Fall 2017

Methods for Analyzing Attribute-Level Best-Worst Discrete Choice Experiments

Amanda Faye Working
Old Dominion University

METHODS FOR ANALYZING ATTRIBUTE-LEVEL BEST-WORST DISCRETE CHOICE EXPERIMENTS

by

Amanda Faye Working
B.A. May 2008, Virginia Wesleyan College
B.S. December 2009, Old Dominion University
M.S. May 2012, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
December 2017

Approved by:

Norou Diawara (Director)

N. Rao Chaganty (Member)

Lucia Tabacu (Member)

Cynthia Jones (Member)

ABSTRACT

METHODS FOR ANALYZING ATTRIBUTE-LEVEL BEST-WORST DISCRETE CHOICE EXPERIMENTS

Amanda Faye Working
Old Dominion University, 2017
Director: Dr. Norou Diawara

Discrete choice experiments (DCEs) have applications in many areas such as social sciences, economics, transportation research, health systems, and clinical decisions to mention a few. Usually discrete choice models (DCMs) focus on predicting the product choice; however, these models do not provide information about what attributes of the products are impacting consumers' choices the most. Today, it is common to record the best and worst features of a product (or profile), also called attribute levels, and the goal is to investigate and build models for estimation of attribute and attribute-level impacts on consumer behavior. Attribute-level best-worst DCEs provide information into what consumers find the most important when considering different products. The design of attribute-level best-worst DCEs and the associated theory are discussed by Street and Knox (2012). Attribute-level best-worst discrete choice models can help to market products to the consumers and are often used in health economics research. These experiments help companies to best target consumers with their products or services. The latter can better advertise their products by highlighting and/or downplaying certain key attributes (or attribute-levels) to best earn the interests and business of consumers. We propose a time dependent model that can adapt to changes that occur in areas such as public opinion. A time dependent model accounts for the impact of time in a consumer's perception of a product and adjusts the utility to reflect that. These models are Markov processes and are often found under dynamic programming. Rust (1994), provides time dependent models for the usual DCEs. We extend this time dependent model to the attribute-level best-worst DCEs. Two example studies are presented to examine the dynamic versus static performance of transition matrices for estimation and inference of attributes and attribute level effects with regards to the expected utilities.

This dissertation is dedicated to my mother, *Vicky Kay Working* (1957 - 2015), and my grandmother, *Zalla Faye Working* (1938 - 2016). I regret that they are unable to be here today to see me complete this work. I hope that I make them proud.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance of my committee, the support of my friends, and the love of my family. There are no words for how truly grateful I am to all of the wonderful people in my life.

I wish to thank, first and foremost, my advisors Dr. Dayanand Naik and Dr. Norou Diawara. I consider it a great honor to have worked with Dr. Naik. He was an amazing person whose loss is felt by many. I am forever thankful to have known and worked with him. My deepest gratitude goes to my advisor over the course of this dissertation, Dr. Diawara. His guidance and support is the reason there is a dissertation. He pushed me to complete my dissertation when the loss of my mother, Vicky Working, and my grandmother, Zalla Working, seemed overwhelming. Without his encouragement and support, I would not be where I am today. There truly are no words to express my heartfelt gratitude.

A special thanks to my dissertation committee, Dr. N. Rao Chaganty, Dr. Lucia Tabacu, and Dr. Cynthia Jones, for their guidance in the completion of my dissertation. They provided invaluable feedback that only helped to improve the quality of my work.

I would like to thank the Department of Mathematics and Statistics and all its amazing faculty and staff for providing a wonderful work environment and financial support during my graduate studies. I have met many great people and made lifelong friendships during my studies here. A very special thanks to Miriam Venable and Sheila Hegwood for always supporting me and giving me an earful when needed. I would also like to thank the Department of Modeling and Simulation for awarding me a fellowship to help financially support me during my research. A special thanks to my boss, Dr. Stephen Tonelson, for encouraging and supporting me to complete my dissertation.

Most importantly, I would like to thank my amazing family and friends. I am grateful to my parents, Dennis and Vicky Working, who worked hard and sacrificed to provide for their children. My parents were always there with their love, guidance, and support. I know that I could not have completed this dissertation without them. I would like to thank my grandma, Zalla Working, for her unconditional love and unwavering belief in me. Although my mom and grandma passed while I worked on my dissertation, I still feel their love and support bolstering me to complete this work. I would like to thank my grandpa, Larry Working, and grandma, Iris Jean England. Although they passed away many years ago, they helped shape me into who I am today and always loved and supported me. My heartfelt gratitude to my sisters and best-friends, Nicky Fuentes and Brandi Working. I feel so very blessed to have them for sisters. No matter what happens in life, we always have each other. They listen to my worries and offer advice when needed. I know that I would not be where I am, or who I am, without them. Thank you to my nephews, Kenneth Tyrrell, Ricky Cutshawshy, and Brayden Yocum, and my bother-in-law Jose Fuentes for their love and support. Thank you to Dory Working for making my dad so happy and for her love and support of me. I would also like to thank my friends for their love and support. A special thanks to Corinne Wilson, who is more like a sister than a friend. Her unwavering belief that I could and would complete my dissertation has always motivated me. I am truly grateful to have a friend like her.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Discrete choice experiments (DCEs) and their modeling describe consumers' behaviors. In these experiments, consumers are given a questionnaire or survey of a series of choice sets. Their task is to choose one alternative from a set of alternatives that benefits them the most. The alternatives, or products, being modeled play the role of the explanatory variables in regression modeled. They may be goods, services, policies, and/or scenarios. DCEs have applications in a multitude of fields including but not limited to, health systems research, public policy, transportation research, and economics. These experiments provide valuable information to businesses on the impact the features of a product have on the likelihood of the product being chosen over competitive alternatives. Best-worst scaling experiments are modified DCEs to elicit further information about the best and worst product, or best and worst attributes and attribute-levels of a product. Our research in the area DCEs is on the attribute-level best-worst DCEs and their models. We make extensions to the traditionally defined utility function and look at the impact of time on expected utility using Markov decision processes (MDPs).

1.1 DCEs PROBABILITY

In DCEs, there exists a set of alternatives which are a set of products, services, or scenarios from which respondents choose one alternative as their preference. Thurstone (1927) presented the idea to quantify people's reaction to a set of alternatives, or stimuli, that are described by a set of attributes from a product. This set is called the choice set and includes all possible alternatives (Train, 2009). In some experiments, the alternatives are divided into multiple choice sets from which each respondent chooses one alternative. Let there n respondents and J alternatives in the experiment. The response variable representing the choices made in the experiment is binary data and denoted as:

$$Y_{ij} = \begin{cases} 1, & \text{if } i^{th} \text{ respondent chooses } j^{th} \text{ alternative,} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$.

In the experiment, each of the alternatives is described by a set of K attributes or characteristics of the product or scenario being modeled. The attributes describing the product are presented in a profile, $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jK})$, where x_{jk} be the value for the k^{th} attribute for the j^{th} alternative, where $1 \leq k \leq K$ and $1 \leq j \leq J$. Using this information about the different alternatives, the respondent chooses their preferred alternative. These characteristics enter into the estimation of the gains or benefits a particular product offers consumers through the utility function.

Many discrete choice models (DCMs) have been proposed under random utility theory (RUT), which assigns a utility value to each alternative. In this setting, individuals choose an alternative with the highest utility. Marschak (1960) developed RUT where each alternative has an associated random utility function. Under RUT, the utility for i^{th} subject choosing the j^{th} choice is defined as:

$$U_{ij} = V_{ij} + \epsilon_{ij}, \tag{2}$$

where

U_{ij} is the utility the i^{th} subject receives from the j^{th} choice,

V_{ij} is the systematic component,

and ϵ_{ij} is the unobserved or error component of the utility function,

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$.

Consumers act in a way to maximize their utilities. The probability that an alternative is chosen given as:

$$P(U_{ij} - U_{ik} > 0, \forall j \neq k) \quad = \quad P(U_{ij} > U_{ik}, \forall j \neq k)$$

$$= P(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, \forall j \neq k)$$

$$= P(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall j \neq k),$$

which is the cumulative distribution function for $\epsilon_{ijk} = \epsilon_{ik} - \epsilon_{ij}$. Therefore, the probability is then dictated by the distribution of the error terms.

The systematic, or observable, component of the utility is defined as $V_{ij} = \mathbf{x}'_j\boldsymbol{\beta}$, where $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jK})$ is the set of explanatory variables, or attributes, corresponding to the j^{th} alternative, $1 \leq j \leq J$, and $\boldsymbol{\beta}$ is the associated regression parameters. Van der Pol et al. (2014) explained that the utility function may be expressed as a linear function, as seen here, a quadratic function, or may possibly be written as a stepwise functions of the explanatory variables. The explanatory variables are the attributes, or characteristics, defining the product in the experiment. One would imagine that information about the consumer would also be among the explanatory variables. Consumer specific information such as gender, salary, and education, are all factors that motivate consumer behavior. However, consumer specific information is often not included in choice models for the reasons that follow.

