \geq 2 \). The race is assumed to finish when either of the processes, represented by \( X_1(t) \) and \( X_2(t) \), reaches its corresponding threshold value \( K_1 \) or \( K_2 \). Again, assume the rate for these two processes are \( \lambda_1 \) and \( \lambda_2 \), respectively. Let \( T = \min(T_1, T_2) \) be the stopping time of the race. Let \( N_1 = X_1(T) \) and \( N_2 = X_2(T) \). Let \( S \) be the total number of hits from \( X_1(t) \) and \( X_2(t) \) when the race between \( a_1 \) and \( a_2 \) is finished. Then, \( S = N_1 + N_2 = \{\min(K_1, K_2), \ldots, K_1 + K_2 - 1\} \). With this notation, the following proposition can be proved.

**Proposition 2.5.1.** Given \( S = s \), define the probability of choosing \( a_1 \) from the set \( \{a_1, a_2\} \) to be \( p(1|S = s) \). Then,
(a) \[ p(1|S = s) = \begin{cases} 0 & \text{if } s < K_1 \\ \frac{(s-1)^{s-K_1}}{(\lambda_1 + \lambda_2)^{K_1}} & \text{if } s \geq \max(K_1, K_2) \\ \frac{(s-1)^{s-K_1}}{(\lambda_1 + \lambda_2)^{K_1}} + \frac{(s-1)^{s-K_2}}{(\lambda_1 + \lambda_2)^{K_2}} & \text{if } K_2 > s \geq K_1 \end{cases} \]

(b) For \( \lambda_1 \geq \lambda_2 \), when \( K_1 \leq K_2 \) and \( K_2 > 1 \), \( P(1|S = s) \) is monotonically decreasing in \( s \).

(c) For \( \lambda_1 \leq \lambda_2 \), when \( K_1 \geq K_2 \) and \( K_1 > 1 \), \( P(1|S = s) \) is monotonically increasing in \( s \).

(d) For \( K_1 = K_2 = 1 \), \( p(1|S = s) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \).

Proof. (a): If \( s = N_1 + N_2 < K_1 \), then \( N_1 < K_1 \), that is, the counter \( X_1(t) \) has not reached its threshold value \( K_1 \) and \( X_2(t) \) must have reached \( K_2 \) counts so that the race stops. Therefore, \( p(1|S = s) = 0 \).

When \( K_2 \geq s > K_1, N_2 < K_2 \), \( X_2(t) \) has not reached its threshold value and it must be \( X_1(t) \) that reaches its threshold value first. Therefore, \( p(1|S = s) = 1 \).

When \( s \geq \max(K_1, K_2) \), either of the counters might have reached its threshold though they can not reach threshold values at the same time. We have the following:

\[ P(S = s) = p(N_1 = K_1, N_2 = s - K_1) + p(N_1 = s - K_2, N_2 = K_2) \]

\[ = p(1, S = s) + p(2, S = s) \]

\[ = \left( \frac{s-1}{K_1 - 1} \right) \frac{\lambda_1}{\lambda_1 + \lambda_2} + \left( \frac{s-1}{K_2 - 1} \right) \frac{\lambda_2}{\lambda_1 + \lambda_2} \]
where \( p(i, S = s), \ i = 1, \ldots, 2, \) is the probability that the process for alternative \( a_i \) reaches its threshold value first and \( S = s \). For example, it is easy to show that

\[
p(1, S = s) = \begin{cases} 
0 & \text{if } s < K_1 \\
\frac{s-1}{K_1 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_1 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_1} & \text{if } K_1 \leq s < K_2 \\
\frac{s-1}{K_2 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_2 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_2} & \text{if } K_2 \leq s < K_1 \\
\frac{s-1}{K_1 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_1 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_1} + \frac{s-1}{K_2 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_2 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_2} & \text{if } s \geq \max(K_1, K_2) 
\end{cases}
\]

In summary, the distribution of \( S \) is given by:

\[
p(S = s) = \begin{cases} 
\frac{s-1}{K_1 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_1 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_1} & \text{if } K_1 \leq s < K_2 \\
\frac{s-1}{K_2 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_2 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_2} & \text{if } K_2 \leq s < K_1 \\
\frac{s-1}{K_1 - 1} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) K_1 \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{s-K_1} \left( \frac{s}{\lambda_1} \right)^s & \text{if } s \geq \max(K_1, K_2) 
\end{cases}
\]

Then, the conditional distribution in (a) can be obtained easily by the following:

\[
p(1|S = s) = \frac{p(1, S = s)}{p(S = s)}.
\]

(b): First rewrite the expression of \( p(1|S = s) \) in (a) as:

\[
p(1|S = s) = \begin{cases} 
0 & \text{if } s < K_1 \\
\frac{1}{1+c^s} & \text{if } s \geq \max(K_1, K_2) \\
1 & \text{if } K_2 > s \geq K_1 
\end{cases}
\]

where

\[
c^s = \frac{(s-K_1)!(K_1-1)!}{(s-K_2)!(K_2-1)!} \left( \frac{\lambda_2}{\lambda_1} \right)^{K_1+K_2} \left( \frac{\lambda_1}{\lambda_2} \right)^s = \frac{(s-K_1)!}{(s-K_2)!} \left( \frac{\lambda_2}{\lambda_1} \right)^s,
\]

and \( c = \frac{(K_1-1)!}{(K_2-1)!} \left( \frac{\lambda_2}{\lambda_1} \right)^{K_1+K_2} \) does not depend on \( s \).

In the special case where \( \lambda_1 = \lambda_2 \) and \( K_1 = K_2 > 1 \), \( p(1|S = s) = 0.5 \). The monotonicity result holds because \( p(1|S = s) = 0.5 \) is constant in \( s \) (notice this case is not strictly monotonically decreasing).
For the case that $\lambda_1 \geq \lambda_2$, $K_1 \leq K_2$ and $K_2 > 1$, $s = \{K_1, \ldots, K_2, \ldots, K_1+K_2-1\}$ and

$$p(1|S = s) = \begin{cases} 1 & \text{if } K_1 \leq s < K_2 \\ \frac{1}{1+c^*} & \text{if } K_2 \leq s . \end{cases}$$

Both $\left(\frac{\lambda_1}{\lambda_2}\right)^s$ and $\frac{(s-K_1)!}{(s-K_2)!}$ are non-decreasing in $s$. Therefore, $c^*$ is increasing in $s$ and $(1+c^*)^{-1}$ is decreasing in $s$. As a result, $p(1|S = s)$ is monotonically decreasing in $s$.

(c): the proof is similar to the proof of (b). When $\lambda_1 \leq \lambda_2$, $K_1 \geq K_2$ and $K_1 > 1$, we have $s = \{K_2, \ldots, K_1, \ldots, K_1+K_2-1\}$ and

$$p(1|S = s) = \begin{cases} 0 & \text{if } K_2 \leq s < K_1 \\ \frac{1}{1+c^*} & \text{if } K_1 \leq s . \end{cases}$$

Again, $p(1|S = s)$ is monotonically increasing in $s$ since $c^*$ is decreasing in $s$ and $(1+c^*)^{-1}$ is non-decreasing in $s$.

(d): For $K_1 = K_2 = 1$, $S = 1$ and the distribution of $S$ is degenerate. Therefore, $p(1|S = s)$ no longer depends on $s$ since

$$p(1|S = s) = p(1|A_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} .$$

With the same notation as above, we can derive the distribution of the stopping time of the race between the processes for alternative $a_1$ and $a_2$ only, as is given below.

**Proposition 2.5.2.** The stopping time $T = \min(T_1, T_2)$ given $S = s$ is distributed as $\text{Gamma}(s, \lambda_1 + \lambda_2)$.

**Proof.** For any integer $s$ in the range $\{\min(K_1, K_2), \ldots, K_1+K_2-1\}$, define the event $\{S = s\}$ when the race between two processes stops to be

$$\{S = s\} = \{N_1 = K_1, N_2 = s - K_1\} \cup \{N_2 = K_2, N_1 = s - K_2\} .$$
The events in the right hand side are disjoint since we assume the race stops if either of processes finishes and the processes would not reach the threshold values at the same time. Let \( N = N_1 + N_2 \). Now consider the conditional pdf \( f(t|S = s) \):

\[
\begin{align*}
f(t|S = s) & = f(t\{N_1 = K_1, N_2 = s - K_1\} \cup \{N_2 = K_2, N_1 = s - K_2\}) \\
& = \frac{f(t, \{N_1 = K_1, N_2 = s - K_1\} \cup \{N_2 = K_2, N_1 = s - K_2\})}{P(\{N_1 = K_1, N_2 = s - K_1\} \cup \{N_2 = K_2, N_1 = s - K_2\})} \\
& = \frac{f(t, \{N_1 = K_1, N_2 = s - K_1\} + f(t, N_2 = K_2, N_1 = s - K_2)}{P(\{N_1 = K_1, N_2 = s - K_1\} \cup \{N_2 = K_2, N_1 = s - K_2\})} \\
& = \frac{f(t|N = s)p(N_1 = K_1, N_2 = s - K_1) + f(t|N = s)p(N_2 = K_2, N_1 = s - K_2)}{P(\{N_1 = K_1, N_2 = s - K_1\} \cup \{N_2 = K_2, N_1 = s - K_2\})} \\
& = f(t|N = s) \sim \text{Gamma}(s, \lambda_1 + \lambda_2).
\end{align*}
\]

The two independent Poisson processes can be combined together to form a new process with rate \( \lambda_1 + \lambda_2 \) by Proposition 3.9 proposed by Townsend and Ashby (1983). Thus, the time taken for the new process to reach \( N \) counts is distributed as \( \text{Gamma}(s, \lambda_1 + \lambda_2) \).

**Proposition 2.5.3.** For a choice set \( A_m = \{a_1, a_2, \ldots, a_m\} \), the probability that either \( a_1 \) or \( a_2 \) is selected from \( A_m \) given \( S = s \), denoted by \( p(C_{1,2}(A_m)|S = s) \), is (a) monotonically decreasing in \( s \) for \( s > 1 \) and (b) monotonically decreasing in \( m \).

**Proof.** (a) For \( s \neq 1 \), according to Proposition 2.5.2, the distribution of the response time \( T_i \) given \( S = s \) \( f(t|S = s) \) is \( \text{Gamma}(s, \lambda_1 + \lambda_2) \). Let \( \lambda = \lambda_1 + \lambda_2 \), then the pdf \( f(t|S = s) \) can be written as:

\[
f(t|S = s) = \frac{\lambda^s t^{s-1}}{\Gamma(s)} e^{-t\lambda}, \quad t > 0.
\]

For \( s_2 > s_1 > 1 \), the likelihood ratio is given by

\[
\frac{f(t|s_2)}{f(t|s_1)} = \frac{\lambda^{s_2} t^{s_2-1}}{\Gamma(s_2)} e^{-t\lambda} = \frac{\Gamma(s_1)}{\Gamma(s_2)} \lambda^{s_2-s_1} t^{s_2-s_1}, \quad t > 0.
\]

55
Since \( s_2 > s_1, \lambda > 0 \) and \( t > 0, \) \( \frac{f(t|s_2)}{f(t|s_1)} \) is an increasing function of \( t \) and this family has monotone likelihood ratio (MLR). The Monotone likelihood ratio property implies that the family of cdfs is stochastically ordered in \( s \), which is defined as following (Casella & Berger, p44, 2002): A cdf \( F(t|s_1) \) is stochastically smaller than a cdf \( F(t|s_2) \) if \( F(t|s_1) \geq F(t|s_2) \) for all \( t \) and \( F(t|s_1) > F(t|s_2) \) for some \( t \). Also, a family of cdfs \( \{ F(t|s), s > 1 \} \) is stochastically increasing in \( s \) if \( s_1 < s_2 \) implies that \( F(t|s_1) \) is stochastically smaller than a cdf \( F(t|s_2) \) (Casella & Berger, p134, 2002). Equivalently speaking, for fixed \( t, \) \( F(t|s) \) is an decreasing function in \( s \).

To see this, let \( t_2 > t_1 \), then

\[
\frac{f(t_2|s_2)}{f(t_2|s_1)} \geq \frac{f(t_1|s_2)}{f(t_1|s_1)} .
\]

That is

\[
f(t_2|s_2)f(t_1|s_1) \geq f(t_1|s_2)f(t_2|s_1) .
\]

Integrate over \( t_1 \) and \( t_2 \) on both side, we have

\[
\int_{t_1}^{\infty} \int_{0}^{t} f(t_2|s_2)f(t_1|s_1)dt_1dt_2 \geq \int_{t}^{\infty} \int_{0}^{t} f(t_1|s_2)f(t_2|s_1)dt_1dt_2 .
\]

Then,

\[
(1 - F(t|s_2))F(t|s_1) \geq F(t|s_2)(1 - F(t|s_1)) ,
\]

where \( F(t|s_1) \) and \( F(t|s_2) \) are the cdfs of \( f(t|s_1) \) and \( f(t|s_2) \). Simplify the expression, we will get

\[
F(t|s_1) \geq F(t|s_2), \quad s_1 < s_2 . \quad (2.5.2)
\]

Therefore, \( F(t|s_1) \) is stochastically smaller than \( F(t|s_2) \) and the family of cdfs \( F(t|s) \) is stochastically increasing in \( s \). In another word, for fixed \( t, \) \( F(t|s) \) is an decreasing function as \( s \) increases.

An intuition to understand that the stochastic ordering of the gamma family of cdfs \( F(t|s) \sim gamma(s, \lambda) \) is to assume three gamma random variables such that
$t_1 \sim \text{Gamma}(s_1, \lambda)$, $t_3 \sim \text{Gamma}(s_2 - s_1, \lambda)$. Then, $t_2 = t_1 + t_3 \sim \text{Gamma}(s_2, \lambda)$. Since $t_3 > 0$, $t_1$ is stochastically smaller than $t_2$.

Next, let the stopping time of the processes associated with $a_1$ and $a_2$ be denoted by $T = \min(T_1, T_2)$. Let the stopping time of the processes associated with other alternatives in the set $A_m$ denoted by $\tau = \min(T_3, \ldots, T_m)$. $T$ and $\tau$ are independent since we assume all the processes are independent. $\tau$ is non-increasing since $\tau$ is the minimum of a set of response times associated with alternatives. Then, the probability that either $a_1$ and $a_2$ is chosen from the set $A_m$ is the probability that $T$ is smaller than $\tau$, that is,

$$p(C_{1,2}(A_m)|S = s) = p(T < \tau|A_m, S = s) = p(F(\tau|s)) .$$

From (2.5.2), $F(\tau|s)$ is a decreasing function of $s$. Therefore, $p(C_{1,2}(A_m)|S = s)$ is also a decreasing function of $s$.

(b): As the number of alternatives in the set $A_m$ increases and $\tau$ is non-increasing, the cdf function $F(\tau|s)$ is monotonically decreasing and $p(C_{1,2}(A_m)|S = s)$ is monotonically decreasing in $m$ as well. \hfill \square

Proposition 2.5.3 leads to the proof of next proposition concerning the stochastic ordering of the family of cdfs for $p(S \leq s|C_{12}(A_m))$.

**Proposition 2.5.4.** Let $A_m = \{a_1, \ldots, a_m\}$ be a choice set and let $C_{1,2}(A_m)$ denote the event that either $a_1$ or $a_2$ is chosen from the set $A_m$. Then, (a) for $s > 1$, the cdf $p(S \leq s|C_{1,2}(A_m))$ is stochastically smaller than $p(S \leq s)$, and (b) for $s > 1$ and $2 \leq m' < m$, the cdf $p(S \leq s|C_{1,2}(A_m))$ is stochastically smaller than $p(S \leq s|C_{1,2}(A_{m'}))$.

**Proof.** (a) By definition of conditional probability,

$$p(S = s|C_{1,2}(A_m)) = \frac{p(C_{1,2}(A_m)|S = s)P(S = s)}{p(C_{1,2}(A_m))} .$$

57
That is,
\[ p(C_{1,2}(A_m)|S = s)P(S = s) = p(S = s|C_{1,2}(A_m))p(C_{1,2}(A_m)) . \] (2.5.3)

From Proposition 2.5.3, \( p(C_{1,2}(A_m)|S = s) \) is a decreasing function of \( s \) for \( s > 1 \). Then, for \( 1 < s_1 < s_2 \), we have
\[ p(C_{1,2}(A_m)|S = s_1) \geq p(C_{1,2}(A_m)|S = s_2) . \]

Replace \( p(C_{1,2}(A_m)|S = s_1) \) and \( p(C_{1,2}(A_m)|S = s_2) \) with (2.5.3), we have
\[ \frac{p(S = s_1|C_{1,2}(A_m))}{p(S = s_1)} \geq \frac{p(S = s_2|C_{1,2}(A_m))}{p(S = s_2)} , \]
which can then be re-organized as
\[ p(S = s_1|C_{1,2}(A_m))p(S = s_2) \geq p(S = s_1)p(S = s_2|C_{1,2}(A_m)) . \]

Summing over \( s_1 \) and \( s_2 \) for both sides of the above inequality, we have,
\[ \sum_{s_2=s+1}^{K_1+K_2-1} \sum_{s_1=min(K_1,K_2)}^{s} p(S = s_1|C_{1,2}(A_m))p(S = s_2) \]
\[ \geq \sum_{s_2=s+1}^{K_1+K_2-1} \sum_{s_1=min(K_1,K_2)}^{s} p(S = s_1)p(S = s_2|C_{1,2}(A_m)) . \]
That is,
\[ p(S \leq s|C_{1,2}(A_m))(1 - p(S \leq s)) \geq p(S \leq s)(1 - p(S \leq s|C_{1,2}(A_m))) , \]
which leads to
\[ p(S \leq s|C_{1,2}(A_m)) \geq p(S \leq s) \]
by canceling \( p(S \leq s|C_{1,2}(A_m))p(S \leq s) \) in both sides. This implies that \( p(S \leq s|C_{1,2}(A_m)) \) is stochastically smaller than \( p(S \leq s) \). In a special case with two alternatives, \( p(S \leq s|C_{1,2}(A_2)) = p(S \leq s) \) as discussed in Proposition 2.5.1.
(b) For any $2 \leq m' < m$,

$$p(C_{1,2}(A_m), C_{1,2}(A_{m'})|S = s_1) = p(C_{1,2}(A_m)|S = s_1)$$

given $A_{m'} \subset A_m$. That is, if $a_1$ or $a_2$ is chosen from set $A_{m'}$, $a_1$ or $a_2$ must be chosen from the larger set $A_m$. Next, let $T = \min(T_1, T_2)$ and assume $T_i, i = 1, \ldots, m$ are independent, then, for $s_1 < s_2$,

$$p(C_{1,2}(A_m)|S = s_1, C_{1,2}(A_{m'})$$

$$= \frac{p(C_{1,2}(A_m), C_{1,2}(A_{m'})|S = s_1)}{p(C_{1,2}(A_{m'})|S = s_1)}$$

$$= \frac{p(C_{1,2}(A_m)|S = s_1)}{p(C_{1,2}(A_{m'})|S = s_1)}$$

$$= \frac{p(T < \min(T_{3, \ldots, T_{m'}})|S = s_1)p(T < \min(T_{m, \ldots, T_m})|S = s_1)}{p(T < \min(T_{3, \ldots, T_{m'}})|S = s_1)}$$

$$= p(T < \min(T_{m', \ldots, T_m})|S = s_1)$$

$$= p(C_{12}(A_{m-m'+1})|S = s_1), \quad (2.5.4)$$

by the independence of $T_i$. Here $A_{m-m'+1}$ is a choice set with alternatives $a_i, i = 1,2,m', \ldots, m$.

From Proposition 2.5.3 (a), $p(C_{12}(A_{m-m'+1})|S = s)$ is monotonically decreasing in $s$. As a result, $p(T < \min(T_{m', \ldots, T_m})|S = s_1)$ is monotonically decreasing in $s$ as well. Then, for $s_1 < s_2$,

$$p(T < \min(T_{m', \ldots, T_m})|S = s_1) \geq p(T < \min(T_{m', \ldots, T_m})|S = s_2). \quad (2.5.5)$$

Note that $p(T < \min(T_{m', \ldots, T_m})|S = s_2)$ can be written as $p(C_{1,2}(A_m)|S = s_2, C_{1,2}(A_{m'}))$. Therefore, $(2.5.5)$ can be written as

$$p(C_{1,2}(A_m)|S = s_1, C_{1,2}(A_{m'}) \geq p(C_{1,2}(A_m)|S = s_2, C_{1,2}(A_{m'})). \quad (2.5.6)$$
Next, notice that
\[
p(C_{1,2}(A_m)|S = s, C_{1,2}(A_{m'})) = \frac{p(S = s, C_{1,2}(A_{m'})|C_{1,2}(A_m))p(C_{1,2}(A_m))}{p(S = s, C_{1,2}(A_{m'}))} = \frac{p(S = s|C_{1,2}(A_m))p(C_{1,2}(A_m))}{p(S = s|C_{1,2}(A_{m'}))p(C_{1,2}(A_{m'}))}.
\]
In the second equality, \(p(S = s, C_{1,2}(A_{m'})|C_{1,2}(A_m))\) can be written as \(p(S = s|C_{1,2}(A_m))\) since if \(a_1\) or \(a_2\) is chosen from set \(A_{m'}\), \(a_1\) or \(a_2\) must be chosen from set \(A_m\).

With this expression, we can rewrite (2.5.6) and obtain the following inequality:
\[
p(S = s_1|C_{1,2}(A_{m'}))p(S = s_2|C_{1,2}(A_m)) \leq p(S = s_2|C_{1,2}(A_{m'}))p(S = s_1|C_{1,2}(A_m)).
\]
Now, summing over \(s_1\) and \(s_2\) for both side of the above inequality, we have
\[
\sum_{s_2=s+1}^{K_1+K_2-1} \sum_{s_1=\min(K_1,K_2)}^{s} p(S = s_1|C_{1,2}(A_{m'}))p(S = s_2|C_{1,2}(A_m)) \leq \sum_{s_2=s+1}^{K_1+K_2-1} \sum_{s_1=\min(K_1,K_2)}^{s} p(S = s_2|C_{1,2}(A_{m'}))p(S = s_1|C_{1,2}(A_m)).
\]
That is,
\[
p(S \leq s|C_{1,2}(A_{m'}))(1 - p(S \leq s|C_{1,2}(A_m))) \leq p(S \leq s|C_{1,2}(A_m))(1 - p(S \leq s|C_{1,2}(A_{m'}))).
\]
Cancelling \(p(S \leq s|C_{1,2}(A_{m'}))p(S \leq s|C_{1,2}(A_m))\) on both sides,
\[
p(S \leq s|C_{1,2}(A_m)) \geq p(S \leq s|C_{1,2}(A_{m'})).
\]
In conclusion, \(p(S \leq s|C_{1,2}(A_m))\) is stochastically smaller than \(p(S \leq s|C_{1,2}(A_{m'}))\) for \(m > m' \geq 2\).

With the above proposition, we can then prove the following theorem which indicates violation of IIA for giving conditions of \(\lambda_i\) and \(K_i\) (\(i = 1,2\)).
Theorem 2.5.5. With the notation specified above, when $\lambda_1 \leq \lambda_2$ and $K_1 \geq K_2$ and $K_1 > 1$, $p(1|C_{1,2}(A_m))$ is monotonically decreasing in $m$, that is, as more alternatives are allowed in the choice set. When $\lambda_1 \geq \lambda_2$, $K_1 \leq K_2$ and $K_2 > 1$, $p(1|C_{1,2}(A_m))$ is monotonically increasing in $m$, that is, as more alternatives are allowed in the choice set.

Proof. Let the $\min(K_1, K_2) = s_{\min}$ and $K_1 + K_2 - 1 = s_{\max}$. For $\lambda_1 \leq \lambda_2$, $K_1 \geq K_2$ and $K_1 > 1$, from Proposition 2.5.1 (c), $p(1|S = s)$ is monotonically increasing in $s$. Note that $S = s$ is the sum of hits for alternative $a_1$ or $a_2$ when either $a_1$ or $a_2$ is selected from the set $A_m$ or $A_m'$. Therefore,

$$p(1|S = s) = p(1|S = s, C_{1,2}(A_m)) = p(1|S = s, C_{1,2}(A_m'))$$

(2.5.7)

since $A_{m'} \subset A_m$ for $m' < m$. Therefore, $p(1|S = s, C_{1,2}(A_m'))$ and $p(1|S = s, C_{1,2}(A_m))$ are both monotonically increasing in $s$.

Next, $p(1|C_{1,2}(A_m))$ can be written as

$$p(1|C_{1,2}(A_m)) = \sum_{s=s_{\min}}^{s_{\max}} p(S = s|C_{1,2}(A_m))p(1|S = s, C_{1,2}(A_m)),$$

where $p(1|S = s, C_{1,2}(A_m))$ is a monotonically increasing function. The above expression is actually the expectation of $p(1|S = s, C_{1,2}(A_m))$ over the distribution of $S$.
given \( C_{1,2}(A_m) \), which can be expressed as

\[
p(1|C_{1,2}(A_m))
\]

\[
= \left[ p(S \geq s_{\text{min}}|C_{1,2}(A_m)) - p(s \geq s_{\text{min}} + 1|C_{1,2}(A_m)) \right] p(1|S = s_{\text{min}}, C_{1,2}(A_m))
\]

\[
+ \left[ p(S \geq s_{\text{max}} - 1|C_{1,2}(A_m)) - p(s \geq s_{\text{max}}|C_{1,2}(A_m)) \right] p(1|S = s_{\text{max}} - 1, C_{1,2}(A_m))
\]

\[
p(S \geq s_{\text{max}}|C_{1,2}(A_m)) p(1|S = s_{\text{max}}, C_{1,2}(A_m))
\]

(2.5.8)

The second equality above is generated by collecting the terms \( p(S \geq s_{\text{min}}) \), \( p(S \geq s_{\text{min}} + 1) \), \ldots, \( p(S \geq s_{\text{max}}) \) together. Notice in (*), \( p(S \geq S_{\text{max}}|C_{12}(A_m)) = p(S = s_{\max}|C_{12}(A_m)) \).

Similarly, the expectation of \( p(1|S = s, C_{1,2}(A_{m'}) \) over the distribution of \( S \) given \( C_{1,2}(A_{m'}) \) can be written as

\[
p(1|C_{1,2}(A_{m'}))
\]

\[
= \sum_{s = \text{min}(K_1, K_2)}^{K_1 + K_2 - 1} p(S = s|C_{1,2}(A_{m'})) p(1|S = s, C_{1,2}(A_{m'}))
\]

\[
= p(S \geq s_{\text{min}}|C_{1,2}(A_{m'})) p(1|S = s_{\text{min}}, C_{1,2}(A_{m'})) + p(S \geq s_{\text{min}} + 1|C_{1,2}(A_{m'}))
\]

\[
[ p(1|S = s_{\text{min}} + 1, C_{1,2}(A_{m'})) - p(1|S = s_{\text{min}}, C_{1,2}(A_{m'})) ] + \cdots +
\]

\[
p(S \geq s_{\text{max}}|C_{1,2}(A_{m'})) [ p(1|S = s_{\text{max}}, C_{1,2}(A_{m'})) - p(1|S = s_{\text{max}} - 1, C_{1,2}(A_{m'})) ]
\]

(2.5.9)
From Proposition 2.5.4, \( p(S \leq s|C_{1,2}(A_m)) \geq p(S \leq s|C_{1,2}(A_{m'})) \) for \( 2 \leq m' < m \).

Compare (2.5.8) and (2.5.9), since the terms in the square brackets are the same by (2.5.7) and \( p(S \geq s|C_{1,2}(A_m)) \leq p(S \geq s|C_{1,2}(A_{m'})) \) for all \( s \),

\[ p(1|C_{1,2}(A_m)) \leq p(1|C_{1,2}(A_{m'})) . \]

Since \( m' < m \), \( p(1|C_{1,2}(A_m)) \) is decreasing in \( m \).

Similarly, we can prove the second part of the theorem. The proof is omitted here.

In this dissertation, there is another situation found to have violation of IIA. For special case of the Poisson race model when all the Poisson processes have the same mean, that is, \( \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} = \cdots = \frac{K_m}{\lambda_m} \), we can show that the IIA does not hold in some situations. The result will be stated in Theorem 2.5.8. To prove this theorem, Lemma 2.5.6 and Proposition 2.5.7 are to be proved first.

**Lemma 2.5.6.** For \( a \geq 1 \), \( \ln \frac{a}{a+1} + \frac{1}{a} > 0 \) and \( \ln \frac{a+1}{a} - \frac{1}{a} < 0 \).

**Proof.** For and \( a > 1 \),

\[
\ln \frac{a}{a+1} + \frac{1}{a} = -\ln(1 + \frac{1}{a}) + \frac{1}{a} = \frac{1}{a} - \left( \frac{1}{a} - \frac{1}{2a^2} + \frac{1}{3a^3} - \frac{1}{4a^4} + \ldots \right) \quad \text{by Taylor series expansion of } \ln(1 + \frac{1}{a})
\]

\[
= \frac{1}{2a^2} - \frac{1}{3a^3} + \frac{1}{4a^4} - \frac{1}{5a^5} \ldots \tag{2.5.10}
\]

The series in (2.5.10) is an absolutely converged series as the limit exists and is finite. Therefore, the terms can be grouped in pairs, that is,

\[
\ln \frac{a}{a+1} + \frac{1}{a} = \left( \frac{1}{2a^2} - \frac{1}{3a^3} \right) + \left( \frac{1}{4a^4} - \frac{1}{5a^5} \right) + \ldots > 0 \quad \text{since every sum in parentheses are positive}.
\]

63
For the second part, we have

\[ \ln \frac{a + 1}{a} - \frac{1}{a} = \ln \left(1 + \frac{1}{a}\right) - \frac{1}{a} \]

by Taylor series expansion of \( \ln \left(1 + \frac{1}{a}\right) \)

\[ = -\frac{1}{2a^2} + \frac{1}{3a^3} - \frac{1}{4a^4} + \frac{1}{5a^5} \cdots \]

\[ = -\left(\frac{1}{2a^2} - \frac{1}{3a^3}\right) - \left(\frac{1}{4a^4} - \frac{1}{5a^5}\right) - \cdots \] since the series is absolutely converged.

\[ < 0. \]

\[ \square \]

**Proposition 2.5.7.** When \( \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} = \cdots = \frac{K_m}{\lambda_m} \), the following is true.

(a)

\[
p(1|S = s) = \begin{cases} 
0 & \text{if } s < K_1, \\
\frac{1}{1 + \left(\frac{s - K_1}{s - K_2}\right)!\left(K_2 - 1\right)!\left(K_1\right)^{K_1 + K_2}\left(K_1\right)^s} & \text{if } s \geq \max(K_1, K_2), \\
1 & \text{if } K_2 > s \geq K_1.
\end{cases}
\]

(b) When \( K_1 \leq K_2 \) and \( K_2 > 1 \), \( p(1|S = s) \) is monotonically decreasing in \( s \).

(c) When \( K_1 \geq K_2 \) and \( K_1 > 1 \), \( p(1|S = s) \) is monotonically increasing in \( s \).

(d) \( K_1 = K_2 = 1 \), \( p(1|S = s) = 0.5 \). The distribution of \( S \) is degenerate.

**Proof.** For (a) and (d): since \( \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} \), we have \( \frac{\lambda_1}{\lambda_2} = \frac{K_1}{K_2} \). Replace \( \frac{\lambda_1}{\lambda_2} \) in Proposition 2.5.2 part (a) and simplify the expression, we obtain the expression in (a). If \( K_1 = K_2 = 1 \), \( \lambda_1 = \lambda_2 \) from \( \frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} \). From (a), \( p(1|S = s) = 0.5 \). Also, \( S \) is the total number of hits from processes \( X_1(t) \) and \( X_2(t) \) when either \( a_1 \) and \( a_2 \) is chosen from the choice set. It follows that \( S = 1 \). Therefore, the distribution of \( S \) is degenerate.
(b): When $K_1 = K_2$, $p(1|S = s) = 0.5$ from (a). The monotonicity result holds for any $s$ ($K_1 \leq s \leq K_1 + K_2 - 1$). When $K_1 < K_2$, rewrite the expression in part (a) as:

$$p(1|S = s) = \begin{cases} 
0 & \text{if } s < K_1, \\
(1 + C^*)^{-1} & \text{if } s \geq \max(K_1, K_2), \\
1 & \text{if } K_2 > s \geq K_1, 
\end{cases}$$

where

$$C^* = C'(\frac{K_2}{K_1})^{K_1 + K_2} \frac{(K_1 - 1)!}{(K_2 - 1)!}$$

and

$$C' = \frac{(s - K_1)!}{(s - K_2)!} \left( \frac{K_1}{K_2} \right)^s = (s - K_1) \cdots (s - K_2 + 1) \left( \frac{K_1}{K_2} \right)^s.$$ 

$C^*$ depends on $s$ only through $C'$. To show $p(1|S = s)$ is an increasing function of $s$ is equivalent to showing that the logarithm of $C'$ is an increasing function of $s$. Taking logarithm of $C'$, we have

$$\ln C' = \ln(s - K_1) + \cdots + \ln(s - K_2 + 1) + s(\ln K_1 - \ln K_2).$$

Now take the first and second derivative of $\ln C'$ with respect to $s$:

$$(\ln C')' = \frac{1}{s - K_1} + \frac{1}{s - K_1 - 1} + \cdots + \frac{1}{s - K_2 + 1} + \ln K_1 - \ln K_2$$

and

$$(\ln C')'' = -\frac{1}{(s - K_1)^2} - \frac{1}{(s - K_1 - 1)^2} - \cdots - \frac{1}{(s - K_2 + 1)^2} < 0.$$ 

Since $(\ln C')'' < 0$, $(\ln C')'$ is decreasing in $s$. If we can show $\min(\ln C')' > 0$, then $(\ln C')' > 0$ and $\ln C'$ and $C'$ are increasing functions of $s$. The minimum of $(\ln C')'$
must be at \( s_{\text{max}} = K_2 + K_1 - 1 \) since \((lnC')'\) is decreasing in \( s \).

\[
\min(lnC')'
= (lnC')'|_{s=K_2+K_1-1}
= \frac{1}{K_2 - 1} + \cdots + \frac{1}{K_1} + \ln K_1 - \ln K_2
= \frac{1}{K_2 - 1} + \cdots + \frac{1}{K_1} + \ln(\frac{K_1}{K_1 + 1} + 1) + \cdots + \frac{1}{K_2 - 1} + \frac{1}{K_2 - 1}
> 0 \quad \text{from Lemma 2.5.6}
\]

Thus, \( lnC' \) and \( C' \) are increasing in \( s \) and \( p(1|S = s) \) is decreasing in \( s \) as required.

(c): The case that \( K_1 = K_2 \) is proved in (b).

For the case that \( K_1 > K_2 \),

\[
C' = \frac{(s - K_1)!}{(s - K_2)!} \left( \frac{K_1}{K_2} \right)^s = \frac{1}{(s - K_2)} \cdots \frac{1}{(s - K_1 + 1)} \left( \frac{K_1}{K_2} \right)^s.
\]

Taking logarithm of \( C' \), we have

\[
\ln C'^* = -\ln(s - K_2) - \cdots - \ln(s - K_1 + 1) + s(\ln K_1 - \ln K_2).
\]

Now take the first and second derivative of \( \ln C' \) with respect to \( s \):

\[
(lnC')' = -\frac{1}{s - K_2} - \cdots - \frac{1}{s - K_1 + 1} + \ln K_1 - \ln K_2
\]

and

\[
(lnC')'' = \frac{1}{(s - K_2)^2} + \frac{1}{(s - K_2 - 1)^2} + \cdots + \frac{1}{(s - K_1 + 1)^2} > 0.
\]

Therefore, \((lnC')'\) is increasing in \( s \).
Next, we will show $\max(\ln C')' < 0$.

\[
\begin{align*}
\max(\ln C')' \\
= & \ (\ln C')'|_{s=K_2+K_1-1} \\
= & \ -\frac{1}{K_1-1} - \cdots - \frac{1}{K_2} + \ln K_1 - \ln K_2 \\
= & \ -\frac{1}{K_2-1} - \cdots - \frac{1}{K_1} + \ln \left( \frac{K_1}{K_1-1} \frac{K_1-1}{K_1-2} \cdots \frac{K_2+1}{K_2} \right) - \ln(K_2) \\
= & \ (\ln \frac{K_1}{K_1-1} - \frac{1}{K_1-1} + (\ln \frac{K_1-1}{K_1-2} - \frac{1}{K_1-2}) + \cdots + (\ln \frac{K_2+1}{K_2} - \frac{1}{K_2}) \\
< & \ 0 \quad \text{apply Lemma 2.5.6}
\end{align*}
\]

Thus, $\ln C'$ and $C'$ are decreasing in $s$ and $p(1|S = s)$ is increasing in $s$. \hfill \Box

We are now ready to prove the IIA does not hold in some situations when all Poisson processes have the same mean.