Let z_i, $i = 1, 2, \ldots, n$, represent subject specific information in the experiment. With the inclusion of subject specific information in the utility function, the utility would then be

$$U_{ij} = \mathbf{x}'_j\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\theta} + \epsilon_{ij},$$

for $i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, J$, with $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ the associated parameter vectors for the choice and subject, respectively. When taking the difference in utilities for two alternatives, as is done when computing the probability in Equation (3), z_i's cancel out. Train (2009) mentions the inclusion of subject specific information into the model in the form of z_{ij} must be tied to the alternatives, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$. Here, the subject's information is also choice specific. Models for including consumer information are captured under the multinomial logit (MNL) or mixed multinomial logit (MMNL).

1.2 COMMON MODELS

DCMs look to model consumer behavior using statistical models. There is an ever growing set of models to apply to DCEs. We provide a few of the most common type of models. However, for the research done in this dissertation, the conditional logit model is used. The conditional logit model has some restrictions, namely the independence between choices. Probit models and generalized extreme value models are commonly utilized in cases where independence is unable to be assumed. Much of the details about these models provided here come from Anderson et al. (1992) and Train (2009) textbook on the simulations of DCEs.

In a book review, Thisse and Norman (1994) examined Anderson et al. (1992) discrete choice model of product differentiation. Psychometric and quantitative science literature have used DCMs integrating indicators in latent class constructs, connecting attributes of a respondent with unobserved behaviors. This is seen in Kamakura and Russell (1989) and Hurtubia et al. (2014) in a transportation and travel discrete choice model study done in Nice, France and rural areas of Switzerland.

However, many authors have ignored the applicability in time dependent and less restrictive model assumption cases. The models built here assume consumer choices are homogeneous, characterizing patterns of choice substitutions by brand switching or price sensitivities.

1.2.1 CONDITIONAL LOGIT

The conditional logit is a popular statistical model applied to consumer behavior. Marschak (1960) states that the probability that an alternative is chosen can be done under the logit model. McFadden (1974) proves that if the error terms are distributed as type I extreme value distribution the probability an alternative is chosen is given by the logit. One of the assumptions for this model is the independence from irrelevant alternatives (IIA). The IIA axiom, or principle, also known as Luce's choice axiom (Luce, 1959), states that the choice probability is not altered by the inclusion or removal of alternatives from the set of alternatives and is important in later models proposed to predict consumer choice behavior. Let the error terms be independently and identically distributed type I extreme value distribution. The probability density function and cumulative distribution function are then:

$$f(\epsilon_{ij}) = e^{-\epsilon_i} e^{-e^{-\epsilon_{ij}}} \tag{3}$$

and

$$F(\epsilon_{ij}) = e^{-e^{-\epsilon_{ij}}}, \tag{4}$$

respectively.

In Equation (3), the probability that the j^{th} alternative is chosen over the k^{th} alternative depends on their respective error terms in their utilities, $\epsilon_{ijk} = \epsilon_{ik} - \epsilon_{ij}$, where $i = 1, 2, \ldots, n$, $j, k = 1, 2, \ldots, J$, and $j \neq k$. The difference of extreme value variables are known to have a logistic distribution. The choice probability derived by McFadden under Luce's IIA axiom is then given by:

$$P(U_{ij} > U_{ik}, \forall j \neq k) = \frac{e^{\mathbf{x}'_{ij}\beta}}{\Sigma_{k=1}^{J} e^{\mathbf{x}'_{ik}\beta}}, \tag{5}$$

where $i = 1, 2, \ldots, n$, $j, k = 1, 2, \ldots, J$, and $j \neq k$.

Referring to Train (2009), the derivation of Equation (5) is as follows:

$$\begin{aligned} P(U_{ij} - U_{ik} > 0, \forall j \neq k) &= P(U_{ij} > U_{ik}, \forall j \neq k) \\ &= P(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, \forall j \neq k) \\ &= P(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall j \neq k). \end{aligned} \tag{6}$$

The probability is then the cumulative distribution function of ϵ_{ik}, and using the independence and identically distributed condition on the error terms the probability is expressed as:

$$\begin{aligned} P(U_{ij} > U_{ik}, \forall j \neq k) &= P(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall j \neq k) \\ &= \int_{-\infty}^{\infty} \prod_{j \neq k} e^{-e^{-(\epsilon_{ij} + V_{ij} - V_{ik})}} e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} d\epsilon_{ij} \end{aligned}$$

$$= \int_{-\infty}^{\infty} \prod_{k=1}^{J} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} e^{-\epsilon_{ij}} d\epsilon_{ij}$$
$$= \int_{-\infty}^{\infty} e^{-e^{-\epsilon_{ij} \sum_{k=1}^{J} -(V_{ij}-V_{ik})}} e^{-\epsilon_{ij}} d\epsilon_{ij}.$$

We substitute $t = e^{-\epsilon_{ij}}$ with $d_t = -e^{-\epsilon_{ij}} d\epsilon_{ij}$, producing:

$$\begin{aligned} P(U_{ij} > U_{ik}, \forall j \neq k) &= -\int_{0}^{\infty} e^{-t \sum_{k=1}^{J} -(V_{ij}-V_{ik})} dt \\ &= \frac{1}{\sum_{k=1}^{J} e^{-(V_{ij}-V_{ik})}} e^{-t}\Big|_{0}^{\infty} \\ &= \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}}. \end{aligned}$$

This derivation may also be found in the Anderson et al. (1992).

We add one more noticeable clarification. In the literature, people use the conditional logit and multinomial logit interchangeably. Agresti (2007) among others are very clear that there is in fact a difference between the two models. The multinomial logit is a special case of the conditional logit model where the explanatory variables are subject specific information, and in the conditional logit the explanatory variables are the attributes about the product (Hoffman and Duncan, 1988). Agresti (2007) states that there is a distinction between the two models; even though, Agresti (1990) states that they are the same. Anas (1983) presents the MNL as a special case approach at the intersection of information theory (entropy function) and the multinomial logit. McFadden and Train (2000) present the MMNL model that includes both subject specific and choice specific information in the set of explanatory variables. The MMNL models is more complex than the conditional logit model because of its dependence on a mixing distribution and associated parameters. However, the model does allow for heterogeneous preferences of the subjects in the model. There exist other models that allow for heterogeneous preferences, or dependence among the error terms. One such model would be the probit model.

1.2.2 GEV

Generalized extreme value (GEV) models are a family of models wherein the unobserved terms of utility have an extreme value distribution. The popular multinomial logit and the nested logit models are among this family of models. McFadden (1978) presented theory behind these distributions and their application to modeling consumer behavior. GEV models are considered because they allow for dependence among the error terms of the utility. Ben-Akiva and Bierlaire (1999) and Train (2009) provide the choice probability and necessary conditions for GEV models that are given below.

Let G be a non-negative and differentiable real-valued function of the systematic component of the utility, that is $G = G(e^{V_1}, e^{V_2}, \ldots, e^{V_J})$. The probability of the j^{th} alternative being chosen is given as

$$P_j = \frac{e^{V_j} \frac{\partial G}{\partial e^{V_j}}}{G}, \tag{7}$$

if the following four properties are satisfied for $j = 1, 2, \ldots, J$.

Properties:

- $G \geq 0$ for all positive values of $e^{V_j}, \forall j = 1, 2, \ldots, J$.
- G is homogeneous of degree $\mu > 0$.
- $G \to \infty$ as $e^{V_j} \to \infty$ for any j.
- For any $\mathrm{j}_1, j_2, \ldots, j_k \in 1, 2, \ldots, J$ the following condition is satisfied

$$(-1)^k \frac{\partial^k G}{\partial e^{V_{i_1}} \ldots \partial e^{V_{i_1}}} \leq 0 \quad \forall \quad e^{V_{i_1}} \in \mathrm{I\!R}^{\mathrm{J}}. \tag{8}$$

Small (1987) extended the GEV models to ordinal data providing the ordered generalized extreme value distribution (OGEV). Swait (2001) presented the Generation logit model termed GenL and demonstrated how the GEV models may be utilized to generate choice sets for DCEs.

1.2.3 MULTIONMIAL PROBIT

Hausman and Wise (1978) and Anderson et al. (1992) presented the probit model as an alternative statistical model for choice data. The multinomial probit model overcomes the limitation of independence of the unobserved terms in the utility that exists in the conditional logit. In some DCEs, the assumption of independence seems contradictory to the experiment. In DCEs, where multiple choices are made by each individual, such as panel data, correlations among the choices would be a typical assumption.

The multinomial probit model assumes that the error terms, $\epsilon_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, J$, have a multivariate normal distribution with a mean of zero and a variance-covariance matrix Σ. Referring to Equation (3), the probability that the j^{th} choice was chosen is the cumulative distribution function of $\epsilon_{ijk} = \epsilon_{ik} - \epsilon_{ij}$. The multivariate normal distribution has some desirable properties. Some properties are that each ϵ_{ij} has a normal distribution and the difference between normally distributed variables also have normal distributions. There are $J-1$ differences between the j^{th} alternative and the other alternatives in the set. These $J-1$ variables have a multivariate normal distribution. Train (2009) defines these differences as $\tilde{U}_{ijk} = U_{ik} - U_{ij}$, $\tilde{V}_{ijk} = V_{ik} - V_{ij}$, and $\tilde{\epsilon}_{ijk} = \epsilon_{ik} - \epsilon_{ij}$.

Furthermore,

$$\begin{aligned} P(U_{ij} > U_{ik}, \forall j \neq k) &= P(U_{ij} - U_{ik} > 0, \forall j \neq k) \\ &= P(U_{ik} - U_{ij} < 0, \forall j \neq k) \\ &= P(\tilde{U}_{ijk} < 0, \forall j \neq k) \\ &= P(\tilde{V}_{ijk} + \tilde{\epsilon}_{ijk} < 0, \forall j \neq k) \\ &= \int I(\tilde{V}_{ijk} + \tilde{\epsilon}_{ijk} < 0, \forall j \neq k)\phi(\tilde{\epsilon}_{ijk})d\tilde{\epsilon}_{ijk}, \end{aligned} \tag{9}$$

where ϕ is the probability density function of the multivariate normal distribution with a mean vector of zero and a $(J-1) \times (J-1)$ variance-covariance matrix Σ_j, where $j = 1, 2, \ldots, J$. The density of the ϵ_{ijk} is given as,

$$\phi(\tilde{\epsilon}_{ijk}) = \frac{1}{(2\pi)^{\frac{1}{2}}(J-1)|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}\tilde{\epsilon}'_{ijk}\Sigma_j^{-1}\tilde{\epsilon}_{ijk}}, \tag{10}$$

where $i = 1, 2, \ldots, n$, $j, k = 1, 2, \ldots, J$, and $j \neq k$.

1.3 OUTLINE OF DISSERTATION

This dissertation presents methods for modeling attribute-level best-worst discrete choice experiments (DCEs) with the inclusion of time. In Chapter 2, attribute-level best-worst DCEs are introduced. These models are a special case of the broader best-worst scaling type models. A literature review of attribute-level best-worst DCEs from their development to methods for modeling this data using models such as conditional logit, multinomial logit, and probit is provided. Model definition as well as probability theory and associated properties are defined. The scope of the work covers the statistical methodology and a synthesized analysis and case studies using simulated and aggregated data.

In Chapter 3, we extend the existing work done on partial profile models for pairwise comparison of choices. Relevant literature about the models are presented. An extension of these models with regards to attribute-level best-worst DCEs is done. Attribute-level best-worst data are presented as indicator functions demonstrating the equivalence of these models to the traditional attribute-level best-worst models. The indicator functions are then generalized providing an alternative method for defining attribute-level models. The functional form of the data definition provides an adaptive model able to conform to changes in the profile, or set of attributes, over time.

In Chapter 4, Markov decision processes (MDPs) are considered with regards to time sensitive attribute-level best-worst DCEs. MDPs are sequential decision making processes that determine an optimal decision policy that maximizes discounted reward over time. Existing work done with MDPs with regards to DCEs have not been extended to best-worst scaling type experiments. We define the necessary properties for MPDs to function with regards to attribute-level best-worst DCEs. We present static and dynamic transition matrices. Utility, acting as the rewards, are defined in multiple ways: 1) using the traditional definition provided in Chapter 2, and 2) using the more malleable form in terms of the systematic component from Chapter 3.

In Chapter 5, estimation of consumer behavior models are provided using numerical maximization under a criteria function is described. We also discuss different possible variant of scenarios. Two different examples with mixed scenarios are presented. A hypothetical situation is simulated and analyzed. The scenarios are then used to model MDPs under stationary and dynamic transition probabilities. We use simulation methods to replicate a real world situation.

In Chapter 6, we provide concluding remarks and future work.

CHAPTER 2

ATTRIBUTE-LEVEL BEST-WORST DISCRETE CHOICE EXPERIMENTS

2.1 LITERATURE REVIEW

Discrete choice experiments (DCEs) and their modeling describe consumer consumers' behaviors. Given a set of descriptors about a product, one can estimate the probability an alternative is chosen provided a statistical model appropriate to the data. However, these models are limited in the information they provide. According to Lancsar et al. (2013), there exist only two ways to gain more information from traditional DCEs through the increase in sample size and/or the number of choice sets evaluated by respondents adding to the burden on respondents in the experiments. Louviere and Woodworth (1991) and Finn and Louviere (1992) presented best-worst scaling experiments that are modified DCEs designed to elicit more information about choice behavior than the pick one approach implemented in the traditional DCEs without the added burden on the respondents.

There exist three cases of best-worst scaling experiments. The cases are: 1) best-worst object scaling, 2) best-worst attribute scaling, and 3) best-worst discrete choice experiment (Lancsar et al., 2013; Flynn, 2010). The first case was introduced in Finn and Louviere (1992). In case 1, a list objects, scenarios, or attributes are given to respondents and they choose the best and worst alternative. Unlike in traditional DCEs, no information about the object is provided to the respondents. In case 2, profiles composed of attribute levels for each attribute describing a product are determined. From the profiles, respondents are tasked with choosing the best and the worst attribute-level pair. Case 3 most closely resembles the traditional DCEs where a set of attributes describing the product are provided to the respondents and the respondents choose the best and worst product from the choice set. In case 3, researchers may ask respondents to sequentially choose the best and worst options repeatedly from the set until none are left to choose from (Lancsar et al., 2013).

These experiments overcome many limitations that exist in other experiments designed to elicit consumer preferences. Ranking experiments or use of rating scales are alternatives methods to gain information on the ordering in product preferences are burdensome on the respondents and rating scales are unreliable due to bias and do not translate across countries (Adamsen et al., 2013; Auger et al., 2007; Flynn, 2010; Loose and Lockshin, 2013; Massey et al., 2015). According to Yoo and Doiron (2013) a full ranking is obtained with best and worst choices, and are more reliable (Louviere et al., 2013). In general, it is often easier to state what you love or hate about something than to choose just one object from a set of objects. Marley and Louviere (2005) states that a single response, or choice, from a best-worst scaling type of experiments provides more information than traditional DCEs, and tend to be easier on respondents.

Although the experiments were presented in the early 1990's, it was not until Marley and Louviere (2005) that mathematical probabilities and properties were formally determined and published. Marley and Louviere (2005), Marley et al. (2008), and Marley and Pihlens (2012) provided the probability and properties to best-worst scaling experiments for the three cases. Additionally, Lancsar et al. (2013) provided the probability and utility definition for case 3 experiments that include the sequential best-worst choice from a set of choices. Other work found in the literature with regards to these experiments are the design of the experiments and dealing with taste heterogeneity. Louviere and Woodworth (1983) stated that orthogonal, main effects, and fractional factorial designs provide better parameter estimates than other designs. In application to best-worst scaling experiments, balance incomplete block designs (BIBD)(Louviere et al. 2013; Parvin et al. 2016) and orthogonal main effects plans (OMEPs) are popular designs (Flynn et al., 2007; Knox et al., 2012; Street and Knox, 2012). These designs and their properties are examined by Street and Burgess (2007). Louviere et al. (2013) looked at the design of experiments for best-worst scaling experiments and stated that it is possible to determine individual parameter estimates for the respondents.

In the Introduction, descriptions of common DCMs were described. While the conditional logit model is popular in application, it does not allow for taste heterogeneity, or heterogeneity in the error terms. Models accounting for taste heterogeneity has been explored in the literature for best-worst scaling experiments. Mueller et al. (2009) and Goodman et al. (2006) dealt with taste heterogeneity in their study by segmenting or forming more homogeneous clusters of products and consumers. Lancsar et al. (2013) accounted for taste heterogeneity by including subject specific information in the definition of the utility. As mentioned in the introduction, subject specific information vary with the alternatives to be included in the model. Lancsar et al. (2013) accomplished this by including interactions with subject specific variables in the model. However,Agresti (2007) pointed out that the inclusion of such interactions is problematic due to the number of parameters in the model.