**Theorem 2.5.8.** For any choice set $A_m$ with $m$ alternatives, under the condition that

\[
\frac{K_1}{\lambda_1} = \frac{K_2}{\lambda_2} = \cdots = \frac{K_m}{\lambda_m}, \text{ if } K_1 \leq K_2 \text{ and } K_2 > 1, \text{ then } p(1|C_{1,2}(A_m)) \text{ is monotonically increasing in } m \text{ (that is, as more alternatives are allowed in the choice set); if } K_1 \geq K_2 \text{ and } K_1 > 1, \text{ then } p(1|C_{1,2}(A_m)) \text{ is monotonically decreasing in } m.
\]

The proof of Theorem 2.5.8 is similar to the proof of Theorem 2.5.5 and is not repeated here.

The results in Theorem 2.5.5 and Theorem 2.5.8 clearly indicate that for the Poisson race model, the ratio of choice probabilities for any two alternatives (i.e. $a_1$ and $a_2$) changes as more alternatives are added to the choice set. It is strong evidence that the independent Poisson race model can violate IIA. Moreover, these two theorems state the directions of change for given conditions on the parameters. Such properties suggest that the Poisson race models might be used to model behavior that traditional logit models cannot capture.
2.6 Transitivity and its violations

In a choice decision problem, the transitivity of preferences means that if an alternative \( a_1 \) is preferred to \( a_2 \), and \( a_2 \) is preferred to \( a_3 \), then \( a_1 \) is preferred to \( a_3 \). For ease of expression, let \( p_{1,2} \) be the “preference” probability that \( a_1 \) is preferred to \( a_2 \). Under a probability model for choices, the transitivity of preferences is also called the stochastic transitivity (Tversky, 1969). There are three forms of stochastic transitivity: weak, moderate and strong, as defined below:

- Weak stochastic transitivity (WT): if \( p_{12} > 0.5 \), \( p_{23} > 0.5 \), then \( p_{13} > 0.5 \).

- Moderate stochastic transitivity (MST): if \( p_{12} > 0.5 \), \( p_{23} > 0.5 \), then \( p_{13} > \min\{p_{12}, p_{23}\} \).

- Strong stochastic transitivity (SST): if \( p_{12} > 0.5 \), \( p_{23} > 0.5 \), then \( p_{13} > \max\{p_{12}, p_{23}\} \).

As mentioned in Section 2.4, the MNL model exhibits strong stochastic transitivity. In fact, Luce’s choice axiom (Luce, 1959) requires strong stochastic transitivity. Assume three alternatives \( a_1 \), \( a_2 \) and \( a_3 \) with utility \( u(1) \), \( u(2) \) and \( u(3) \), respectively. By Luce’s Choice Axiom, the odds of choosing one alternative over another as the ratio of their utilities. Therefore, for \( p_{12} > 0.5 \), \( p_{23} > 0.5 \), we have

\[
p_{12}/p_{21} = u(1)/u(2) > 1
\]

and

\[
p_{23}/p_{32} = u(2)/u(3) > 1
\]

Then,

\[
p_{13}/p_{31} = u(1)/u(3) = u(1)/u(2) \times u(2)/u(3) > \max\{u(1)/u(2), u(2)/u(3)\}.
\]
This implies that \( p_{13} > \max\{p_{12}, p_{23}\} \) as the product of two positive numbers greater than 1 must be greater than any of the two numbers.

Strong stochastic transitivity is a severe requirement and frequently not met by human preference behaviors (Navarick & Fantino, 1972). The MNL and Gamma(\( K \)) models cannot capture such violation. On the other hand, they are only special cases of the general Poisson race models. Without the restriction that all \( K_i \) are equal, the Poisson race model might exhibit different behavior in terms of transitivity. In this section, the conditions for transitivity and intransitivity under the general Poisson race model are discussed. Examples are given to demonstrate the violation of transitivity.

### 2.6.1 Sufficient condition for strong stochastic transitivity

Consider a choice set \( A_2 = \{a_1, a_2, \ldots, a_m\} \). Under the assumptions for the independent Poisson race model discussed in Section 2.2.1, there is an independent Poisson process associated with each alternative \( a_i \). Again, let \( \lambda_i \) be the exponential rate of the Poisson process, \( K_i \) be the threshold value and \( T_i \) be the minimum time taken for the counter to accumulate \( K_i \) counts for the \( i \)th alternative, where \( i = 1, \ldots, m \). The race between any two alternatives is equivalent to the comparison of corresponding response times. Recall that \( T_i \sim \text{Gamma}(K_i, \lambda_i) \). We derived some sufficient conditions for transitivity for choice sets with three alternatives.

**Theorem 2.6.1.** Let \( A_3 = \{a_1, a_2, a_3\} \), under the independence Poisson race model, the strong stochastic transitivity of preferences holds in the following situations:

(a) When any two of the gamma random variables are stochastically ordered. This includes the case that \( K_1 = K_2 \), or \( \lambda_1 = \lambda_2 \) and \( K_1 < K_2, \lambda_1 > \lambda_2 \).
(b) When \( K_1/\lambda_1 = K_2/\lambda_2 = K_3/\lambda_3 \).

The proof of the Theorem 2.6.1 requires the following lemma.

**Lemma 2.6.2.** Assume two family of distributions with cdfs \( F_1(t) \) and \( F_2(t) \) and \( F_1(t) \geq F_2(t) \); that is, \( F_1 \) is stochastically greater than \( F_2 \). Let \( h(t) \) be a monotonically increasing function, then,

\[
\int_0^\infty h(t)dF_1(t) \leq \int_0^\infty h(t)dF_2(t) .
\]

**Proof.** To prove this lemma, first write the integration as a summation. Without loss of generality, the proof is carried out with continuous distributions, that is, the cdfs \( F_i(t) (i = 1, 2) \) are differentiable and pdfs exist. For discrete distributions, the proof is easier as the integration is directly written as a summation (the proof is actually carried out in Theorem 2.5.5 with \( F_i(t) \) and \( h(t) \) being known.

The positive real line \((0, \infty)\) can be represented by \( \lim_{t_1 \to 0, t_n \to \infty} (t_1, t_2, \ldots, t_n) \), assuming \( n - 1 \) intervals each with length \( \delta \), where \( n \to \infty \) and \( \delta \to 0 \). Also, let \( G(t) = 1 - F(t) \). Then,

\[
\int_0^\infty h(t)dF_1(t) = \lim_{t_1 \to 0} \lim_{t_n \to \infty} \sum_{t_1}^{t_n} h(t)f_1(t)
\]

\[
= \lim_{\delta \to 0} \lim_{t_1 \to 0} \lim_{t_n \to \infty} \frac{1}{\delta} \{ G_1(t_1)h(t_1) + G_1(t_2)[h(t_2) - h(t_1)] + \cdots \\
+ G_1(t_n)[h(t_n) - h(t_{n-1})] \}
\]

\[
\leq \lim_{\delta \to 0} \lim_{t_1 \to 0} \lim_{t_n \to \infty} \frac{1}{\delta} \{ G_2(t_1)h(t_1) + G_2(t_2)[h(t_2) - h(t_1)] + \cdots \\
+ G_2(t_n)[h(t_n) - h(t_{n-1})] \}
\]

\[
= \int_0^\infty h(t)dF_2(t) .
\]
Proof of the Theorem 2.6.1. (a) The gamma family has pdf of the form:

\[ f(t|\lambda_i, K_i) = \frac{\lambda_i^{K_i}t^{K_i-1}}{\Gamma(K_i)} e^{-t\lambda_i}. \]

When \( K_1 = K_2 = K \), assume without loss of generality that \( \lambda_1 > \lambda_2 \). Then, the likelihood ratio is given by

\[ \frac{f(t|\lambda_1, K)}{f(t|\lambda_2, K)} = \frac{\lambda_1^Ke^{-t\lambda_1}}{\lambda_2^Ke^{-t\lambda_2}}, \quad t > 0 \]

so the likelihood ratio is monotonically decreasing in \( t \). Then, for \( t_2 > t_1 \),

\[ \frac{f(t_2|\lambda_1)}{f(t_2|\lambda_2)} \leq \frac{f(t_1|\lambda_1)}{f(t_1|\lambda_2)}. \]

That is

\[ f(t_2|\lambda_2)f(t_1|\lambda_1) \geq f(t_1|\lambda_2)f(t_2|\lambda_1). \]

If we integrate over \( t_1 \) and \( t_2 \) on both sides, we have

\[ \int_t^\infty \int_0^t f(t_2|\lambda_2)f(t_1|\lambda_1)dt_1dt_2 \geq \int_t^\infty \int_0^t f(t_1|\lambda_2)f(t_2|\lambda_1)dt_1dt_2. \]

Then,

\[ (1 - F_2(t))F_1(t) \geq F_2(t)(1 - F_1(t)), \]

where \( F_1(t) \) and \( F_2(t) \) are the cdfs of \( f(t|\lambda_1) \) and \( f(t|\lambda_2) \). On Simplifying the expression, we get

\[ F_1(t) \geq F_2(t). \]

Therefore, \( F_1(t) \) is stochastically smaller than \( F_2(t) \) and \( p_{12} = p(T_1 < T_2) > 0.5 \) or \( p_{21} = p(T_1 > T_2) < 0.5 \).

Next, from the formulation of preference probability in Proposition 2.2.1, we have

\[ p_{31} = p(T_1 > T_3) = \int_0^\infty F_3(t)dF_1(t) \]

71
and
\[ p_{32} = p(T_2 > T_3) = \int_0^\infty F_3(t) dF_2(t). \]

where \( F_3(t) \) is a non-decreasing function. Using the result that \( F_1(t) \geq F_2(t) \) (so that \( p_{12} \geq 0.5 \) and Lemma 2.6.2, let \( h(t) = F_3(t) \), we have
\[
\int_0^\infty F_3(t) dF_1(t) \leq p_{32} = p(T_2 > T_3) = \int_0^\infty F_3(t) dF_2(t).
\]

That is,
\[ p_{31} \leq p_{32}. \]

Equivalently,
\[ p_{13} = 1 - p_{31} \geq 1 - p_{32} = p_{23}. \]

Thus, given \( p_{12} > 0.5 \), if \( p_{23} > 0.5 \), it follows that
\[ p_{13} \geq p_{23} > 0.5. \]

In words, given \( a_1 \) is preferred to \( a_2 \), if \( a_2 \) is preferred to \( a_3 \), then \( a_1 \) is preferred to \( a_3 \).

The transitivity of preferences probability holds here. The proof above only requires that the cdfs \( F_i(t) \) are stochastic ordered, e.g., \( F_1(t) \leq F_2(t) \). Thus, any conditions of \( K_i \) and \( \lambda_i \) that leads to this ordering of the cdfs have strong stochastic transitivity.

Two conditions of \( K_i \) and \( \lambda_i \) are discussed below.

First, when \( K_1 < K_2 \) and \( \lambda_1 = \lambda_2 \). From the proof of Proposition 2.5.2 in section 2.5, the gamma family has MLR in \( K \) and the cdf is stochastically increasing in \( K \), that is, for \( K_1 < K_2 \), \( F_1(t) \geq F_2(t) \). Second, for the case that \( K_1 < K_2 \) and \( \lambda_1 > \lambda_2 \), since the rate of Poisson process for \( a_1 \) is greater than that for \( \lambda_2 \) and the threshold value for process 1 is smaller than 2, it is more likely that that process one is finished first with a shorter response time \( T_1 \). That is, \( T_1 \) is stochastically smaller than \( T_2 \).
(b) For the case that \( K_1/\lambda_1 = K_2/\lambda_2 = K_3/\lambda_3 \), from Proposition 2.3.4, when the mean \((K_i/\lambda_i)\) of the three gamma random variables are equal, the preference probability \( p_{12} \) is increasing as \( K_2 \) increases. \( p_{12} = 0.5 \) when \( K_1 = K_2 \). Now, if \( p_{12} > 0.5 \), then, \( K_1 < K_2 \). Similarly, if \( p_{23} > 0.5 \), \( K_2 < K_3 \). Therefore, \( K_1 < K_3 \) and \( p_{13} > 0.5 \). Thus the strong transitivity of probability holds.

\[\square\]

### 2.6.2 Violation of stochastic transitivity

Consider Lemma 2.6.1 and its proof, it is clear that the necessary condition for possible violation of transitivity might appear is that \( K_1 < K_2 < K_3 \) and \( \lambda_1 < \lambda_2 < \lambda_3 \).

It is possible to construct an example that violates even the weak form of stochastic transitivity. Let \( K_2 = \lambda_2 = 1 \) and find \((K_1, \lambda_1)\) and \((K_3, \lambda_3)\) such that \( p_{12} = p(T_1 < T_2) = 0.5 \) and \( p_{23} = p(T_2 < T_3) = 0.5 \). From the expression of preference probability in Proposition 2.2.1,

\[
p_{12} = \int_0^\infty (1 - F_2(t_1))dF_1(t_1) = \int_0^\infty e^{-t_1}dF_1(t_1) = \left( \frac{\lambda_1}{1 + \lambda_1} \right)^{K_1} = 0.5
\]

and

\[
p_{23} = \int_0^\infty F_2(t_3)dF_3(t_3) = \int_0^\infty (1 - e^{-t_3})dF_1(t_3) = 1 - \left( \frac{\lambda_3}{1 + \lambda_3} \right)^{K_3} = 0.5.
\]

Solving these two equations, the sets of \((\{K_1, \lambda_1\})\) and \((\{K_3, \lambda_3\})\) that gives \( p_{12} = p_{23} = 0.5 \) are obtained. \( p_{13} \) can be computed with these sets of values. The exact probabilities \( p_{13} \) are calculated by Proposition 2.2.2 whenever one of the \( K_i \) is an integer. In the case that both \( K_i \) are not integers, the cdf of the beta distribution is evaluated using Proposition 2.2.6 and the computation of the probabilities are highly accurate. If transitivity holds, \( p_{13} = 0.5 \). Otherwise, intransitivity is present, as
shown in Table 2.6. In this table, the row headings are the parameters \((K_1, \lambda_1)\) for \(T_1\) and the column headings are parameters for \(T_3\) such that \(p_{12} = p_{23} = 0.5\). Note that the \(\lambda_i\) are given with only two decimal places but in the actual calculation many more decimal places are used. The entries in the table are the preference probability \(p_{13}\). Some of the entries differ from 0.5. The seemingly small variations from 0.5 cannot be explained by random errors alone since the exact probabilities are calculated whenever one of the \(K_i\) is an integer.

<table>
<thead>
<tr>
<th>((K_i, \lambda_i))</th>
<th>(0.5, 0.33)</th>
<th>(0.7, 0.59)</th>
<th>(0.9, 0.86)</th>
<th>(1, 1)</th>
<th>(2, 2.44)</th>
<th>(3, 3.85)</th>
<th>(4, 5.29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.5, 0.33)</td>
<td>0.5</td>
<td>0.4996</td>
<td>0.4998</td>
<td>0.5</td>
<td>0.5013</td>
<td>0.502</td>
<td>0.5023</td>
</tr>
<tr>
<td>(0.7, 0.59)</td>
<td>0.5004</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5005</td>
<td>0.5008</td>
<td>0.501</td>
</tr>
<tr>
<td>(0.9, 0.86)</td>
<td>0.5002</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5001</td>
<td>0.5002</td>
<td>0.5002</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(2, 2.44)</td>
<td>0.4987</td>
<td>0.4995</td>
<td>0.4999</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4997</td>
<td>0.4995</td>
</tr>
<tr>
<td>(3, 3.85)</td>
<td>0.498</td>
<td>0.4992</td>
<td>0.4998</td>
<td>0.5</td>
<td>0.5003</td>
<td>0.5</td>
<td>0.4997</td>
</tr>
<tr>
<td>(4, 5.29)</td>
<td>0.4977</td>
<td>0.499</td>
<td>0.4998</td>
<td>0.5</td>
<td>0.5005</td>
<td>0.5003</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 2.6: Violation of weak stochastic transitivity. The row headings are the parameters \((K_1, \lambda_1)\) for \(T_1\) and the column headings are parameters for \(T_3\) such that \(p_{12} = p_{23} = 0.5\). The entries in the table are the preference probability \(p_{13}\). Some of the entries differ from 0.5. The seemingly small variations from 0.5 cannot be explained by random errors alone since the exact probabilities are calculated whenever one of the \(K_i\) is an integer.

Another example of violation of transitivity is illustrated in Table 2.7. Setting \(K_2 = \lambda_2 = 1\), using the same set of \((K_3, \lambda_3)\) and \(K_1\) as in Table 2.6, but adding 0.01 to \(\lambda_1\). Notice that \(p_{23} = 0.5\) since \(K_1, K_2, \lambda_1\) and \(\lambda_2\) do not change. \(p_{12}\) and \(p_{13}\) can be computed again with the same methods used for Table 2.6 and the values for \(p_{13}\)
are shown in Table 2.7. The values $p_{12}$ are the same as the values under column (1, 1) in the table. All $p_{12}$ are greater than 0.5. There are a few entries in the table that are smaller than 0.5, which indicates violation of weak stochastic transitivity since $p_{12} > 0.5$ would imply that $p_{13} > 0.5$ if weak stochastic transitivity holds.

<table>
<thead>
<tr>
<th>$(K_i, \lambda_i)$</th>
<th>(0.5, 0.33)</th>
<th>(0.7, 0.59)</th>
<th>(0.9, 0.86)</th>
<th>(1, 1)</th>
<th>(2, 2.44)</th>
<th>(3, 3.85)</th>
<th>(4, 5.29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.5, 0.34)</td>
<td>0.5047</td>
<td>0.5048</td>
<td>0.5053</td>
<td>0.5056</td>
<td>0.5073</td>
<td>0.5081</td>
<td>0.5085</td>
</tr>
<tr>
<td>(0.7, 0.60)</td>
<td>0.5033</td>
<td>0.5033</td>
<td>0.5036</td>
<td>0.5037</td>
<td>0.5046</td>
<td>0.5051</td>
<td>0.5053</td>
</tr>
<tr>
<td>(0.9, 0.87)</td>
<td>0.5023</td>
<td>0.5025</td>
<td>0.5027</td>
<td>0.5028</td>
<td>0.5033</td>
<td>0.5036</td>
<td>0.5037</td>
</tr>
<tr>
<td>(1,1,01)</td>
<td>0.5019</td>
<td>0.5022</td>
<td>0.5024</td>
<td>0.5025</td>
<td>0.5029</td>
<td>0.5031</td>
<td>0.5032</td>
</tr>
<tr>
<td>(2, 2.45)</td>
<td>0.4995</td>
<td>0.5005</td>
<td>0.5016</td>
<td>0.5012</td>
<td>0.5014</td>
<td>0.5012</td>
<td>0.5013</td>
</tr>
<tr>
<td>(3, 3.86)</td>
<td>0.4986</td>
<td>0.4999</td>
<td>0.5006</td>
<td>0.5008</td>
<td>0.5014</td>
<td>0.5012</td>
<td>0.501</td>
</tr>
<tr>
<td>(4, 5.30)</td>
<td>0.4981</td>
<td>0.4995</td>
<td>0.5003</td>
<td>0.5006</td>
<td>0.5013</td>
<td>0.5012</td>
<td>0.501</td>
</tr>
</tbody>
</table>

Table 2.7: Another example with violation of strong stochastic transitivity. The entries in the table are the preference probability $p_{13}$. The row headings are the parameters $(K_1, \lambda_1)$ for $T_1$ and the column headings are parameters for $T_3$ such that $p_{23} = 0.5$ and $p_{12}$ is the values in the column under (1, 1). When transitivity holds, $p_{12} > 0.5$ also implies that $p_{13} > 0.5$. However, there are a few entries in the table that are smaller than 0.5, which indicates intransitivity as well.

The two examples above shows that intransitivity does exists for the case that $K_1 > K_2 > K_3$ and $\lambda_1 > \lambda_2 > \lambda_3$ (or vice versa). However, they are special cases involve at least one pair that is indifferent in terms of choice. It is not easy to find a combination of parameter values to demonstrate the violation of weak stochastic transitivity such that $p_{13} < 0.5$ given $p_{12} > 0.5$ and $p_{23} > 0.5$. 

75
2.6.3 Ranking intransitivity

A Poisson race model might exhibit ranking intransitivity. That is, the estimates of the Poisson race model do not preserve preference ordering while the estimates from the MNL model do. To demonstrate this phenomenon, consider the example of baseball games given in Section 2.4. Using the preference probabilities in Table 2.1, the teams can be ordered from the strongest to the weakest. For example, $p_{12} = 0.44$, that is, team 1 won 44% of games that it played with team 2. Therefore, team 2 is ranked before team 1. In this way, all twelve teams can be ranked according to the percentage in the first row. Similarly, another set of rankings can be obtained by the second row of data, and so on. There are totally 12 set of rankings. If the rankings are transitive, it does not matter which row is used to rank the teams because the same rank order would be obtained.

Table 2.8 gives the ranking orders of the twelve team based on the observed winning percentages, the prediction by the MNL model (Gamma(1) model) and the prediction by the general Poisson race model(Gamma($K_1, \ldots, K_{12}$)model). The top part of the table is the rankings given by the observed data. The rankings are not transitive as they vary from row to row. The parentheses indicate teams with the same percentage of winning games. It seems that team 1, 7 are among the best teams while team 5, 6, 11 and 12 are among the weakest teams. In the middle of Table 2.8 are the rankings given by the MNL model. From the first row to the last row, the rankings are the same. The bottom part of the table shows the rankings is from the estimates of the general Poisson race model. The rankings are not the same from row to row. Compared with the top part of the table, the variation of the team rankings
from row to row is smaller. Both models predict team 1 and 7 are the strongest while
team 5 and 6 are the the weakest teams.

An interesting form of ranking intransitivity, referred as "Voter’s paradox", describes a situation where personal preference ranking of three or more alternatives can result in a group ranking that is not transitive (Arrow, 1950). The ranking order estimated by Gamma($K_1, \ldots, K_{12}$) model in Table 2.8 can demonstrate Voter’s paradox. In row 3, $9 > 2 > 8$ (the notation means that 9 is preferred to 2 and 2 is preferred to 8); in row 8, $2 > 8 > 9$; in row 10, $8 > 9 > 2$. With majority vote, 9 is preferred to 2 and 2 is preferred to 8 but 8 is preferred to 9. This is clearly a violation of transitivity. Notice that this is also a violation of the IIA assumption as the preference order of any two of the alternatives changes with the addition of the third alternative.

Traditional models of choice usually predict preferences between alternatives by computing their utility values, then identifying the alternative with greatest value. Rational choice theory is based on such procedure. The concept of rationality is defined through a number of principles and transitivity and IIA are two of them. Violations of these principles make the traditional models such as MNL invalid estimation methods. Alternative models have been used to replace MNL model, such as the mixed logit, conditional logit or multinomial probit models. But these models often have their own assumptions that are difficult to meet or sometimes are computationally infeasible. The Poisson race model, with MNL model as its special case, can capture to some extent these violations of rationality. It might be potentially a good alternative model for human choice behaviors.
<table>
<thead>
<tr>
<th>By the (i)th team</th>
<th>Ranking of the teams by actual outcome of the games</th>
<th>Strongest (\rightarrow) weakest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 1 4 (7 9) (3 5 8 12) 11 10 6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 5 2 (7 10) (1 4) 11 6 12 9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(1 11) 5 (7 8) (3 4 12) (9 10) 6 2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(7 10) (8 9) (1 2) (3 4) 5 11 6 12</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(12 1) 7 (6 8) 4 5 9 10 11 2 3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(1 8 (9) 11) (2 3 4) (6 7) 12 5 10</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(1 2 6 7 9) (10 11) 3 (4 5) 12 8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(7 1 8 9 10) 11 (2 3 4 5) 12 6</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(2 1 3) (5 7 8 9) (10 11) 4 12 6</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(1 6 12) 3 (7 9) (2 5 8 10) (4 11)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>(1 10 2 4) (7 8 9 12) (5 11) (3 6)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>(7 1 2 8) 4 9 6 (3 12) 11 10 5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ranking of the teams estimated by Gamma(1) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (\rightarrow) 12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ranking of the teams estimated by Gamma((K_1, \ldots, K_{12})) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

Table 2.8: The ranking of teams from the strongest to the weakest according to the percent of games that the \(i\)th team defeats other teams.
2.7 Conclusion

This chapter serves two purposes. The first one is to derive the necessary expressions for the choice probabilities of Poisson race models; the second one is to understand the properties of the models under various situations. This work is not only a summary but more importantly an extension of previous work on Poisson race models. It provides both theoretical and empirical understanding of the model properties which might help to build better models of human choice.
CHAPTER 3

MODELING DEPENDENCE IN BINARY CHOICE DATA

3.1 Introduction

As mentioned in Section 1.2, a basic assumption in modeling the choice between profiles is that a respondent subconsciously maps the profiles in the choice set into a utility space and chooses whichever profile has maximum utility (Gustafsson, Herrmann, & Huber 2000). Underlying this assumption is Luce’s choice axiom (Luce, 1959) which implies a scale utility value for every product profile and assumes choice probabilities proportional to these utility values. The multinomial logit (MNL) model (McFadden, 1974) is the most commonly used choice model and is based on Luce’s choice axiom. In the MNL model, the probabilistic nature of utility is captured through error terms associated with alternative-specific utilities. These error terms give rise to ‘smooth’ probabilities bounded away from the extremes of 0 and 1. The psychology literature, however, suggests that extreme probabilities do exist, especially in the case of dominance of one profile over another (Rumelhart & Greeno, 1971; Huber & Murphy, 1979). For two alternatives $a_1$ and $a_2$ having the same attributes, $a_1$ is said to dominate $a_2$ if $a_1$ is at least as good as $a_2$ on all
attributes and better on some attributes. In this case of *dominance in attributes*, the
decision maker can identify the superior alternative without an intermediate step of
forming overall utilities. Moreover, if there is a scale ordering of levels within at-
tributes, then the superior alternative can be identified with certainty. Under these
conditions, extreme choice probabilities very near 0 and 1 are likely to occur.

The two credit card profiles in Table 3.1 show an example of dominance in at-
tributes. Card $a_2$ in the right column of Table 3.1 is as good as, or better than, card
$a_1$ in the left column on every attribute; it charges the smaller interest rate, no annual
fee, offers a larger cash reward, and neither card offers travel points. Thus, a rational
decision maker would choose card $a_2$ with probability 1.0, even if the perceived utility
difference between the levels of each attribute is very small.

If $a_1$ is better than $a_2$ in some attributes but not in others, then neither $a_1$ nor
$a_2$ dominates the other and a “trade-off relationship” occurs. In a trade-off relation-
ship, a decision maker needs some means to consolidate the relative advantages and
disadvantages of the alternatives into an overall measure of attractiveness or utility.
As a result of this consolidation process, the two alternatives may turn out to be of
similar, or of very different, utility. The latter case may be characterized as *utility
dominance*, and it translates into extreme choice probabilities of 0 and 1 only in the
limit as the model parameter values tend to infinity. If $a_1$ is slightly better than
$a_2$ in all attributes, then the expected utility for $a_1$ is only slightly larger than that
for $a_2$. As a result of a utility-based comparison, the probability of choosing $a_1$ is
only slightly higher than that of choosing $a_2$ and both are quite close to 0.5, whereas
dominance in attributes leads to corresponding probabilities of 0 or 1.
Table 3.1: An example of dominance with credit cards. The second card is better or at least as good as the first card in all attributes.

<table>
<thead>
<tr>
<th></th>
<th>Credit card $a_1$</th>
<th>Credit card $a_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest rate</td>
<td>17.99</td>
<td>14.99</td>
</tr>
<tr>
<td>Annual fee</td>
<td>$30</td>
<td>$0</td>
</tr>
<tr>
<td>Travel points</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cash reward</td>
<td>No</td>
<td>3%</td>
</tr>
</tbody>
</table>

The recognition of dominance in attributes leads to the comparison of alternatives, attribute by attribute, with a well-behaved preference relationship among the levels of each alternative (Bouyssou & Pirlotm, 2004). Current conjoint analysis methods pay little attention to the idea of dominance in attributes (French, 1988) and, similar to Luce(1959), the current practice in the design of conjoint choice experiments is to avoid dominance in attributes in the construction of choice sets. However, for the purpose of making predictions about future choices of an individual, the choice sets are usually at least partially predetermined by the marketplace. Then, given the preferences of an individual consumer, dominance relationships may well exist in a choice set. Unless the model can include possibility of dominance, any resulting inferences will be biased.

One possibility for dealing with dominance relationships post hoc is to introduce a dominance operator that would, conditional on posterior knowledge about consumers’ preferences, identify dominance relationships and then set the corresponding choice
probabilities to 0 and 1. The disadvantage of such a strategy is that it is not possible to test whether consumers realize particular dominance relationships and to what extent they act upon them in a deterministic way. For example, if one alternative dominates another as a result of an advantage in a single attribute which is of minor importance to that respondent, then the choice probabilities will be at the extremes of 0 and 1 if the individual is motivated to identify the best alternative, but may be close to 0.5 if the individual is satisfied with an alternative that is relatively good enough.

In this chapter, a new model capable of handling dominance relationships is developed. The model belongs to the class of race models that have mostly been employed in experimental psychology to model response time (Townsend & Ashby, 1983). Section 1.1 describes the basic idea of a race model and Chapter 2 discusses the Poisson race model in detail. Briefly, in a race model, a decision maker accumulates evidence, over time, in favor of individual alternatives within a choice set. A choice occurs when the amount of evidence for any one alternative in the choice set exceeds a specific threshold value. In a Poisson race model, evidence in favor of an alternative is assumed to accrue in discrete units that are sometimes called “hits” and the decision maker tracks the number of hits in favor of each alternative in specific counter. As soon as any one counter reaches its threshold, the corresponding alternative is chosen and the race terminates.

In traditional Poisson race models, the alternative-specific counters are, conditional on model parameters, assumed to operate independently and in parallel, but in order to handle dominance, we will introduce dependence into the model via the notion of a shared counter. As an example, consider the two credit cards in Table 3.1.
Suppose, for simplicity, that a particular individual is concerned only about the interest rate, so that the levels of this attribute are the only ones to trigger any hits. If the 17.99% interest rate triggers a hit on the counter for card $a_1$, dominance can be preserved by counting this also as a hit for card $a_2$ which has the lower interest rate of 14.99%. Thus, all the hits triggered by 17.99% interest rate will be counted as evidence for both alternatives and recorded on a shared counter, whereas a hit triggered by the 14.99% interest rate will be counted as evidence in favor of card $a_2$ only and recorded on the unique counter for $a_2$. Evidence accumulated in favor of an alternative is the sum of the hits on its unique counter and the shared counter. Shared counters can be introduced for every attribute in the product profile in a similar way.

The introduction of shared counters gives rise to the possibility of tied races where two or more alternatives reach their threshold values simultaneously. This will occur, in particular, if the unique counters accrue hits at a rate that is small relative to that of the shared counter. When the race results in a tie and the consumer’s goal is to find an alternative that is “good enough”, he or she may pick at random among the tied alternatives. But, if the goal is to identify the best alternative in the set, a second race may start where the influence of the shared counts is downgraded. We model this ‘discounted rate’ as a parameter which enables us to move continuously from the case where ties are broken at random to the case where slightly dominant alternatives are picked with certainty.

This chapter deals with choice sets of size two, but the general ideas can be extended to larger choice sets with appropriate tie breaking rules. In Section 3.2, the dependent Poisson race model is formulated in a general choice context. Bayesian estimation of model parameters using MCMC methods is discussed in Section 3.3.
Section 3.4 presents the results of a simulation study and, in Section 3.6, the model is fit to real data originating from a conjoint experiment on credit cards. A discussion of our results is given in Section 3.7.

### 3.2 Model formulation

Consider a binary choice set $A = \{a_1, a_2\}$. The standard Poisson race model assumes that the two alternatives in $A$ are associated with two independent Poisson processes, denoted by $X_1(t)$ and $X_2(t)$, having rates $\lambda_1$ and $\lambda_2$, respectively. Under this model, information accrues to the counters in discrete steps, or hits. The race terminates when one of the processes reaches the finish line, specified by a positive integer threshold parameter, $K$. This happens at the smallest $t$ for which $X_1(t) = K$ or $X_2(t) = K$. The alternative whose counter first accumulates $K$ hits is chosen. The standard model is described by Townsend & Ashby (1983), LaBerge (1994), Van Zandt, Colonius & Proctor (2000), and elsewhere. This model is modified by replacing the independent Poisson processes with dependent Poisson processes. The structure of the dependence enables us to model attribute-based processing.

#### 3.2.1 The dependent Poisson race model

First, for simplicity, we assume that each of the two alternatives is characterized by the levels of a single attribute. Hits in favor of each alternative arrive according to the corresponding Poisson processes $X_1(t)$ and $X_2(t)$. Unique counters store the hits providing exclusive evidence for each of the two alternatives and a shared counter tracks hits shared by both alternatives. When the sum of the hits on the shared counter and one of the unique counters reaches a threshold value $K$, a choice occurs.
Thus, each of $X_1(t)$ and $X_2(t)$ can be decomposed into a shared process and a process unique to the corresponding alternative. Define the rate for the shared process to be

$$\lambda_s = \min(\lambda_1, \lambda_2).$$

(3.2.1)

The remaining rate for $X_1(t)$, unique to process 1, is

$$\lambda_{u1} = \lambda_1 - \lambda_s.$$ 

(3.2.2)

Similarly, process 2 has a unique rate

$$\lambda_{u2} = \lambda_2 - \lambda_s.$$ 

(3.2.3)

This decomposition of $X_1(t)$ and $X_2(t)$ results in three independent Poisson processes. When there is a hit stimulated by the alternatives, the hit is contributed either by the shared process or by one of the unique processes. The probability that the hit is contributed by any of the three processes is proportional to their rates (c.f., Karlin & Taylor, 1975, Chapter 5).

The Poisson race model discussed above described the structure of a race between two dependent Poisson processes $X_1(t)$ and $X_2(t)$ and is called the dependent Poisson race model throughout this dissertation.

Consider two single-attribute alternatives whose levels can be rank ordered. If $a_1$ dominates $a_2$, then $\lambda_s = \lambda_2$, $\lambda_{u2} = 0$ and, whenever there is a hit in favor of $a_2$, the hit is accrued on the shared counter so that it represents a hit also in favor of $a_1$. On the other hand, a hit generated by unique process 1 is accrued only on the unique counter for alternative 1. Such mechanism ensures that the total number of hits in favor of $a_1$ is never smaller than that in favor of $a_2$ so that the alternative $a_1$ is always chosen. If the two alternatives have the same attribute levels, the two unique rates
are zero, the hits are accumulated only on the shared counter and the result is a tie between the two alternatives.

The dependent Poisson race model can easily be extended to encompass multi-attribute alternatives. We assume there is a shared process and two unique processes associated with each of $J$ attributes. For the $j$th attribute, we denote the shared and unique process rates by $\lambda_{sj}, \lambda_{u1j}, \lambda_{u2j}, j = 1, \ldots, J$, respectively. These $3J$ Poisson processes are independent. The overall shared and unique rates for the Poisson processes associated with the two alternatives, aggregated over all attributes, are

$$
\lambda_s = \sum_{j=1}^{J} \lambda_{sj}, \quad \lambda_{u1} = \sum_{j=1}^{J} \lambda_{u1j}, \quad \lambda_{u2} = \sum_{j=1}^{J} \lambda_{u2j}.
$$

(3.2.4)

The shared process, with rate $\lambda_s$, models how close the two alternatives are in their various attributes. When they are similar, the shared rate is large and the possibility of ties is high. In the case of dominance in attributes, one of the unique rates is zero and the better alternative is always selected. In the case of non-dominance, one alternative is better than the other in some, but not all, attributes, and trade-off effects are reflected by the relative magnitudes of $\lambda_{u1}$ and $\lambda_{u2}$.

In the following theorem, we derive the choice probabilities and the probability of ties for a choice set $A = \{a_1, a_2\}$ under the dependent Poisson race model.