Although the research in the area of best-worst scaling experiments has increased, there is a lack of literature on some topics. Louviere et al. (2013) revealed that no sample size estimation methods exist for these experiments. Additionally, Lancsar et al. (2013) noted that literature on the experiments is sparse and does not include a "comprehensive guide" for implementing these experiments.

This dissertation focuses on best-worst attribute scaling, also known as attribute-level best-worst DCEs. These experiments seek to determine the extent to which attributes and their associated attribute-levels impact consumer behavior. Louviere and Timmermans (1990) introduced hierarchical information integration (HII) for the examination of the valuation of attributes in DCEs. Under HII, the impact of an attribute necessitates discerning the various levels of the attribute. An experiment must be designed in a way to measure the different levels varying across products to determine such an impact. In attribute-level best-worst DCEs, the levels of attributes are well defined and vary across profiles, or products, providing sufficient information to measure their impact. Attribute-level best-worst discrete choice experiments provide more information into consumer's choices of products than the usual discrete choice experiments and add more value to the understanding of the data (Marley and Louviere, 2005). Those models outperform the standard logit modeling in terms of goodness of fit as mentioned in Hole (2011) in the context of attribute attendance.

Traditional DCEs provide information about attributes but not the overall impact of the attributes and their levels. In traditional DCEs, attributes are not scaled, or broken down into categorical levels; therefore, the impact they have is subjective (Flynn et al., 2007; Flynn et al., 2008; Lancsar et al., 2007). The parameter estimates in attribute-level best-worst DCEs provide information about the impact of the attributes and attribute-levels where as in traditional DCEs the parameters describe the change in level of an attribute (Yoo and Doiron, 2013). Lancsar et al. (2007) presented five methods for determining the relative impact of attributes. Four of the five models are possible with the traditional DCEs, and the last model is the attribute-level best-worst scaling experiment. The methods for traditional DCEs are: partial log-likelihood, marginal rates substitution, Hicksan welfare measure (appropriate in willingness to pay type problems), and probability analysis. Lancsar et al. (2007) examined the all five models in a study in a health systems research study dealing with cardiac patients and found the attribute impacts between the different methods comparable. However, it was determined that the attribute-level best-worst scaling provided greater precision in determining the impacts.

Other comparative studies have been done to examine the differences between best-worst scaling experiments and traditional DCEs. Both Potoglou et al. (2011) and Whitty et al. (2014) did empirical comparisons of traditional DCEs to attribute-level best-worst DCEs. Potoglou et al. (2011) noted that direct comparisons of attribute estimates are not possible due to the scaling of the attributes but found no indication that traditional DCEs outperformed attribute-level best-worst DCEs in the quality of life study. Whitty et al. (2014) study on preference in health care technology found differences in preference weights between the best-worst scaling experiments and traditional DCEs and questioned the use of one or both methods in priority settings.

Understanding the impact attribute and attribute-levels have on utility is desirable. The guiding ideology in DCEs is that consumers behave in a way to maximize utility. Understanding the impacts attributes and attribute-levels have on consumer behavior provides information with regards to developing and advertising a product, service, or policy to consumers. A preponderance of the literature on attribute-level best-worst DCEs are empirical studies often in the area of health systems research and marketing. Examples include Flynn et al. (2007) on seniors' quality of life, Coast et al. (2006) and Flynn et al. (2008) on dermatologist consultations, Marley and Pihlens (2012) on cellphones, Knox et al. (2012) and (2013) on choices in contraceptives for women.

While there exist literature on attribute-level best-worst DCEs, it is rather scarce compared to the work done on traditional DCEs. In this chapter, we provide utility definition and the resulting choice probabilities and properties. In Chapter 3 of this dissertation, we use the utility definition and choice probabilities to extend the work done by Großman et al. (2009) to fit models on a function of the attributes and attribute-levels. In Chapter 4, we extend Markov Decision Processes (MDPs) to these experiments. We utilize the model developed in Chapter 3 to apply in Chapter 4.

2.2 METHODOLOGY OF ATTRIBUTE-LEVEL BEST-WORST

Traditional DCEs and their models are built around the work done by Thurstone (1927) that a set of features or attributes describe a product, service, policy, or scenario by which respondents use to judge alternatives. Consumers choose the alternative that provides them with the greatest utility compared to the other alternatives available. Attribute-level best-worst scaling DCEs are modified DCEs designed to elicit more information about the impact the attributes and attribute-levels have on the utility of a product. As mentioned by Louviere and Timmermans (1990), an experiment must be designed in a way to evaluate combinations of attribute-levels to obtain information about attribute impacts on utility. Best-worst attribute-level DCEs provide such an experimental design to attain these impacts.

Following the setup as described by Street and Knox (2012), there are K attributes, or characteristics, that describe the products of interest, where the attributes are denoted by A_k for $k = 1, 2, \ldots, K$. Each attribute consists of l_k levels for $k = 1, 2, \ldots, K$. A design is said to be a balanced design if $l_k = l_{k'}, \forall k \neq k'$. In the study on women's contraceptives published by Knox et al. (2012) and (2013) an unbalanced design was used. In their contraceptive data, there were $K = 7$ attributes, with attribute levels $l_1 = 8, l_2 = 3, l_3 = 4, l_5 = 4, l_6 = 8, l_7 = 9$, and $l_8 = 6$. The 2^{nd} attribute is the contraceptive's effect on acne, and the levels associated with this attribute are no effect, improves, or worsens acne symptoms. Flynn et al. (2007) used a balanced design for their quality of life experiment with five attributes (attachment, security, role, enjoyment, and control) each with four attribute levels (none, little, lot, and all) for attachment, security, and enjoyment and (none, few, many, all) for role and control.

In the attribute-level best-worst DCEs, each product is represented by a profile $\mathbf{x} = (x_1, x_2, \ldots, x_K)$, where x_k is the attribute-level for the k^{th} attribute, A_k, that makes up the product for $k = 1, 2, \ldots, K$. The attribute-levels take values from 1 to l_k for $k = 1, 2, \ldots, K$. The number of possible profiles is given by $\Pi_{k=l}^{K} l_i$. Referring back to the contraceptive data, their experiment would produce a total of $165,888$ profiles, or products. In Knox et al. (2012), they specify there is a total of $27,648$ profiles, but that is per the 6 attribute-levels of the attribute contraceptive effectiveness. In Flynn et al. (2007), the total possible profiles is 1024. Full factorial designs are generally not used due to the number of profiles. Often, OMEP designs are efficient and optimal designs promoted in the literature that provide sufficient information to estimate parameters (Louviere and Woodworth,1983; Street and Burgess, 2007; Street and Knox, 2012). Using OMEP design, Knox et al. (2012) was able to reduce the number of profiles to 32 profiles per each of the 8 product types, producing a total 256 profiles considered. Similarly, Flynn et al. (2007) considered 32 profiles using an OMEP design.

In these experiments, respondents are tasked with choosing a pair of attribute-levels that contains that best and the worst attribute-level for a given profile. For every profile the choice set is then,

$$\mathcal{C}_x = \{(x_1, x_2), \ldots, (x_1, x_K), (x_2, x_3), \ldots, (x_{K-1}, x_K), (x_2, x_1), ..., (x_K, x_{K-1})\},$$

where the first attribute-level is considered to be the best and the second is the worst. From the profile $\mathcal{C}_x$, the respondent determines from the $\tau = K(K-1)$ choices given which is the best-worst pair.

In our setup, we extend the state of choices as follows. Let there be G profiles and the associated profiles are given as,

$$\begin{aligned}
x_1 &= (x_{11}, x_{12}, ..., x_{1K}) \\
x_2 &= (x_{21}, x_{22}, ..., x_{2K}) \\
&\vdots \\
x_G &= (x_{G1}, x_{G2}, ..., x_{GK}).
\end{aligned}$$

The corresponding choice sets for the G profiles are given in Figure 1. To simplify the notation, let

$$\mathcal{C}_1 = \mathcal{C}_{x_1}, \mathcal{C}_2 = \mathcal{C}_{x_2} \ldots, \mathcal{C}_G = \mathcal{C}_{x_G}$$

.

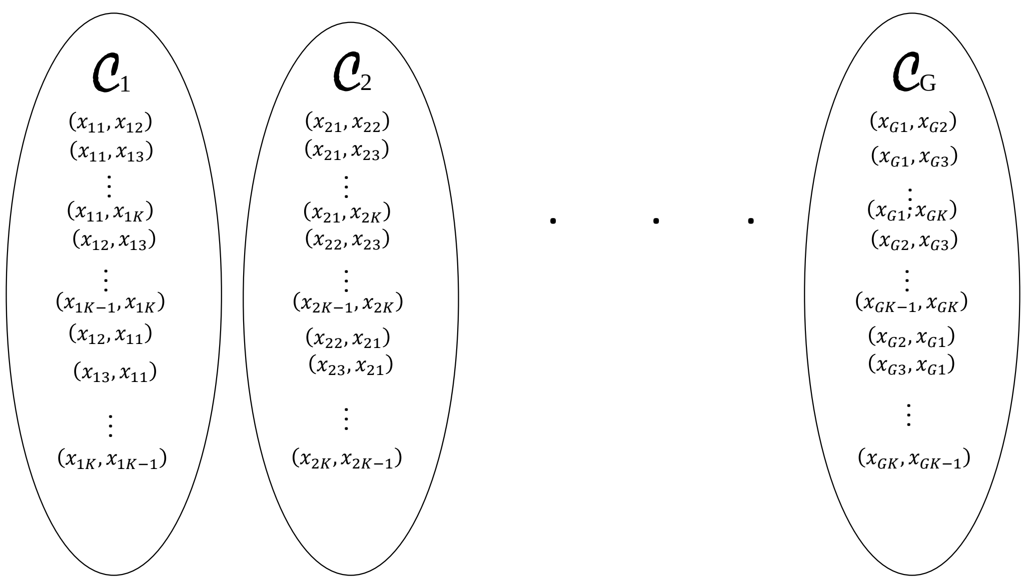


Figure 1: The G choice sets in an experiment with corresponding choice pairs

The total number of attribute-levels is $L = \sum_{i=1}^{k} l_i$, and $J = \sum_{k=1}^{K} l_k(L - l_k)$ is the total number of unique attribute-level pairs in the experiment (Street and Knox, 2012). Within each of the G choice sets there are $\tau = K(K-1)$ choice pairs.

2.2.1 DATA

In the experiment, there is a total of $J = \sum_{k=1}^{K} l_k(L - l_k)$ alternatives. However, within a choice set there is a total of $\tau = K(K-1)$ choices in each of the G choice sets evaluated. Each respondent will have made G choices within the experiment. The response variable representing the choices within each of the choice sets for the experiment are binary data and denoted as:

$$Y_{isj} = \begin{cases} 1, & \text{if } s^{th} \text{ respondent chooses } j^{th} \text{ alternative in the } i^{th} \text{ choice set,} \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

for $i = 1, 2, \ldots, G$, $s = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, \tau$.

For the attribute-level best-worst DCEs, the data, $\mathbf{X}$ is composed of indicators for the best and worst attributes and attribute-levels. Consider the choice pair $(x_{ij}, x_{ij'})$ from the choice set $\mathcal{C}_i$, for $i = 1, 2, \ldots, G$, $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $1 \leq x_{ij} \leq l_j$. Let X be the $J \times p$ design matrix, where $p = K + \sum_{k=1}^{K}$. The rows of X correspond to the possible choice pairs. Let $X_{A_1}, X_{A_2}, \ldots, X_{A_K}$ be the data corresponding to the attributes A_k, $k = 1, 2, \ldots, K$. Then,

$$X_{A_k} = \begin{cases} 1, \text{ if } x_{ij} \in A_k \text{ for } k = 1, 2, \ldots, K, \\ -1, \text{ if } x_{ij'} \in A_k \text{ for } k = 1, 2, \ldots, K, \\ 0, \text{ otherwise.} \end{cases} \tag{12}$$

Let $X_{A_k x_{ik}}$ be the data for the attribute-level $1 \leq x_{ik} \leq l_k$ within attribute A_k, $\forall k = 1, 2, \ldots, K$. Referring to the choice pair $(x_{ij}, x_{ij'})$, the corresponding data for the attribute-levels are given by,

$$X_{A_k x_{ik}} = \begin{cases} 1, \text{if } x_{ij} = x_{ik} \in A_k \text{ is the best attribute-level,} \\ -1, \text{if } x_{ij'} = x_{ik} \in A_k \text{ is the worst attribute-level,} \\ 0, \text{otherwise.} \end{cases} \tag{13}$$

2.2.2 EXAMPLE

Let us consider an example with an unbalanced design, as was done in Street and Knox (2012), where there are $K = 3$ attributes with attribute-levels,

$l_1 = 2, l_2 = 3$, and $l_3 = 4$. There are $2 \times 3 \times 4 = 24$ possible profiles, or products, in this experiment. The total number of attribute-levels is $L = \sum_{i=1}^{k} l_i = 9$, and the total number of choice pairs is $J = \sum_{k=1}^{K} l_k(L - l_k) = 52$.

The design matrix X is a $J \times p$ matrix. In this experiment, the number of columns in X, or the number of explanatory variables in the experiment, is $p = K + \sum_{k=1}^{K} l_k = 12$. So, X would be a 52×12 design matrix in this example.

Let us consider the profile $\mathbf{x}_i = \{x_{i1}, x_{i2}, x_{i3}\} = \{1, 1, 1\}$. The first level for each of the $K = 3$ attributes define the profile. The corresponding choice set to profile $\mathbf{x}_i$ would then be,

$$\mathcal{C}_i = \{(x_{i1}, x_{i2}), (x_{i1}, x_{i3}), (x_{i2}, x_{i3}), (x_{i2}, x_{i1}), (x_{i3}, x_{i1}), (x_{i3}, x_{i2})\}.$$

Suppose an individual chooses the pair (x_{i1}, x_{i2}). The data for this choice pair would consist of: $X_{A_1} = 1$, $X_{A_2} = -1$, $X_{A_3} = 0$, $X_{A_1 1} = 1$, $X_{A_1 2} = 0$, $X_{A_2 1} = -1$, $X_{A_2 2} = X_{A_2 3} = 0$, $X_{A_3 1} = X_{A_3 2} = X_{A_3 3} = X_{A_3 4} = 0$.

The data corresponding to the (x_{i1}, x_{i2}) choice pair in the design matrix X would appear as $(1, -1, 0, 1, 0, -1, 0, 0, 0, 0, 0, 0)$.

2.2.3 PROPERTIES AND PROBABILITY

For the purpose of this dissertation, we consider the conditional logit model for the data. The details regarding the use of this model in the traditional DCEs were provided in the Introduction. Here, we look at the probability theory and necessary properties for the application of this model to attribute-level best-worst DCEs.

Marley and Louviere (2005) developed the probability theory for best-worst scaling experiments including attribute-level best-worst DCEs. In attribute-level best-worst DCEs, there are two components being modeled, the best choice and the worst choice of attribute levels from a profile $\mathbf{x}_i$, where $i = 1, 2, \ldots, G$. Under random utility theory (Marschak, 1960), there are random utilities U_{ij} corresponding to the K attribute-levels in the profile. Consider the choice pair $(x_{ij}, x_{ij'})$, for $i = 1, 2, \ldots, G$, $j, j' = 1, 2, \ldots, K$, and $j \neq j'$. According to Marley and Louviere (2005), the definition of utility consistent with random utility theory satisfies, $U_{ij} = -U_{ij'}$ and

$U_{ijj'} = U_{ij} - U_{ij'}$ for $i = 1, 2, \ldots, G$, $j, j' = 1, 2, \ldots, K$, and $j \neq j'$. The associated probabilities satisfy:

$$B_{\mathbf{x}_i}(x_{ij}) = P(U_{ij} > U_{iq}, \forall q \in \mathbf{x}_i), \tag{14}$$

$$W_{\mathbf{x}_i}(x_{ij'}) = P(U_{ij'} < U_{iq}, \forall q' \in \mathbf{x}_i), \tag{15}$$

and

$$BW_{x_i}(x_{ij}, x_{ij'}) = P(U_{ij} > U_{iq} > U_{ij'}, \forall q, q' \in \mathbf{x}_i). \tag{16}$$

Referring to Equation 2, utility consists of a systematic component, V_{ij}, and an error term, ϵ_{ij} producing $U_{ij} = V_{ij} + \epsilon_{ij}$. The definition of utility associated with the best-worst choice pair under random utility theory is given by:

$$U_{ijj'} = U_{ij} - U_{ij'} = V_{ij} - V_{ij'} + \epsilon_{ij} - \epsilon_{ij'}$$

for $i = 1, 2, \ldots, G$, $j, j' = 1, 2, \ldots, K$, and $j \neq j'$. The definition of the utilities under the random utility model are unable to be modeled under the conditional logit model due to the definition of the error components (Marley and Louviere, 2005).

To use the conditional logit model, the choice probabilities have to satisfy the inverse random utility theory (Marley and Louviere, 2005), where the utilities are for the best and worst attribute levels are defined as:

$$U_{ij} = V_{ij} + \epsilon_{ij}, \tag{17}$$

$$U_{ij'} = -V_{ij} + \epsilon_{ij}, \tag{18}$$

and

$$U_{ijj'} = V_{ij} - V_{ij'} + \epsilon_{ijj'} = V_{ijj'} + \epsilon_{ijj'}, \tag{19}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$.