**Theorem 3.2.1.** Assume that the dependent Poisson race model holds with shared rate $\lambda_s \geq 0$ and unique rates $\lambda_{u1}, \lambda_{u2} \geq 0$. Assume that at least one rate is positive. Define

$$
p_s = \frac{\lambda_s}{\lambda_s + \lambda_{u1} + \lambda_{u2}}, \quad p_1 = \frac{\lambda_{u1}}{\lambda_s + \lambda_{u1} + \lambda_{u2}}, \quad \text{and} \quad p_2 = \frac{\lambda_{u2}}{\lambda_s + \lambda_{u1} + \lambda_{u2}},
$$

(3.2.5)

87
and let the positive integer $K$ be the threshold value for both alternatives in the choice set $A = \{a_1, a_2\}$. (i) The probability of a tie is

$$P(\text{tie}) = \sum_{i=1}^{K} \frac{(2K - i - 1)!}{(i-1)!(K-i)!(K-i)!} p_1^{i}p_2^{K-i}.$$  

(3.2.6)

(ii) The probability that accumulated evidence in favor of $a_1$ reaches its threshold before that of $a_2$ is

$$P(1) = \sum_{k=0}^{K-1} \frac{(K+k-1)!}{k!(K-1)!} p_1^{k}p_2^{K-k} + \sum_{i=1}^{K-1} \left[ \sum_{k=0}^{i-1} \left( \frac{(K+k-1)!}{(i-1)!(K-i)!(K-i)!} + \frac{(K+k-1)!}{i!(K-i-1)!(K-i)!} \right) p_1^{i}p_2^{K-i} \right].$$

(3.2.7)

(iii) The probability that accumulated evidence in favor of $a_2$ reaches its threshold before that of $a_1$ is

$$P(2) = \sum_{k=0}^{K-1} \frac{(K+k-1)!}{k!(K-1)!} p_1^{k}p_2^{K-k} + \sum_{i=1}^{K-1} \left[ \sum_{k=0}^{i-1} \left( \frac{(K+k-1)!}{(i-1)!(K-i)!(K-i)!} + \frac{(K+k-1)!}{i!(K-i-1)!(K-i)!} \right) p_1^{i}p_2^{K-i} \right].$$

(3.2.8)

When $K = 1$, these expressions reduce to $P(1) = p_1$, $P(2) = p_2$, and $P(\text{tie}) = p_s$.

Proof. The aggregation of three independent Poisson processes with individual rates $\lambda_s$, $\lambda_{u_1}$, $\lambda_{u_2}$ is itself a Poisson process with rate $\lambda_s + \lambda_{u_1} + \lambda_{u_2}$ (see, for example, Karlin and Taylor, 1975, Chapter 5). The individual processes can be recovered by assigning each ‘hit’ on the aggregated process to one of the individual processes. Each assignment is made at random with probabilities $p_s$, $p_1$ and $p_2$ which are defined in (3.2.5).

(i) In order for a tie to occur, the hits on the two unique counters must be the same, there must be at least one hit on the shared counter, and the last hit must be on the shared counter. Additionally, the sum of the hits on the shared counter and either of the unique counters must be $K$. The probability that exactly $i$ hits are on
the shared counter \((1 \leq i \leq K)\), that exactly \(K - i\) hits on each unique counter and that the last hit is shared is
\[
\frac{(2K - i - 1)!}{(i - 1)!(K - i)!(K - i)!} p^i_s p_1^{K-i} p_2^{K-i}.
\]
Summing over all the possible values of \(i\) yields the probability of a tie in (3.2.6).

(ii) In order for the first alternative \(a_1\) to be selected, the last hit must be either on the shared counter or on unique counter 1. The probability that there is no hit on the shared counter and that unique counter 1 reaches the threshold \(K\) before unique counter 2 is given by
\[
\sum_{k=0}^{K-1} \frac{(K + k - 1)!}{k!(K - 1)!} p_1^K p_2^k.
\] (3.2.9)

When there are \(i\) hits on the shared counter \((1 \leq i \leq K - 1)\), the alternative \(a_1\) is selected if the sum of hits on the shared counter and on unique counter 1 is \(K\) while the sum of hits on the shared counter and on unique counter 2 is at most \(K - 1\). Consider first the case where the last hit is on the shared counter. The probability that this case obtains and that alternative \(a_1\) is selected is
\[
\sum_{i=1}^{K-1} \sum_{k=0}^{K-i-1} \frac{(K + k - 1)!}{(i - 1)!(K - i)!k!} p^i_s p_1^{K-i} p_2^k.
\] (3.2.10)

Consider now the case where the last hit is on unique counter 1. The probability that this case obtains and that alternative \(a_1\) is selected is
\[
\sum_{i=1}^{K-1} \sum_{k=0}^{K-i-1} \frac{(K + k - 1)!}{i!(K - i - 1)!k!} p^i_s p_1^{K-i} p_2^k.
\] (3.2.11)

The sum of (3.2.9), (3.2.10) and (3.2.11) gives formula (3.2.7).

(iii) Expression (3.2.8) is obtained in a similar fashion to 3.2.7.

The second term in each of (3.2.7) and (3.2.8) accounts for the presence of hits on the shared counter, where the last hit occurs either on the unique counter or the
shared counter (see the proof of Theorem 3.2.1 for details). If the shared rate \( \lambda_s \) is set to zero, then \( p_s = 0 \) and (3.2.7) and (3.2.8) give the choice probabilities for \( a_1 \) and \( a_2 \) in the independent race model. Thus, the independent Poisson race model is a special case of the dependent race model. Note that \( P(1) \) and \( P(2) \) have the desirable property that they have the same order as \( \lambda_{u_1} \) and \( \lambda_{u_2} \); that is, if \( \lambda_{u_1} = \lambda_{u_2} \), then \( P(1) = P(2) \); if \( \lambda_{u_1} > \lambda_{u_2} \), then \( P(1) > P(2) \).

In a conjoint choice study with two alternatives, exactly one alternative must be chosen from each choice set. However, in the framework of the dependent race model, there is a positive probability of a tie and we allocate \( P(tie) \) proportionally to each of the two alternatives via a tie-breaking probability. Let \( p_{tb} \) be the proportion of \( P(tie) \) allocated to \( a_1 \); that is, \( p_{tb} \) is the probability that \( a_1 \) is chosen conditional on having a tied race. Then, the choice probabilities for the alternatives in a choice set \( A = \{a_1, a_2\} \) are given by

\[
P_A(1) = P(1) + p_{tb}P(tie) \quad \text{and} \quad P_A(2) = P(2) + (1 - p_{tb})P(tie),
\]

where \( P(1), P(2) \) and \( P(tie) \) are given in Theorem 3.2.1.

We adopt the following definition of \( p_{tb} \) indexed by a parameter \( \epsilon \in [0, \infty) \) which allows us to capture a range of models for tie-breaking:

\[
p_{tb} = \frac{\lambda_{u_1} + \epsilon \lambda_s}{\lambda_{u_1} + \lambda_{u_2} + 2\epsilon \lambda_s} = \frac{p_1 + \epsilon p_s}{p_1 + p_2 + 2\epsilon p_s}. \tag{3.2.13}
\]

As a result of this tie-breaking rule, the ratio of the choice probabilities \( P_A(1)/P_A(2) \) is implicitly a function of \( \epsilon \). It also has the properties described in the following theorem.
Theorem 3.2.2. Under the dependent Poisson race model with shared rate $\lambda_s \geq 0$ and unique rates $\lambda_{u_1}, \lambda_{u_2} \geq 0$, where at least one is rate is positive, the following properties hold.

1. When $\lambda_{u_1} = \lambda_{u_2}$, then $P_A(1)/P_A(2) = P(1)/P(2) = 1$.

2. When $\lambda_{u_1} > \lambda_{u_2}$, then $1 < P_A(1)/P_A(2) \leq P(1)/P(2)$. The last comparison is an equality if and only if $\min(\lambda_s, \epsilon) = 0$ and either $\lambda_{u_2} = 0$ or $\lambda_{u_2} > 0$ and $K = 1$.

3. When $\lambda_{u_1} > \lambda_{u_2}$ and $\lambda_s > 0$, then $P_A(1)/P_A(2)$ is a decreasing function of $\epsilon$.

4. The choice probabilities $(P_A(1), P_A(2))$ are a continuous function of the triple $(\lambda_s, \lambda_{u_1}, \lambda_{u_2})$.

Proof. 1. When $\lambda_{u_1} = \lambda_{u_2}$, $P(1) = P(2)$ and $p_{tb} = 0.5$, therefore, $P_A(1)/P_A(2) = P(1)/P(2) = 1$.

2. When $\lambda_{u_1} > \lambda_{u_2}$, to show that

$$\frac{P_A(1)}{P_A(2)} = \frac{P(1) + p_{tb} P(tie)}{P(2) + (1 - p_{tb}) P(tie)} \leq \frac{P(1)}{P(2)} \tag{3.2.14}$$

with the interpretation that $P(1)/0 = \infty$, it is sufficient to show that

$$p_{tb} P(2) - (1 - p_{tb}) P(1) \leq 0.$$

Replacing $p_{tb}$ with the expression in (3.2.13), we have that

$$p_{tb} P(2) - (1 - p_{tb}) P(1) = \frac{[(\lambda_{u_1} + \epsilon \lambda_s) P(2) - (\lambda_{u_2} + \epsilon \lambda_s) P(1)]/[\lambda_{u_1} + \lambda_{u_2} + 2 \epsilon \lambda_s]}.$$

This has the same sign as

$$\lambda_{u_1} P(2) - \lambda_{u_2} P(1) + \epsilon \lambda_s [P(2) - P(1)] \tag{3.2.15}.$$
It suffices to show that both $P(1) - P(2) > 0$ and $\lambda u_2 P(1) - \lambda u_1 P(2) \geq 0$. The first inequality follows from the properties of $P(1)$ and $P(2)$. To establish the second inequality, we write

\[
\lambda u_2 P(1) - \lambda u_1 P(2) = (\lambda_s + \lambda u_1 + \lambda u_2) \sum_{k=0}^{K-1} \frac{(K + k - 1)!}{k!(K-1)!} (p_1^{K-k+1} p_2^k - p_1^{k+1} p_2^K)
\]

\[
+ (\lambda_s + \lambda u_1 + \lambda u_2) \sum_{i=1}^{K-1} \sum_{k=0}^{K-i-1} \left[ \frac{(K + k - 1)!}{(i-1)! (K-i)! k!} + \frac{(K + k - 1)!}{i!(K-i-1)! k!} \right] p_i^i (p_1^{K-i} p_2^{k+1} - p_1^{k+1} p_2^{K-i}).
\]

In the first term of (3.2.16), the coefficients are all positive and

\[
p_1^{K-k+1} p_2^k - p_1^{k+1} p_2^K = p_1^{k+1} p_2^{k+1} (p_1^{K-k-1} - p_2^{K-k-1})
\]

\[
\geq 0 \quad \text{since} \quad p_1 > p_2 \quad \text{and} \quad K \geq k + 1.
\]

Similarly, in the second term of (3.2.16), the coefficients are all positive and

\[
p_1^{K-i} p_2^{k+1} - p_1^{k+1} p_2^{K-i} = p_1^{k+1} p_2^{k+1} (p_1^{K-i-k-1} - p_2^{K-i-k-1})
\]

\[
\geq 0 \quad \text{since} \quad p_1 > p_2 \quad \text{and} \quad K - i \geq k + 1.
\]

Therefore, $\lambda u_2 P(1) - \lambda u_1 P(2) \geq 0$.

To show that $P_A(1)/P_A(2) > 1$, we note that $P(1)/P(2) > 1$ and $p_{tb} > 0.5$. The conditions for equality follow from inspection of (3.2.14).

3. From (3.2.13),

\[
p_{tb} = \frac{\lambda u_1 + \epsilon \lambda_s}{\lambda u_1 + \lambda u_2 + 2 \epsilon \lambda_s}.
\]
Since $\lambda_{u_1} > \lambda_{u_2}$ and $\lambda_s > 0$, $p_{tb}$ is decreasing in $\epsilon$. Also, $1 - p_{tb}$ is increasing in $\epsilon$. From (3.2.12),

$$\frac{P_A(1)}{P_A(2)} = \frac{P(1) + p_{tb}P(tie)}{P(2) + (1 - p_{tb})P(tie)}.$$  

Both numerator and denominator are positive. The numerator is decreasing in $\epsilon$ while the denominator is increasing in $\epsilon$. Hence the ratio is decreasing in $\epsilon$.

4. From (3.2.5), $(p_s, p_1, p_2)$ is continuous in $(\lambda_s, \lambda_{u_1}, \lambda_{u_2})$. From (3.2.6)–(3.2.8), $(P(1), P(2), P(tie))$ is continuous in $(p_s, p_1, p_2)$. From (3.2.13), $p_{tb}$ is continuous in $(\lambda_s, \lambda_{u_1}, \lambda_{u_2})$. Since the composition of continuous functions is continuous, (3.2.12) implies the result.

The parameters of the model determine the choice probabilities. We note in passing that the rates themselves do not determine the choice probabilities. First, only the relative rates are relevant. The expressions (3.2.6)-(3.2.13) rely on $p_s, p_1$ and $p_2$ defined in (3.2.5). Setting time aside, the choices are determined by a sequence of independent multinomial trials that have categories “hit on both processes”, “hit on process 1 alone”, and “hit on process 2 alone”. The threshold parameter, $K$, determines when the race ends in a fashion reminiscent of the negative binomial. As $K$ increases, with $p_1/p_2 > 1$ ($\lambda_{u_1} > \lambda_{u_2}$), $P(1)$ increases; as $K \to \infty$, $P(1) \to 1$. This can be proved with arguments similar to the proof of Proposition 2.3.3 and Proposition 2.3.8. In this limiting case, the rates can be interpreted as utilities in the axiomatic sense, and the alternative with the greater utility is chosen with probability 1.

When $\epsilon = 0$, then

$$p_{tb} = \frac{\lambda_{u_1}}{\lambda_{u_1} + \lambda_{u_2}}$$

93
and the proportions of \( P(tie) \) allocated to the two alternatives are proportional to their unique rates. This tie-breaking rule implies that, when there are ties between the two dependent races, the Poisson processes continue beyond the threshold value \( K \) and the first unique counter to take the lead wins. In the case of dominance in attributes with \( a_1 \) better than \( a_2 \), this tie-breaking rule implies that \( P_A(1) = 1 \) since \( \lambda_{u_2} = 0, P(2) = 0 \) and \( p_{tb} = 1 \). Similarly, if \( a_2 \) is better than \( a_1 \), then \( P_A(1) = 0 \) and \( P_A(2) = 1 \) since \( \lambda_{u_1} = 0, P(1) = 0 \) and \( p_{tb} = 0 \).

In the limit as \( \epsilon \to \infty \), with \( \lambda_s > 0 \), then (3.2.13) tends to \( p_{tb} = 0.5 \) and ties are broken at random. In the case of dominance of \( a_1 \) over \( a_2 \), when \( \lambda_s \) is small relative to \( \lambda_{u_1} \) and \( \lambda_{u_2} \), \( P(tie) \) is small, \( P(1) \approx 1 \), and \( P_A(1) \approx 1 \). When \( \lambda_s \) is large relative to \( \lambda_{u_1} \) and \( \lambda_{u_2} \), \( P(tie) \approx 1 \). This rule then implies that \( P_A(1) \approx P(1) + 0.5P(tie) \approx 0.5 \).

When \( \epsilon \) is a small positive number greater than 0, the effect of the tie-breaking rules falls between the two extreme cases when \( \lambda_s > 0 \). That is, there is a positive probability of an individual making a “wrong” choice even when one alternative dominates the other in every attribute. This tie-breaking rule provides some flexibility in modeling choice data with a pair of alternatives with dominance relationship since some respondents might choose the dominated alternatives.

Figure 3.1 shows an example demonstrating these properties. In this example, there are two alternatives, \( a_1 \) and \( a_2 \), each with two attributes. The rate of the Poisson process associated with each attribute of the corresponding alternatives are shown in Table 3.2.1. The rates for \( a_1 \) are fixed with \( \lambda_{11} = 55 \) and \( \lambda_{12} = 60 \) for the two attributes. For the second attribute, \( \lambda_{21} = 40 \) and \( \lambda_{22} \) starts at 70 and changes in steps of size 1. In an independent model, the combined rates for \( a_1 \) and \( a_2 \) are \( \lambda_1 = 55 + 60 = 115 \) and \( \lambda_2 = 110, 111, \ldots, 119 \). Throughout, assume
$K = 2$. $P(1)/P(2)$ for the independent model is plotted in Figure 3.1 (solid line with ‘*’ markers) against the values of $\lambda_2$. In a dependent model with a shared counter, for the first attribute, the shared rate $\lambda_{s1} = \min(\lambda_{11}, \lambda_{21}) = 40$ and similarly $\lambda_{s2} = 40$. Therefore, $\lambda_s = \lambda_{s1} + \lambda_{s2} = 40 + 60 = 100$. The unique rates are $\lambda_{u1} = \lambda_{u11} + \lambda_{u12} = 15 + 0 = 15$, and $\lambda_{u2} = 10, 11, \ldots, 19$. With these given shared and unique rates, the ratio of the choice probabilities $P(1)/P(2)$ is plotted against values of $\lambda_{u2}$ in the same Figure 3.1 (solid line with ‘o’ markers). Additionally, the ratio $P_A(1)/P_A(2)$ is computed for the dependent model given $\epsilon = 0, 0.5, \infty$. This figure demonstrates the properties of the tie-breaking rule listed in Theorem 3.2.2. It also shows that the ratio of choice probabilities $P(1)/P(2)$ and $P_A(1)/P_A(2)$ under the dependent model are further from 1 than they are in the independent model with the same rates for the two alternatives. Large values of $\epsilon$ in the dependent model produce performance more similar to the independent model than do small values of $\epsilon$.

<table>
<thead>
<tr>
<th>Attribute 1</th>
<th>Attribute 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>55</td>
</tr>
<tr>
<td>$a_2$</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 3.2: Two alternatives $a_1$ and $a_2$. The entries in the table are rates associated with each attribute in the corresponding alternative.
Figure 3.1: Ratio of Choice Probabilities plotted against values of $\lambda_{u2}$. 
3.2.2 Modeling the Poisson process rates

Consider a conjoint choice experiment where each respondent evaluates \( M \) choice sets \( A_1, A_2, \ldots, A_M \), each having \( H = 2 \) alternatives; that is, \( A_m = \{a_{1m}, a_{2m}\} \), for \( m = 1, \ldots, M \). Let \( x_{jm} = \{x_{1jm}, x_{2jm}\} \) denote the levels of the \( j \)th attribute in alternatives \( a_{1m} \) and \( a_{2m} \) in choice set \( A_m \), with \( x_{hjm} \geq 0 \) and where the \textit{a priori} worst level of an attribute is always coded as 0 (\( h = 1, 2; j = 1, \ldots, J; m = 1, \ldots, M \)).

As in Section 3.2.1, the \( j \)th attribute in the two alternatives is associated with the dependent Poisson processes \( X_{1j}(t) \) and \( X_{2j}(t) \), respectively.

We require that \( \lambda_{hjm} \) be a non-negative, increasing function of \( x_{hjm} \) for \( h = 1, 2 \), so that preferred levels of an attribute make a greater contribution to the rate. An implicit assumption is that the attribute levels are ordered according to some sensible preference order, for example, “low interest rate” is preferred to “high interest rate” for a consumer. Once this (consumer-specific) function \( \lambda_{hjm} \) is established, the unique and shared rates are determined as in (3.2.1) and (3.2.2). For the \( j \)th attribute and the \( m \)th choice set, we have:

\[
\lambda_{sjm} = \min(\lambda_{1jm}, \lambda_{2jm}), \quad \lambda_{u_{1jm}} = \lambda_{1jm} - \lambda_{sjm}, \quad \text{and} \quad \lambda_{u_{2jm}} = \lambda_{2jm} - \lambda_{sjm}.
\]

The rates are aggregated over the \( J \) attributes to yield the overall shared and unique rates for the \( m \)th choice set.

There are a variety of functions which appear natural for modeling conjoint choice data. The dependent Poisson race model developed in this chapter uses the notion of a linear utility, additive across the attributes. In the analysis appearing in Section 3.6, the levels of the attributes are ascaled so that the priori worst level of each attribute across the entire experiment is coded as 0 and other levels of the attribute are recoded.
linearly between the best level and the worst level. The attribute-specific rates are modeled as

\[ \lambda_{1jm} = e^{x_{1jm}\beta_j} - 1 \quad \text{and} \quad \lambda_{2jm} = e^{x_{2jm}\beta_j} - 1, \]

\( j = 1, \ldots, J; \ m = 1, \ldots, M, \) with \( \beta_j \geq 0 \) for all \( j \) and at least one \( \beta_j > 0 \). With appropriately scaled \( x_{jhjm} \), for \( h = 1, 2 \), the parameters \( \beta_1, \ldots, \beta_J \) reflect the importance of the attributes to the consumer. The exponential transformation ensures non-negativity of the rates and echoes traditional formulations such as the multinomial logit (MNL) model (McFadden, 1974), while 1 is subtracted so that \( \lambda_{hjm} \) is zero when \( x_{hjm} = 0 \).

The aggregated shared and unique rates for the \( m \)th choice set are

\[ \lambda_{sm} = e^{\beta_0} + \sum_{j=1}^{J} \lambda_{sjm}, \]

\[ \lambda_{u1m} = \sum_{j=1}^{J} [e^{x_{1jm}\beta_j} - 1 - \lambda_{sjm}], \]

\[ \lambda_{u2m} = \sum_{j=1}^{J} [e^{x_{2jm}\beta_j} - 1 - \lambda_{sjm}]. \]

where \( \beta_0 \) is an intercept parameter for the shared rate. The intercept accounts for two features of the experiment. First, it accounts for the common shared rate across alternatives that is due to the worst level of every attribute. Second, it accounts for attributes of the alternatives that have not been described as part of the choice task. Under the assumption that the levels of these attributes are the same for all alternatives, these additional aspects contribute only to the shared rate. For example, in the credit card experiments of Section 3.5 and Section 3.6, the issuer went undescribed. When this formulation of the model is to be used, subjects should
be instructed to compare the alternatives assuming that all undescribed attributes of
the alternatives are identical.

3.2.3 Alternative models

In Sections 3.4 and 3.6, the performance of the dependent race model is compared
to that of several other models. The first alternative model is an independent Poisson
race model with no shared counter (so $\lambda_{am} = 0$) and two unique rates given by

$$
\lambda_{u1m} = e^{\beta_0} + \sum_{j=1}^{J} e^{x_{1jm}\beta_j} - 1 \quad \text{and} \quad \lambda_{u2m} = e^{\beta_0} + \sum_{j=1}^{J} e^{x_{2jm}\beta_j} - 1,
$$

where $\beta_0$ is an intercept parameter, and $\beta_j \geq 0$ for all $j = 1, \ldots, J$ and at least one
$\beta_j > 0$, and the threshold value is a positive integer $K \geq 1$. In the special case $K = 1,$
this is a logit model with an additive rate structure.

The next model is the multinomial logit model (MNL) which can be specified as a
Poisson race model with threshold value $K = 1$, no shared counter and unique rates
as follows:

$$
\lambda_{u1m} = e^{\sum_{j=1}^{J} x_{1jm}\beta_j} \quad \text{and} \quad \lambda_{u2m} = e^{\sum_{j=1}^{J} x_{2jm}\beta_j}.
$$

Notice that this model has a multiplicative rate structure. It does not have an inter-
cept term due to the identifiability issue. The comparison of the two logit models
will show if, and to what extent, the difference between an additive and a multi-
pllicative rate structure drives the results as compared with the impact of the shared
counter.
3.3 Estimation

Throughout this section, let $Y_i = (y_{i,1}, \ldots, y_{i,M})$ be the vector of responses for the $i$th respondent ($i = 1, \ldots, I$), where $y_{i,m} = 1$ if alternative $a_{1m}$ is chosen in the $m$th choice set $A_m = (a_{1m}, a_{2m})$, and $y_{i,m} = 0$ otherwise.

3.3.1 Hierarchical Bayesian Model

First, we consider the dependent race model. For respondent $i$, let the threshold value be $K_i$ for all alternatives, and the set of attribute importance parameters be $\beta_i = \{\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,J}\}$. There is also a parameter $\epsilon_i$ associated with the tie-breaking rule. Given $\Theta_i = \{K_i, \beta_i, \epsilon_i\}$, the likelihood function for respondent $i$ is

$$L(Y_i|\Theta_i) = \prod_{m=1}^{M} P_{A_m}(1|\Theta_i)^{y_{i,m}}[1 - P_{A_m}(1|\Theta_i)]^{1-y_{i,m}}, \quad (3.3.1)$$

where $P_{A_m}(1|\Theta_i)$ is the probability of choosing $a_{1m}$ from choice set $A_m$ given $\Theta_i$.

We formulate a hierarchical model to take advantage of the fact that, usually, there are relatively few observations per respondent compared with the large number of respondents in a typical conjoint experiment. Setting $\gamma_i = \log(\beta_i)$ with $\beta_i > 0$, we choose the hierarchical prior distribution of $\gamma_i$ to be multivariate normal with mean $\mu_{\gamma}$ and covariance matrix $\Sigma_\gamma$; that is,

$$[\gamma_i|\mu_\gamma, \Sigma_\gamma] \sim MVN(\mu_\gamma, \Sigma_\gamma). \quad (3.3.2)$$

We specify the hyper-prior distribution for $(\mu_\gamma, \Sigma_\gamma)$ to be Normal-Inverse-Wishart

$$[\mu_\gamma, \Sigma_\gamma] \sim NIW(A, d, a, c), \quad (3.3.3)$$
where $a$ is a vector of length $(J + 1)$, $c$ and $d$ are positive constants and $A$ is a $(J + 1) \times (J + 1)$ matrix. The joint prior distribution in (3.3.3) can be written as

$$[\Sigma_\gamma] \sim IW(A, d), \quad (3.3.4)$$

$$[\mu_\gamma|\Sigma_\gamma] \sim N(a, c^{-1}\Sigma_\gamma). \quad (3.3.5)$$

With this specification, $E[\Sigma_\gamma] = (d - (J + 1) - 1)^{-1}A$, where $J + 1$ is the dimension of $\Sigma_\gamma$ (see O’Hagan, 1994, Chapter 10).

The prior distribution of the threshold value $K_i$ is:

$$[K_i|\eta] \sim Poisson(\eta) + 1, \quad (3.3.6)$$

where $K_i, i = 1, \ldots, I$ are iid random variables. Although a hyper-prior could be placed on $\eta$ to quantify heterogeneity in $K_i$, we assume, for simplicity, that the distribution of heterogeneity for $K_i$ is known with $\eta$ fixed. $\eta$ is set equal to 1 in our application which corresponds to a prior mean of $E[K_i] = 2, i = 1, \ldots, I$.

In the tie-breaking rule (3.2.13) for the dependent race model, the prior distribution of $\epsilon_i$ is

$$log(\epsilon_i) \sim Normal(\mu_\epsilon, \sigma_\epsilon^2), \quad (3.3.7)$$

where $\epsilon_i$ are iid random variables. For simplicity, we also assume the distribution of heterogeneity for $\epsilon_i$ is known with $\mu_\epsilon = -6$ and $\sigma_\epsilon = 3$. This prior assigns mass to small $\epsilon_i$ which, in our experience, provides good results to predictive performance of the models.

To compare the different tie-breaking mechanisms discussed in Section 3.2.1, fixed values of $\epsilon_i$ for all $I$ respondents are used instead of specifying a prior distribution. In the simulation study in Section 3.4, values of $\epsilon_i$ equal to 0.0001, 0.001, 0.01,
0.1, 1 and ∞ are used. The case where $\epsilon_i = 0$ is not investigated since the likelihood would be zero if a respondent were to choose a dominated alternative erroneously and the posterior distribution of the parameters could not then be obtained. In Section 3.6, the use of a prior distribution of $\epsilon_i$ as well as the fixed values $\epsilon_i = 0.001$ and $\epsilon_i = \infty$ for $i = 1, \ldots, I$ are compared.

The same prior distributions for $\gamma_i$ and $K_i$ listed from (3.3.2) to (3.3.6) are used for the independent race model. For the MNL model and the independent model with $K_i = 1$, $K_i$ does not need to be estimated. The MNL model has one fewer dimension in the $\gamma_i$ vector since there is no intercept term and, consequently, the dimensions of $\mu_\gamma$ and $\Sigma_\gamma$ are each reduced by one.

### 3.3.2 Posterior distributions and MCMC procedures

The model parameters are estimated using Markov chain Monte Carlo (MCMC) methods (see Chen, Shao & Ibrahim (2000) for a recent presentation of MCMC methods). The parameter vector $\mu_\gamma$ and matrix $\Sigma_\gamma$ are generated using Gibbs sampling methods while $\epsilon_i$, $\gamma_i$ and $K_i$ are estimated using Metropolis-Hastings algorithms. The details are discussed below.

Let $I$ be the total number of respondents. Uncertainty about the distribution of $\gamma_i$ in the population is captured by the prior distribution on $(\mu_{\gamma}, \Sigma_{\gamma})$. Given these parameters, the model is one of conditional independence across individuals. The parameters $\gamma_i$, $K_i$ and $\epsilon_i$, $i = 1, \ldots, I$ are respondent-specific. If $(\mu_{\gamma}, \Sigma_{\gamma})$ were known, $\gamma_i$, $K_i$ and $\epsilon_i$ would be updated using the data from the $i$th respondent alone.
The prior distributions of parameters \( \gamma_i, K_i \) and \( \epsilon_i \) are described in Section 3.3.1. For all I respondents, the joint prior distribution can be written as

\[
[\gamma | \mu_\gamma, \Sigma_\gamma][\mu_\gamma, \Sigma_\gamma][\epsilon][K],
\]

where \( \gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_I\} \), \( K = \{K_1, K_2, \ldots, K_I\} \), and \( \epsilon = \{\epsilon_1, \ldots, \epsilon_I\} \). The probability density function \( [\mu_\gamma, \Sigma_\gamma] \) is given in (3.3.3), the densities \( [\gamma | \mu_\gamma, \Sigma_\gamma] \) and \([\epsilon]\) are the product of densities in (3.3.2) and (3.3.7) across the I respondents, and the probability mass function \( [K] \) is the product of (3.3.6) across the I respondents.

The data are \( Y = \{Y_1, Y_2, \ldots, Y_I\} \), and they contribute a likelihood which is the product of the likelihoods (3.3.1) over \( i \). This leads to the posterior distribution, up to the unknown normalizing constant,

\[
[\gamma, K, \epsilon, \mu_\gamma, \Sigma_\gamma | Y] \propto [Y | \gamma, K, \epsilon, \mu_\gamma, \Sigma_\gamma][\gamma | \mu_\gamma, \Sigma_\gamma][\mu_\gamma, \Sigma_\gamma][\epsilon][K].
\]

To compare the effectiveness of various models for out-of-sample prediction, we replace the full likelihood for an individual with a portion of the likelihood, updating with only some of the choice sets. The form of the posterior distribution based on this reduced sample is the same as that based on the full sample.

This posterior distribution is analytically intractable, and so we make use of MCMC methods both to explore the posterior and to compute estimates of various quantities. The algorithm that we use combines a Gibbs step, where the form of the conditional generation is known and Metropolis-Hastings steps for distributions are not available in closed form. The algorithm is as follows:

1. Initialize the sampler. Draw \( \mu_\gamma^{(0)} \) and \( \Sigma_\gamma^{(0)} \) as a random draw from its prior distribution, the \( NIW(A, d, a, c) \). For \( \gamma_i^{(0)} \), a random vector is drawn from
Set the initial values for $K_i^{(0)} = 1$, $i = 1, \ldots, I$. For $\epsilon_i^{(0)}$, initial values are drawn from the prior distribution, the lognormal distribution with parameters $\mu_\epsilon$ and $\sigma_\epsilon^2$.

2. Generate the hyperparameters with a Gibbs step. For the $n$th iteration, generate $\Sigma_{\gamma}^{(n)}|\gamma^{(n-1)}$ from (3.3.8) and $\mu_{\gamma}^{(n)}|\Sigma_{\gamma}^{(n)}$ from (3.3.9).

For our model, $(\mu_\gamma, \Sigma_\gamma)$ is conditionally independent of $Y$ and the rest of the parameters, given $\gamma$. Following O’Hagan (1994, Chapter 10),

$$[\mu_\gamma, \Sigma_\gamma | \gamma] \sim NIW(A^*, d^*, a^*, c^*),$$

where

$$A^* = A + IS + cI(c + I)^{-1}(a - \bar{\gamma})(a - \bar{\gamma})' , \quad S = 1/I \sum_{i=1}^{I} (\gamma_i - \bar{\gamma})(\gamma_i - \bar{\gamma})' ,$$

$$d^* = d + I , \quad a^* = (c + I)^{-1}(ca + I\bar{\gamma}) , \quad c^* = c + I .$$

The conditional distributions are

$$[\Sigma_\gamma | \gamma] \sim IW(A^*, d^*) , \quad (3.3.8)$$

$$[\mu_\gamma | \Sigma_\gamma] \sim N(a^*, c^*^{-1}\Sigma_\gamma) . \quad (3.3.9)$$

3. Generate respondent-specific parameters, for $i = 1, \ldots, I$, with a series of Metropolis-Hastings steps.

(i) Define $\epsilon^* = log(\epsilon)$. Generate $\epsilon_i^{*(q)}$, a candidate value for $\epsilon_i^*$, from a random walk chain, such that

$$\epsilon_i^{*(q)} \sim N\left(\epsilon_i^{*(n-1)}, \nu \sigma_\epsilon^2\right) ,$$
where $\nu$ is a small positive number (we took $\nu = 0.4$ in our examples).

The acceptance probability is computed as

$$\alpha_{\epsilon_i} = \min \left( \frac{[Y_i|\epsilon_i^{(q)}, K_i^{(n-1)}, \gamma_i^{(n-1)}, \mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}]|\epsilon_i^{(q)}|\mu_\epsilon, \sigma_\epsilon^2}{[Y_i|\epsilon_i^{(n-1)}, K_i^{(n-1)}, \gamma_i^{(n-1)}, \mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}]|\epsilon_i^{(n-1)}|\mu_\epsilon, \sigma_\epsilon^2}, 1 \right).$$

Then, with $u \sim U(0, 1)$, set

$$\epsilon_i^{(n)} = \begin{cases} 
\epsilon_i^{(q)} & \text{if } u < \alpha_{\epsilon_i}, \\
\epsilon_i^{(n-1)} & \text{if } u > \alpha_{\epsilon_i}.
\end{cases}$$

(ii) Generate $\gamma_i^{(q)}$, a candidate value for $\gamma_i$, from a random walk chain, such that

$$\gamma_i^{(q)} \sim N \left( \gamma_i^{(n-1)}, \nu \Sigma_\gamma^{(n)} \right),$$

where $\nu$ is a small positive number (we took $\nu = 0.4$ in our examples).

The acceptance probability is computed as

$$\alpha_{\gamma_i} = \min \left( \frac{[Y_i|\gamma_i^{(q)}, K_i^{(n-1)}, \epsilon_i^{(n)}, \mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}]|\gamma_i^{(q)}|\mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}}{[Y_i|\gamma_i^{(n-1)}, K_i^{(n-1)}, \epsilon_i^{(n)}, \mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}]|\gamma_i^{(n-1)}|\mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}}}, 1 \right).$$

Then, with $u \sim U(0, 1)$, set

$$\gamma_i^{(n)} = \begin{cases} 
\gamma_i^{(q)} & \text{if } u < \alpha_{\gamma_i}, \\
\gamma_i^{(n-1)} & \text{if } u > \alpha_{\gamma_i}.
\end{cases}$$

(iii) Generate $K_i^{(q)}$, a candidate value for $K_i$, from a mixture of a random walk chain and an independence proposal distribution, such that

$$K_i^{(q)} \sim K^{(n-1)} \pm 1, \text{ with } p = 0.1,$$

$$K_i^{(q)} \sim \text{Poisson}(\eta) + 1, \text{ with } p = 0.9.$$

If $K^{(n-1)} = 1$ and $K_i^{(q)} - 1 = 0$, $K_i^{(q)}$ is set to 1 since $K_i^{(q)}$ has to be positive.