If we assume that the random error terms are independently and identically distributed type I extreme value distribution, or the Gumbel distribution, then the choice probability comes directly from the conditional logit. The choice probability is then,

$$\begin{aligned} BW_{x_i}(x_{ij}, x_{ij'}) &= P(U_{ijj'} > U_{iqq'}, \forall q, q' \in \mathcal{C}_i) \\ &= P(V_{ijj'} + \epsilon_{ijj'} > V_{iqq'} + \epsilon_{iqq'}, \forall q, q' \in \mathcal{C}_i) \\ &= P(\epsilon_{iqq'} - \epsilon_{ijj'} < V_{ijj'} + V_{iqq'}, \forall q, q' \in \mathcal{C}_i), \end{aligned} \tag{20}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$.

Since the error terms come from the type I extreme value distribution, their difference is a logistic distribution. It follows from McFadden (1974) that the best, worst, and best-worst attribute-level choice probabilities are defined by the conditional logit as:

$$B_{x_i}(x_{ij}) = \frac{exp(V_{ij})}{\sum\limits_{x_{iq} \in \mathbf{x}_i} exp(V_{iq})}, \tag{21}$$

$$W_{x_i}(x_{ij'}) = \frac{exp(-V_{ij'})}{\sum\limits_{x_{iq'} \in \mathbf{x}_i} exp(-V_{iq'})}, \tag{22}$$

and

$$BW_{x_i}(x_{ij}, x_{ij'}) = \frac{exp(V_{ijj'})}{\sum\limits_{(x_{iq}, x_{iq'}) \in \mathcal{C}_i} exp(V_{iqq'})}, \tag{23}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$.

Marley et al. (2008) provide essential properties to the above probabilities. They define the choice probability as:

$$BW_{\mathbf{x}_i}(x_{ij}, x_{ij'}) = \frac{\frac{b(x_{ij})}{b(x_{ij'})}}{\sum\limits_{\forall \left(x_{ij}, x_{ij'}\right) \in \mathcal{C}_{x_i}, j \neq j'} \frac{b(x_{ij})}{b(x_{ij'})}}, \tag{24}$$

where x_{ij} is chosen as the best attribute-level, and $x_{ij'}$ is the worst attribute-level, and b is some positive scale function or impact of attribute for $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$. Under the conditional logit, the scale function is defined as $b(x_{ij}) = exp(V_{ij})$, and the probability is as given in Equation (23).

Essential properties of probability hold for Equation (24),

$$BW_{x_i}(x_{ij}, x_{ij'}) \geq 0, \quad \forall i, j \tag{25}$$

and

$$\sum_{\forall (x_{ij}, x_{ij'}) \in \mathcal{C}_{x_i}, j \neq j'} BW_{x_i}(x_i, x_j) = 1, \tag{26}$$

where $j, j' = 1, 2, \ldots, K$, $j \neq j'$, and $\forall i = 1, 2, \ldots, G$.

With such assumptions, the consumer is expected to select choices with higher BW_{x_i} values. We denote $BW_{x_i}(x_{ij}, x_{ij'})$ as $P^i_{jj'}$. Attribute-level best-worst models are called maxdiff models because they maximize the difference in utility.

Associated properties of the maxdiff model mentioned in Marley et al. (2008) are:
2-invertibility For profile **i**,

$$P^i_{jj'} P^i_{j'j} = P^i_{qq'} P^i_{q'q},$$

where $1 \leq j, j', q, q' \leq k$ and $j \neq j'$ and $q \neq q'$.

We show that this property holds for the conditional logit and the associated choice probability in Equation (23). For profile **i**,

$$\begin{aligned} P^i_{jj'} P^i_{j'j} &= \frac{\exp(V_{ij} - V_{ij'})}{\sum_{(x_{ik}, x_{ik'}) \in \mathcal{C}_i} \exp(V_{ik} - V_{ik'})} \frac{\exp(V_{ij'} - V_{ij})}{\sum_{(x_{ik}, x_{ik'}) \in \mathcal{C}_i} \exp(V_{ik} - V_{ik'})} \\ &= \frac{\exp(V_{ij} - V_{ij'} + V_{ij'} - V_{ij})}{\left(\sum_{(x_{ik}, x_{ik'}) \in \mathcal{C}_i} \exp(V_{ik} - V_{ik'}) \right)^2} \\ &= \frac{1}{\left(\sum_{(x_{ik}, x_{ik'} \in \mathcal{C}_i} \exp(V_{ik} - V_{ik'}) \right)^2} \\ &= P^i_{qq'} P^i_{q'q}, \end{aligned}$$

where $j \neq j'$ and $j, j' = 1, 2, \ldots, K$. We see that $P^i_{jj'} P^i_{j'j}$ does not depend on the choice pair $(x_{ij}, x_{ij'})$ as this information cancels out.

3-reversibility: For profiles $\mathbf{i}, \mathbf{i}'$, and $\mathbf{i}''$,

$$P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'} = P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j}$$

where $x_{ij'} = x_{i'q}$, $x_{i'q'} = x_{i''r}$, and $x_{ij} = x_{i''r'}$, and $j \neq j', q \neq q'$, and $r \neq r'$

We show that this property holds for the conditional logit and the associated choice probability in Equation (23). For profiles **i, i',i"**, the conditions $x_{ij'} = x_{i'q}$, $x_{i'q'} = x_{i''r}$, and $x_{ij} = x_{i''r'}$ place similar conditions on the systematic components $V_{ij'} = V_{i'q}$, $V_{i'q'} = V_{i''r}$ and $V_{ij} = V_{i''r'}$ and $V_{ij} = V_{i''r'}$, where $j \neq j', q \neq q'$, and $r \neq r'$. Then,

$$\begin{aligned} P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'} &= \frac{\exp(V_{ij} - V_{ij'})}{\sum\limits_{(x_{ia}, x_{ia'}) \in \mathcal{C}_i} \exp(V_{ia} - V_{ia'})} \times \frac{\exp(V_{i'q} - V_{i'q'})}{\sum\limits_{(x_{i'b}, x_{i'b'}) \in \mathcal{C}_{i'}} \exp(V_{i'b} - V_{i'b'})} \times \\ &\quad \frac{\exp(V_{i''r} - V_{i''r'})}{\sum\limits_{(x_{i''c}, x_{i''c'}) \in \mathcal{C}_{i''}} \exp(V_{i''c} - V_{i''c'})}, \end{aligned}$$

and

$$\begin{aligned} P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j} &= \frac{\exp(V_{i''r'} - V_{i''r})}{\sum\limits_{(x_{i''c}, x_{i''c'}) \in \mathcal{C}_{i''}} \exp(V_{i''c} - V_{i''c'})} \times \frac{\exp(V_{i'q'} - V_{i'q})}{\sum\limits_{(x_{i'b}, x_{i'b'}) \in \mathcal{C}_{i'}} \exp(V_{i'b} - V_{i'b'})} \times \\ &\quad \frac{\exp(V_{ij'} - V_{ij})}{\sum\limits_{(x_{ia}, x_{ia'}) \in \mathcal{C}_i} \exp(V_{ia} - V_{ia'})}, \end{aligned}$$

where $j \neq j', \quad q \neq q'$, and $r \neq r'$.

Then applying the condition that $V_{ij'} = V_{i'q}$, $V_{i'q'} = V_{i''r}$ and $V_{ij} = V_{i''r'}$:

$$\begin{aligned} P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'} &= \frac{\exp(V_{ij} - V_{ij'})}{\sum\limits_{(x_{ia}, x_{ia'}) \in \mathcal{C}_i} \exp(V_{ia} - V_{ia'})} \times \frac{\exp(V_{ij'} - V_{i'q'})}{\sum\limits_{(x_{i'b}, x_{i'b'}) \in \mathcal{C}_{i'}} \exp(V_{i'b} - V_{i'b'})} \times \\ &\quad \frac{\exp(V_{i'q'} - V_{ij})}{\sum\limits_{(x_{i''c}, x_{i''c'}) \in \mathcal{C}_{i''}} \exp(V_{i''c} - V_{i''c'})} \\ &= \frac{1}{\sum\limits_{(x_{ia}, x_{ia'}) \in \mathcal{C}_i} \exp(V_{ia} - V_{ia'})} \times \frac{1}{\sum\limits_{(x_{i'b}, x_{i'b'}) \in \mathcal{C}_{i'}} \exp(V_{i'b} - V_{i'b'})} \times \end{aligned}$$

$$\frac{1}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})}$$

and

$$\begin{aligned}
P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j} &= \frac{\exp(V_{i''r'}-V_{i''r})}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})}\times\frac{\exp(V_{i''r}-V_{i'q})}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})}\times \\
&\quad \frac{\exp(V_{i'q}-V_{i''r'})}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_{i}}\exp(V_{ia}-V_{ia'})} \\
&= \frac{1}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})}\times\frac{1}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})}\times \\
&\quad \frac{1}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_{i}}\exp(V_{ia}-V_{ia'})}
\end{aligned}$$

where $j \neq j', q \neq q'$, and $r \neq r'$. Terms cancel in the numerator resulting in:

$$\begin{aligned}
P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'} &= \frac{1}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_{i}}\exp(V_{ia}-V_{ia'})}\times\frac{1}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})}\times \\
&\quad \frac{1}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})} \\
&= P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j},
\end{aligned}$$

where $j \neq j', q \neq q'$, and $r \neq r'$.