The acceptance probability for $K_i^{(q)}$ if drawn from the random walk chain is

$$\alpha_{K_i} = \min \left( \frac{[Y_i|K_i^{(q)}, \gamma_i^{(n)}, \epsilon_i^{(n)}, \mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}]|K_i^{(q)}|\eta}{[Y_i|K_i^{(n-1)}, \gamma_i^{(n)}, \epsilon_i^{(n)}, \mu_\gamma^{(n)}, \Sigma_\gamma^{(n)}]|K_i^{(n-1)}|\eta}}, 1 \right).$$
The acceptance probability for $K_i^{(q)}$ if drawn from the independence chain is

$$\alpha_{K_i} = \min \left( \frac{[Y_i|K_i^{(q)}, \gamma_i^{(n)}, \epsilon_i^{(n)}, \mu_{\gamma}^{(n)}, \Sigma_{\gamma}^{(n)}]}{[Y_i|K_i^{(n-1)}, \gamma_i^{(n)}, \epsilon_i^{(n)}, \mu_{\gamma}^{(n)}, \Sigma_{\gamma}^{(n)}]}, 1 \right).$$

Then, with $u \sim U(0, 1)$, set

$$K_i^{(n)} = \begin{cases} K_i^{(q)} & \text{if } u \leq \alpha_{K_i}, \\ K_i^{(n-1)} & \text{if } u > \alpha_{K_i}. \end{cases}$$

Repeat steps 2 and 3 for $N$ iterations.

### 3.3.3 Calculation of choice probabilities-a short cut

The formulae (3.2.5)–(3.2.8) used to calculate $P(tie)$, $P(1)$ and $P(2)$ and the likelihood in (3.3.1) must be evaluated at each iteration of the MCMC chain for each respondent. The most time consuming part of this calculation, computation of the many coefficients in the sums, is repeated many, many times across respondents and over iterations. The technique of pre-computation can be used to streamline the MCMC routine. The idea of pre-computation is to handle expensive, repetitive calculations a single time (typically before the iterative algorithm begins), storing the needed results for future, quick use. Such a simplification is described below.

For any given value $K_i = K$, there are $K(K + 2)$ possible combinations of hits on the shared counter and the two unique counters, as follows. Exactly $K$ combinations correspond to ties, with $k$ hits on the shared counter and $K - k$ hits on each unique counter. There are $K + (K - 1) + \ldots + 1$ combinations that lead to the counter for $a_1$ having reached $K$ first and the same number of combinations that lead to the counter for $a_2$ having reached $K$ first. Thus there are a total of $K(K + 2)$ cases.
Now let $\tilde{N} = \{\tilde{N}_s, \tilde{N}_1, \tilde{N}_2\}$ be a matrix of size $K(K+2) \times 3$, with $\tilde{N}_s$, $\tilde{N}_1$, and $\tilde{N}_2$ denoting vectors of possible hits at the time when the threshold is reached. The $j$th row of $\tilde{N}$ is denoted by $\{\tilde{N}_s(j), \tilde{N}_1(j), \tilde{N}_2(j)\}$. Additionally, a matrix $\tilde{I} = \{\tilde{I}_s, \tilde{I}_1, \tilde{I}_2\}$ of indicators is used to record whether a row of $\tilde{N}$ corresponds to a tie or a chosen alternative. For example, if a row $j$ in $\tilde{N}$ corresponds to a tie, then the $j$th row of $\tilde{I}$ is $\tilde{I}_s(j) = 1$, $\tilde{I}_1(j) = 0$ and $\tilde{I}_2(j) = 0$.

The hits on the three counters that precede the threshold-reaching hit could have arrived in any sequence. The number of different sequences is given by the multiplicative coefficients within the summation terms in Theorem 3.2.1. A vector $\tilde{C}$ of length $K(K+2)$ is used to store these coefficients for every combinations of hits. The $j$th element of $\tilde{C}$ is set to be

$$\tilde{C}(j) = \begin{cases} 
\frac{(\tilde{N}_s(j) + \tilde{N}_1(j) + \tilde{N}_2(j) - 1)!}{(\tilde{N}_s(j) - 1)!\tilde{N}_1(j)!\tilde{N}_2(j)!} & \text{if } \tilde{I}_s(j) = 1 \\
\frac{(\tilde{N}_s(j) + \tilde{N}_1(j) + \tilde{N}_2(j) - 1)!}{(\tilde{N}_s(j) - 1)!\tilde{N}_1(j)!\tilde{N}_2(j)!} I(\tilde{N}_s(j) \neq 0) & \text{if } \tilde{I}_1(j) = 1 \\
\frac{(\tilde{N}_s(j) + \tilde{N}_1(j) + \tilde{N}_2(j) - 1)!}{\tilde{N}_1(j)!\tilde{N}_2(j)!} & \text{if } \tilde{I}_2(j) = 1 \\
\end{cases}$$

where $j = 1, \ldots, K(K+2)$ and $I(\tilde{N}_s(j) \neq 0) = 1$ if the condition in parenthesis is true. For any given $K$, the vectors $\tilde{C}$, $\tilde{N}$, $\tilde{I}$ are the same. Thus, they can be calculated in advance and stored in a multidimensional array.

In each MCMC iteration, for each respondent, when $K$, $p_s$, $p_1$ and $p_2$ are specified, a vector

$$\tilde{p} = \{p_s^{\tilde{N}_s}, p_1^{\tilde{N}_1}, p_2^{\tilde{N}_2}\}$$

of length $K(K+2)$ is computed whose $j$th element is $\{p_s^{\tilde{N}_s(j)}, p_1^{\tilde{N}_1(j)}, p_2^{\tilde{N}_2(j)}\}$. 

107
Then, the expression in Theorem 3.2.1 can be written as

\[ P(tie) = \tilde{C}'[\tilde{p} \cdot \tilde{I}_s], \tag{3.3.10} \]
\[ P(1) = \tilde{C}'[\tilde{p} \cdot \tilde{I}_1], \tag{3.3.11} \]
\[ P(2) = \tilde{C}'[\tilde{p} \cdot \tilde{I}_2]. \tag{3.3.12} \]

where the operator ‘•’ is defined as the element-wise product of two vectors or matrices.

### 3.4 Simulation study

A simulation study is used to examine the performance of the MCMC algorithm and the difference between the models. For the design of this simulation study, every respondent evaluates \( M = 6 \) choice sets with two alternatives each. The alternatives have \( J = 4 \) attributes. The design matrix is given in Table 3.3. In this design, sets 5 and 6 are dominant sets since their second alternative dominates their first alternative. We assume that \( I = 51 \) respondents participate in the study. The true model is assumed to be a dependent model with \( \epsilon_i = 0.001, i = 1, \ldots, I. \)

To simulate a set of choice data \( Y = \{y_1, \ldots, y_{51}\} \), the true parameter values need to be set first. The true threshold values \( K_i \) for the respondents are all simulated from a shifted Poisson distribution (3.3.6) with \( \eta = 1. \) For the true \( \gamma = \{\gamma_1, \ldots, \gamma_{51}\}, \) assume \( \gamma_i \sim MVN(\mu_\gamma, \Sigma_\gamma) \), we first set \( \mu_\gamma = [0.69, 1.18, 0.40, -2.10, 0.018] \). Purposely, the first attribute is set to have the largest coefficient which is 1.18 and the third attribute to have the smallest coefficient which is -2.10. Also, we assume \([\mu_\gamma, \Sigma_\gamma] \sim NIW(A, d, a, c)\). Given \( \mu_\gamma \), a random realization of \( \Sigma_\gamma \) is simulated from the conditional distribution \([\Sigma_\gamma | \mu_\gamma] \sim IW(A_{\mu_\gamma}, d + 1)\) and is used as the true \( \Sigma_\gamma \).
Here $A_{\mu} = A + c(\mu - a)(\mu - a)'$ with $a = 0$, $c = 1/9$, $d = 20$, and $A = 14I_5$. The parameters $a, c, d, A$ are set arbitrarily based on the dimensions of $\mu$, $\Sigma$. Then, we have a realization as the true $\Sigma$ and

$$\Sigma_{\gamma} = \begin{pmatrix} 1.3205 & -0.2197 & 0.6317 & 0.4022 & 0.3287 \\ -0.2197 & 0.7244 & -0.2925 & -0.2305 & -0.1106 \\ 0.6317 & -0.2925 & 1.2386 & 0.5403 & 0.2131 \\ 0.4022 & -0.2305 & 0.5403 & 1.3148 & 0.1943 \\ 0.3287 & -0.1106 & 0.2131 & 0.1943 & 1.0839 \end{pmatrix}. $$

Given $\mu, \Sigma$, a random matrix $\gamma = \{\gamma_1, \ldots, \gamma_I\}$ is simulated from multivariate normal distribution $MVN(\mu, \Sigma_{\gamma})$ and used as the true $\gamma$ to simulate the choice data. Let $\bar{\gamma}_{true}$ denote the average of the true $\gamma_i, i = 1, \ldots, 51$. Also, $\bar{\gamma}_{true} = [0.632, 1.047, 0.300, -1.972, -0.163]$. These numbers suggest that, in our simulated data, the first attribute is the most important attribute and the third attribute is the least important attribute.
Given values for the vectors $K_i$ and $\gamma_i$, $i = 1, \ldots, I$, the probabilities of choosing $a_{1m}$ and $a_{2m}$ for choice set $A_m$, $m = 1, \ldots, M$, are computed by (3.2.16), Theorem 3.2.1 and, for the dependent Poisson race model, tie-breaking rule (3.2.13) with fixed $\epsilon = 0.001$. Since there are 51 respondents each evaluating 6 choice sets, we let a $51 \times 6$ matrix $P_{\text{true}}$ contain the probabilities for the 51 respondents of choosing $a_{1m}$ in choice sets $A_m$, $m = 1, \ldots, 6$. With these probabilities, the response data $Y = \{y_1, \ldots, y_{51}\}$ are generated as a realization of Bernoulli trials. The simulated data are summarized in Table 3.4. In this data set, although a large proportion of respondents choose the dominant alternative $a_{2m}$ in the dominant sets 5 and 6, some respondents do choose the non-dominant alternative $a_{1m}$.

<table>
<thead>
<tr>
<th></th>
<th>Choosing $a_1$</th>
<th>Choosing $a_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>40 (0.78)</td>
<td>11 (0.22)</td>
</tr>
<tr>
<td>Set 2</td>
<td>33 (0.65)</td>
<td>18 (0.35)</td>
</tr>
<tr>
<td>Set 3</td>
<td>43 (0.84)</td>
<td>8 (0.16)</td>
</tr>
<tr>
<td>Set 4</td>
<td>31 (0.61)</td>
<td>20 (0.39)</td>
</tr>
<tr>
<td>Set 5</td>
<td>2 (0.04)</td>
<td>49 (0.96)</td>
</tr>
<tr>
<td>Set 6</td>
<td>10 (0.20)</td>
<td>41 (0.80)</td>
</tr>
</tbody>
</table>

Table 3.4: Simulated Data. Each row shows the number out of 51 respondents who choose alternatives $a_1$ and $a_2$, with corresponding proportions of respondents in parentheses.

This set of simulated data is analyzed with the dependent Poisson race models with the same fixed value of $\epsilon_i = \epsilon$ for all respondents where $\epsilon = 0.0001, 0.001, 0.01, 0.1, 1$.
and ∞ (denoted by Dep(ε)), the independent Poisson race model (denoted by Indep), its special case with $K = 1$ (denoted by Indep($K = 1$)), and the MNL model.

The dependent Poisson race models and the independent Poisson race models use the same proper but weakly informative priors given in Section 3.3.1. The prior distribution for $[\mu_\gamma, \Sigma_\gamma]$ is $NIW(A, d, a, c)$ with $a = 0$, $c = 1/9$, $d = 20$, $A = 14I_5$, which corresponds to prior expectations of $E[\mu_\gamma | \Sigma_\gamma] = 0$ and $E[\Sigma_\gamma] = I_5$ (see (3.3.4) and (3.3.5)). Our choice of $c = 1/9$ ensures that the prior plays a minimal role in the estimation of $\mu_\gamma$, for all but minimal amounts of heterogeneity, $\Sigma_\gamma$. The prior for $K_i$ is $Po(1)+1$. For the Indep($K = 1$) model the same priors are used but the thresholds $K_i$ are fixed to 1. For the MNL model we set $E[\Sigma_\gamma] = I_4$ and so use $A = 15I_4$. Each of the 10 independent MCMCs used to quantify simulation error is run for 55000 iterations. Posterior summaries are based on the last 50000 iterations.

Let $\bar{P}_{im}$ be the posterior mean of the choice probabilities for alternative $a_{1m}$ for the $i$th respondent and the $m$th choice set. Then, the mean squared error ($MSE$) over all respondents for the $m$th choice set is calculated by

$$MSE_m = \frac{1}{51} \sum_{i=1}^{51} (\bar{P}_{im} - P_{true,im})^2$$

for $m = 1, \ldots, 6$. Summing over $m = 1, \ldots, 6$ gives the overall $MSE$. These statistics are given in Table 3.5, together with the corresponding standard errors of the $MSE$ statistics over 10 iterations.

Overall, the dependent race models have smaller $MSE$, and so recover the data generating probabilities better than the three independence models (see the last column of Table 3.5). The comparison between Indep and Indep($K = 1$) suggests that allowing for $K_i > 1$ is important. The Indep($K = 1$) model with an additive rate structure performs worse than the MNL model in all choice sets. This indicates that
Table 3.5: Mean squared errors for the Dep($\epsilon = 0.0001, 0.001, 0.01, 0.1, 1, \infty$) model, Indep model, Indep($K = 1$) model and the MNL model, with standard errors over 10 MCMC chains in parentheses.

<table>
<thead>
<tr>
<th>Model</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
<th>Set 4</th>
<th>Set 5</th>
<th>Set 6</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep ($\epsilon = 0.0001$)</td>
<td>0.0197</td>
<td>0.0508</td>
<td>0.0139</td>
<td>0.0483</td>
<td>0.0117</td>
<td>0.0147</td>
<td>0.1593</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0010)</td>
<td>(0.0002)</td>
<td>(0.0006)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.00018)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.001$)</td>
<td>0.0204</td>
<td>0.0536</td>
<td>0.0122</td>
<td>0.0499</td>
<td>0.0087</td>
<td>0.0144</td>
<td>0.1592</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0006)</td>
<td>(0.0001)</td>
<td>(0.0005)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.01$)</td>
<td>0.0231</td>
<td>0.0553</td>
<td>0.0111</td>
<td>0.0497</td>
<td>0.0065</td>
<td>0.0184</td>
<td>0.1642</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0008)</td>
<td>(0.0001)</td>
<td>(0.0006)</td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.1$)</td>
<td>0.0298</td>
<td>0.0532</td>
<td>0.0112</td>
<td>0.0477</td>
<td>0.0081</td>
<td>0.0268</td>
<td>0.1768</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0001)</td>
<td>(0.0002)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 1$)</td>
<td>0.0385</td>
<td>0.0496</td>
<td>0.0110</td>
<td>0.0463</td>
<td>0.0120</td>
<td>0.0381</td>
<td>0.1954</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0003)</td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>Dep ($\epsilon \rightarrow \infty$)</td>
<td>0.0413</td>
<td>0.0494</td>
<td>0.0115</td>
<td>0.0469</td>
<td>0.0142</td>
<td>0.0420</td>
<td>0.2052</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0001)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Indep</td>
<td>0.0520</td>
<td>0.0437</td>
<td>0.0104</td>
<td>0.0449</td>
<td>0.0159</td>
<td>0.0604</td>
<td>0.2273</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0002)</td>
<td>(0.0004)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>0.0596</td>
<td>0.0840</td>
<td>0.0169</td>
<td>0.0696</td>
<td>0.0170</td>
<td>0.0749</td>
<td>0.3219</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0004)</td>
<td>(0.0001)</td>
<td>(0.0002)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.0306</td>
<td>0.0810</td>
<td>0.0163</td>
<td>0.0665</td>
<td>0.0115</td>
<td>0.0373</td>
<td>0.2433</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0006)</td>
</tr>
</tbody>
</table>

The significant differences in the $MSE$ statistics are not due to the general superiority of an additive rate structure. The independent Poisson race model (Indep) has slightly smaller overall $MSE$ than the MNL model but the improvement in overall $MSE$ is relatively small compared with the improvement shown by the dependent Poisson race models.

The range of $\epsilon$ values studied helps to assess the effect of tie-breaking rules. The overall $MSE$ becomes smaller as $\epsilon$ decreases. The models with $\epsilon = 0.0001$ and 0.001 have almost the same overall $MSE$ but the values of the $MSE$ for the different choice.
sets differ slightly. The models with $\epsilon = 0.01$ and 0.1 have larger $MSE$ but are still considerably better than the MNL and independent race models. The models with $\epsilon = 1$ are very close to the model with $\epsilon = \infty$. It appears that the tie-breaking rule has a strong effect on the $MSE$ statistics. As $\epsilon$ increases, the dependent race model becomes closer to the independent race model and the MNL model. The simulation also suggests that there exists a range of tie-breaking rules that will lead to reasonably good performance of the dependent Poisson race model.

If we look at the two dominant choice sets (sets 5 and 6), the dependent model with $\epsilon \leq 0.1$ generally has smallest $MSE$ (except for set 5 when $\epsilon = 0.0001$), followed by the MNL model and then the dependent model with $\epsilon = 1$ or $\infty$. The two independent models have the worst $MSE$s. For the non-dominant choice set 1, the group of dependent models have the best $MSE$s, MNL model falls in the middle while the independent race models perform least well. For sets 2, 3, 4, the independent race model with different $K_i$, values has the best $MSE$s; the MNL model has the second worst $MSE$s (slightly better than the Indep($K = 1$) model). The group of dependent race models falls in the middle and the models with large $\epsilon$ usually have smaller $MSE$s since they are more similar to the independent race model.

The results of the overall $MSE$ does show that the dependent race model with $\epsilon = 0.001$ has the smallest $MSE$, which is consistent with the model from which the data set is simulated. What we learn from the simulation study is that there are significant differences in the $MSE$ statistics between the dependent Poisson race models, the independent Poisson race model and the MNL model. This illustrates that the dependent race model implies a response surface that cannot be recovered
well by models that do not account for the dependence in the alternative specific counters.

Table 3.6 shows the posterior estimated $\mu_\gamma$ from different models. All models suggest that the first attribute has the largest coefficient while the third attribute has the smallest coefficient. However, not even the dependent race model with $\epsilon = 0.001$ recover the $\bar{\gamma}_{true}$ well, which is not surprising since the size of the simulated data set is fairly small. Note that the $\mu_\gamma$ for all these models are not directly comparable since $K_i, i = 1, \ldots, 51$ are different for different models and since the MNL model has one fewer dimension than the other models.

We conclude that the dependent race model with a specific tie breaking mechanism may be distinguished empirically from other data generating mechanisms. Based on the sizeable $MSE$ differences among the models in the simulation study, we expect substantial gains from the application of the proposed model to real data should our model be closer to the data generating mechanism. In order to test the proposed model in a practical setting, we conducted a conjoint experiment on credit cards as described in Sections 3.5 and 3.6.

### 3.5 A choice data set constructed from rating data as a pilot study

Choice-experiments are usually designed with the explicit goal of avoiding dominance relationships. As a pilot study to compare the dependent race model to the independent race models and the two logit models, a set of choice data was constructed from a set of rating data which was readily available (Liu, 2006). Note that this type of data presents a serious challenge to the dependent Poisson race model since respondents may be less likely to realize dominance relationships in rating tasks.
Table 3.6: Estimated $\mu, \gamma$ for all models. The numbers in the parentheses are the standard errors over 10 MCMC chains. $\gamma_{\text{true}}$ is the average of $\gamma_i, i = 1, \ldots, 51,$ of the true model from which the choice data are simulated from. It seems that the first attribute has the largest coefficient while the third attribute has the smallest coefficient.

than in choice tasks. If evidence for respondents’ recoganition of dominance relationship in choice sets is found, it will provide strong empirical support for the practical relevance of accurately modeling dominance relationships. Since the data collection involved ratings of 12 different credit cards described on a single sheet of paper, respondents may or may not have engaged in attribute-wise processing to facilitate the consistency of their ratings.
3.5.1 Data description

The respondents were asked to rate the credit cards on a scale from 0 to 10, with 0 being extremely unappealing and 10 being extremely appealing. The design of the twelve credit cards is given in Table 3.7. There are 55 respondents who participated in the study, and all data records are complete. To construct choice data, the twelve cards are put into six pairs and that the card with the higher rating is assumed to be chosen, had this been a choice study.

<table>
<thead>
<tr>
<th>Card</th>
<th>Interest Rate</th>
<th>Annual Fee</th>
<th>Travel Points</th>
<th>Cash Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>9.99</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>9.99</td>
<td>$30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>17.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>17.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>17.99</td>
<td>$30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>14.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>14.99</td>
<td>$30</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>14.99</td>
<td>$30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>12.99</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>12.99</td>
<td>$30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>12.99</td>
<td>$30</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 3.7: Design matrix for the credit card data.

Exploratory data analysis is used to identify any respondents with highly unexpected rating patterns. The ratings of individual respondents are plotted against the mean ratings given by all respondents. Several respondents are identified as suspicious because their ratings decrease as the mean ratings increase, as shown in Figure 3.2 for the 52th respondent. It is very likely that this respondent reversed the scale of
the ratings and the data should not be used. In this way, four suspicious respondents are identified and deleted. The data for the remaining 51 respondents are used.
Figure 3.2: A suspicious consumer. His/her ratings for the cards decrease as the mean ratings increase.
For the design in Table 3.7, it is obvious that a dominance relationship does exist between some cards. For example, card 10 dominates card 9 in every attribute and card 12 dominates card 11 in every attribute. Respondents' ratings seem to be affected by dominance relationships among cards as illustrated in Figure 3.3. To create this plot, first the twelve cards are ordered by mean ratings in descending order. Then, we consider all \( \binom{12}{2} = 66 \) pairs of cards. In each pair, the cards are ordered such that the first card has a higher mean rating than the second card. Therefore, the difference in mean ratings between any two cards is positive. For all 66 possible combinations of any two cards, the fraction of respondents who rate the first card higher than the second card is plotted against the mean of the difference between the two cards. Among these pairs, some have dominance relationship between two cards. The points corresponding to dominant pairs are marked in Figure 3.3. For larger differences in mean ratings \((\geq 3.5)\), the fraction of respondents who give higher rating to the first card on average is large for both dominant and non-dominant sets. This is expected since the first card in each pair has a higher mean rating. However, for medium to small difference in mean rating \((\leq 3.5)\), the fraction of respondents who give higher ratings to the first card is larger in the dominant sets than in the non-dominant sets. It seems that, on average the respondents might recognize the dominance relationships in the pair of cards even in the case where one card is only slightly more favorable than the other. This finding supports the idea that respondents at least, in part, engaged in attribute-wise processing, possibly to ensure the internal consistency of their ratings.
Figure 3.3: Evidence of dominance. For all 66 possible combinations of any two cards, the fraction of respondents who rate the first card higher than the second card is plotted against the mean of the difference between the two card ratings.
A set of choice data is constructed from the rating data as follows. For each respondent, the twelve credit cards are separated into 6 choice sets of size 2. The attributes are recoded with the worst level denoted by zero and the best level denoted by 1. The attribute with intermediate levels are all quantitative and the levels are scaled linearly between the best and the worst levels. The design matrix of the choice sets is given in Table 3.3.

Some respondents give the same ratings to the two cards in a choice set. We treat these ties as missing data. For the remaining data, one choice set is held out randomly for each respondent and this holdout sample is not used to fit the models. Two holdout samples are constructed. For the first holdout sample, three people chose the non-dominant alternative in the dominant sets. For the second, no respondent chose the non-dominant alternatives in the dominant sets.

We fit the models described in Section 3.3 with the MCMC method, without the holdout samples. The first 5000 iterations are discarded as burn-in while the next 50000 iterations are used in estimation. For the dependent Poisson race model with $\epsilon = 0.0001, 0.001, 0.01, 0.1, 1, \infty$ and the independent Poisson race model, the prior distribution for $[\mu, \Sigma]$ is $NIW(A, d, a, c)$ with $a = 0$, $c = 1/9$, $d = 20$, $A = 14I_5$. The prior distribution for $\gamma$ is $MVN(\mu_\gamma, \Sigma_\gamma)$. The prior distribution for $K$ is $Po(1) + 1$. For the Indep($K = 1$) model, the prior distribution for $\gamma$ is the same as above, but $K = 1$. The prior distribution of $\gamma$ for the MNL model has only four dimensions, and thus $A = 15I_4$.

The models can be compared in terms of statistics for the holdout samples. For predictive power, we expect a good model to have small mean squared error (MSE)
and high correct choice rate and correct choice probabilities (i.e. probability of selecting a priori better card). A third measure is the joint likelihood for the holdout sample given the predicted choice probabilities. Note that in the holdout sample, some are dominant sets while others are not. It is interested to compare these measures separately for the dominant sets and for the non-dominant sets with a trade-off relationship as well.

3.5.2 Measures of holdout performance

Before we define the measures used to quantify the models’ performance on the holdout data, some notation is needed. Several binary choice sets for each respondent are held out to form a holdout sample, which is not used in fitting the models but in validating the model’s predictive performance. Let \( y = \{y_1, \ldots, y_I\} \) represent the holdout sample. Here \( I = 51 \). Let the total number of choice sets held out for each respondent be \( N_{ho} \). This is the same for all respondents in this study. Denote the collection of the indexes of choice sets by \( \mathcal{H}_i, i = 1, \ldots, N_{ho} \). The holdout choice sets can be selected randomly for each respondent so that \( \mathcal{H}_i \) might vary from respondent to respondent. Then, \( y_i = \{y_{im}, m \in \mathcal{H}_i\} \) is a vector representing the choice data held out for the \( i \)th respondent, where \( y_{im} = 1 \) if the respondent chooses the first credit card for the \( m \)th choice set and \( y_{im} = 0 \) otherwise. For sets \( m = 1, \ldots, 12 \), we use \( \mathcal{D} \) to denote the collection of indexes for dominant pairs and \( \mathcal{E} = \{m \notin \mathcal{D}\} \) for the collection of indexes for non-dominant pairs. From the design of the this study, \( \mathcal{D} = \{5, 6\} \). Also, let \( N_{dom} \) and \( N_{nondom} \) be the total number of dominant and non-dominant choice sets, respectively, in the holdout sample for all respondents together. Notice that \( I \times N_{ho} = N_{dom} + N_{nondom} \).
For a given model, let \( P_{A_{im}}(1|\Theta_{in}) \) be the probability of choosing \( a_{1m} \) from set \( A_{m} \) for the \( i \)th consumer on the \( n \)th MCMC iteration, where \( m \) indexes choice sets, \( i = 1, \ldots, I \), \( n = 1, \ldots, N \) and \( N \) is total number of MCMC iterations after burn-in. This is the predicted choice probability for the first card in the \( m \)th choice set for the \( i \)th respondent. \( \bar{P}_{A_{im}}(1) \) is the posterior mean of \( P_{A_{im}}(1|\Theta_{in}) \).

Next, the holdout statistics used to measure the predictive performance of the models are computed for the holdout sample \( y \).

First, two mean squared error statistics. Small values of mean squared error indicate better prediction of the choice probabilities for the holdout sample. These are now defined.

Mean squared error: \( MSE_1 \) is based on the posterior mean \( \bar{P}_{A_{im}}(1) \) and defined as follows for the holdout data and the qualitatively different subsets thereof:

\[
MSE_1(\text{overall}) = \frac{1}{I \times N_{ho}} \sum_{i=1}^{I} \sum_{m \in H_i} (\bar{P}_{A_{im}}(1) - y_{im})^2,
\]

\[
MSE_1(\text{dom}) = \frac{1}{N_{dom}} \sum_{i=1}^{I} \sum_{m \in (D \cap H_i)} (\bar{P}_{A_{im}}(1) - y_{im})^2,
\]

\[
MSE_1(\text{nondom}) = \frac{1}{N_{nondom}} \sum_{i=1}^{I} \sum_{m \in (E \cap H_i)} (\bar{P}_{A_{im}}(1) - y_{im})^2.
\]

Note that \( m \in (D \cap H_i) \) and \( m \in (E \cap H_i) \) collect holdout choice sets with and without a dominating alternative, respectively. A second \( MSE \) statistic, \( MSE_2 \), takes full account of the uncertainty in the posterior distribution and is defined as follows:

\[
MSE_2(\text{overall}) = \frac{1}{N \times I \times N_{ho}} \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{m \in H_i} (P_{A_{im}}(1|\Theta_{in}) - y_{im})^2.
\]

For brevity, both here and in the following, the additional expressions for the choice sets that contain a dominant alternative and those that do not are omitted. \( MSE_2 \)
reflects the variation of $P_{A_{im}}(1|\Theta_{in})$ across iterations of the MCMC and is therefore larger than $MSE_1$.

**Correct choice rate:** $R_{cor}$ is the percent of times that a model correctly predicts an observed response applying a zero-one loss. The prediction is given by

$$
\hat{y}_{im} = \begin{cases} 
1 & \text{if } P_{A_{im}}(1) \geq 0.5, \ m \in H_i, \\
0 & \text{if } P_{A_{im}}(1) < 0.5, \ m \in H_i.
\end{cases}
$$

The correct choice rate $R_{cor}$ is then obtained as the normalized count of the number of times that the elements in $\hat{Y}$ are equal to those in $Y$:

$$
R_{cor} = \frac{1}{I \times N_{ho}} \sum_{i=1}^{I} \sum_{m \in H_i} I(y_{im}, \hat{y}_{im}),
$$

where $I(k, j) = 1$ if $k = j$ and $I(k, j) = 0$ otherwise.

**Correct choice probability:** $P_{cor}$ is defined as the average predicted choice probability of the alternatives chosen in the holdout sample, and computed as follows:

$$
P_{cor} = \frac{1}{I \times N_{ho}} \sum_{i=1}^{I} \sum_{m \in H_i} [y_{im} \bar{P}_{A_{im}}(1) + (1 - y_{im})(1 - \bar{P}_{A_{im}}(1))].
$$

**Joint predictive likelihood:** $L_{\text{holdout}}$ of the holdout sample, is the predictive probability of the holdout data given the data used to fit the model; that is,

$$
L_{\text{holdout}} = \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{I} \prod_{m \in H_i} [P_{A_{im}}(1|\Theta_{in})^{y_{im}}(1 - P_{A_{im}}(1|\Theta_{in}))(1 - y_{im})].
$$

(3.5.1)

The estimated probabilities for the dominant and non-dominant sets are computed for $m$ multiplying over $m \in D \cap H_i$ and $m \in E \cap H_i$, respectively. However, as the size of the holdout sample increases, the joint likelihood approaches zero quickly. The standard error of $L$ is large compared to the mean of $L$ over multiple MCMC chains. In such cases, comparisons of model performance based on $L$ may not be reliable.
Quasi-likelihood of the holdout data: $QL_{\text{holdout}}$ is based on the predictive probability for each choice task, $\bar{P}_{Aim}(1)$. The predictive probabilities for the holdout data are summarized by their geometric mean or, alternatively, by their product (an unnormalized version of the geometric mean):

$$QL_{\text{holdout}} = \prod_{i=1}^{I} \prod_{m \in \mathcal{H}_i} [\bar{P}_{Aim}(1)^{y_{im}}(1 - \bar{P}_{Aim}(1))^{(1 - y_{im})}].$$

The quasi-likelihood is far easier to estimate accurately than is the joint predictive density of the holdout data. The quasi-likelihood can be used to compute quasi-Bayes factors, and it can be interpreted in much the same way as the joint predictive density.

With the above notation, the normalization is the same for all models to be compared.

3.5.3 Results of the pilot study

For the set of data described in Section 3.5.1, $MSE_1$, $MSE_2$, $R_{cor}$ and $L_{\text{holdout}}$ are computed. The summary of $MSE_1$ and $MSE_2$ for the first holdout sample is listed in Table 3.8. The dependent Poisson race models with $\epsilon \leq 0.01$ have smaller $MSE_1(\text{overall})$ than the MNL model. The dependent Poisson race models with $\epsilon \geq 0.1$ are similar in terms of $MSE_1(\text{overall})$ and they are worse than the MNL model. The two independent race models have the largest MSEs. For $MSE_1(\text{dom})$, the same pattern is observed. For $MSE_1(\text{nondom})$, the MNL model and the $\text{Dep}(\epsilon = 0.0001)$ model seem to have the smallest values, followed by the dependent race models with $\epsilon = 0.001, 0.01$. The other dependent race models with $\epsilon \geq 0.1$ and the independent race models have relatively large $MSE_1$ values. For $MSE_2(\text{overall})$ and $MSE_2(\text{dom})$, again, the dependent race model with $\epsilon \leq 0.01$ have smaller values than the MNL model which is followed by the other dependent race models and then the independent
race models. For $MSE_2(nondom)$, the MNL model and the Dep($\epsilon = 0.0001$) model have the same smallest value. Other dependent race models have larger values than the MNL model but smaller than the independent race models.

Table 3.9 summarizes $P_{cor}$ for the first holdout sample. The dependent Poisson race model with $\epsilon \geq 0.01$ have the largest correct choice probabilities overall, followed by the MNL model, the other dependent race models and the independent race models. For the dominant sets, $P_{cor}(dom)$ for the dependent race models with $\epsilon \leq 0.1$ are the best followed by MNL and then the other models. For the non-dominant sets, the correct choice probability of the Dep($\epsilon = 0.001$) model is about the same as that of the MNL model, followed by other dependent race models and the independent models.

The likelihood ratio based on the first holdout sample is also listed in Table 3.9. By comparing the likelihood values of all models with that of the independent race model, we obtain a set of likelihood ratios. These likelihood ratios are summarized in Table 3.10. For the overall likelihood ratio, the smaller the $\epsilon$, the larger the ratio, which indicates that the predictive performance is improved dramatically. The performance of the MNL model falls in between the dependent models with $\epsilon = 0.01$ and 0.1. For dominant sets, the dependent models with $\epsilon \leq 0.01$ are considerably better than the other models which have likelihood ratios close to 1. For non-dominant sets, the dependent race models with smaller $\epsilon(\leq 0.001)$ are slightly better than the MNL model. The likelihood ratios for the non-dominant sets are quite close to 1. Among all the models, the Indep($K = 1$) model seems to have the poorest predictive power.
Table 3.8: MSE statistics for the first holdout sample, with standard errors of the statistics over 10 iterations in parentheses. The dependent Poisson race model with \( \epsilon \leq 0.01 \) has smaller \( MSE_1(overall) \) and \( MSE_2(overall) \) than the MNL model. The smaller \( \epsilon \) is, the better the performance of the dependent model. The Independent models are the worst models in terms of \( MSE \).
Table 3.9: Correct choice probability and likelihood of the first holdout sample, with standard errors of the statistics over 10 iterations in parentheses. The dependent Poisson race model with $(\epsilon \geq 0.01)$ have the largest correct choice probabilities and likelihood for the holdout sample.
Table 3.10: Likelihood ratio of the first holdout sample for all models vs. Indep model. The dependent Poisson race model with ($\epsilon \leq 0.01$) have the largest likelihood ratio.
The above analysis is repeated with a second holdout sample. The results are listed in Table 3.11, Table 3.12 and Table 3.13. For $MSE_1$ (overall), all the dependent race models beat the MNL model which is followed by the independent race models. For $MSE_1$ (dom), the dependent race models with $\epsilon \leq 0.01$ have smaller MSE than the MNL model. The independent models again perform least well. For the non-dominant sets, the MNL model and the indep($K = 1$) model have the largest MSE while the dependent race models with small $\epsilon$ perform best again. A similar pattern is observed for all $MSE_2$ and $P_{cor}$ statistics. For the likelihood and the likelihood ratio, the dependent race model with small $\epsilon$ has a remarkably large overall likelihood and likelihood ratio compared with other models. For the dominant sets, the MNL model also has a relatively large likelihood and likelihood ratio. For the non-dominant sets, the dependent race model with small $\epsilon$ has moderately improved performance over that of the independent race models.

The statistics computed for the two holdout samples suggest that the dependent race models provide better predictive performance than the traditional MNL models and the independent race models. When the dominance relationship among alternatives is recognized by respondents and reflected in the data, a model that can take into account the dependence relationship leads to large improvement of predictive performance of the model. Among all the statistics, the two independent race models almost always have the worst performance, as they do not take into account the dependence structure among the alternatives. The MNL model seems to perform better on average than the independent race model, especially in the non-dominant sets. But it still can not capture the dominance relationship as the dependent race models do.