4-reversibility: For profiles $\mathbf{i}, \mathbf{i}', \mathbf{i}''$, and $\mathbf{i}'''$,,

$$P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'}P^{i'''}_{ww'} = P^{i'''}_{w'w}P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j}$$

where $x_{ij'} = x_{i'q}$, $x_{i'q'} = x_{i''r}$, $x_{i''r} = x_{i'''w}$, and $x_{ij} = x_{i'''w'}$, and $j \neq j', q \neq q', r \neq r'$ and $w \neq w'$.

We show that this property holds for the conditional logit and the associated choice probability in Equation (23). For profiles $\mathbf{i}, \mathbf{i}', \mathbf{i}''$, and $\mathbf{i}'''$, the conditions $x_{ij'} = x_{i'q}$, $x_{i'q'} = x_{i''r}$, $x_{i''r} = x_{i'''w}$, and $x_{ij} = x_{i'''w'}$ place similar conditions on the systematic components $V_{ij'} = V_{i'q}$, $V_{i'q'} = V_{i''r}$, $V_{i''r} = V_{i'''w}$, and $V_{ij} = V_{i'''w'}$, where $j \neq j', q \neq q', r \neq r'$ and $w \neq w'$. Then,

$$
\begin{aligned}
P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'}P^{i'''}_{ww'} = & \frac{\exp(V_{ij}-V_{ij'})}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_i}\exp(V_{ia}-V_{ia'})} \times \frac{\exp(V_{i'q}-V_{i'q'})}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})} \times \\
& \frac{\exp(V_{i''r}-V_{i''r'})}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})} \times \frac{\exp(V_{i'''w}-V_{i'''w'})}{\sum\limits_{(x_{i'''d},x_{i'''d'})\in\mathcal{C}_{i'''}}\exp(V_{i'''d}-V_{i'''d'})},
\end{aligned}
$$

and

$$
\begin{aligned}
P^{i'''}_{w'w}P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j} = & \frac{\exp(V_{i'''w'}-V_{i'''w})}{\sum\limits_{(x_{i'''d},x_{i'''d'})\in\mathcal{C}_{i'''}}\exp(V_{i'''d}-V_{i'''d'})} \times \frac{\exp(V_{i''r'}-V_{i''r})}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})} \times \\
& \frac{\exp(V_{i'q'}-V_{i'q})}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})} \times \frac{\exp(V_{ij}-V_{ij'})}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_i}\exp(V_{ia}-V_{ia'})},
\end{aligned}
$$

where $j \neq j', q \neq q'$, and $r \neq r'$.

Then applying the condition that $V_{ij'} = V_{i'q}$, $V_{i'q'} = V_{i''r}$, $V_{i''r} = V_{i'''w}$, and $V_{ij} = V_{i'''w'}$:

$$
\begin{aligned}
P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'}P^{i'''}_{ww'} = & \frac{\exp(V_{ij}-V_{ij'})}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_i}\exp(V_{ia}-V_{ia'})} \times \frac{\exp(V_{ij'}-V_{i'q'})}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})} \times \\
& \frac{\exp(V_{i'q'}-V_{i''r'})}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})} \times \frac{\exp(V_{i''r'}-V_{ij})}{\sum\limits_{(x_{i'''d},x_{i'''d'})\in\mathcal{C}_{i'''}}\exp(V_{i'''d}-V_{i'''d'})},
\end{aligned}
$$

and

$$
\begin{aligned}
P^{i'''}_{w'w}P^{i''}_{r'r}P^{i'}_{q'q}P^{i}_{j'j} = & \frac{\exp(V_{i'''w'}-V_{i'''w})}{\sum\limits_{(x_{i'''d},x_{i'''d'})\in\mathcal{C}_{i'''}}\exp(V_{i'''d}-V_{i'''d'})} \times \frac{\exp(V_{i'''w'}-V_{i''r})}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}}\exp(V_{i''c}-V_{i''c'})} \times \\
& \frac{\exp(V_{i''r'}-V_{i'q})}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})} \times \frac{\exp(V_{i'q}-V_{i'''w'})}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_i}\exp(V_{ia}-V_{ia'})},
\end{aligned}
$$

where $j \neq j', q \neq q'$, and $r \neq r'$. We cancel terms in the numerator resulting in:

$$
P^{i}_{jj'}P^{i'}_{qq'}P^{i''}_{rr'}P^{i'''}_{ww'} = \frac{1}{\sum\limits_{(x_{ia},x_{ia'})\in\mathcal{C}_i}\exp(V_{ia}-V_{ia'})} \times \frac{1}{\sum\limits_{(x_{i'b},x_{i'b'})\in\mathcal{C}_{i'}}\exp(V_{i'b}-V_{i'b'})} \times
$$

$$\frac{1}{\sum\limits_{(x_{i''c},x_{i''c'})\in\mathcal{C}_{i''}} \exp(V_{i''c}-V_{i''c'})} \frac{1}{\sum\limits_{(x_{i'''d},x_{i'''d'})\in\mathcal{C}_{i'''}} \exp(V_{i'''d}-V_{i'''d'})}$$
$$= P^{i'''}_{w'w} P^{i''}_{r'r} P^{i'}_{q'q} P^{i}_{j'j},$$

where $j \neq j', q \neq q'$, and $r \neq r'$.

We defined the utility for the choice pair $(x_{ij}, x_{ij'})$ under inverse utility theory in Equation (19) as

$$U_{ijj'} = V_{ijj'} + \epsilon_{ijj'},$$

where $V_{ijj'}$ is the systematic component and $\epsilon_{ijj'}$ is the error term for $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$. The systematic component can be expressed as,

$$V_{ijj'} = V_{ij} - V_{ij'} = (\mathbf{x}_{ij} - \mathbf{x}_{ij'})'\beta, \tag{27}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$. The data x_{ij} are indicators of the attribute $x_{ij} \in A_j$ and the attribute-level x_{ij}. The systematic component V_{ij} can be written as,

$$V_{ij} = \mathbf{x}_{ij}\beta = \beta_{A_j} + \beta_{x_{ij}A_j}, \tag{28}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$.

Under the conditional logit, the probability that $(x_{sj}, x_{sj'})$ is chosen is given as,

$$P^{i}_{jj'} = \frac{\exp(V_{ijj'})}{\sum\limits_{x_j, x_{j'} \in \mathcal{C}_s, x_j \neq x_{j'}} \exp(V_{ijj'})}, \tag{29}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$.

Then Equation (24) with the choice of the scale function $b(x_{ij}) = \exp(\beta_{A_j} + \beta_{x_{ij}A_j}) = \exp(V_{ij})$ becomes Equation (29). This is seen by,

$$\frac{b(x_{ij})}{b(x_{ij'})} = \exp(V_{ij} - V_{ij'}) = \exp(V_{ijj'}), \tag{30}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$.

We assume the error terms come from a type I extreme value distribution and use the conditional logit model to estimate the $p \times 1$ parameter vector,

$$\beta' = (\beta_{A_1}, \beta_{A_2}, \ldots, \beta_{A_k}, \beta_{A_1 1}, \beta_{A_1 2}, \ldots, \beta_{A_1 l_1}, \ldots, \beta_{A_k 1}, \ldots, \beta_{A_k l_k}). \tag{31}$$

The likelihood for estimating the model parameters based on a random sample n individuals as in Equation (11) is given as:

$$L(\boldsymbol{\beta}, \mathbf{Y}) = \prod_{s=1}^{n} \prod_{i=1}^{G} \prod_{j \neq j'} P_{ijj'}^{Y_{isj}}. \tag{32}$$

Estimation of the parameters is done in SAS maximizing the likelihood given in Equation 2.2.3.

Attribute and attribute-level data in the experiments are a series of $1's$ and $0's$, indicating the attributes and attribute-levels in the choice pair. When fitting a conditional logit model to the data, parameter estimates for the last attribute and last attribute-level for each attribute are not retrievable due to singularity issues. According to Flynn et al. (2007), these parameter estimates are needed to determine the impact of attribute, which is the essential purpose for experiments of this design. To estimate these parameters, the following identifiability condition defined on the parameters of the attribute-levels must be met,

$$\Sigma_{i=1}^{l_k} \beta_i = 0 \tag{33}$$

or

$$\beta_{l_k} = -\Sigma_{j=1}^{l_i - 1} \beta_j \tag{34}$$

for all $k = 1, 2, \ldots, K$ (Street and Burgess, 2007; Flynn et al., 2007; Graßhoff et al., 2003).