130
The performance of the dependent race models depends heavily on the tie-breaking rule, expressed through $\epsilon$. For data with small $\epsilon$, the choice probabilities become more extreme, as discussed in Section 3.2.1. Thus, smaller $\epsilon$ values are preferred for modeling data with dominance relationships. There is a range of $\epsilon$ values that can lead to reasonably good models. For large $\epsilon$ values, the dependent race models become similar to the independent race models. The predictive performance of the dependent Poisson race models thus gets closer to that of the independent models and the MNL models.
<table>
<thead>
<tr>
<th>Model</th>
<th>(MSE_1) (overall)</th>
<th>(MSE_1) (dom)</th>
<th>(MSE_1) (nondom)</th>
<th>(MSE_2) (overall)</th>
<th>(MSE_2) (dom)</th>
<th>(MSE_2) (nondom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep ((\epsilon = 0.0001))</td>
<td>0.1098</td>
<td>0.0117</td>
<td>0.1841</td>
<td>0.1564</td>
<td>0.0364</td>
<td>0.2475</td>
</tr>
<tr>
<td></td>
<td>(2.7E-04)</td>
<td>(2.3E-04)</td>
<td>(3.7E-04)</td>
<td>(4.9E-04)</td>
<td>(3.6E-04)</td>
<td>(6.7E-04)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>0.1116</td>
<td>0.0116</td>
<td>0.1875</td>
<td>0.1522</td>
<td>0.0331</td>
<td>0.2426</td>
</tr>
<tr>
<td></td>
<td>(2.8E-04)</td>
<td>(3.4E-04)</td>
<td>(5.7E-04)</td>
<td>(1.1E-03)</td>
<td>(8.5E-04)</td>
<td>(1.4E-03)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.01))</td>
<td>0.1162</td>
<td>0.0159</td>
<td>0.1923</td>
<td>0.1538</td>
<td>0.0339</td>
<td>0.2447</td>
</tr>
<tr>
<td></td>
<td>(3.8E-04)</td>
<td>(5.2E-04)</td>
<td>(8.3E-04)</td>
<td>(1.4E-03)</td>
<td>(1.0E-03)</td>
<td>(1.7E-03)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.1))</td>
<td>0.1256</td>
<td>0.0283</td>
<td>0.1994</td>
<td>0.1672</td>
<td>0.0469</td>
<td>0.2584</td>
</tr>
<tr>
<td></td>
<td>(6.1E-04)</td>
<td>(4.6E-04)</td>
<td>(1.3E-03)</td>
<td>(1.4E-03)</td>
<td>(1.2E-03)</td>
<td>(1.8E-03)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 1))</td>
<td>0.1339</td>
<td>0.0404</td>
<td>0.2048</td>
<td>0.1804</td>
<td>0.0607</td>
<td>0.2711</td>
</tr>
<tr>
<td></td>
<td>(9.2E-04)</td>
<td>(5.9E-04)</td>
<td>(1.4E-03)</td>
<td>(5.6E-04)</td>
<td>(7.4E-04)</td>
<td>(9.4E-04)</td>
</tr>
<tr>
<td>Dep ((\epsilon \to \infty))</td>
<td>0.1366</td>
<td>0.0461</td>
<td>0.2052</td>
<td>0.1873</td>
<td>0.0687</td>
<td>0.2772</td>
</tr>
<tr>
<td></td>
<td>(7.3E-04)</td>
<td>(4.2E-04)</td>
<td>(1.2E-03)</td>
<td>(5.3E-04)</td>
<td>(7.7E-04)</td>
<td>(7.3E-04)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.1405</td>
<td>0.0574</td>
<td>0.2036</td>
<td>0.1987</td>
<td>0.0833</td>
<td>0.2862</td>
</tr>
<tr>
<td></td>
<td>(1.1E-03)</td>
<td>(6.2E-04)</td>
<td>(1.6E-03)</td>
<td>(3.8E-04)</td>
<td>(6.3E-04)</td>
<td>(8.6E-04)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.1557</td>
<td>0.0591</td>
<td>0.2290</td>
<td>0.2203</td>
<td>0.0885</td>
<td>0.3204</td>
</tr>
<tr>
<td></td>
<td>(4.3E-04)</td>
<td>(6.3E-04)</td>
<td>(2.9E-04)</td>
<td>(1.2E-03)</td>
<td>(3.6E-04)</td>
<td>(2.2E-03)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.1402</td>
<td>0.0255</td>
<td>0.2271</td>
<td>0.1813</td>
<td>0.0393</td>
<td>0.2891</td>
</tr>
<tr>
<td></td>
<td>(2.9E-04)</td>
<td>(4.3E-04)</td>
<td>(2.6E-04)</td>
<td>(8.7E-05)</td>
<td>(4.7E-04)</td>
<td>(3.3E-04)</td>
</tr>
</tbody>
</table>

Table 3.11: \(MSE\) statistics for the second holdout sample, with standard errors over 10 iterations in parentheses. Overall, the dependent race models have better performance in \(MSE\)s than the MNL and the independent models.
Table 3.12: Correct choice probabilities and likelihood of the second holdout sample, with standard errors over 10 iterations in parentheses. Overall, the dependent race models have better performance $P_{cor}$ than the MNL and the independent models, especially in the dominant sets. For the likelihood, the dependent race model with small $\epsilon$ has a remarkably large overall likelihood compared with other models.
<table>
<thead>
<tr>
<th>Model</th>
<th>$LR(\text{overall})$</th>
<th>$LR(\text{dom})$</th>
<th>$LR(\text{nondom})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep ($\epsilon = 0.0001$)</td>
<td>100.49</td>
<td>21.37</td>
<td>5.48</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.001$)</td>
<td>107.37</td>
<td>20.62</td>
<td>4.74</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.01$)</td>
<td>47.52</td>
<td>13.36</td>
<td>4.03</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.1$)</td>
<td>9.81</td>
<td>5.50</td>
<td>2.09</td>
</tr>
<tr>
<td>Dep ($\epsilon = 1$)</td>
<td>4.05</td>
<td>2.41</td>
<td>1.54</td>
</tr>
<tr>
<td>Dep ($\epsilon \rightarrow \infty$)</td>
<td>2.14</td>
<td>1.70</td>
<td>1.50</td>
</tr>
<tr>
<td>Indep</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>0.07</td>
<td>1.26</td>
<td>0.11</td>
</tr>
<tr>
<td>MNL</td>
<td>0.50</td>
<td>6.38</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table 3.13: Likelihood ratio of the second holdout sample of all models vs. Indep model. The dependent race model with small $\epsilon$ has a remarkably large overall likelihood ratio compared with other models. The MNL model also has a relatively large likelihood ratio in the dominant sets. For the non-dominant sets, the dependent race model with small $\epsilon$ has moderately improved performance over that of the independent race models.
3.6 A Conjoint choice data Set

A conjoint choice study is designed to collect a set of data on consumer preference of credit card types. In this study, there are 18 choice sets with two credit cards each. Each credit card is described by four attributes: interest rate, annual fee, travel points and cash reward.

3.6.1 Design and data description

Our choice of designs for this study is in keeping with Busemeyer & Wang (2000) suggestion that model comparisons should be done via design points that differ qualitatively. Here, we included choice sets that pose a trade-off problem to the decision maker and also choice sets where one alternative dominates. Any accurate account of human choice should not only be capable of handling dominant alternatives but should also be able to extrapolate from trade-off situations to dominance relationships and vice versa.

With this goal in mind, we constructed the first 12 choice sets using the SAS macro %choiceff by Kuhfeld(2005). This macro produces D-efficient designs for fixed-effects MNL models and takes a point summary of prior information about the vector of coefficients as input. As input to the design procedure, we used the results of the pilot study that suggest that interest rate is the most important attribute, followed by annual fee and cash reward, and that travel points are relatively less important to respondents. The chosen design is given in the first 12 choice sets in Table 3.14 and is a reasonable compromise to an efficient design when using the prior information and when all coefficients are assumed to be zero in the MNL model. Although this design generating procedure is a crude approximation to an optimal design in the presence of
heterogeneity amongst individuals, this choice of design points should level the effect of experimental design for model comparisons, possibly favoring the MNL model. Note that the design choice includes set 4, where the first alternative is dominant. If the MNL model is the data generating mechanism, dominance is not an issue because the model assumes strictly alternative-based processing. The remaining 6 choice sets are not based on a particular design criterion but are included for a cross-validation study. They are selected from the design used for the pilot study. Among these, sets 12, 17 and 18 each contain a dominant alternative.
<table>
<thead>
<tr>
<th>Set</th>
<th>Interest rate(%)</th>
<th>Annual fee $</th>
<th>Travel points</th>
<th>Cash reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.99</td>
<td>30</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>14.99</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>12.99</td>
<td>30</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>12.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>17.99</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>9.99</td>
<td>30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>14.99</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>17.99</td>
<td>30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>14.99</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>12.99</td>
<td>30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>14.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>17.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>9.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>17.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>17.99</td>
<td>30</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>17.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>9.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>9.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>12.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>14.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>14.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>14.99</td>
<td>30</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>14.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>14.99</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>9.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>9.99</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>9.99</td>
<td>30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>17.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>17.99</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>17.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>16</td>
<td>14.99</td>
<td>0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>14.99</td>
<td>30</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>14.99</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>12.99</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>18</td>
<td>12.99</td>
<td>30</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>12.99</td>
<td>30</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 3.14: Design matrix of the conjoint choice data experiment. First 12 sets are constructed by SAS and the last 6 sets are from the pilot study.
A total of 90 undergraduate marketing students with no prior exposure to conjoint analysis participated in the study. The survey used to collect information from the respondents is presented in Appendix B. A set of demographic questions is asked at the end of the study and used to assess whether a respondent gives considered answers. Based on their response to which attribute they valued most and least, and their choice decisions for the dominant pairs, two of the respondents seem to make choices randomly and are eliminated from further analysis. For respondents who pick the “worse” card in a dominant pair, if there is no obvious evidence that they make random selection during the study, we keep their data. The resulting data are summarized in Table 3.15. From the data summary, we see that several respondents do choose the “worse” card in sets 4, 12, 17 and 18 with dominance relationship.
Table 3.15: Summary of the conjoint choice data.

<table>
<thead>
<tr>
<th>Set</th>
<th>Counts</th>
<th>Proportion</th>
<th>Set</th>
<th>Counts</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>0.216</td>
<td>10</td>
<td>85</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>0.784</td>
<td></td>
<td>3</td>
<td>0.034</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>0.625</td>
<td>11</td>
<td>32</td>
<td>0.364</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>0.375</td>
<td></td>
<td>56</td>
<td>0.636</td>
</tr>
<tr>
<td>3</td>
<td>48</td>
<td>0.545</td>
<td>12</td>
<td>3</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.455</td>
<td></td>
<td>85</td>
<td>0.966</td>
</tr>
<tr>
<td>4</td>
<td>86</td>
<td>0.977</td>
<td>13</td>
<td>57</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.023</td>
<td></td>
<td>31</td>
<td>0.352</td>
</tr>
<tr>
<td>5</td>
<td>77</td>
<td>0.875</td>
<td>14</td>
<td>52</td>
<td>0.591</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.125</td>
<td></td>
<td>36</td>
<td>0.409</td>
</tr>
<tr>
<td>6</td>
<td>82</td>
<td>0.932</td>
<td>15</td>
<td>80</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.068</td>
<td></td>
<td>8</td>
<td>0.091</td>
</tr>
<tr>
<td>7</td>
<td>59</td>
<td>0.670</td>
<td>16</td>
<td>57</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.330</td>
<td></td>
<td>31</td>
<td>0.352</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>0.625</td>
<td>17</td>
<td>2</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>0.375</td>
<td></td>
<td>86</td>
<td>0.977</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>0.125</td>
<td>18</td>
<td>2</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>0.875</td>
<td></td>
<td>86</td>
<td>0.977</td>
</tr>
</tbody>
</table>
3.6.2 Model comparison and results

A total of six models are evaluated: the dependent model with $\epsilon_i, i = 1, \ldots, I$ ($I = 88$), being estimated, denoted by Dep($\epsilon$ estimated)); the dependent race model with $\epsilon_i = \epsilon = 0.001$ for all $i = 1, \ldots, 88$, denoted by Dep($\epsilon = 0.001$); the dependent race model with $\epsilon_i \to \infty$, denoted by Dep($\epsilon \to \infty$); the independent model, denoted by Indep; the independent model with $K_i = 1, i = 1, \ldots, 88$, which is the logit model with an additive rate structure, denoted by Indep($K = 1$); the most widely used MNL model with a multiplicative rate structure. These models are analyzed with the MCMC method described in Section 3.3. In the analysis, the attributes are recoded with a priori worst level denoted by zero and a priori best level denoted by 1. The intermediate levels are scaled linearly between the best and the worst levels. For the dependent Poisson race models and the independent Poisson race model, the prior distribution for $[\mu_\gamma, \Sigma_\gamma]$ is $NIW(A, d, a, c)$ with $a = 0, c = 1/9, d = 20, A = 14I_5$. The prior distribution for $\gamma_i$ is $MVN(\mu_\gamma, \Sigma_\gamma)$. The prior distribution for $K_i, i = 1, \ldots, 88$ is $Po(1)+1$. For the Dep($\epsilon$ estimated) model, the prior for $\epsilon_i$ is given by (3.3.7). For the Indep($K = 1$) model, the prior distribution for $\gamma_i$ is the same as above but $K_i = 1, i = 1, \ldots, 88$. The prior distribution for $\gamma_i$ in the MNL model has only four dimensions and thus $A = 15I_4$.

To evaluate the predictive performance, the data set described in Section 3.6.1 is divided into a calibration sample and a holdout sample. The calibration sample is used to fit the models and the holdout sample is used to compare the predictive performance of the models. The calibration sample and the holdout sample are chosen in four different ways. (i) The first holdout sample consists of six randomly selected choice sets for each respondent, with the remaining twelve sets are used as
the calibration sample. Both dominant and non-dominant pairs may be present in the calibration and holdout samples. (ii) The second holdout sample consists of all dominant pairs plus two randomly chosen non-dominant pairs. The remaining twelve non-dominant pairs are used as the calibration sample. Since only non-dominant choice sets are used to fit the models, as is common in most choice studies, the results might provide insight on the performance of the models in predicting choice probabilities under standard conjoint designs. (iii) In the third holdout sample, only the last six choice sets are used for all respondents as the calibration sample and use the first twelve sets as the holdout sample. The last six choice sets are the same sets we used in the pilot study in Section 3.5. This third split is another cross validation exercise since both calibration and holdout data contain choice sets with dominant alternatives. In contrast to the first split, however, the amount of calibration data is relatively small. (iv) The fourth holdout sample consists of the last six choice sets. The first twelve sets are used to fit the models. Since the first twelve sets are a D-efficient design for the MNL model, it is interesting to see whether the MNL model outperforms other models in predicting the holdout sample.

We run 10 independent MCMCs for each data set to quantify simulation error. The number of iterations in each of the independent MCMCs is 55000 where the first 5000 iterations are discarded as burn-in. Since the third split described above results in a relatively small calibration sample, we use 110000 iterations per MCMC run and discard 10000 run as burn-in in this case. A model with relatively better performance is characterized by smaller $MSE_1$ and $MSE_2$, and larger hit probability, $P_{cor}$, correct choice rate $R_{cor}$, and quasi-likelihood $QL_{holdout}$. The joint likelihoods $L_{holdout}$ were computed for all models, but the computational results might not be stable since the
magnitude of the standard errors are large relative to the likelihoods. In this case, the more stable alternative measure, the logarithm of the quasi-likelihood and its standard error are computed.

For the first holdout sample with six randomly selected choice sets for each respondent, the summaries of $MSE_1$, $MSE_2$, $P_{cor}$ and $R_{cor}$ are listed in Table 3.16. The Dep($\epsilon$ estimated) model has smallest $MSE_1$ and $MSE_2$. The mean of $\epsilon_i$ over all $I = 88$ respondents is 0.20. As $\epsilon_i$ goes to infinity, the $MSE$ statistics approach, but are smaller than, those of the Indep model. The Indep($K = 1$) model is the worst among all models. For the the MNL model, the results are mixed. Its $MSE_1$ (overall) is slightly better than the Dep($\epsilon \to \infty$) model but its $MSE_2$ (overall) is worse than the Indep model. For both $MSE_1$ (dom) and $MSE_2$ (dom), the MNL model is almost as good as the two best dependent models, but its $MSE_1$ (nondom) and especially $MSE_2$ (nondom) statistic for the non-dominant pairs are not as good. The values in the parentheses are standard errors of these $MSE$ statistics over 10 MCMC chains. Since these standard errors are very small, the difference in the $MSE$s between different models are statistically significant.

For $P_{cor}$ and $R_{cor}$, the Dep($\epsilon$ estimated) model and Dep($\epsilon = 0.001$) model seems to have the highest correct choice probabilities and hit rates. The MNL model performs quite well as compared to these two models. Relatively, the MNL model has slightly better performance in the dominant sets than in the non-dominant sets. The Dep($\epsilon \to \infty$) model follows the MNL model but its performance is closer to the two independent models which have smaller $P_{cor}$ and $R_{cor}$.
<table>
<thead>
<tr>
<th>Model</th>
<th>(MSE_1)</th>
<th>(MSE_2)</th>
<th>(P_{cor})</th>
<th>(R_{cor})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall</td>
<td>dom</td>
<td>nondom</td>
<td>overall</td>
</tr>
<tr>
<td>Dep ((\epsilon) estimated)</td>
<td>0.108</td>
<td>0.045</td>
<td>0.128</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>(9.7E-05)</td>
<td>(2.2E-05)</td>
<td>(9.8E-05)</td>
<td>(9.8E-05)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>0.109</td>
<td>0.047</td>
<td>0.128</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>(4.1E-05)</td>
<td>(1.6E-05)</td>
<td>(5.3E-05)</td>
<td>(4.8E-05)</td>
</tr>
<tr>
<td>Dep ((\epsilon \rightarrow \infty))</td>
<td>0.120</td>
<td>0.051</td>
<td>0.141</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>(5.0E-05)</td>
<td>(5.1E-05)</td>
<td>(5.8E-05)</td>
<td>(7.7E-05)</td>
</tr>
<tr>
<td>Indep</td>
<td>0.128</td>
<td>0.062</td>
<td>0.149</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>(6.8E-05)</td>
<td>(4.4E-05)</td>
<td>(8.0E-05)</td>
<td>(7.8E-05)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.144</td>
<td>0.074</td>
<td>0.166</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>(4.4E-05)</td>
<td>(8.2E-05)</td>
<td>(3.9E-05)</td>
<td>(7.7E-05)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.118</td>
<td>0.046</td>
<td>0.141</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>(3.3E-05)</td>
<td>(9.5E-06)</td>
<td>(4.4E-05)</td>
<td>(1.2E-04)</td>
</tr>
</tbody>
</table>

Table 3.16: \(MSE_1\), \(MSE_2\), \(P_{cor}\), and \(R_{cor}\) for the first holdout sample. The numbers in the parentheses are the standard errors over 10 independent MCMC chains. The Dep(\(\epsilon\) estimated) model has smallest \(MSE_1\) and \(MSE_2\). The Indep(\(K = 1\)) model is the worst among all models. The MNL model is generally worse than the dependent models. For \(P_{cor}\) and \(R_{cor}\), the Dep(\(\epsilon\) estimated) model and Dep(\(\epsilon = 0.001\)) model seems to have the highest correct choice probabilities and correct choice rates. The MNL model performs quite well as compared to these two models. Relatively, the MNL model has slightly better performance in the dominant sets than in the non-dominant sets. The Dep(\(\epsilon \rightarrow \infty\)) model follows the MNL model but its performance is closer to the two independent models which have smaller \(P_{cor}\) and \(R_{cor}\).
The logarithm of Quasi-likelihood for each model is given in Table 3.17. The three dependent models have the largest \( \log(QL_{\text{holdout}}) \). The MNL model is better than the Indep model but worse than the dependent models. The Indep\((K = 1)\) model is the worst among all. The quasi-likelihood ratio based on the quasi-likelihood of all models vs. the Indep model shows the magnitude of the remarkable difference between the models. For the overall quasi-likelihood ratio, the smaller the \( \epsilon \), the larger the quasi-likelihood ratio, which indicates that the predictive performance is improved dramatically. The Indep\((K = 1)\) model performs worse than the Indep model. The MNL model falls between the dependence and independent models.

The estimated posterior means of \( \mu_\gamma \), listed in Table 3.18, provide an indication of the importance of the attribute. A large positive number suggests that the attribute is weighted heavily in the respondents’ choice decisions, while a small positive or a negative value suggests that the attribute has relatively small influence in the decision making. From the table, it seems that interest rate is considered as the most important attribute as it has the largest positive coefficient in all models. Annual fee is the second most important attribute. The attribute “cash rewards” has the third largest coefficients, but both Dep\((\epsilon \text{ estimated})\) and Dep\((\epsilon = 0.001)\) models suggest that it is not important for most respondents since the coefficients are negative. all the dependent models suggest that “travel points” is not important while the estimates by other models suggest that it does play some role in the decision making. It is important to note that these coefficients are not directly comparable across models as the \( K_i \) for each model is different.
Table 3.17: The logarithm of quasi-likelihood ($\log QL_{\text{holdout}}$) and quasi-likelihood ratio ($QLR_{\text{holdout}}$) for the first holdout sample. The numbers in the parentheses are the standard errors over 10 independent MCMC chains. The three dependent models have the largest $\log(QL_{\text{holdout}})$. The MNL model is better than the Indep model but worse than the dependent models. The Indep($K = 1$) model is the worst among all. The quasi-likelihood ratio based on the quasi-likelihood of all models vs. the Indep model shows the magnitude of remarkable difference between the models.
Table 3.18: Posterior means of $\mu_\gamma$ in all models for the first holdout sample. The numbers in the parentheses are the standard errors over 10 MCMC chains. It seems that interest rate is considered as the most important attribute as it has the largest positive coefficient in all models. Annual fee is the second most important attribute. The attributes, “cash rewards” and “travel points” are relatively less important.
Next, the holdout statistics are computed based on the second holdout sample consisting of all dominant pairs plus two randomly chosen non-dominant pairs. In Table 3.19, for both $MSE_1$ and $MSE_2$, in all cases, the Dep($\epsilon$ estimated) and Dep ($\epsilon = 0.001$) models have similar values and are the best among all models. In the Dep($\epsilon$ estimated) model, the mean of $\epsilon_i$ over all $I = 88$ respondents is 0.18. As $\epsilon_i$ goes to infinity, $MSE$ increases but is still smaller than the Indep model. The Indep($K = 1$) model is worse than all other models. The MNL model has smaller $MSE_1(overall)$ but larger $MSE_2(overall)$ than the Dep($\epsilon \to \infty$) model. It is also interesting to note that the MNL model has small $MSE_1$ and $MSE_2$ statistics in predicting the choice probabilities in the dominant pairs but the same statistics for the non-dominant pairs are not as good. Again, the dependent model has higher correct choice probability $P_{cor}$ and higher correct choice rate $R_{cor}$. The Indep and the Indep($K = 1$) models perform less well than the group of dependent models. The $P_{cor}(overall)$ and $P_{cor}(dom)$ of the MNL model are higher than the Dep($\epsilon \to \infty$) model, but these two statistics are much smaller than that of the Dep($\epsilon \to \infty$) in predicting the choice probability for the non-dominant sets. For the correct choice rate $R_{cor}$, the MNL model is only slightly better than the Indep($K = 1$) model.

Table 3.20 lists the $\log(QL_{holdout})$ statistic for the second set of holdout data. In all three cases, the Dep($\epsilon$ estimated) is the best model with the largest $\log(QL_{holdout})$ value, followed by the Dep($\epsilon = 0.001$) model and then the Dep($\epsilon \to \infty$) model. The Indep model and Indep($K = 1$) model have small $QL_{holdout}$ values. The $QL_{holdout}(overall)$ of the MNL model is larger than the Dep($\epsilon \to \infty$) model. It is also interesting to note that the MNL model has larger $QL_{holdout}$ value than the Dep($\epsilon = 0.001$) model in the dominant sets but smaller than that of the Indep model in the non-dominant
sets. The likelihood ratios based on the quasi-likelihood of the all models vs. the Indep model again show the remarkable magnitude of the difference between different models.

The estimated posterior means of $\mu_\gamma$ are listed in Table 3.21. Again, in all models, interest rate is considered the most important attribute as it has the largest positive coefficients. Annual fee is the second most important attribute. The attribute, cash rewards, has the third largest coefficients, but the Dep($\epsilon$ estimated) and Dep($\epsilon = 0.001$) models suggest that it is not important as the coefficients are nearly zero or negative. For travel points, all models except the Indep($K = 1$) and MNL model suggest that it is not important.
<table>
<thead>
<tr>
<th>Model</th>
<th>$MSE_1$ overall</th>
<th>$MSE_1$ dom</th>
<th>$MSE_1$ nondom</th>
<th>$MSE_2$ overall</th>
<th>$MSE_2$ dom</th>
<th>$MSE_2$ nondom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep ($\epsilon$ estimated)</td>
<td>0.054</td>
<td>0.025</td>
<td>0.111</td>
<td>0.068</td>
<td>0.030</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>(2.2E-05)</td>
<td>(1.1E-05)</td>
<td>(6.3E-05)</td>
<td>(2.9E-05)</td>
<td>(2.2E-05)</td>
<td>(6.3E-05)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.001$)</td>
<td>0.054</td>
<td>0.026</td>
<td>0.109</td>
<td>0.068</td>
<td>0.027</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>(1.4E-05)</td>
<td>(6.2E-06)</td>
<td>(4.1E-05)</td>
<td>(4.5E-05)</td>
<td>(2.4E-05)</td>
<td>(1.0E-04)</td>
</tr>
<tr>
<td>Dep ($\epsilon \to \infty$)</td>
<td>0.073</td>
<td>0.048</td>
<td>0.122</td>
<td>0.094</td>
<td>0.058</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>(4.7E-05)</td>
<td>(6.3E-05)</td>
<td>(4.1E-05)</td>
<td>(3.3E-05)</td>
<td>(6.4E-05)</td>
<td>(7.2E-05)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>0.086</td>
<td>0.065</td>
<td>0.128</td>
<td>0.107</td>
<td>0.075</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>(7.7E-05)</td>
<td>(9.5E-05)</td>
<td>(6.6E-05)</td>
<td>(6.2E-05)</td>
<td>(9.1E-05)</td>
<td>(6.1E-05)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>0.100</td>
<td>0.072</td>
<td>0.156</td>
<td>0.141</td>
<td>0.092</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>(6.6E-05)</td>
<td>(9.4E-05)</td>
<td>(2.4E-05)</td>
<td>(4.8E-05)</td>
<td>(9.1E-05)</td>
<td>(8.1E-05)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.064</td>
<td>0.027</td>
<td>0.138</td>
<td>0.099</td>
<td>0.032</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>(6.6E-06)</td>
<td>(1.6E-05)</td>
<td>(3.1E-05)</td>
<td>(2.6E-05)</td>
<td>(3.1E-05)</td>
<td>(1.2E-04)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$P_{cor}$ overall</th>
<th>$P_{cor}$ dom</th>
<th>$P_{cor}$ nondom</th>
<th>$R_{cor}$ overall</th>
<th>$R_{cor}$ dom</th>
<th>$R_{cor}$ nondom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep ($\epsilon$ estimated)</td>
<td>0.882</td>
<td>0.953</td>
<td>0.741</td>
<td>0.932</td>
<td>0.974</td>
<td>0.847</td>
</tr>
<tr>
<td></td>
<td>(7.6E-05)</td>
<td>(7.5E-05)</td>
<td>(1.6E-04)</td>
<td>(5.2E-04)</td>
<td>(3.7E-17)</td>
<td>(1.6E-03)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.001$)</td>
<td>0.890</td>
<td>0.967</td>
<td>0.735</td>
<td>0.936</td>
<td>0.974</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td>(5.4E-05)</td>
<td>(8.2E-05)</td>
<td>(1.1E-04)</td>
<td>(5.6E-04)</td>
<td>(3.7E-17)</td>
<td>(1.7E-03)</td>
</tr>
<tr>
<td>Dep ($\epsilon \to \infty$)</td>
<td>0.804</td>
<td>0.850</td>
<td>0.712</td>
<td>0.930</td>
<td>0.974</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>(2.3E-04)</td>
<td>(2.8E-04)</td>
<td>(1.5E-04)</td>
<td>(2.5E-04)</td>
<td>(3.7E-17)</td>
<td>(7.6E-04)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>0.775</td>
<td>0.809</td>
<td>0.706</td>
<td>0.923</td>
<td>0.974</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td>(2.2E-04)</td>
<td>(2.7E-04)</td>
<td>(1.5E-04)</td>
<td>(2.5E-04)</td>
<td>(3.7E-17)</td>
<td>(7.6E-04)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>0.735</td>
<td>0.781</td>
<td>0.643</td>
<td>0.910</td>
<td>0.974</td>
<td>0.780</td>
</tr>
<tr>
<td></td>
<td>(1.9E-04)</td>
<td>(2.6E-04)</td>
<td>(6.6E-05)</td>
<td>(2.9E-04)</td>
<td>(3.7E-17)</td>
<td>(8.7E-04)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.855</td>
<td>0.933</td>
<td>0.699</td>
<td>0.921</td>
<td>0.974</td>
<td>0.814</td>
</tr>
<tr>
<td></td>
<td>(1.2E-04)</td>
<td>(1.6E-04)</td>
<td>(5.0E-05)</td>
<td>(3.8E-04)</td>
<td>(3.7E-17)</td>
<td>(1.1E-03)</td>
</tr>
</tbody>
</table>

Table 3.19: Two mean squared error statistics, correct choice probabilities and correct choice rates for the second holdout sample. The numbers in the parentheses are the standard errors over 10 independent MCMC chains. The Dep($\epsilon$ estimated) and Dep ($\epsilon = 0.001$) models have similar values and are the best among all models. The Indep($K = 1$) model is worse than all other models. The MNL model has smaller $MSE_1$ (overall) but larger $MSE_2$ (overall) than the Dep($\epsilon \to \infty$) model. For $P_{cor}$ and $R_{cor}$, the dependent model are the best. The Indep and the Indep($K = 1$) models perform worse than the group of dependent models. The $P_{cor}$ (overall) and $P_{cor}$ (dom) of the MNL model are better than the Dep($\epsilon \to \infty$) model, but its performance deteriorates dramatically in the non-dominant sets.
<table>
<thead>
<tr>
<th>Model</th>
<th>( \log(QL_{\text{holdout}}) )</th>
<th>( QLR_{\text{holdout}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall dom nondom</td>
<td>overall dom nondom</td>
</tr>
<tr>
<td>Dep ((\epsilon \text{ estimated}))</td>
<td>-104.59 (0.09) -42.56 (0.09) -62.03 (0.03)</td>
<td>1.26E+23 3.06E+19 4.20E+03</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>-117.93 (0.61) -55.69 (0.61) -62.23 (0.02)</td>
<td>9.47E+17 2.76E+14 3.41E+03</td>
</tr>
<tr>
<td>Dep ((\epsilon \to \infty))</td>
<td>-138.32 (0.09) -69.95 (0.08) -68.38 (0.02)</td>
<td>2.80E+08 3.86E+07 7.33E+00</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>-157.80 (0.12) -87.43 (0.09) -70.37 (0.03)</td>
<td>1.00E+00 1.00E+00 1.00E+00</td>
</tr>
<tr>
<td>MNL</td>
<td>-123.91 (0.02) -47.54 (0.04) -76.36 (0.02)</td>
<td>4.96E+14 2.04E+17 2.48E-03</td>
</tr>
</tbody>
</table>

Table 3.20: The logarithm of the quasi-likelihood, \( \log(QL_{\text{holdout}}) \) and quasi-likelihood ratio, \( QLR_{\text{holdout}} \) for the second hold out sample. The numbers in the parentheses are the standard errors over 10 independent MCMC chains. The Dep(\(\epsilon \text{ estimated}\)) model is the best model, followed by the Dep(\(\epsilon = 0.001\)) model and the MNL model and then the Dep(\(\epsilon \to \infty\)) model. The Indep model and Indep\((K = 1)\) model are the worst. The MNL model has larger \( QL_{\text{holdout}} \) value than the Dep(\(\epsilon = 0.001\)) model in the dominant sets but smaller than that of the Indep model in the non-dominant sets.
<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept</th>
<th>Interest rate</th>
<th>Annual fee</th>
<th>Travel points</th>
<th>Cash reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep($\epsilon$ estimated)</td>
<td>-3.082</td>
<td>0.601</td>
<td>0.323</td>
<td>-0.804</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.217)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.013)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Dep($\epsilon = 0.001$)</td>
<td>-3.532</td>
<td>0.542</td>
<td>0.245</td>
<td>-0.909</td>
<td>-0.073</td>
</tr>
<tr>
<td></td>
<td>(0.439)</td>
<td>(0.014)</td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Dep($\epsilon \to \infty$)</td>
<td>-3.495</td>
<td>0.862</td>
<td>0.669</td>
<td>-0.370</td>
<td>0.442</td>
</tr>
<tr>
<td></td>
<td>(0.159)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.015)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Indep</td>
<td>-3.316</td>
<td>1.089</td>
<td>0.933</td>
<td>-0.086</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.012)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Indep($K = 1$)</td>
<td>-2.459</td>
<td>1.624</td>
<td>1.518</td>
<td>0.455</td>
<td>1.341</td>
</tr>
<tr>
<td></td>
<td>(0.200)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.010)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>MNL</td>
<td>NA</td>
<td>1.919</td>
<td>1.356</td>
<td>0.480</td>
<td>1.030</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.21: Averages of the posterior means of $\mu_\gamma$ for all models for the second calibration sample. The numbers in the parentheses are the standard errors over 10 MCMC chains. Again, interest rate and annual fee are the two most important attributes.
For the third holdout sample with the first twelve sets in Table 3.14., the two $MSE$ statistics, $P_{cor}$ and $R_{cor}$ are listed in Table 3.22. For $MSE_1(overall)$, $MSE_1(dom)$ and $MSE_1(nondom)$, the dependent model performs well with small $\epsilon_i$. As $\epsilon_i$ approaches infinity, the $MSE_1$ values increase but are still smaller than those of the Indep model. The same pattern is observed for $MSE_2(overall)$, $MSE_2(dom)$ and $MSE_2(nondom)$. The $Indep(K = 1)$ model has the largest $MSE$ values. For the MNL model, $MSE_1(overall)$ is close to that of the Dep($\epsilon \rightarrow \infty$) model. Relatively, its $MSE_1$ is slightly better in the dominant sets than that in the non-dominant sets. The $MSE_2$ of the MNL model is larger than the Indep model for all choice sets (overall) and for the dominant sets, while for the dominant sets, the MNL model is the best of all models. Among all choice sets, $P_{cor}$ of the Dep($\epsilon = 0.001$) model is the best, followed closely by the Dep($\epsilon$ estimated) model. The Indep model and the MNL model fall in the second best group and are slightly better than the Dep($\epsilon \rightarrow \infty$) model. The $Indep(K = 1)$ model is the worst. For the dominant sets only, the MNL model has highest $P_{cor}$ among all models. For the non-dominant sets, the Indep model has the highest $P_{cor}$ while the MNL model has the second smallest $R_{cor}$. For $R_{cor}$, the dependent models have larger values than the independent models in all three situations. The MNL model has virtually the same $R_{cor}$ as the Dep($\epsilon \rightarrow \infty$) model.

Table 3.23 lists the $log(QL_{holdout})$ statistic for the second set of holdout data. The first two dependent models are the best models. The MNL model is very similar to the the Dep($\epsilon \rightarrow \infty$) model and Indep models and the $Indep(K = 1)$ model are the worst two. The likelihood ratios based on the quasi-likelihood of the all models vs. the Indep model are listed in Table 3.23 which shows the huge magnitude of difference
between the models. Again, the MNL model performs better in the dominant pairs than in the non-dominant pairs.

The estimated posterior means of $\mu$, are listed in Table 3.24. In all models, interest rate is the most important attribute as it has the largest positive coefficient. Annual fee is the second important attribute. The attribute, cash rewards, has the third largest coefficients, but the Dep(\(\epsilon\) estimated) and Dep(\(\epsilon = 0.001\)) models suggest that it has little importance as the coefficients are either small or negative. For travel points, all the dependent models suggest that it is not important, while the rest of the models give positive estimates.
### Table 3.22: \(MSE_1\), \(MSE_2\), \(P_{cor}\) and \(R_{cor}\) for the third holdout sample. The numbers in the parentheses are the standard errors over 10 independent MCMC chains.

<table>
<thead>
<tr>
<th>Model</th>
<th>(MSE_1)</th>
<th>(MSE_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall</td>
<td>dom</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dep (estimated)</td>
<td>0.111</td>
<td>0.030</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(4.9E-05)</td>
<td>(5.4E-05)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>0.112</td>
<td>0.029</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(1.1E-04)</td>
<td>(6.9E-05)</td>
</tr>
<tr>
<td>Dep ((\epsilon \to \infty))</td>
<td>0.125</td>
<td>0.063</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(1.9E-04)</td>
<td>(2.8E-04)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.149</td>
<td>0.083</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(1.2E-04)</td>
<td>(5.0E-05)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.120</td>
<td>0.028</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(3.9E-05)</td>
<td>(7.1E-06)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>(P_{cor})</th>
<th>(R_{cor})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall</td>
<td>dom</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dep (estimated)</td>
<td>0.766</td>
<td>0.938</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(5.0E-04)</td>
<td>(2.0E-04)</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>0.737</td>
<td>0.948</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(5.9E-04)</td>
<td>(4.4E-04)</td>
</tr>
<tr>
<td>Dep ((\epsilon \to \infty))</td>
<td>0.752</td>
<td>0.845</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(4.6E-04)</td>
<td>(2.9E-04)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.747</td>
<td>0.824</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(5.9E-04)</td>
<td>(1.3E-04)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.690</td>
<td>0.805</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(5.6E-05)</td>
<td>(9.7E-05)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.761</td>
<td>0.964</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>(5.9E-05)</td>
<td>(1.5E-04)</td>
</tr>
<tr>
<td>Model</td>
<td>$\log(QL_{\text{holdout}})$</td>
<td>$QLR_{\text{holdout}}$</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td>overall</td>
<td>dom</td>
</tr>
<tr>
<td>Dep ($\epsilon$ estimated)</td>
<td>-380.50</td>
<td>-28.24</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Dep ($\epsilon = 0.001$)</td>
<td>-386.39</td>
<td>-30.02</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>Dep ($\epsilon \to \infty$)</td>
<td>-417.87</td>
<td>-42.63</td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>-464.85</td>
<td>-51.14</td>
</tr>
<tr>
<td></td>
<td>(2.35)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Indep ($K = 1$)</td>
<td>-483.96</td>
<td>-55.21</td>
</tr>
<tr>
<td></td>
<td>(0.53)</td>
<td>(0.31)</td>
</tr>
<tr>
<td>MNL</td>
<td>-409.92</td>
<td>-30.43</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.11)</td>
</tr>
</tbody>
</table>

Table 3.23: The logarithm of quasi-likelihood, $\log(QL_{\text{holdout}})$, and quasi-likelihood ratio $QLR_{\text{holdout}}$ for the third holdout sample. The numbers in the parentheses are the standard errors over 10 independent MCMC chains. The first two dependent models are the best models. The MNL model is very similar to the the Dep($\epsilon \to \infty$) model and Indep models and the Indep($K = 1$) model are the worst.
<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept</th>
<th>Interest rate</th>
<th>Annual fee</th>
<th>Travel points</th>
<th>Cash reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Dep}(\epsilon \text{ estimated})$</td>
<td>-2.711</td>
<td>0.720</td>
<td>0.473</td>
<td>-1.142</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.199)</td>
<td>(0.023)</td>
<td>(0.025)</td>
<td>(0.029)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>$\text{Dep}(\epsilon = 0.001)$</td>
<td>-2.250</td>
<td>0.934</td>
<td>0.657</td>
<td>-0.965</td>
<td>0.143</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.032)</td>
<td>(0.034)</td>
<td>(0.044)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>$\text{Dep}(\epsilon \rightarrow \infty)$</td>
<td>-2.926</td>
<td>1.592</td>
<td>1.464</td>
<td>-0.100</td>
<td>1.206</td>
</tr>
<tr>
<td></td>
<td>(0.173)</td>
<td>(0.026)</td>
<td>(0.025)</td>
<td>(0.035)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>$\text{Indep}$</td>
<td>-2.151</td>
<td>2.796</td>
<td>2.642</td>
<td>1.124</td>
<td>2.366</td>
</tr>
<tr>
<td></td>
<td>(0.183)</td>
<td>(0.073)</td>
<td>(0.073)</td>
<td>(0.078)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>$\text{Indep}(K = 1)$</td>
<td>-2.167</td>
<td>3.426</td>
<td>3.265</td>
<td>1.747</td>
<td>2.987</td>
</tr>
<tr>
<td></td>
<td>(0.432)</td>
<td>(0.045)</td>
<td>(0.046)</td>
<td>(0.049)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>$\text{MNL}$</td>
<td>NA</td>
<td>2.158</td>
<td>1.884</td>
<td>0.862</td>
<td>1.478</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.009)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.24: Posterior means of $\mu_\gamma$ for all models for the third calibration sample. The numbers in parentheses are the standard errors over 10 MCMC chains. The attribute “interest rate” is the most important attribute and annual fee is the second important attribute. The other two attributes, cash rewards and travel points, are of little importance.
For the fourth holdout sample with the last six choice sets, Table 3.25 shows that the Dep(ε estimated) and Dep(ε = 0.001) models outperform the MNL model in terms of the $MSE_1$ and $MSE_2$. The Dep(ε → ∞) model is better than the $MNL$ model in terms of the overall $MSE_1$ and $MSE_2$ for the sets with a trade-off relationship. The Indep model performs similar to the MNL model, and the Indep($K = 1$) model is the worst among all models. The MNL model does relatively better in the dominant sets but worse in the sets with a trade-off relationship. Similar patterns are observed in $P_{cor}$, $R_{cor}$ and quasi-likelihood in Table 3.26. Overall, even in the case of an efficient design for the MNL model, the MNL model does not seem to outperform the dependent model. For the parameter estimates given in Table 3.27, it seems that interest rate is the most important attribute, followed by annual fee and then cash reward. Travel points are the least important.
<table>
<thead>
<tr>
<th>Model</th>
<th>( MSE_1 )</th>
<th>( MSE_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall</td>
<td>dom</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>0.087</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(3.4E-05)</td>
<td>(1.9E-05)</td>
</tr>
<tr>
<td>Dep ((\epsilon \rightarrow \infty))</td>
<td>0.085</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>(2.4E-05)</td>
<td>(2.4E-05)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.115</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>(2.6E-05)</td>
<td>(4.1E-05)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.114</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(2.3E-05)</td>
<td>(2.4E-05)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>( P_{cor} )</th>
<th>( R_{cor} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall</td>
<td>dom</td>
</tr>
<tr>
<td>Dep ((\epsilon = 0.001))</td>
<td>0.793</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>(9.1E-05)</td>
<td>(1.1E-04)</td>
</tr>
<tr>
<td>Dep ((\epsilon \rightarrow \infty))</td>
<td>0.794</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>(7.0E-05)</td>
<td>(1.6E-04)</td>
</tr>
<tr>
<td>Indep ((K = 1))</td>
<td>0.734</td>
<td>0.828</td>
</tr>
<tr>
<td></td>
<td>(6.4E-05)</td>
<td>(6.8E-05)</td>
</tr>
<tr>
<td>MNL</td>
<td>0.716</td>
<td>0.777</td>
</tr>
<tr>
<td></td>
<td>(5.9E-05)</td>
<td>(9.9E-05)</td>
</tr>
</tbody>
</table>

Table 3.25: \( MSE_1, MSE_2, P_{cor} \) and \( R_{cor} \) for the fourth holdout sample. The numbers in parentheses are the standard errors over 10 independent MCMC chains. The Dep(\( \epsilon \) estimated) and Dep(\( \epsilon = 0.001 \)) models outperform the MNL model. The Dep(\( \epsilon \rightarrow \infty \)) model is close to the MNL model and the Indep model. The Indep(\( K = 1 \)) model is the worst among all models.
Table 3.26: The quasi-loglikelihood, $\log(QL_{\text{holdout}})$, and quasi-likelihood ratio $QLR_{\text{holdout}}$ for the fourth holdout sample. The numbers in parentheses are the standard errors over 10 independent MCMC chains. The Dep($\epsilon$ estimated) and Dep($\epsilon = 0.001$) models outperform the MNL model. The Dep($\epsilon \to \infty$) model is close to the MNL model. The Indep($K = 1$) model is the worst among all models.
<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept</th>
<th>Interest rate</th>
<th>Annual fee</th>
<th>Travel points</th>
<th>Cash reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep(ε estimated)</td>
<td>-2.069</td>
<td>0.325</td>
<td>-0.225</td>
<td>-1.225</td>
<td>-0.404</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Dep(ε = 0.001)</td>
<td>-0.228</td>
<td>0.452</td>
<td>-0.080</td>
<td>-1.102</td>
<td>-0.278</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Dep(ε → ∞)</td>
<td>-3.621</td>
<td>0.824</td>
<td>0.428</td>
<td>-0.288</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Indep</td>
<td>-3.597</td>
<td>1.095</td>
<td>0.740</td>
<td>0.088</td>
<td>0.740</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Indep(K = 1)</td>
<td>-2.896</td>
<td>1.665</td>
<td>1.381</td>
<td>0.732</td>
<td>1.411</td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>MNL</td>
<td>NA</td>
<td>1.958</td>
<td>1.140</td>
<td>0.311</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

Table 3.27: Posterior means of $\mu_\gamma$ for all models for the fourth holdout sample. The numbers in parentheses are the standard errors over 10 MCMC chains. It seems that interest rate is the most important attribute, followed by annual fee and then cash reward. Travel points are the least important.
The statistics computed for the four holdout samples suggest first that the dependent race model with a reasonable tie-breaking rule provides better predictive performance than the traditional MNL model and the independent race models. For all four holdout samples, in the Dep(\(\epsilon\) estimated) model, each respondent has his or her own \(\epsilon_i\) value, but the predictive power of the model is not necessarily better than the Dep(\(\epsilon = 0.001\)) model. This is consistent with the result from the simulated data in Section 3.4. In other words, as long as \(\epsilon\) is reasonably small, it does not matter whether we assume that all respondents have the same value of \(\epsilon_i\). When \(\epsilon_i \to \infty\), the predictive performance of the model usually gets worse and approaches that of the independent model. This is also consistent with the properties of the tie-breaking rule described in Section 3.2.1.

It seems that the dominance relationship among alternatives is recognized by respondents and reflected in the data. The performance of the dependent models relative to other models supports our belief that a model that can take into account dependence relationships leads to large improvement in predictive performance. Over all the statistics, the two independent race models usually have the worst performance, as they do not take into account the dependence structure among the alternatives. The logit model Indep(\(K = 1\)) is almost always the worst model, which strongly suggests that the additive rate structure is not the underlying reason that the dependent race models outperform the MNL model.

The MNL model usually performs as well as (and sometimes better than) the Dep(\(\epsilon \to \infty\)) model and the independent race model for all choice sets. The dependent models with a small \(\epsilon_i\) for each respondent beats the MNL model consistently in almost all statistics considered, no matter whether they are for all holdout choice
sets, for the dominant choice sets only, or for non-dominant choice sets. An interesting observation is that when we look at the dominant sets and non-dominant sets separately, while the MNL model has fairly good performance in predicting the dominant sets, its performance in predicting the non-dominant sets is usually not so good. Such conclusions can be drawn from all holdout samples. This is even true when the MNL model is fitted to the second calibration sample where no dominant sets are present. It suggests that the MNL model cannot adapt well to both dominant sets and non-dominant sets at the same time. If the coefficients estimated for a respondent result in good predictive choice probabilities in the dominant sets, the same estimates can not predict the choice probabilities in the non-dominant sets well. On the other hand, we have seen the reverse phenomena in the pilot study. That is, while the performance in predicting the non-dominant sets is good, the performance in predicting the dominant sets is poorer. Taken together the studies suggest that the MNL model cannot capture the relationship between alternatives as naturally as the dependent model does.

3.7 Conclusion

In this chapter, a dependent Poisson race model for modeling consumer choice in the presence of possibly dominated alternatives has been proposed. This model assumes that there is a shared process that tracks the information shared by the alternatives in a choice set. By defining the rate of the shared process, a model that can take into account the dominance relationship among alternatives is built up. The formulae for the choice probabilities are derived in a general context. Through
simulation, it is shown that the dependent race model can account for dominance successfully and its application to conjoint choice data is illustrated through an example on consumer preference of credit cards. In terms of predictive power, the dependent Poisson race models with proper tie-breaking rules provide a better fit to the data than the MNL model and independent Poisson race model.

The introduction of a shared process takes into account dependence between alternatives caused by an overlap in attribute levels or similarities in levels. Similarity among alternatives is a more general concept than the overlap (e.g. a 10% interest rate is overlapping with a 7% interest rate on a one-dimension real line) in attribute levels since the former concept is relevant to both nominal and ordinal attributes, while the latter is applicable to attributes with a clear direction of preference. This chapter has focused on the latter concept in formulating the dependent race model. In future research, we will develop a more general model to account for product similarity which will allow for attributes that do not have a preferred direction, such as color of credit card, type of card, sponsor of card (bank, airline, etc.). In the next chapter, the models are extended to handle choice data with more than two alternatives in each set.
CHAPTER 4

DEPENDENT RACE MODEL FOR CONJOINT CHOICE EXPERIMENTS WITH THREE ALTERNATIVES

4.1 Introduction

Conjoint choice experiments usually have more than two alternatives in a choice set. An example of a choice set with three alternatives is given in Table 4.1. Respondents are asked to select the card that they most prefer. As in the binary case, the multinomial logit model is often the method used to analyze this type of data. However, the limitations of the MNL model extend to cases with multiple alternatives as well. In Chapter 3, the Poisson race model with a dependence structure has been shown to have better predictive performance than the independent Poisson race model and the MNL model. This chapter applies the dependent Poisson race model to larger choice sets and to compare its performance with other models.

Conceptually, Poisson race models can be adapted naturally to the case with more than two alternatives. Each alternative can be represented by a Poisson process and the hits in favor of an alternative are accumulated on the associated counter. Once any process accumulates the required number of hits, a choice is made. Again, since these alternatives all have the same attributes, some of the hits are shared by two or more alternatives. That is, whenever a hit is generated by some level of one attribute,
the same hit is accumulated on other alternatives with the same or better level of this
attribute as well. Equivalently, for any alternative, its Poisson process has a shared
counter with one or many other processes. Then, the same approach as that proposed
in Chapter 3 can be applied.

However, as the size of the choice set increases, the relationship among alterna-
tives becomes increasingly complicated. Alternatives can have the same level of an
attribute, or some may have the same level while others have different levels. One
alternative may be better than another on one attribute but worse on another at-
ttribute. The relationships among alternatives can have many more variations than
the relationship between only two alternatives. If we use \( A_3 = \{a_1, a_2, a_3\} \) as an
example, the possible relationships among alternatives are:

1. A three-way trade-off relationship, denoted by \( (1, 2, 3) \).

2. \( a_1 \) dominates \( a_2 \), but both \( a_1 \) and \( a_2 \) form trade-off relationships with \( a_3 \), denoted
   by \( (1 > 2), 3 \).

3. \( a_1 \) and \( a_2 \) have a trade-off relationship, but both dominate \( a_3 \), denoted by
   \( (1, 2) > 3 \).

4. \( a_1 \) dominates both \( a_2 \) and \( a_3 \), but \( a_2 \) and \( a_3 \) have a trade-off relationship, denoted
   by \( 1 > (2, 3) \).

5. \( a_1 \) dominates \( a_2 \), and \( a_2 \) dominates \( a_3 \), denoted by \( 1 > 2 > 3 \).

For each alternative, its associated Poisson process might be completely shared
with one other alternative but only partially shared with the third one. As a result,
the number of shared counters increases dramatically as the number of alternatives
Table 4.1: An example of dominance with three credit cards.

<table>
<thead>
<tr>
<th></th>
<th>Credit card $a_1$</th>
<th>Credit card $a_2$</th>
<th>Credit card $a_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual fee</td>
<td>$0</td>
<td>$30</td>
<td>$0</td>
</tr>
<tr>
<td>Travel points</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cash reward</td>
<td>No</td>
<td>3%</td>
<td>No</td>
</tr>
</tbody>
</table>

increases. For example, Table 4.1 lists three credit cards $a_1$, $a_2$ and $a_3$ with the relationship ($3 > 1$), 2. $a_1$ differs from $a_3$ only in terms of interest rate. $a_1$ differs from $a_2$ in terms of all attributes except travel points. Therefore, the Poisson process associated with $a_1$ should share more hits with the process for $a_3$ than with that for $a_2$. In fact, the process for $a_1$ is completely shared with $a_3$ since $a_1$ is dominated by $a_3$. $a_2$ is not dominated by either $a_1$ or $a_3$. Thus, the process for $a_2$ is only partially shared with the processes for $a_1$ and $a_3$ and not all hits are shared. To handle the increased complexity of problem, the approach proposed for binary choice sets must be modified. This chapter discusses the formulation of the model, the computation of the choice probabilities and shows how to apply the model to a trinary choice set (that is a choice set containing three alternatives).

4.2 Dependent Poisson race model for choice sets with more than two alternatives

4.2.1 Formulation of Poisson rates

Consider a trinary choice set $A_3 = \{a_1, a_2, a_3\}$, where each alternative is assumed to be associated with a Poisson process, denoted by $X_i(t)$ with rate $\lambda_i$, where
The alternative whose counter first accumulates \( K \) hits is the choice that is made. For simplicity, first assume that each of the three alternatives is characterized by the levels of a single attribute. Assume respondents have a direction of preference for the levels of this attribute and that this direction is the same for all respondents. Hits in favor of each alternative arrive according to the corresponding Poisson processes \( X_i(t), i = 1, 2, 3 \).

Each of these three Poisson processes can be decomposed into non-overlapping, independent components shared with one or more other processes and a component unique to itself. With three alternatives, this decomposition of \( X_1(t), X_2(t) \) and \( X_3(t) \) results in seven independent Poisson processes, with one three-way shared process, three two-way shared processes and three processes unique to the alternatives, denoted by \( X_{123}(t), X_{12}(t), X_{13}(t), X_{23}(t), X_{u1}(t), X_{u2}(t), X_{u3}(t) \), respectively. Then,

\[
X_i(t) = X_{ui}(t) + X_{ij}(t) + X_{ik}(t) + X_{123}(t), \quad i, j, k = 1, 2, 3, i \neq j \neq k.
\]

These seven processes generate hits independently. When the sum of the hits in favor of one alternative reaches a threshold value \( K \), that is, \( X_i(t) \geq K \), a choice occurs. The probability that a hit is contributed by any of the processes is proportional to their rates (Karlin & Taylor, 1975). We define the rate for the process shared by all three alternatives to be

\[
\lambda_{123} = \min(\lambda_1, \lambda_2, \lambda_3).
\]  

The rate shared by alternatives \( a_1 \) and \( a_2 \) is

\[
\lambda_{12} = \min(\lambda_1, \lambda_2) - \lambda_{123}.
\]
The rates \( \lambda_{13} \) and \( \lambda_{23} \) can be formulated in a similar way. The rate for \( X_1(t) \), unique to process 1, is

\[
\lambda_{u_1} = \lambda_1 - \lambda_{12} - \lambda_{13} - \lambda_{123}.
\]  

(4.2.3)

The unique rates for process 2 and process 3 are obtained in a similar fashion.

Consider three single-attribute alternatives whose levels can be rank ordered; for example, three credit cards presented with interest rate only, and as given in Table 3.1. It is reasonable to say that a rational respondent would choose a low interest rate over a high interest rate. Then, \( a_3 \) dominates \( a_2 \), and \( a_2 \) dominates \( a_1 \). According to (4.2.1), (4.2.2) and (4.2.3), \( \lambda_{123} = \lambda_1 \), \( \lambda_{23} = \lambda_2 - \lambda_1 \), \( \lambda_{13} = 0 \), \( \lambda_{12} = 0 \), \( \lambda_{u_3} = \lambda_3 - \lambda_2 \), \( \lambda_{u_2} = 0 \) and \( \lambda_{u_1} = 0 \). With this definition of rates, whenever there is a hit in favor of \( a_1 \), the hit is accrued on the shared counter for all three processes \( X_{123}(t) \), so that it also represents a hit in favor of both \( a_1 \) and \( a_2 \). A hit in favor of \( a_2 \) but not \( a_1 \), is generated by process \( X_{23}(t) \), and is also a hit in favor of \( a_3 \). A hit in favor of \( a_3 \) alone is generated by \( X_{u_3}(t) \), the unique process for \( a_3 \). This mechanism ensures that the total number of hits in favor of \( a_3 \) is never smaller than that in favor of \( a_2 \) or \( a_1 \).

For alternatives with \( J > 1 \) attributes, we assume each attribute is associated with a set of shared and unique processes. There are then a total of \( 7J \) independent latent processes. For the \( j \)th attribute, denote the shared and unique process rates by \( \lambda_{123,j} \), \( \lambda_{12,j} \), \( \lambda_{13,j} \), \( \lambda_{23,j} \), \( \lambda_{u_1,j} \), \( \lambda_{u_2,j} \), \( \lambda_{u_3,j} \), \( j = 1, \ldots, J \), respectively. The overall shared and unique rates associated with the three alternatives, aggregated over all attributes, are

\[
\lambda_{123} = \sum_{j=1}^{J} \lambda_{123,j}, \quad \lambda_{12} = \sum_{j=1}^{J} \lambda_{12,j}, \quad \lambda_{u_1} = \sum_{j=1}^{J} \lambda_{u_1,j}.
\]  

(4.2.4)

Other two-way shared rates and unique rates can be formulated in a similar way.
The shared processes with the above defined rates track how close these alternatives are in their attributes. When all three alternatives are close in all attributes, the shared rate $\lambda_{123}$ should be large, and there is a high probability of a three-way tie. When the similarity of any two of the alternatives, e.g., $a_1$ and $a_2$, are of interest, the extent of similarity should be evaluated by adding $\lambda_{12}$ and $\lambda_{123}$ together. If $\lambda_{123} + \lambda_{12}$ is large, there is big chance of having a two-way tie between these two alternatives. Additionally, especially in the case of a trade-off relationship among alternatives, the unique rates alone do not reflect the relative advantage of one alternative over another alternative. The related two-way shared rates and the unique rates have to be summed before comparing the alternatives. For example, the relative magnitude of $\lambda_{u_1} + \lambda_{13}$ and $\lambda_{u_2} + \lambda_{23}$ reflects the trade-off between alternatives $a_1$ and $a_2$.

For a choice set of size $m$, the $m$ dependent Poisson processes associated with $m$ alternatives can be decomposed into $\sum_{i=1}^{m} \binom{m}{i}$ independent latent processes, with $m$ unique processes, $\binom{m}{2}$ processes shared by two alternatives, $\binom{m}{3}$ processes shared with three alternatives, and so on. The rates associated with these independent latent processes can be formulated in a fashion similar to the case with three alternatives.

### 4.2.2 Computation of choice probabilities

In the trinary case, the expression for the choice probabilities under the dependent race model can not be written out directly, but we can compute the choice probabilities accurately with a method similar to that used to compute the choice probabilities in the binary case. We begin by describing the objects needed to implement the calculations.
Let \( \tilde{N} = \{ \tilde{N}_{u_1}, \tilde{N}_{u_2}, \tilde{N}_{u_3}, \tilde{N}_{12}, \tilde{N}_{13}, \tilde{N}_{23}, \tilde{N}_{123} \} \) be a matrix collecting all possible combinations of hits on all shared and unique counters at the race’s stopping time, where \( \tilde{N}_1, \tilde{N}_2, \) and \( \tilde{N}_3 \) are the vectors of hits on the three unique counters, \( \tilde{N}_{12}, \tilde{N}_{13}, \tilde{N}_{23} \) are vectors of hits on the counters shared by any two alternatives, and \( \tilde{N}_{123} \) is vector of hits on the counter shared by all three alternatives. A matrix \( \tilde{I} = \{ \tilde{I}_1, \tilde{I}_2, \tilde{I}_3, \tilde{I}_{12}, \tilde{I}_{13}, \tilde{I}_{23}, \tilde{I}_{123} \} \) of indicators is used to record whether a row of \( \tilde{N} \) corresponds to a choice of one alternative, a two-way tie or three-way tie between alternatives. For example, if a row \( j \) in \( \tilde{N} \) corresponds to a three-way tie, then in the \( j \)th row of \( \tilde{I}, \tilde{I}_{123}(j) = 1 \) and all other entries in the same row are zeros.

The hits on the three counters can arrive in any sequence for each given combination of hits at the race’s stopping time. A vector \( \tilde{C} \), with the same length as \( \tilde{N} \), is used to store the number of sequences for every combinations of hits. The \( j \)th element of \( \tilde{C} \) is calculated as follows:

\[
\tilde{C}(j) = I(\tilde{I}_d(j) = 1) \sum_p \left[ \frac{(\sum_q \tilde{N}_q(j) - 1)!}{(\tilde{N}_p(j) - 1)! \prod_{q \neq p} \tilde{N}_q(j)!} I(\tilde{N}_p(j) > 0) \right]. \tag{4.2.5}
\]

Here, \( I(*) \) is an indicator function and \( I(*) = 1 \) when \( * \) holds. To clarify the notations in (4.2.5), consider a race where \( a_1 \) hits its threshold at least as quickly as \( a_2 \) or \( a_3 \). Let \( \tilde{I}_d(j) = 1 \) indicates the result of the race as follows. Suppose that \( d = \{1\} \), then alternative \( a_1 \) is the choice made. The last hit has to be generated either by the unique process \( X_{u_1}(t) \) or by one of the processes shared by \( a_1 \), i.e., \( X_{12}(t), X_{13}(t), \) and \( X_{123}(t) \). The total number of hits in favor of \( a_1 \) is equal to \( K = N_{u_1} + N_{12} + N_{13} + N_{123} \). With \( d = \{1\} \), in (4.2.5), \( p \) is an element of the set of indexes for the counters whose hits might contribute to the total number of hits in favor of \( a_1 \) so that \( p \in \{1, 12, 13, 123\} \). Also, \( q \) belongs to the set of indexes for all counters. Similarly, if \( d = \{12\} \), then there is a tie between \( a_1 \) and \( a_2 \) when the race
Table 4.2: Number of rows in $\tilde{N}$ for given $K$. The number increases rapidly as $K$ increases.

<table>
<thead>
<tr>
<th>K</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>7</td>
<td>62</td>
<td>283</td>
<td>921</td>
<td>2436</td>
<td>5586</td>
<td>11536</td>
<td>21980</td>
<td>39283</td>
<td>66636</td>
</tr>
</tbody>
</table>

stops. The last hit has to be on counter $X_{12}(t)$ or $X_{123}(t)$. Then, $p \in \{12, 123\}$. Also, notice that $N_{u_1} + N_{12} + N_{13} + N_{123} = N_{u_2} + N_{12} + N_{23} + N_{123} = K$. If $d = \{123\}$, then the last hit has to be on the counter for $X_{123}(t)$ and $p = \{123\}$.

The vectors $\tilde{C}$, $\tilde{N}$, and $\tilde{I}$ each depends on $K$, but they do not depend on the rates. They can be calculated in advance and stored in a multidimensional array. Notice that the number of rows for $\tilde{N}$ depends on $K$ only and increases rapidly as $K$ increases. Table 4.2 gives the number of rows in $\tilde{N}$ for given $K \leq 10$. As the size of $\tilde{N}$, $\tilde{I}$ and $\tilde{C}$ become large, the storage of these objects can become an issue. Fortunately, as we have seen Chapter 3, the value of $K$ is usually quite small.

To compute the choice probabilities, we need the rates. First, for given $K$, let

$$ p_q = \frac{\lambda_q}{\sum_q \lambda_q}, \quad q \in \{1, 2, 3, 12, 13, 23, 123\}. \quad (4.2.6) $$

Then, a vector

$$ \tilde{p} = \{\prod_q p_q^{N_q}\}, \quad q \in \{1, 2, 3, 12, 13, 23, 123\} \quad (4.2.7) $$

of the same length as $\tilde{N}$ is computed. Its $j$th element is $\prod_q p_q(j)^{N_q(j)}$.

With this notation, the probability of a three-way tie between $a_1$, $a_2$ and $a_3$ can be written as

$$ P(123) = \tilde{C}'[\tilde{p} \bullet \tilde{I}_{123}], \quad (4.2.8) $$

171
where the operator \( ' \bullet ' \) is defined as the element-wise product of two vectors or matrices. Similarly, the probability of having a two-way tie between \( a_1 \) and \( a_2 \) is

\[
P(12) = \tilde{C}'[\tilde{p} \bullet \tilde{I}_{12}],
\]

and the probability of \( a_1 \) winning the race is

\[
P(1) = \tilde{C}'[\tilde{p} \bullet \tilde{I}_1].
\]

Other probabilities can be written in the same manner.

From this discussion, it is obvious that if all the shared rates \( \lambda_q, q \in \{12, 23, 13, 123\} \), are set to zero, then the above expressions give the choice probabilities for \( a_1 \), \( a_2 \) and \( a_3 \) in the independent race model.

### 4.2.3 Tie-breaking rule

In order to model conjoint choice data where exactly one alternative is chosen from a given choice set, the ties have to be handled in a systematic way. In the binary case, when \( a_1 \) and \( a_2 \) are tied, the ties are allocated to \( a_1 \) and \( a_2 \) proportionally via a tie-breaking rule (see Section 3.2.1. In a similar fashion, the two-way ties and three-way ties for the trinary sets are also allocated proportionally via a set of tie-breaking rules.

The probability of a three-way tie \( P_{123} \) is allocated to \( a_1 \), \( a_2 \) and \( a_3 \) with proportions \( t_{1,123}, t_{2,123} \) and \( t_{3,123} \), respectively. Thus, \( t_{1,123} + t_{2,123} + t_{3,123} = 1 \). For simplicity, let \( t_{123} = t_{1,123}, t_{213} = t_{2,123} \) and \( t_{312} = t_{3,123} \), where the first subscript indicates the alternative which the tie-probability is allocated to and the last two subscripts indicates other alternatives involved in the tie besides the one indicated by the first subscript. With this notation, \( t_{123} = t_{132}, t_{213} = t_{231} \) and \( t_{312} = t_{321} \). Similarly, the
probability of a two-way tie, e.g. $P_{12}$, is allocated to $a_1$ and $a_2$ with proportions $t_{12}$ and $t_{21}$, where $t_{12} + t_{21} = 1$. Accordingly, $t_{13}$ and $t_{31}$ are proportions to divide $P_{13}$, $t_{23}$ and $t_{32}$ are proportions to divide $P_{23}$. With these proportions, the probabilities of choosing $a_1$, $a_2$ and $a_3$ from set $A = \{a_1, a_2, a_3\}$ are

$$P_A(1) = P(1) + t_{12}P_{12} + t_{13}P_{13} + t_{123}P_{123},$$

$$P_A(2) = P(2) + t_{21}P_{12} + t_{13}P_{13} + t_{213}P_{123},$$

$$P_A(3) = P(3) + t_{32}P_{23} + t_{31}P_{13} + t_{312}P_{123}.$$  \hfill (4.2.11)

The proportion $t_{ij}$ is defined to be:

$$t_{ij} = \frac{\lambda_{u_i} + \lambda_{ik} + \epsilon(\lambda_{ij} + \lambda_{ijk})}{\lambda_{u_i} + \lambda_{u_j} + \lambda_{ik} + \lambda_{jk} + 2\epsilon(\lambda_{ij} + \lambda_{ijk})},$$  \hfill (4.2.12)

where $\epsilon \geq 0$, $i \neq j$ and $i, j = 1, 2, 3$. The proportion $t_{ijk}$ is given by

$$t_{ijk} = \frac{\lambda_{u_i} + \epsilon\lambda_{ijk}}{\lambda_{u_i} + \lambda_{u_j} + \lambda_{u_k} + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} + 6\epsilon\lambda_{ijk}}$$

$$+ \frac{\lambda_{ij} + \epsilon\lambda_{ijk}}{\lambda_{u_i} + \lambda_{u_j} + \lambda_{u_k} + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} + 6\epsilon\lambda_{ijk}}$$

$$\times \frac{\lambda_{u_i} + \lambda_{ik} + \epsilon(\lambda_{ij} + \lambda_{ijk})}{\lambda_{u_i} + \lambda_{ik} + \lambda_{u_j} + \lambda_{jk} + 2\epsilon(\lambda_{ij} + \lambda_{ijk})}$$

$$+ \frac{\lambda_{ik} + \epsilon\lambda_{ijk}}{\lambda_{u_i} + \lambda_{u_j} + \lambda_{u_k} + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} + 6\epsilon\lambda_{ijk}}$$

$$\times \frac{\lambda_{u_i} + \lambda_{ij} + \epsilon(\lambda_{ik} + \lambda_{ijk})}{\lambda_{u_i} + \lambda_{ij} + \lambda_{u_k} + \lambda_{jk} + 2\epsilon(\lambda_{ik} + \lambda_{ijk})},$$  \hfill (4.2.13)

where $\epsilon \geq 0$, and $(i, j, k) = (1, 2, 3), (2, 1, 3)$ or $(3, 1, 2)$.

To understand the tie-breaking rule in (4.2.11), first consider the case where $\epsilon = 0$ and a two-way tie. Without loss of generality, a two-way tie between $a_1$ and $a_2$ is
considered. When this occurs, the race is assumed to continue between only $a_1$ and $a_2$ (and $a_3$ drops out of the race). The alternative which first pulls ahead in this continued race wins and the outcome depends on the total number of hits on the counters related to $a_1$ and $a_2$. For $a_1$, the associated counters are the ones tracking hits from the processes $X_{u_1}(t)$ and $X_{13}(t)$ since these two processes generate hits not shared by $a_2$. For $a_2$, the associated counters are the ones tracking hits from processes $X_{u_2}(t)$ and $X_{23}(t)$. The hits generated by $X_{12}(t)$ need not be counted as these hits are shared by both $a_1$ and $a_2$. Therefore, the chance that $a_1$ wins the “continued” race is proportional to $\lambda_{u_1} + \lambda_{13}$ and the chance that $a_2$ wins is proportional to $\lambda_{u_2} + \lambda_{23}$.

With these considerations, we have
\[ t_{12} = \frac{\lambda_{u_1} + \lambda_{13}}{\lambda_{u_1} + \lambda_{u_2} + \lambda_{13} + \lambda_{23}}. \tag{4.2.14} \]

This is exactly the same interpretation tie-breaking rule proposed in Chapter 3 for the choice sets with two alternatives by viewing $\lambda_{u_1} + \lambda_{13}$ as the “unique rate” for $a_1$ and $\lambda_{u_2} + \lambda_{23}$ as the “unique rate” for $a_2$. In (4.2.12), the parameter $\epsilon$ is introduced to capture a range of tie-breaking rules. When $\epsilon = 0$, a dominant alternative is always chosen over the dominated alternative. The alternative with the greater overall rate is chosen with probability greater than 0.5. When $\epsilon \to \infty$, $t_{12} \to 0.5$ and the ties are broken at random. For $\epsilon > 0$, there is always a positive probability that the ‘worse’ alternative of the two considered will be chosen.

In the case of a three-way tie, the race is assumed to continue with all three alternatives in the running. Again, first consider the case of $\epsilon = 0$. The three-way race continues either until one alternative pulls ahead (this alternative is the winner of the race and is chosen) or until an alternative falls behind. If an alternative falls behind, it drops out, and the race becomes a two-way race. In this event, the two-way
tie is broken according to (4.2.12). Consequently, we have the following:

\[ t_{123} = \frac{\lambda_{u_1}}{\lambda_{u_1} + \lambda_{u_2} + \lambda_{u_3} + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{12}} + \frac{\lambda_{12}}{\lambda_{u_1} + \lambda_{u_2} + \lambda_{u_3} + \lambda_{12} + \lambda_{13} + \lambda_{23}} t_{12} + \frac{\lambda_{13}}{\lambda_{u_1} + \lambda_{u_2} + \lambda_{u_3} + \lambda_{12} + \lambda_{13} + \lambda_{23}} t_{13}. \]  

(4.2.15)

In this case, if \( a_1 \) dominates \( a_2 \) and \( a_3 \), \( a_1 \) is always chosen. Again, in (4.2.13), \( \epsilon \) is introduced to allow a positive probability of choosing a dominated alternative. When \( \epsilon \to \infty \), \( t_{ijk} = 1/3 \), and the ties are broken at random.

4.3 Modeling conjoint choice data

4.3.1 Dependent Poisson race model

Consider a conjoint choice experiment where each respondent evaluates \( M \) choice sets \( A_1, A_2, \ldots, A_M \), each having \( H = 3 \) alternatives; that is, \( A_m = \{a_{1m}, a_{2m}, a_{3m}\} \), for \( m = 1, \ldots, M \). Let \( x_{jm} = \{x_{1jm}, x_{2jm}, x_{3jm}\} \) denote the levels of the \( j \)th attribute for all alternatives in choice set \( A_m \), with \( x_{hjm} \geq 0 \) and where the a priori worst level of an attribute is always coded as 0 (\( h = 1, 2, 3; j = 1, \ldots, J; m = 1, \ldots, M \)). As in Section 4.2, the \( j \)th attribute in these three alternatives is associated with the dependent Poisson processes \( X_{1j}(t), X_{2j}(t), \) and \( X_{3j}(t) \). Rate \( \lambda_{hjm} \) is a non-negative, increasing function of \( x_{hjm} \) for \( h = 1, 2, 3 \), so that preferred levels of an attribute make a greater contribution to the rate.

As in the binary case, the levels of the attributes are scaled so that the worst level of each attribute across the entire experiment is coded as 0. Further levels of the attribute are recoded linearly between the best level and the worst level. Then, as in Section 3.2.2, the attribute-specific rates are modeled as

\[ \lambda_{1jm} = e^{x_{1jm}\beta_j} - 1, \lambda_{2jm} = e^{x_{2jm}\beta_j} - 1 \quad \text{and} \quad \lambda_{3jm} = e^{x_{3jm}\beta_j} - 1, \]  

(4.3.1)
where \( j = 1, \ldots, J \); \( m = 1, \ldots, M \), with \( \beta_j \geq 0 \) for all \( j \) and at least one \( \beta_j > 0 \).

With appropriately scaled \( x_{hjm} \), for \( h = 1, 2, 3 \), the parameters \( \beta_1, \ldots, \beta_J \) reflect the importance of the attributes. The exponential transformation ensures non-negativity of the rates and echoes traditional formulations such as the multinomial logit (MNL) model (McFadden, 1974), while 1 is subtracted so that \( \lambda_{hjm} \) is zero when \( x_{hjm} = 0 \).

Then, for the \( j \)th attribute and the \( m \)th choice set, the shared and unique rates are:

\[
\begin{align*}
\lambda_{123,jm} &= \min(\lambda_{1jm}, \lambda_{2jm}, \lambda_{3jm}), \\
\lambda_{12,jm} &= \min(\lambda_{1jm}, \lambda_{2jm}) - \lambda_{123,jm}, \\
\lambda_{u1,jm} &= \lambda_{1jm} - \lambda_{12,jm} - \lambda_{13,jm} - \lambda_{123,jm}.
\end{align*}
\]

The rates for \( \lambda_{23,jm}, \lambda_{13,jm}, \lambda_{u2,jm}, \) and \( \lambda_{u3,jm} \) are formulated similarly. Then, the rates are aggregated over the \( J \) attributes to yield the overall shared and unique rates for the \( m \)th choice set,

\[
\begin{align*}
\lambda_{123,m} &= e^{\beta_0} + \sum_{j=1}^{J} \lambda_{123,jm}, \\
\lambda_{12,m} &= \sum_{j=1}^{J} \lambda_{12,jm} \quad \text{and} \quad \lambda_{u1,m} = \sum_{j=1}^{J} \lambda_{u1,jm},
\end{align*}
\]

where \( \beta_0 \) is an intercept parameter for the shared rate. As in the binary case, the intercept accounts for both the common shared rate across alternatives that is due to the worst level of every attribute and unseen attributes of the alternatives. Under the assumption that the levels of these attributes are the same for all alternatives, these additional aspects contribute only to the three-way shared rate.
4.3.2 Alternative models

In Section 4.5, we compare the performance of the dependent race model to that of several other models. The first alternative model is an independent Poisson race model with no shared counters and three unique rates as follows:

\[
\begin{align*}
\lambda_{123,m} &= \lambda_{12,m} = \lambda_{13,m} = \lambda_{23,m} = 0 \\
\lambda_{uh,m} &= e^{\beta_0} + \sum_{j=1}^{J} [e^{x_{hm}\beta_j} - 1], \ h = 1, 2, 3, \\
\end{align*}
\]

(4.3.3)

where \(\beta_0\) is an intercept parameter, and \(\beta_j \geq 0\) for all \(j = 1, \ldots, J\), with at least one \(\beta_j > 0\), and the threshold value is a positive integer \(K \geq 1\). In the special case \(K = 1\), this is a logit model with an additive rate structure.

Another alternative model is the multinomial logit model (MNL) which can be specified as a Poisson race model with threshold value \(K = 1\), no shared counter and unique rates as follows:

\[
\lambda_{uh,m} = e^{\sum_{j=1}^{J} x_{hm}\beta_j}, \ h = 1, 2, 3.
\]

(4.3.4)

This model has a multiplicative rate structure. It does not have an intercept term, as including an intercept would create a non-identifiable model and the data would contain no information about the intercept. The MNL model will be compared with the independent Poisson race model with \(K = 1\). The comparison of the two logit models will show if, and to what extent, the difference between an additive and a multiplicative rate structure drives our results as compared with the impact of the shared counter.
4.4 Estimation

Let \( Y_i = (y_{i1}, \ldots, y_{iM}) \) be the matrix of responses for the \( i \)th respondent (\( i = 1, \ldots, I \)) across \( M \) choice sets, where \( y_{im} = (y_{1im}, y_{2im}, y_{3im}) \) for \( m = 1, \ldots, M \). If alternative \( a_{1m} \) is chosen in the \( m \)th choice set \( A_m = \{a_{1m}, a_{2m}, a_{3m}\} \), then \( y_{im} = (1, 0, 0) \). First, we consider the dependent race model. For respondent \( i \), the threshold value is \( K_i \) for all alternatives, and the set of attribute importance parameters is \( \beta_i = \{\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,J}\} \). There is also a parameter \( \epsilon_i \) associated with the tie-breaking rule. Given \( \Theta_i = \{K_i, \beta_i, \epsilon_i\} \), the likelihood function for respondent \( i \) is

\[
L(Y_i|\Theta_i) = \prod_{m=1}^{M} P_{A_m}(1|\Theta_i)^{y_{1im}}P_{A_m}(2|\Theta_i)^{y_{2im}}P_{A_m}(3|\Theta_i)^{y_{3im}},
\]

(4.4.1)

where \( P_{A_m}(h|\Theta_i) \) is the probability of choosing \( a_{hm}, h = 1, 2, 3 \) from choice set \( A_m \) given \( \Theta_i \).

Similar to the binary case in Section 3.3, a hierarchical model is formulated. The hierarchical model and the prior distributions in the trinary case have exactly the same forms as those in the binary case, as listed from (3.3.2) to (3.3.7). On the other hand, we also condition on common fixed values for the parameter \( \epsilon_i, i = 1, \ldots, I \), and investigate the corresponding posterior distribution by comparing models with different fixed values. These fixed values for \( \epsilon_i, i = 1, \ldots, I \) range from 0.001 to \( \infty \).

MCMC methods are used to simulate from the posterior distributions. These methods are described in Section 3.3.2. Different from the binary case, the data, \( Y = \{Y_1, Y_2, \ldots, Y_I\} \), represents the trinary data for totally \( I \) respondents. They contribute a likelihood which is the product of the likelihoods (3.3.1) over \( i \). This leads to the posterior distribution, up to the unknown normalizing constant,

\[
[\gamma, K, \epsilon, \mu_\gamma, \Sigma_\gamma|Y] \propto [Y|\gamma, K, \epsilon, \mu_\gamma, \Sigma_\gamma][\gamma|\mu_\gamma, \Sigma_\gamma][\mu_\gamma, \Sigma_\gamma][\epsilon][K].
\]

178
In this expression, \([Y|\gamma, K, \epsilon, \mu_\gamma, \Sigma_\gamma]\) is given (4.4.1).

4.5 A conjoint choice study with three alternatives

4.5.1 Design and data description

To study whether the dependent race model might improve prediction for choice data, a conjoint choice experiment on credit cards was conducted. Each credit card was described by the four attributes used in the binary study in Section 3.6. The first attribute, \(x_1\), is interest rate, described by four levels (9.99%, 12.99%, 14.99% and 17.99%). The second attribute, \(x_2\), is annual fee with two levels $30 and $0. The third and fourth attributes are travel points \((x_3)\) and cash reward \((x_4)\). Both have two levels (Yes and No). The attributes are recoded so that the worst level is 0 and the best level is 1. Interest rate is considered to be a continuous variable, and its levels are rescaled proportionally.

The design is constructed based on several criteria. First, since a primary goal of the experiment is to compare the dependent race model to the MNL model, the design should give fair treatment (or favor) to the MNL model. To implement this, we decided that the final design should have a reasonably large D-value as compared to the D-efficient design for the MNL model given by \(\%Choiceff\) macro of Kuhfeld in SAS (Kuhfeld, 2005). Second, the trinary sets should contain all of the types of relationships described in Section 4.1. Third, the design should be based on a set of non-zero estimated parameter values for all attributes. That is, the design is constructed based on a priori information about the importance of each attribute.

Under the MNL model with \(\beta\) being the vector of coefficients for interest rate \((x_1)\), annual fee \((x_2)\), travel pints \((x_3)\) and cash reward \((x_4)\), a priori values for \(\beta\) are set
to be [3.96, 1.82, 0.18, 1.24] for the four attributes, respectively. These parameters are set so that interest rate is the most important attribute, annual fee is the second most important attribute, and cash reward is the third most important attribute. A candidate design with 12 trinary sets was generated by the SAS macro %Choiceff. The candidate design had a D-value of 1.46. By switching the levels of the attributes in the candidate design, new designs with six trade-off sets and six sets exhibiting a dominance relationship were generated and their D-values were computed again with the %Choiceff macro. From these modified designs, a design with small number of switches and reasonably large D-value was selected, as shown in the left part of Table 4.3. This design has a D-value of 1.05.

Some binary sets were included in the study for two reasons. First, to study whether we can extrapolate from trinary sets to binary sets, or vice versa. Second, to study whether we observe violations of IIA in the experimental data. In adding binary sets, there was a tension between the desire to have enough binary sets to fit the model on these sets alone, while not overburdening the subject with too many choice tasks (as he/she might tire of evaluation and resort to random selection or fail to complete the survey). With these thoughts in mind, eight binary choice sets were constructed from the trinary sets by retaining two of the three alternatives. Three binary sets with a trade-off relationship are from trinary sets where one alternative dominates one other alternative (e.g., (1 > 2) , 3 type). Four binary sets are constructed from trinary sets without any dominance relationships. One binary set is generated from a trinary set with the dominance relationship such as 3 > (1, 2). No binary sets are constructed from trinary sets with relationships such as (1, 2) > 3 or 1 > 2 > 3. The binary sets are shown to the right side of the corresponding trinary sets in Table 4.3.
Table 4.3: Design matrix of the conjoint choice experiment and summary of the raw data. $x_1$-$x_4$ represent the levels of interest rate, annual fee, travel points and cash reward, respectively. The number (N) and percent (100%) of subjects selecting the corresponding credit card are also listed.
<table>
<thead>
<tr>
<th>Categories</th>
<th>N</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>52</td>
<td>0.63</td>
</tr>
<tr>
<td>Number of credit cards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0.12</td>
</tr>
<tr>
<td>1</td>
<td>33</td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>0.30</td>
</tr>
<tr>
<td>More than 2</td>
<td>15</td>
<td>0.18</td>
</tr>
<tr>
<td>Charges incurred</td>
<td>28</td>
<td>0.34</td>
</tr>
<tr>
<td>Most important</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest rates</td>
<td>48</td>
<td>0.58</td>
</tr>
<tr>
<td>Annual fee</td>
<td>28</td>
<td>0.34</td>
</tr>
<tr>
<td>Travel points</td>
<td>2</td>
<td>0.02</td>
</tr>
<tr>
<td>Cash reward</td>
<td>5</td>
<td>0.06</td>
</tr>
<tr>
<td>Least Important</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest rates</td>
<td>18</td>
<td>0.22</td>
</tr>
<tr>
<td>Annual fee</td>
<td>7</td>
<td>0.08</td>
</tr>
<tr>
<td>Travel points</td>
<td>35</td>
<td>0.42</td>
</tr>
<tr>
<td>Cash reward</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>Travel a lot</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>Considered choice</td>
<td>82</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 4.4: Summary of the demographic questions

A total of 84 students in a Marketing class at The Ohio State University participated in the experiment in Spring 2006. The students were given a questionnaire and presented with choice sets following the design in Table 4.3. The survey instrument appears in Appendix C. Of the 84 students, one did not complete the survey and two indicated that they made random selections during the tasks. The responses of these three are removed from the data set. The rest of the data are used in the subsequent analyses.

The data are summarized in Table 4.3. The column under “N” and the column under % give the number and the percent of students (out of 81) who chose each card.
in a choice set, respectively. The left panel lists the results for the trinary sets and
the right panel lists the results for the binary sets.

A comparison of choice sets 2 and 14, 6 and 16 suggest a violation of the IIA
assumption. For example, in choice set 14 with only two cards, the observed ratio of
choice probabilities is about $42/39 = 1.08$. However, when another card is added to
the choice set (see set 2), the ratio of choice probabilities for the two cards in set 14
is $4/36 = 0.11$. An easy explanation of this phenomenon is that the new alternative
is very similar to the preferred alternative in set 14, but slightly better as it includes
a cash reward. Most individuals jump to this dominant alternative and a few switch
away from the non-preferred alternative in set 14. From this pair of choice sets, it
is clear that, aggregated over all students, the marketing students do not exhibit
IIA behavior. Possible explanations include (substantial) heterogeneity of preference
among the students or violation of IIA at the individual level. The data sport both
explanations. In a similar case, the ratio changes from 3.76 in set 16 to 77.0 in set 6.
The same phenomenon can be observed for other pairs of trinary and binary sets.

Table 4.4 summarizes students’ answers to a set of demographic questions attached
to the end of the survey. Note that students did not see these questions until they
finished all choice tasks, so that answers to these questions did not influence their
choices. About 58% of students consider interest rate to be the most important
attribute while 34% consider annual fee to be the most important attribute. Also,
while about 22% of students consider interest rates to be the least important attribute,
only 8% consider annual fee to be the least important attribute. From this table, we
would expect interest rate and annual fee play important roles in respondents’ choice
decisions with large differences between individuals’ assessment of these attributes.
The effect of travel points and cash rewards might be relatively small, with less heterogeneity among respondents.

4.5.2 Measures of holdout performance

The data described in Section 4.5.1 is divided into a calibration sample and a holdout sample. The calibration sample is used to fit the models and the holdout sample is used to compare the predictive performance of the models. Measures of the holdout performance discussed in Section 3.5.2 are defined for the trinary sets in the same way. Several choice sets for each respondent are held out to form a holdout sample represented by \( Y = \{y_1, \ldots, y_I\} \). Here \( I = 81 \) and these holdout sets can be either trinary or binary. Let \( \mathcal{H}_i, i = 1, \ldots, N_{ho} \) denote the collection of the indexes of the held out choice sets. \( \mathcal{H}_i \) might vary from respondent to respondent. Then, \( y_i = \{y_{im}, m \in \mathcal{H}_i\} \) is the holdout data for the \( i \)th respondent, where \( y_{im} = \{y_{him}\} \) \((h = 1, \ldots, H)\) is a vector of length three for trinary sets \((H = 3)\) or length two \((H = 2)\) for binary sets indicating which card is selected in a choice set. Let the total number of choice sets held out for each respondent be \( N_{ho} \). In our analyses, \( N_{ho} \) will be the same for all respondents. For sets \( m = 1, \ldots, 20 \), we use \( \mathcal{D} \) to denote the collection of indexes for choice sets with some form of dominance relationship and \( \mathcal{E} = \{m \notin \mathcal{D}\} \) to denote the collection of indexes for choice sets with only trade-off relationship. From the design of the study, \( \mathcal{D} = \{2, 4, 5, 6, 7, 9\} \). Also, let \( N_{dom} \) and \( N_{nondom} \) be the total number of choice sets with and without dominance relationships in the holdout sample across all respondents, respectively. For a given model, let \( P_{A_{im}}(h|\Theta_{in}) \) be the probability of choosing \( a_{hm} \) from set \( A_m \) for the \( i \)th consumer on the \( n \)th MCMC iteration after burn-in, where \( i = 1, \ldots, I, n = 1, \ldots, N \) and \( m \)
indexes choice sets. Then, \( \hat{P}_{A_{im}}(h) \) is the posterior mean of \( P_{A_{im}}(h|\Theta_{in}) \) over a total of \( N \) MCMC iterations after burn-in.

With this notation, we redefine \( MSE_1, MSE_2, R_{corr}, P_{corr}, QL \) and \( L \) for holdout samples with a mixture of trinary and binary sets:

\[
MSE_1(overall) = \frac{1}{I \ast N_{ho}} \sum_{i=1}^{I} \sum_{m \in \mathcal{H}_i} \sum_{h=1}^{H} (\hat{P}_{A_{im}}(h) - y_{him})^2, \tag{4.5.1}
\]

\[
MSE_2(overall) = \frac{1}{N \ast I \ast N_{ho}} \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{m \in \mathcal{H}_i} \sum_{h=1}^{H} (P_{A_{im}}(h|\Theta_{in}) - y_{him})^2. \tag{4.5.2}
\]

\[
P_{corr} = \frac{1}{I \ast N_{ho}} \sum_{i=1}^{I} \sum_{m \in \mathcal{H}_i} \sum_{h=1}^{H} y_{him} \hat{P}_{A_{im}}(h|\Theta_i). \tag{4.5.3}
\]

\[
R_{corr} = \frac{1}{I \ast N_{ho}} \sum_{i=1}^{I} \sum_{m \in \mathcal{H}_i} \{y_{im}, \hat{y}_{im}\}. \tag{4.5.4}
\]

where \( \hat{y}_{im} = \{\hat{y}_{him}\} \) is the estimated response of \( i \)th respondent for the \( m \)th holdout choice set. The estimated response is \( \hat{y}_{him} = 1 \) if \( P_{A_{im}}(h) = max\{P_{A_{im}}(h), h = 1, \ldots, H\} \), and \( \hat{y}_{him} = 0 \) otherwise.

\[
L_{\text{holdout}} = \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{I} \prod_{m \in \mathcal{H}_i} \prod_{h=1}^{H} [P_{A_{im}}(h|\Theta_{in})y_{him}]. \tag{4.5.5}
\]

\[
QL_{\text{holdout}} = \prod_{i=1}^{I} \prod_{m \in \mathcal{H}_i} \prod_{h=1}^{H} [\hat{P}_{A_{im}}(h)y_{him}]. \tag{4.5.6}
\]

In the above expressions, \( m \in \mathcal{H}_i \) can be replaced with \( m \in \mathcal{D} \cap \mathcal{H}_i \) and \( m \in \mathcal{E} \cap \mathcal{H}_i \) to compute the holdout measures for choice sets with and without dominance relationships, respectively.
4.5.3 Results and discussion

A total of six models are evaluated: the dependent race model with $\epsilon_i, i = 1, \ldots, I$ having unspecified values, denoted by $\text{Dep}(\epsilon_i \text{ differ})$; the dependent race model with $\epsilon_i = \epsilon = 0.001$ for all $i = 1, \ldots, I$, denoted by $\text{Dep}(\epsilon = 0.001)$; the dependent race model with $\epsilon \to \infty$, denoted by $\text{Dep}(\epsilon \to \infty)$; the independent race model, denoted by $\text{Indep}$; the logit model with an additive rate structure, which is the independent race model with $K_i = 1, i = 1, \ldots, I$, denoted by $\text{Indep}(K = 1)$; the MNL model with a multiplicative rate structure. These models are analyzed with the MCMC method described in Section 3.3.2. For the analysis in the chapter, the attributes are recoded with the worst level denoted by zero and the best level denoted by 1. The intermediate levels are scaled linearly between the best and the worst levels.

For the dependent Poisson race models and the independent Poisson race model, the prior distribution for $[\mu_\gamma, \Sigma_\gamma]$ is $NIW(A, d, a, c)$ with $a = 0, c = 1/9, d = 20, A = 14I_5$. The prior distribution for $\gamma_i$ is $MVN(\mu_\gamma, \Sigma_\gamma)$. The prior distribution of $K_i, i = 1, \ldots, I$, is $Po(1) + 1$. For the $\text{Dep}(\epsilon \text{ estimated})$ model, the prior for $\epsilon_i$ is given by (3.3.7). For the $\text{Indep}(K = 1)$ model, the prior distribution for $\gamma_i$ is the same as above but with $K_i = 1, i = 1, \ldots, I$. The prior distribution for $\gamma_i$ for the MNL model has only four dimensions and thus $A = 15I_4$.

With this data set, we choose calibration data and holdout data in four different ways. First, all trinary sets are used to fit the models and all eight binary sets are used as the holdout sample. This division enables us to look at the extrapolation of the models from trinary sets to binary sets. Second, the calibration set consists of all binary sets and the holdout sample consists of all trinary sets, which helps to evaluate the extrapolation from binary sets to trinary sets. Third, the twelve trinary sets are
divided evenly into a calibration set and a holdout set. This allows us to evaluate the models with a small calibration data set of trinary choice sets only. Finally, one trinary set with only a trade-off relationship, one trinary set with a dominance relationship and one binary set with a trade-off relationship are held out for each respondent and the rest of the data are used as the calibration data set. This division lets us to evaluate model performance based on a mixed design with both trinary and binary sets.

Simulation errors are quantified by completing 10 independent MCMCs runs for each of the first three data set. There are 11000 iterations in each of the independent MCMC runs. The first 1000 iterations are discarded as burn-in. The fourth division of the data sets results in a relatively small calibration data set, and so we use 105000 iterations per MCMC runs and discard 5000 as burn-in in this case. A model with relatively better performance is characterized by smaller $MSE_1$ and $MSE_2$, and larger correct choice probability, $P_{corr}$, correct choice rate $R_{corr}$, and logarithm of quasi-likelihood ($QL$) and joint likelihood $L$ for the holdout data.

Table 4.5 shows the holdout measures computed for the first the holdout samples. For all measures except $P_{corr}$, the dependent race models with $\epsilon = 0.001$ and with different $\epsilon_i$ are the best. The dependent race model with $\epsilon \to \infty$ and the independent race model with different $K_i$ are very close to each other in all measures. They are slightly better than the independent race model with $K = 1$. The MNL model is the worst except in terms of $P_{corr}$ and $R_{corr}$.

Figure 4.1 plots the estimated mean choice probabilities of the chosen cards in the eight holdout binary sets for all respondents for the MNL model against those of the Dep($\epsilon = 0.001$) model. There is some evidence that the MNL model predicts
choice probabilities that are more extreme (close to 0 or 1) than the Dep($\epsilon = 0.001$) model, as indicated by the clusters of points above the 45 degree line near zero and the clusters below the 45 degree line near 1. This holds especially in sets 13, 14, 16, 17 and 19. Since the correct choice probability $P_{corr}$ is the average of the estimated mean probabilities for the chosen card, the $P_{corr}$ for the MNL model should be relatively large.

In fact, $P_{corr}$ can be artificially inflated in the following fashion. Take any model, and find the card with the largest choice probability, set this probability to 1 and the others to 0. The resulting statistic is $R_{corr}$. When the model is correct $R_{corr}$ will exceed $P_{corr}$. It is only for extremely poor models that we see reversals. We believe that the strong performance of the MNL model for $P_{corr}$ is in part an artifact of extreme probabilities.

In contrast to $P_{corr}$, $\log(QL)$ is more difficult to inflate. $\log(QL)$ is obtained by summing the individual log-predictive likelihood over the holdout cases. No benefit is gained by altering these predictive likelihoods. In keeping with the view that $P_{corr}$ provides an overly optimistic view of the performance of the MNL model, specifically because predictive choice probabilities are too extreme, the $\log(QL)$ of the MNL model is much much smaller than that of the Dep($\epsilon = 0.001$) model. The logarithm of the joint likelihood $L$ behaves in the same way as the $\log(QL)$. By looking at the quasi-likelihood ratio $QLR$ and joint likelihood ratio $LR$ with respect to the Indep model, the first two dependent race models have remarkably better predictive power than the other models. However, the standard errors for $\log(QL)$ and $\log(L)$ for the two independent models are quite large, indicating that the estimates might not be stable for these two models.
Table 4.5: Statistics for the first holdout data set. All binary sets have been held out and all trinary sets have been used to fit the model. The numbers in parentheses are the standard errors over 10 independent MCMC chains.
Figure 4.1: Scatterplots of the estimated mean choice probabilities of the chosen cards in the eight holdout binary sets for all respondents for MNL model vs. Dep(ε = 0.001) model.
The second holdout sample consists of all trinary sets and the eight binary sets are used to fit the models. The holdout statistics are summarized in Table 4.6 and Table 4.7. This analysis suggests that MNL model has smallest $MSE_1$ and $MSE_2$, largest $P_{corr}$, $log(QL)$ and $log(L)$. The $\text{Dep}(\epsilon \to \infty)$ model are the model with second smallest $MSE_1$ and $MSE_2$. The $\text{Dep}(\epsilon_i \text{ differ})$ and the $\text{Dep}(\epsilon = 0.001)$ models have largest $MSE_1$, $MSE_2$ for all sets combined and for the dominant sets only. In the non-dominant sets, the Indep and $\text{Indep}(K = 1)$ models have largest $MSE_1$ and $MSE_2$. The MNL model has the largest $P_{corr}$, followed by the $\text{Dep}(\epsilon \to \infty)$ model. The remaining models have very similar correct choice rates. For $R_{corr}$, the $\text{Dep}(\epsilon_i \text{ differ})$ and the $\text{Dep}(\epsilon = 0.001)$ models are slightly better than other models but the difference is very small. From the likelihood ratio in Table 4.7, the MNL model has best predictive performance, followed by the $\text{Dep}(\epsilon \to \infty)$ model. The independent race models have better performance than the $\text{Dep}(\epsilon_i \text{ differ})$ and the $\text{Dep}(\epsilon = 0.001)$ models. It seems that this analysis support the MNL model. However, it contradicts the results for the first set of holdout sample. The reason behind this contradiction will be discussed later with the results from other analyses.
### Table 4.6: $MSE_1$, $MSE_2$, $P_{corr}$ and $R_{corr}$ for the second holdout sample. The holdout sample consists of all trinary sets. The numbers in parentheses are the standard errors over 10 MCMC chains.
Table 4.7: $\log(QL)$, $\log(L)$ and likelihood ratios for the second holdout sample. The holdout sample consists of all trinary sets. The numbers in parentheses are the standard errors over 10 MCMC chains.
The third holdout sample consists of six randomly chosen trinary sets and the remaining six trinary sets are used to fit the models. The holdout statistics are summarized in Table 4.8 and Table 4.9. In this set of analyses, the MNL model has the best performance in terms of all measures. The Dep($\epsilon_i$ differ) and the Dep($\epsilon = 0.001$) models are the second best models. These two models are very close, with the Dep($\epsilon = 0.001$) model being slightly better. The remaining models are not as good as these three models. The Dep($\epsilon \rightarrow \infty$ model) is better than the two independent race models. These results suggest that the Dependence models have relatively good performance as compared to the MNL model and are clearly better than the independent race models.
<table>
<thead>
<tr>
<th>Model</th>
<th>Dep (εi differ)</th>
<th>Dep (ε = 0.001)</th>
<th>Dep (ε → ∞)</th>
<th>Indep</th>
<th>Indep (K = 1)</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.288 (1E-04)</td>
<td>0.283 (8E-05)</td>
<td>0.365 (1E-04)</td>
<td>0.376 (2E-04)</td>
<td>0.380 (8E-05)</td>
<td>0.258 (1E-04)</td>
</tr>
<tr>
<td>dom</td>
<td>0.175 (2E-04)</td>
<td>0.169 (1E-04)</td>
<td>0.230 (6E-05)</td>
<td>0.238 (2E-04)</td>
<td>0.241 (2E-04)</td>
<td>0.154 (9E-05)</td>
</tr>
<tr>
<td>nodom</td>
<td>0.413 (2E-04)</td>
<td>0.408 (1E-04)</td>
<td>0.515 (2E-04)</td>
<td>0.529 (4E-04)</td>
<td>0.533 (2E-04)</td>
<td>0.372 (2E-04)</td>
</tr>
<tr>
<td>all</td>
<td>0.348 (2E-04)</td>
<td>0.341 (1E-04)</td>
<td>0.404 (8E-05)</td>
<td>0.413 (1E-04)</td>
<td>0.416 (9E-05)</td>
<td>0.341 (1E-04)</td>
</tr>
<tr>
<td>dom</td>
<td>0.207 (2E-04)</td>
<td>0.199 (2E-04)</td>
<td>0.259 (1E-04)</td>
<td>0.269 (3E-04)</td>
<td>0.270 (2E-04)</td>
<td>0.192 (1E-04)</td>
</tr>
<tr>
<td>nodom</td>
<td>0.503 (2E-04)</td>
<td>0.498 (1E-04)</td>
<td>0.564 (4E-04)</td>
<td>0.573 (4E-04)</td>
<td>0.577 (2E-04)</td>
<td>0.504 (2E-04)</td>
</tr>
<tr>
<td>all</td>
<td>0.692 (6E-05)</td>
<td>0.699 (7E-05)</td>
<td>0.629 (1E-04)</td>
<td>0.622 (2E-04)</td>
<td>0.620 (1E-04)</td>
<td>0.745 (1E-04)</td>
</tr>
<tr>
<td>dom</td>
<td>0.826 (4E-05)</td>
<td>0.834 (6E-05)</td>
<td>0.766 (1E-04)</td>
<td>0.754 (1E-04)</td>
<td>0.753 (1E-04)</td>
<td>0.866 (7E-05)</td>
</tr>
<tr>
<td>nodom</td>
<td>0.544 (1E-04)</td>
<td>0.550 (1E-04)</td>
<td>0.478 (2E-04)</td>
<td>0.476 (4E-04)</td>
<td>0.473 (2E-04)</td>
<td>0.612 (2E-04)</td>
</tr>
<tr>
<td>all</td>
<td>0.852 (5E-04)</td>
<td>0.859 (4E-04)</td>
<td>0.779 (4E-04)</td>
<td>0.769 (3E-04)</td>
<td>0.768 (3E-04)</td>
<td>0.874 (3E-04)</td>
</tr>
<tr>
<td>dom</td>
<td>0.919 (4E-17)</td>
<td>0.918 (2E-04)</td>
<td>0.880 (4E-04)</td>
<td>0.872 (5E-04)</td>
<td>0.873 (4E-04)</td>
<td>0.935 (3E-04)</td>
</tr>
<tr>
<td>nodom</td>
<td>0.779 (1E-03)</td>
<td>0.793 (8E-04)</td>
<td>0.667 (6E-04)</td>
<td>0.656 (8E-04)</td>
<td>0.652 (6E-04)</td>
<td>0.806 (6E-04)</td>
</tr>
</tbody>
</table>

Table 4.8: Statistics for the third holdout sample. The holdout sample consists of six trinary sets and the calibration data set consists of the remaining six trinary sets. The numbers in parentheses are the standard errors over 10 MCMC chains.
Table 4.9: $\log(QL)$, $\log(L)$ and likelihood ratio for the third holdout sample. The holdout sample consists of six trinary sets and the calibration data set consists of the remaining six trinary sets. The numbers in parentheses are the standard errors over 10 MCMC chains.
The fourth holdout sample consists of three choice sets for each respondent: set 1 is a trinary set with trade-off relationship, set 7 is a trinary set with a $(1,2) > 3$ type of relationship, and set 16 is a binary set with trade-off relationship. The mean squared errors, correct choice probability and correct choice rates for this analysis are summarized in Table 4.10. Both the overall and set-by-set statistics are listed. The dependent race model with $\epsilon = 0.001$ and with different $\epsilon_i$ seem to have consistently good predictive performance in terms of all holdout measures. The MNL model falls behind the Dep($\epsilon = 0.001$) model and Dep($\epsilon_i$ differ) model, but it has the highest overall correct choice probability. The dependent race model with $\epsilon \rightarrow \infty$ falls between the MNL models and the independent race models. The independent race models seem to be the worst models in terms of the overall holdout measures. It is worth noting that the independent race models have smallest $MSE_1$, $MSE_2$ and a high $P_{corr}$ for choice set 7 but, their predictive power drops substantially for set 1 and 16.

The quasi-likelihood and joint likelihood and the likelihood ratios with respect to the independent race model are given in Table 4.11. From the quasi-likelihood ratio $QLR$ and the joint likelihood ratio $LR$, it is evident that the Dep($\epsilon = 0.001$) model has the best predictive power, followed closely by the Dep($\epsilon_i$ differ) model. The Dep($\epsilon \rightarrow \infty$) model is comparable with the Indep model, and both are slightly better than the independent race model with $K = 1$. The MNL model seems to have good predictive power in set 1 but its performance in set 16 is poor. In Figure 4.2, the posterior mean of the predicted choice probabilities of the chosen card in sets 1, 7 and 16 for the MNL model are plotted against those of the Dep($\epsilon_i$ differ) model, the Indep model, and the Dep($\epsilon \rightarrow \infty$) model, respectively. For choice set 1, the
points scatter quite evenly above and below the 45 degree line. For choice sets 7 and 16, the points fall into two clusters. The cluster with small values lies above the line, showing that the predicted choice probabilities made by the MNL model are generally smaller than the other three models when the predicted choice probabilities are small ($< 0.5$). The cluster with large values locates below the line, showing that the predicted choice probabilities from the MNL model are generally larger when the predicted choice probabilities are large ($> 0.5$). This explains larger correct choice probability $P_{corr}$ for the MNL model in Table 4.10 and its poorer $LR$ and $QLR$ values. This phenomenon is consistent with that observed with the first holdout sample.
<table>
<thead>
<tr>
<th>Set</th>
<th>Dep $(\epsilon_i \text{ differ})$</th>
<th>Dep $(\epsilon = 0.001)$</th>
<th>Dep $(\epsilon \rightarrow \infty)$</th>
<th>Indep $(K = 1)$</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.278 (1E-04)</td>
<td>0.273 (2E-04)</td>
<td>0.325 (1E-04)</td>
<td>0.341 (3E-04)</td>
<td>0.347 (4E-04)</td>
</tr>
<tr>
<td>1</td>
<td>0.374 (4E-04)</td>
<td>0.367 (6E-04)</td>
<td>0.498 (4E-04)</td>
<td>0.504 (3E-04)</td>
<td>0.511 (3E-04)</td>
</tr>
<tr>
<td>7</td>
<td>0.126 (2E-04)</td>
<td>0.122 (3E-04)</td>
<td>0.133 (2E-04)</td>
<td>0.118 (1E-04)</td>
<td>0.119 (2E-04)</td>
</tr>
<tr>
<td>16</td>
<td>0.332 (3E-04)</td>
<td>0.329 (2E-04)</td>
<td>0.344 (3E-04)</td>
<td>0.400 (8E-04)</td>
<td>0.412 (1E-04)</td>
</tr>
</tbody>
</table>

| MSE1 | all | 0.303 (2E-04) | 0.297 (2E-04) | 0.344 (9E-05) | 0.369 (4E-04) |
|      | 1   | 0.421 (4E-04) | 0.412 (6E-04) | 0.527 (3E-04) | 0.550 (4E-04) |
|      | 7   | 0.142 (3E-04) | 0.136 (3E-04) | 0.149 (3E-04) | 0.130 (2E-04) |
|      | 16  | 0.345 (3E-04) | 0.342 (2E-04) | 0.357 (2E-04) | 0.428 (7E-04) |

| MSE2 | all | 0.712 (1E-04) | 0.716 (3E-04) | 0.677 (2E-04) | 0.646 (3E-04) |
|      | 1   | 0.548 (3E-04) | 0.552 (5E-04) | 0.481 (2E-04) | 0.506 (3E-04) |
|      | 7   | 0.835 (3E-04) | 0.839 (3E-04) | 0.806 (4E-04) | 0.865 (4E-04) |
|      | 16  | 0.754 (3E-04) | 0.755 (3E-04) | 0.745 (6E-04) | 0.568 (8E-04) |

| $P_{corr}$ | all | 0.628 (5E-05) | 0.627 (6E-05) | 0.612 (2E-04) | 0.602 (2E-04) |
|            | 1   | 0.682 (8E-04) | 0.681 (1E-03) | 0.648 (2E-04) | 0.639 (2E-04) |
|            | 7   | 0.593 (5E-09) | 0.593 (5E-09) | 0.593 (5E-09) | 0.593 (5E-09) |
|            | 16  | 0.599 (0E+00) | 0.599 (0E+00) | 0.586 (4E-03) | 0.562 (5E-03) |

| $R_{corr}$ | all | 0.628 (5E-05) | 0.627 (6E-05) | 0.612 (2E-04) | 0.602 (2E-04) |
|            | 1   | 0.682 (8E-04) | 0.681 (1E-03) | 0.648 (2E-04) | 0.639 (2E-04) |
|            | 7   | 0.593 (5E-09) | 0.593 (5E-09) | 0.593 (5E-09) | 0.593 (5E-09) |
|            | 16  | 0.599 (0E+00) | 0.599 (0E+00) | 0.586 (4E-03) | 0.562 (5E-03) |

Table 4.10: Holdout statistics for the fourth holdout sample with two trinary sets and one binary set. The numbers in the parentheses are the standard errors over 10 independent MCMC chains.
<table>
<thead>
<tr>
<th>Model</th>
<th>Dep ((\epsilon_i \text{ differ}))</th>
<th>Dep ((\epsilon = 0.001))</th>
<th>Dep ((\epsilon \to \infty))</th>
<th>Indep ((K = 1))</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>-115.5 (7E-02)</td>
<td>-113.8 (1E-01)</td>
<td>-132.5 (9E-02)</td>
<td>-133.8 (1E-01)</td>
<td>-135.8 (1E-01)</td>
</tr>
<tr>
<td>1</td>
<td>-53.1 (4E-02)</td>
<td>-52.4 (8E-02)</td>
<td>-65.4 (4E-02)</td>
<td>-65.5 (4E-02)</td>
<td>-66.5 (5E-02)</td>
</tr>
<tr>
<td>7</td>
<td>-20.9 (3E-02)</td>
<td>-20.3 (5E-02)</td>
<td>-23.5 (3E-02)</td>
<td>-20.9 (4E-02)</td>
<td>-21.0 (5E-02)</td>
</tr>
<tr>
<td>16</td>
<td>-41.6 (6E-02)</td>
<td>-41.2 (4E-02)</td>
<td>-43.6 (7E-02)</td>
<td>-47.3 (8E-02)</td>
<td>-48.3 (1E-01)</td>
</tr>
</tbody>
</table>

**log(QL)**

<table>
<thead>
<tr>
<th>Model</th>
<th>Dep ((\epsilon_i \text{ differ}))</th>
<th>Dep ((\epsilon = 0.001))</th>
<th>Dep ((\epsilon \to \infty))</th>
<th>Indep ((K = 1))</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>-114.8 (0.34)</td>
<td>-113.3 (0.22)</td>
<td>-133.7 (0.18)</td>
<td>-134.6 (0.28)</td>
<td>-136.2 (0.32)</td>
</tr>
<tr>
<td>1</td>
<td>-52.8 (0.14)</td>
<td>-51.8 (0.18)</td>
<td>-65.4 (0.06)</td>
<td>-65.9 (0.18)</td>
<td>-66.9 (0.22)</td>
</tr>
<tr>
<td>7</td>
<td>-20.8 (0.04)</td>
<td>-20.1 (0.05)</td>
<td>-23.4 (0.05)</td>
<td>-21.0 (0.05)</td>
<td>-21.0 (0.07)</td>
</tr>
<tr>
<td>16</td>
<td>-41.2 (0.11)</td>
<td>-40.8 (0.12)</td>
<td>-43.3 (0.16)</td>
<td>-46.3 (0.10)</td>
<td>-47.4 (0.11)</td>
</tr>
</tbody>
</table>

**log(L)**

<table>
<thead>
<tr>
<th>Model</th>
<th>Dep ((\epsilon_i \text{ differ}))</th>
<th>Dep ((\epsilon = 0.001))</th>
<th>Dep ((\epsilon \to \infty))</th>
<th>Indep ((K = 1))</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>8E+07 (1.09)</td>
<td>4E+08 (1.97)</td>
<td>3.56 (0.07)</td>
<td>1 (0.95)</td>
<td>0.13 (0.48)</td>
</tr>
<tr>
<td>1</td>
<td>3E+05 (1.09)</td>
<td>5E+05 (1.97)</td>
<td>1.19 (0.07)</td>
<td>1 (0.95)</td>
<td>0.39 (5E+04)</td>
</tr>
<tr>
<td>7</td>
<td>301.9 (1.09)</td>
<td>433.2 (2.50)</td>
<td>40.35 (1.07)</td>
<td>1 (0.95)</td>
<td>0.35 (6E-05)</td>
</tr>
<tr>
<td>16</td>
<td>154.2 (1.09)</td>
<td>248.2 (2.50)</td>
<td>19.12 (1.07)</td>
<td>1 (0.95)</td>
<td>0.33 (2E-05)</td>
</tr>
</tbody>
</table>

Table 4.11: The quasi-likelihood \((QL)\), joint likelihood \((L)\) and the likelihood ratios for the fourth holdout sample. The numbers in parentheses are the standard errors over 10 independent MCMC chains.
Figure 4.2: Scatterplots of posterior means of choice probabilities for the MNL model vs those of the Dep(\(\epsilon_i\) differ), Indep, Dep(\(\epsilon \to \infty\)) model.
Table 4.12 gives the estimated posterior means for $\mu_\gamma$, the mean of the distribution of $\gamma = \ln \beta$. These parameters are not comparable across models, as $K_i$ is not the same for all models, but the relative size of the numbers within a model indicates the importance of the attributes. For all models, interest rate is the most important attribute, followed by annual fee, with cash reward and the travel points being the least important. This is consistent with what respondents indicated in their answers to the demographic questions.

Note that the intercept for the Dep($\epsilon = 0.001$) model is quite large compared to the estimates for the four attributes. This large intercept results in a large three-way tie probability. Then, the tie-breaking rule in (4.2.13) allocates these three-way tie-probabilities to the three cards. For the Dep($\epsilon_i$ differ) model, the histogram of the posterior mean of $\epsilon_i$, $i = 1, \ldots, 81$, across MCMC iterations is shown in Figure 4.3. It appears to be important to have small $\epsilon$ values for each respondent. From this plot, we can see that for most respondents the $\epsilon_i$ are small (close to zero). The dependent race model with $\epsilon = 0.001$ for all respondents works equally well. In contrast, when $\epsilon \to \infty$, the predictive power of the dependent race model decreases toward that of the independent race models. The threshold values $K_i$, $i = 1, \ldots, 81$, for the dependent race models are mainly less than 3, and the median of a sample of size 1000 from the posterior distribution of $K_i$ is 2. By comparing the estimates of $K_i$ from calibration data with only trinary sets and calibration data with only binary sets (see Figure 4.4), the estimates of $K_i$ are shown to remain small, regardless of the size of the choice set. It seems that the respondents do not increase or decrease their thresholds value substantially when given binary sets instead of trinary sets, or vice versa.
\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Set & Model & Intercept & Interesrate & Annualfee & Travelpoints & Cashreward \\
\hline
1 & Dep \((\epsilon_i \text{ differ})\) & -3.095 & -1.331 & -2.028 & -3.997 & -3.328 \\
& & (0.167) & (0.033) & (0.036) & (0.046) & (0.039) \\
& Dep \((\epsilon = 0.001)\) & -1.055 & -1.631 & -2.346 & -4.318 & -3.583 \\
& & (0.054) & (0.044) & (0.047) & (0.062) & (0.053) \\
& Dep \((\epsilon \to \infty)\) & -3.965 & 0.000 & -0.357 & -2.626 & -1.672 \\
& & (0.193) & (0.009) & (0.009) & (0.077) & (0.026) \\
& Indep \((K = 1)\) & 2.910 & 1.416 & 1.253 & -2.417 & -0.125 \\
& & (0.022) & (0.006) & (0.007) & (0.186) & (0.024) \\
& Indep \((K = 1)\) & 3.257 & 1.696 & 1.537 & -2.296 & -0.100 \\
& & (0.033) & (0.008) & (0.007) & (0.205) & (0.043) \\
& MNL & NA & 1.312 & 0.694 & -1.482 & -0.344 \\
& & NA & (0.003) & (0.002) & (0.019) & (0.004) \\
\hline
4 & Dep \((\epsilon_i \text{ differ})\) & -1.546 & -0.138 & -0.870 & -3.464 & -2.132 \\
& & (0.225) & (0.055) & (0.067) & (0.107) & (0.083) \\
& Dep \((\epsilon = 0.001)\) & -0.388 & -0.204 & -0.922 & -3.537 & -2.193 \\
& & (0.038) & (0.058) & (0.067) & (0.102) & (0.082) \\
& Dep \((\epsilon \to \infty)\) & -2.639 & 0.888 & 0.491 & -2.828 & -0.714 \\
& & (0.291) & (0.011) & (0.012) & (0.202) & (0.024) \\
& Indep \((K = 1)\) & -0.280 & 1.535 & 1.287 & -1.891 & -0.029 \\
& & (0.431) & (0.006) & (0.004) & (0.110) & (0.021) \\
& Indep \((K = 1)\) & -0.173 & 1.828 & 1.584 & -1.845 & 0.052 \\
& & (0.352) & (0.006) & (0.008) & (0.274) & (0.048) \\
& MNL & NA & 2.570 & 1.840 & 0.427 & 1.061 \\
& & NA & (0.013) & (0.014) & (0.017) & (0.014) \\
\hline
\end{tabular}
\caption{Averages of the posterior means of \(\mu_i\) for all models for the first and the fourth calibration sets. The numbers in parentheses are the standard errors over 10 MCMC chains.}
\end{table}
Figure 4.3: Histogram of posterior means of $\epsilon_i$ for the Dep($\epsilon_i$ differ) model.
Figure 4.4: The estimated $K_i$ from the first calibration set vs. $K_i$ from the second calibration set. The first calibration set consists of all trinary sets and the second calibration set consists of all binary sets.
The results for the first and fourth holdout samples provide strong evidence that the dependent race models with a proper tie-breaking rule have better overall predictive power than the MNL model and the independent race models. However, the results based on the second and the third holdout samples do not support the dependent race models. We notice that there are only eight binary sets and six trinary sets for calibration in the second and third sets of analysis, respectively, while twelve and seventeen sets are used to calibrate the first and fourth analyses. It is likely that the second and third calibration sets are not large enough to provide sufficient information for fitting the models. To investigate this phenomenon, we also conducted a case-deleted analysis.

Case-deletion methods were used to cross-validate all the models mentioned previously. To have the most data for calibration, a single choice set (one case) is deleted at a time, and the holdout measures are computed for this case. The same analysis is repeated for every choice set for each respondent. Computation of the holdout measures is briefly explained here. Let \( f(s) \) be the posterior distribution given the full data, and \( f_{-i}(s) \) be the posterior distribution when the \( i \)th case is dropped. The forms of \( f(s) \) and \( f_{-i}(s) \) are known up to a normalizing constant. Let \( s_1, s_2, \ldots, s_N \) be a random sample drawn from the full posterior distribution \( f(s) \). Then, for some real valued function \( g(s) \), e.g., \( MSE_1, R_{corr}, P_{corr} \), and \( log(QL) \), the expectation

\[
E_{-i}[g(s)] = \int g(s)f_{-i}(s)ds
\]

(provided that it exists) can be estimated by

\[
\hat{E}_{-i}[g(s)] = \frac{\sum_{n=1}^{N} w_{-i}(s_n)g(s_n)}{\sum_{n=1}^{N} w_{-i}(s_n)},
\]

(4.5.7)
where \( w_{-i} = q_{-i}(s)/q(s) \) and \( q_{-i}(s) \) and \( q(s) \) are the known parts of the two posterior distributions \( f(s) \) and \( f_{-i}(s) \) without the normalizing constants. The weights \( w_{-i} \) are normalized to have a mean of 1. To ensure asymptotic normality of the estimator, the variance of the weights should be finite. A useful indicator of the size of variance of weights is the Effective sampling size (ESS), which is defined to be

\[
ESS = \frac{N}{1 + \text{var}(w_{-i})}.
\]

If the variance of the \( w_{-i} \) is finite, then the ESS is a relatively large fraction of \( N \).

Table 4.13 lists the holdout measures computed with respect to the case-deleted posterior distributions. These results confirm that the dependent race model with different \( \epsilon_i \) and with the same \( \epsilon = 0.001 \) for all 81 respondents are the best models. The MNL model falls behind these two dependent race models, but is better than the dependent race model with \( \epsilon \to \infty \) and the two independent race models. Its correct choice rate for the dominance sets are the highest among all the models. The independent race model and the independent race model with \( K = 1 \) are the worst models.

The minimum effective sample sizes for all subjects averaged over 10 MCMC chains are reported for each choice set in Table 4.14. The minimum effective sample sizes seem to be reasonably large for most choice sets for the dependent race models and the MNL model. However, the minimum ESS for several cases in the two independent race models are quite small, which might indicate that importance sampling technique is not working for these models.
<table>
<thead>
<tr>
<th>Model</th>
<th>Dep ( (\epsilon_1 \text{ differ}) )</th>
<th>Dep ( (\epsilon = 0.001) )</th>
<th>Dep ( (\epsilon \rightarrow \infty) )</th>
<th>Indep ( (K = 1) )</th>
<th>Indep ( (K = 1) )</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.339 (9E-05)</td>
<td>0.337 (2E-04)</td>
<td>0.390 (3E-04)</td>
<td>0.404 (2E-04)</td>
<td>0.410 (2E-04)</td>
<td>0.351</td>
</tr>
<tr>
<td>( \epsilon ) differed ( (\epsilon = 0.001) )</td>
<td>( (\epsilon \rightarrow \infty) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dom</td>
<td>0.130 (8E-05)</td>
<td>0.129 (2E-04)</td>
<td>0.156 (8E-05)</td>
<td>0.160 (4E-04)</td>
<td>0.164 (2E-04)</td>
<td>0.127</td>
</tr>
<tr>
<td>nodom</td>
<td>0.429 (1E-04)</td>
<td>0.426 (2E-04)</td>
<td>0.490 (5E-04)</td>
<td>0.508 (2E-04)</td>
<td>0.515 (2E-04)</td>
<td>0.447</td>
</tr>
<tr>
<td>( K = 1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.655 (2E-04)</td>
<td>0.658 (2E-04)</td>
<td>0.602 (2E-04)</td>
<td>0.591 (1E-04)</td>
<td>0.588 (1E-04)</td>
<td>0.650</td>
</tr>
<tr>
<td>( P_{corr} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dom</td>
<td>0.684 (2E-04)</td>
<td>0.690 (2E-04)</td>
<td>0.623 (2E-04)</td>
<td>0.658 (3E-04)</td>
<td>0.655 (3E-04)</td>
<td>0.674</td>
</tr>
<tr>
<td>nodom</td>
<td>0.642 (2E-04)</td>
<td>0.645 (2E-04)</td>
<td>0.593 (3E-04)</td>
<td>0.563 (1E-04)</td>
<td>0.560 (1E-04)</td>
<td>0.640</td>
</tr>
<tr>
<td>( R_{corr} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.825 (3E-04)</td>
<td>0.828 (3E-04)</td>
<td>0.799 (8E-04)</td>
<td>0.765 (5E-04)</td>
<td>0.765 (7E-04)</td>
<td>0.817</td>
</tr>
<tr>
<td>( \log(QL) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dom</td>
<td>0.926 (4E-17)</td>
<td>0.928 (4E-04)</td>
<td>0.925 (4E-04)</td>
<td>0.925 (3E-04)</td>
<td>0.926 (2E-04)</td>
<td>0.934</td>
</tr>
<tr>
<td>nodom</td>
<td>0.784 (4E-04)</td>
<td>0.788 (5E-04)</td>
<td>0.747 (1E-03)</td>
<td>0.700 (9E-04)</td>
<td>0.700 (9E-04)</td>
<td>0.770</td>
</tr>
<tr>
<td>( QLR )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>-894.2 (0.329)</td>
<td>-889.3 (0.560)</td>
<td>-1024.7 (0.739)</td>
<td>-1032.6 (0.925)</td>
<td>-1048.1 (0.925)</td>
<td>-945.0</td>
</tr>
<tr>
<td>( \log(QL) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dom</td>
<td>-115.4 (0.151)</td>
<td>-115.2 (0.140)</td>
<td>-152.5 (0.108)</td>
<td>-153.4 (0.507)</td>
<td>-158.1 (0.456)</td>
<td>-122.0</td>
</tr>
<tr>
<td>nodom</td>
<td>-778.8 (0.305)</td>
<td>-774.1 (0.488)</td>
<td>-872.3 (0.660)</td>
<td>-879.2 (0.777)</td>
<td>-890.1 (0.425)</td>
<td>-823.1</td>
</tr>
<tr>
<td>QLR</td>
<td>1E+60 (2E+62)</td>
<td>2697 (1E-07)</td>
<td>1E+38 (1E+38)</td>
<td>1E+07 (1E+38)</td>
<td>1E+07 (1E+38)</td>
<td>1E+07</td>
</tr>
<tr>
<td>dom</td>
<td>3E+16 (4E+16)</td>
<td>3 (1E-02)</td>
<td>5E+13 (5E+13)</td>
<td>1E+05 (3E+24)</td>
<td>1E+05 (3E+24)</td>
<td>1E+05</td>
</tr>
<tr>
<td>nodom</td>
<td>4E+43 (4E+45)</td>
<td>1064 (1E-05)</td>
<td>3E+22 (3E+22)</td>
<td>2E-05 (2E-05)</td>
<td>2E-05 (2E-05)</td>
<td>2E-05</td>
</tr>
</tbody>
</table>

Table 4.13: Holdout statistics based on case-deleted estimates.
<table>
<thead>
<tr>
<th>Set</th>
<th>Dep ($\epsilon_i$ differ)</th>
<th>Dep ($\epsilon = 0.001$)</th>
<th>Dep ($\epsilon \to \infty$)</th>
<th>Indep ($K = 1$)</th>
<th>Indep ($K = 1$)</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5967</td>
<td>5944</td>
<td>5558</td>
<td>308</td>
<td>271</td>
<td>1308</td>
</tr>
<tr>
<td>2</td>
<td>696</td>
<td>4032</td>
<td>3533</td>
<td>91</td>
<td>136</td>
<td>3307</td>
</tr>
<tr>
<td>3</td>
<td>4952</td>
<td>5239</td>
<td>5321</td>
<td>199</td>
<td>441</td>
<td>3332</td>
</tr>
<tr>
<td>4</td>
<td>5799</td>
<td>2365</td>
<td>8062</td>
<td>8065</td>
<td>7557</td>
<td>6369</td>
</tr>
<tr>
<td>5</td>
<td>6401</td>
<td>4192</td>
<td>8476</td>
<td>8831</td>
<td>8653</td>
<td>7742</td>
</tr>
<tr>
<td>6</td>
<td>647</td>
<td>3019</td>
<td>3589</td>
<td>373</td>
<td>230</td>
<td>2789</td>
</tr>
<tr>
<td>7</td>
<td>4114</td>
<td>4974</td>
<td>5296</td>
<td>403</td>
<td>636</td>
<td>805</td>
</tr>
<tr>
<td>8</td>
<td>2468</td>
<td>3511</td>
<td>7194</td>
<td>594</td>
<td>497</td>
<td>5615</td>
</tr>
<tr>
<td>9</td>
<td>371</td>
<td>2412</td>
<td>695</td>
<td>338</td>
<td>168</td>
<td>1222</td>
</tr>
<tr>
<td>10</td>
<td>641</td>
<td>1044</td>
<td>620</td>
<td>235</td>
<td>296</td>
<td>1279</td>
</tr>
<tr>
<td>11</td>
<td>6815</td>
<td>7176</td>
<td>4681</td>
<td>777</td>
<td>677</td>
<td>4022</td>
</tr>
<tr>
<td>12</td>
<td>2552</td>
<td>3100</td>
<td>5575</td>
<td>219</td>
<td>223</td>
<td>3641</td>
</tr>
<tr>
<td>13</td>
<td>3931</td>
<td>5234</td>
<td>1100</td>
<td>509</td>
<td>329</td>
<td>663</td>
</tr>
<tr>
<td>14</td>
<td>1017</td>
<td>1370</td>
<td>299</td>
<td>63</td>
<td>38</td>
<td>360</td>
</tr>
<tr>
<td>15</td>
<td>2422</td>
<td>2922</td>
<td>779</td>
<td>567</td>
<td>787</td>
<td>1983</td>
</tr>
<tr>
<td>16</td>
<td>3884</td>
<td>4313</td>
<td>2563</td>
<td>5150</td>
<td>5948</td>
<td>680</td>
</tr>
<tr>
<td>17</td>
<td>3035</td>
<td>3623</td>
<td>1772</td>
<td>1043</td>
<td>925</td>
<td>611</td>
</tr>
<tr>
<td>18</td>
<td>5931</td>
<td>5633</td>
<td>5677</td>
<td>3138</td>
<td>2882</td>
<td>3941</td>
</tr>
<tr>
<td>19</td>
<td>6683</td>
<td>6914</td>
<td>2944</td>
<td>3243</td>
<td>3557</td>
<td>1778</td>
</tr>
<tr>
<td>20</td>
<td>4737</td>
<td>5393</td>
<td>1828</td>
<td>711</td>
<td>557</td>
<td>1627</td>
</tr>
</tbody>
</table>

Table 4.14: Minimum mean effective sample sizes (average over 10 MCMC chains) for all subjects for each choice set.
4.6 Conclusion

Conjoint studies with more than two alternatives are common in practice. Since traditional choice models and the independence race model can be easily extended to choice studies with multiple alternatives, it is important for the dependence race model we develop in Chapter 3 to have the same ease of extendibility. This chapter formulates the dependent race model in the case of multiple alternatives, derives the expression for the choice probabilities and demonstrates how to apply the models to choice data. With experimental data, we show that the dependent race model with a proper tie-breaking rule has consistently better predictive performance than the MNL model and the independent race models.

As seen in the analyses, when the dependent race models are used to extrapolate from trinary sets to binary sets, the results suggest that the dependent race models are better models in terms of the predictive performance. However, the extrapolation from binary sets to trinary sets fails to reach the same conclusion. The fourth method of analyzing the data as well as the case-deleted analyses suggests that failure is probably due to insufficient information. In future research, it might be interesting to investigate whether trinary sets contain the same amount of information about the respondents’ preferences as the binary sets.

Theoretically, the dependent race model we propose can be extended to choice sets with any number of alternatives. However, when a respondent is given two alternatives, the evaluation of the trade-off or dominance relationship is easily carried out. As the number of alternatives increases, such evaluation becomes increasingly difficult and respondents might resort to various heuristic strategies. Some of these
strategies are discussed by Bettman, Luce and Payne (1998). In this case, the dependent race model might be able to be combined with heuristic strategies to provide a better model for the decision-making process.
CHAPTER 5

CONTRIBUTIONS AND FUTURE WORK

This chapter summarizes the contributions of this research and discusses directions for future work.

5.1 Contributions

First, a range of theoretical properties of the independent Poisson race models have been derived. The expressions for choice probabilities, especially those based on the beta distribution, make it possible to compute the choice probabilities quickly and accurately when the threshold values are not integers without resorting to simulation. The results on the monotonicity and limiting behavior of the choice probabilities are very useful for understanding the range of choice behavior that can described with the model. Most important of all, the discussion on violation of IIA and transitivity provides an in-depth understanding of the implications of the Poisson race model and strong evidence that the class of Poisson race models can be used to model a range of decision behaviors that traditional models might not be able to capture.

The second contribution of this research is a new class of Poisson race models, called the “dependent Poisson race models”. This class of models is developed with the underlying idea that the alternatives in a choice set are compared across attributes
first and then the relative advantages are aggregated to obtain the overall advantage of an alternative. Thus, alternatives with the same level of an attribute should have a shared realization of the Poisson process. An experiment was conducted for this dissertation to produce a data set concerning preferences amongst credit cards. For these data, the dependent Poisson race model is shown to have better out-of-sample predictive performance than the multinomial logit model.

The third contribution of this research is an extension of the new dependent Poisson race model to choice sets with more than two alternatives. Methods were derived for computing the choice probabilities accurately and also proposals were made on how to handle ties in the races. A second conjoint choice study of credit cards was conducted. With this data set, the dependent Poisson race is again shown to have superior predictive performance as compared to the multinomial logit model and the independent Poisson race models. This improved predictive performance was established regardless of whether the held-out choice sets showed a dominance relationship or a trade-off relationship, provided there are sufficient information from the choice sets.

5.2 Future directions

There are many interesting issues for further research. One issue is the formulation of rates associated with each attribute. In the studies discussed in this dissertation, the attributes have a clear direction of preference and the direction is the same for all respondents. Many attributes do not have such properties, for example, the color and style of a car. Some of these attributes are inherently multidimensional. For some, the preference ordering varies from respondent to respondent. For this type of
attribute, some measure of “similarity” might be used to determine how much of the rate is shared and how much is unique to each attribute.

A second issue is whether the linear structure for the combination of rates across attributes is appropriate. While this assumption is at the foundation of conjoint analysis (Krantz, Luce, Suppes & Tversky, 1971), there are many other settings where a collection of attributes is more (or less) valuable than the sum of their values. For such problems, a careful description of an “attribute” is needed. Structures with a component-wise monotonicity could still retain the dependence among the Poisson processes while replacing (4.2.4) with a different expression.

A third issue is the application of the dependent Poisson race model to various types of choice data, for example, conjoint rating studies and partial profile studies. In a conjoint rating study, a respondent would rate the alternatives presented according to some scale. The Poisson race model, assuming the race continues until all the processes reaches their threshold, might be adapted to model conjoint rating data. In a partial profile study, the set of attributes presented in one choice set might be different from the set of attributes seen in the next choice set. For the unseen attributes, no matter whether they have been presented previously or not, the respondent would need construct a latent model for them in order to access the value of the offer. The latent model for the unseen attributes will have strong affect on the dependence structure of the race models.
APPENDIX A

PROOF OF LEMMA 2.3.5

Lemma 2.3.5 is proved by induction method in this appendix. The induction hypothesis is stated as below. For any integers $K_1$ and $K_2$ and $i \geq 0$,

$$\frac{(K_2 + K_1 - 1)!}{(K_2 + K_2 - 1 - i)!} = \sum_{j=0}^{i} c_{j, b}^{(i)} K_2^j$$

(A1)

where $b = K_1 - i$ and $c_{j, b}^{(i)}$ is the coefficient of $K_2^j$. The coefficient is defined through a recursive relationship:

$$c_{j, b}^{(q)} = c_{j-1, b}^{(q-1)} + (b + q - 1)c_{j, b}^{(q-1)}$$

for any integer $q \geq 0$ (A2)

with

$$c_{0, b}^{(1)} = b,$$

$$c_{j, b}^{(q)} = 1 \text{ if } j = q,$$

$$c_{j, b}^{(q)} = 0 \text{ if } j < 0 \text{ or } j > q.$$

Proof. For simplicity, let

$$G(i) = \frac{(K_2 + K_1 - 1)!}{(K_2 + K_2 - 1 - i)!} \text{ and } G'(i) = \sum_{j=0}^{i} c_{j, b}^{(i)} K_2^j.$$

(A3)

For $i = 0$, $G(0) = 1$. For $i \geq 1$, let $b = K_1 - i$, then,

$$G(i) = (K_2 + K_1 - 1) \ldots (K_2 + K_1 - i) \text{ totally } i \text{ terms.}$$

$$= (K_2 + b)(K_2 + b + 1) \ldots (K_2 + b + i - 1)$$

$$= \prod_{j=1}^{i} (K_2 + b + j - 1).$$
Therefore, if $i = 1$, $G(1) = K_2 + K_1 - 1$.

(1) When $i = 0$, $j = 0$, and $b = K_1$. Since $j = q$, from the definition of $c_{j, b}^{(q)}$ in the induction hypothesis, $c_{0, b}^{(0)} = 1$. Therefore,

$$G'(0) = c_{0, b}^{(0)} K_2^0 = 1 = G(0).$$

When $i = 1$, $j = 0$, 1, and $b = K_1 - 1$. From (A2),

$$c_{j=0, b}^{(i=1)} = c_{-1, b}^{(0)} + bc_{0, b}^{(0)} = b$$

and

$$c_{j=1, b}^{(i=1)} = c_{0, b}^{(0)} + bc_{1, b}^{(0)} = 1.$$

Therefore,

$$G'(1) = c_{j=0, b}^{(i=1)} K_2^0 + c_{j=1, b}^{(i=1)} K_2^1 = b + K_2 = K_1 - 1 + K_2 = G(1).$$

(2) When $i = r - 1$, assume $G(r - 1) = G'(r - 1)$ is true. That is,

$$G(r - 1) = (K_2 + K_1 - 1) \ldots (K_2 + K_1 - (r - 1)) \text{ totally } r - 1 \text{ terms}$$

$$= (K_2 + b)(K_2 + b + 1) \ldots (K_2 + b + r - 2)$$

$$= \prod_{j=1}^{r-1} (K_2 + b + j - 1)$$

$$= \sum_{j=0}^{r-1} c_{j, b}^{(r-1)} K_2^j$$

$$= G'(r - 1),$$

where $b = K_1 - i = K_1 - (r - 1)$.  

216
(3) For \(i = r\), \(b = K_1 - i = K_1 - r\). Then, for \(G(r)\), we have
\[
G(r) = (K_2 + K_1 - 1) \ldots (K_2 + K_1 - (r - 1))(K_2 + K_1 - r)
\]
\[
= (K_2 + b)(K_2 + b + 1) \ldots (K_2 + b + r - 2)(K_2 + b + r - 1)
\]
\[
= \left[ \sum_{j=0}^{(r-1)} c_j^{(r-1)} K_2^j \right] (K_2 + b + r - 1) \quad \text{by (A4)}
\]
\[
= \sum_{j=0}^{r-1} c_j^{(r-1)} K_2^{j+1} + (b + r - 1) \sum_{j=0}^{r-1} c_j^{(r-1)} K_2^j
\]
\[
= \sum_{j=0}^{r} c_j^{(r-1)} K_2^j + (b + r - 1) \sum_{j=0}^{r} c_j^{(r-1)} K_2^j
\]
\[
\quad \text{since } c_j^{(q)} = 0 \text{ if } j > q \text{ or } j < 0
\]
\[
= \sum_{j=0}^{r} [c_j^{(r-1)} + (b + r - 1)c_j^{(r-1)}] K_2^j
\]
\[
= \sum_{j=0}^{r} c_j^{(r)} K_2^j \quad \text{by (A2)}
\]
\[
= G'(r) \quad \text{(A5)}
\]

The induction hypothesis is true for the case \(i = r\) and Lemma 2.3.5 is proved. \(\square\)
APPENDIX B

A CONJOINT CHOICE EXPERIMENT WITH TWO CREDIT CARDS

In this appendix, a survey used to conduct the conjoint choice experiment is included. According to the design matrix in Table 3.14, there are 18 choice sets with two credit cards each. There are totally four versions of the survey according to different randomization scheme. The order of the choice sets presented in each page and which credit card appears first in every choice set are randomized. In the administration of the experiment, respondents are given any of the four versions of the survey randomly. Since different versions contain essentially the same information, only one is included in this appendix.
INSTRUCTIONS

Welcome to our marketing study! In this study, you will be presented with 36 different credit card offers in pairs. Please indicate the credit card you prefer. More detailed instructions will be provided when you are ready for the task.

Information on the following characteristics of credit card offers is provided:

- Interest Rate (APR)
- Annual Fee
- Travel Mileage Points
- Cash Rebate

Based upon these product attributes, product configurations such as the followings will be described to you.

- 9.99% Interest Rate
  - $30 Annual Fee
  - Travel Mileage Points – 1 point for every $1 Spent

- 17.99% Interest Rate
  - No Annual Fee
  - 1% Cash Rebate

Your task is to compare each pair of credit cards and select the one that you prefer.
Choice set 1:

Which credit card would you prefer?  1st      2nd

Choice set 2:

Which credit card would you prefer?  1st      2nd
Choice set 3:

Which credit card would you prefer?  1st      2nd

Choice set 4:

Which credit card would you prefer?  1st      2nd
Choice set 5:

Which credit card would you prefer?  1st      2nd

Choice set 6:

Which credit card would you prefer?  1st      2nd
Choice set 7:

Which credit card would you prefer?  1st  2nd

Choice set 8:

Which credit card would you prefer?  1st  2nd
Choice set 9:

Which credit card would you prefer?  1st      2nd

Choice set 10:

Which credit card would you prefer?  1st      2nd
Choice set 11:

Which credit card would you prefer? 1st 2nd

Choice set 12:

Which credit card would you prefer? 1st 2nd
Choice set 13:

Which credit card would you prefer? 1st 2nd

Choice set 14:

Which credit card would you prefer? 1st 2nd
Choice set 15:

Which credit card would you prefer? 1st 2nd

Choice set 16:

Which credit card would you prefer? 1st 2nd
Choice set 17:

Which credit card would you prefer? 1st 2nd

Choice set 18:

Which credit card would you prefer? 1st 2nd
Demographic questions:

1. Gender:
   M         F

2. How many credit cards do you have?
   0   1    2   more than 2

3. During the past year, did you incur any charges due to interest on your credit cards:
   Yes     No

4. Among the four attributes:
   1    interest rate
   2    annual fee
   3    travel mileage points
   4    cash reward

   Which one was the most important to you when making your choices: ___________

   Which one is the least important to you when making your choices: ___________

5. Do you travel a lot?
   Yes           No
APPENDIX C

A CONJOINT CHOICE EXPERIMENT WITH THREE CREDIT CARDS

In this appendix, a survey used to conduct the conjoint choice experiment is included. According to the design matrix in Table 4.3, there are 12 trinary sets and 8 binary sets. In the survey, the trinary sets are presented before the binary sets. Again, there are totally four versions of the survey according to different randomization scheme. The order of the choice sets presented in each page and which credit card appears first in every choice set are randomized. In the administration of the experiment, respondents are given any of the four versions of the survey randomly. Since different versions contain essentially the same information, only one is included in this appendix.
INSTRUCTIONS

In this study, you will be presented with different credit card offers either in pairs or triples. For each pair or triple of cards, you will be asked which card you prefer. Please make a choice for each set of cards--ties are not allowed.

Information on the following attributes of credit card offers is provided:

- Interest Rate (APR)
- Annual Fee
- Travel Mileage Points
- Cash Rebate

Interest Rates range from 9.99% to 17.99%. Annual Fees range from $0 to $30. When Travel Mileage Points are available, the rate is 1 point for every $1 spent. When a Cash Rebate is available, it is a 1% Cash Rebate.

Based upon these product attributes, product configurations such as the following will be described to you:

These three cards have the same Interest Rate. The first card, on the left, has Travel Mileage Points but does not have a Cash Rebate. The second card, in the middle, has a 1% Cash Rebate but does not have Travel Mileage Points. The third card, on the right, has neither Travel Mileage Points nor a Cash Rebate, but there is no Annual Fee.

You will be asked to circle 1st, 2nd, or 3rd to indicate your preferred card.
Choice set 1:

![Choice Set 1 Images]

Which credit card would you prefer?   1st      2nd      3rd

Choice set 2:

![Choice Set 2 Images]

Which credit card would you prefer?   1st      2nd      3rd

Choice set 3:

![Choice Set 3 Images]

Which credit card would you prefer?   1st      2nd      3rd
Choice set 4:

Which credit card would you prefer?  1st  2nd  3rd

Choice set 5:

Which credit card would you prefer?  1st  2nd  3rd

Choice set 6:

Which credit card would you prefer?  1st  2nd  3rd
Choice set 7:

Which credit card would you prefer?  1st      2nd      3rd

Choice set 8:

Which credit card would you prefer?  1st      2nd      3rd

Choice set 9:

Which credit card would you prefer?  1st      2nd      3rd
Choice set 10:

Which credit card would you prefer? 1st 2nd 3rd

Choice set 11:

Which credit card would you prefer? 1st 2nd 3rd

Choice set 12:

Which credit card would you prefer? 1st 2nd 3rd
Choice set 13:

Which credit card would you prefer? 1st 2nd

Choice set 14:

Which credit card would you prefer? 1st 2nd

Choice set 15:

Which credit card would you prefer? 1st 2nd
Choice set 16:

Which credit card would you prefer? 1st 2nd

Choice set 17:

Which credit card would you prefer? 1st 2nd

Choice set 18:

Which credit card would you prefer? 1st 2nd
Choice set 19:

Which credit card would you prefer? 1st 2nd

Choice set 20:

Which credit card would you prefer? 1st 2nd

Did you make considered choices, or did you simply choose cards at random?

Considered Choice Random Choice

Thank you for finishing the survey!
BIBLIOGRAPHY


239