Suppose the last attribute-level for attribute k is chosen as the best, the the other levels of the attribute would be $X_{A_k 1} = X_{A_k 2} = \ldots = X_{A_k l_k - 1} = -1$ and the parameter would be estimated as in Equation (34). Similarly, if the last attribute-level for attribute k is chosen as the worst, the the other levels of the attribute would be $X_{A_k 1} = X_{A_k 2} = \ldots = X_{A_k l_k - 1} = 1$. Similar coding is unable to be applied to the attributes, so the identifiability condition is only applied to the attribute-levels.

CHAPTER 3

FUNCTIONAL FORM OF ATTRIBUTE-LEVEL BEST-WORST DCEs

3.1 INTRODUCTION

A common goal in the analysis of DCEs is to build a functional form of the attributes and estimate the associated parameters that reflect their utility. Generating the design, after combining attribute-levels of the best and worst, requires simulation of choice experiments (sometimes with large number of choice sets) and a model expression that accounts for their weights. We will describe the designs for paired comparisons and its adaptations to DCEs and a simulated example.

3.2 PAIRED COMPARISONS

In recent years, more of the literature on DCEs has been on methods for designing efficient experiments of such types. One major issue is related to the fact that effects could not be independently identified. By identifying the DCE, data become unbiased as described in Street and Burgess (2007). Such conclusion can be found in Fedorov (1972). Louviere and Woodworth (1983) discussed efficient designs of these experiments included orthogonal arrays, main effects plans, and fractional factorial designs. These experiments are the common choices for DCEs. Orthogonal arrays, as presented by Hinkelmann (2011) are fractional factorial designs that are defined by (R, k, l, t). The array is $R \times k$ array with elements of the l symbols, such that any t columns of the array has $t - tuplet$ as a row R/l^t times. For example, if $t = 2$, then any pair of columns would have each pair of symbols occur R/l^2 times. R/l^t is referred to as the index of the array, t is the strengths, k is the number of constraints, and l is the number of levels. Street and Knox (2012) examined the use of OMEPs, which is an orthogonal array of strength $t = 2$, for attribute-level best-worst DCEs to maximize the information while decreasing the choice task for respondents. The work done by Graßhoff et al. (2003), (2004), and (2013), and Großman et al. (2009) look at the conditions for optimal designs when a functional form of the attributes is used and paired comparisons of alternatives are studied. As mentioned in Graßhoff et al. (2013), the quality of the outcome depends heavily on the design.

Our interest lies in the area of paired comparison best-worst attribute-level type designs. Louviere et al. (2013) discuss the limitations inherent in paired comparison type of studies due to the number of situations evaluated by respondents. However, many researchers have developed methods for designing and models for paired comparison data. We consider the initial work done by Graßhoff et al. (2003) for paired comparison experimental designs. Their work presents a functional form of the data entered into the linear model of the form:

$$Y(\mathbf{s}) = \mathbf{f}(\mathbf{s})'\boldsymbol{\beta} + \epsilon(\mathbf{s}), \tag{35}$$

where Y is the observed response of dimension J, $\mathbf{s} = \{s_1, s_2, \ldots, s_K\}$ is a set of K factors or attributes, $\mathbf{f} = \{f_1, f_2, \ldots, f_K\}$ is a set of regression functions, and $\epsilon(\mathbf{s})$ are the error terms in the model.

We define $\mathbf{f}$ in the following way:

$$\mathbf{f}(\mathbf{s})'\boldsymbol{\beta} = \sum_{k=1}^{K} \beta_k f_k(s_k, a_k), \tag{36}$$

where

$a_k \in \mathbb{R}^K$ are impact weights,

$\beta_k \in \mathbb{R}^K$ are the parameters,

$f_k \in \mathbb{R}^K \to \mathbb{R}$ are the activation function of the k^{th} attribute.

Let there be n subjects. The model, $Y = X\beta + \epsilon$, can be written as,

$$X = \mathbf{f}(\mathbf{s}) = \begin{pmatrix} f_{11}(s_1, a_1) & . & . & . & . & f_{K1}(s_K, a_K) \\ f_{12}(s_1, a_1) & . & . & . & . & f_{K2}(s_K, a_K) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ f_{1n}(s_1, a_1) & . & . & . & . & f_{Kn}(s_K, a_K) \end{pmatrix}_{n \times K},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix},$$

and

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

The values of $\boldsymbol{\beta}$ are found by,

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'Y$$

using the X and Y as defined above.

By extending this work to paired comparisons of alternatives, we propose two sets of K factors/attributes, $\mathbf{s} = (s_1, s_2, \ldots, s_K)$ and $\mathbf{t} = (t_1, t_2, \ldots, t_K)$ corresponding to the two alternatives being compared in the model. The paired comparison model is then written as:

$$Y(\mathbf{s}, \mathbf{t}) = (\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{t}))'\boldsymbol{\beta} + \epsilon_{\mathbf{s},\mathbf{t}}, \tag{37}$$

where there are two sets of K factors/attributes, $\mathbf{s}$ and $\mathbf{t}$, corresponding to each alternative, $\mathbf{f}$ is the set of regression functions, and with error function $\epsilon(\mathbf{s}, \mathbf{t})$. It is important to note that the same identifiability conditions we presented in the previous chapter, Equation (34), are also imposed here in their design for the K

factors.

Graßhoff et al. (2003) provided the conditions under which models expressed as in Equation (37) are optimal. As mentioned in Ruseckaite et al. (2017), the main drawback for such design is the fact that although alternatives are not the choice, their utilities are the same; therefore, we bring in a distinction among the alternatives. Specifically, D-optimal designs that maximize the determinant of the information matrix are utilized. The information matrix associated with the paired comparison data is,

$$\mathbf{M}(\mathbf{s},\mathbf{t}) = (\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{t}))(\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{t}))', \tag{38}$$

where $M(t,t) = 0$, no information is gained in comparing an alternative to itself, and $M(s,t) = M(t,s)$ (Graßhoff et al., 2004).

Green (1974) mentions that of the fractional factorial designs, orthogonal arrays present the smallest number of profiles needed to estimate main effects from the experiments. This is one of the reasons OMEPs are often used in best-worst scaling experiments, especially the attribute-level best-worst DCEs as was done in Knox et al. (2012) and Flynn et al. (2007).

There are some disadvantages to paired comparison experiments. For one, Louviere et al. (2013) indicated that these experiments can be burdensome on respondents. The number of comparisons respondents are asked to evaluate can become overwhelming as the number of alternatives increase in the experiment. Louviere et al. (2013) reviewed issues in the designs of experiments in terms of designs becoming so complicated that respondents are burdened cognitively in the experiment. They reference a study done by Deshazo and Fermo (2002) that presented evidence of this burden on the respondents.

3.2.1 PARTIAL PROFILES

Often time, respondents lessen the burden of the choice tasks by considering a subset of attributes when making their decisions. In Louviere and Timmermans (1990), hierarchical information integration is proposed and is based on the belief that respondents assess choices, specifically their attributes, in a hierarchical fashion. Hess and Hensher (2010) state that decisions are sometimes made after ignoring some of the attributes or giving such attributes a lower level of importance. However, as they mentioned, more work is needed in the case where there are multiple attributes, and where respondents could ignore some of the attributes, and include the confounding component or variation in the modeling. In the work done by Hole (2011), a two stage process for completing choice tasks is provided. In the first stage of the process, respondents choose the subset of attributes they find meaningful in the evaluation of each alternative. The second step is to choose the alternative that provides them the highest utility.

There exists evidence that a subset of attributes may be used to evaluate alternatives. In the work done by Großman et al. (2009), these subsets are partial profiles. The work done here extends the generalized form of the model given in Equation (37) to work in the case of partial profiles. Here, there still exists two sets of factors, $\mathbf{s}$ and $\mathbf{t}$. However, the set of factors are not required to have all K factors represented, $\mathbf{s} = (s_1, s_2, \ldots, s_{K_1})$ and $\mathbf{t} = (t_1, t_2, \ldots, t_{K_2})$ where $K_1 \leq K$ and $K_2 \leq K$. The pairs of factors, $(\mathbf{s}, \mathbf{t})$ include only the factors represented in $\mathbf{s}$ and $\mathbf{t}$. Let there be K_3 factors in common to sets $\mathbf{s}$ and $\mathbf{t}$, where $K_3 \leq K$. Then, $(\mathbf{s}, \mathbf{t}) = \{(s_1, t_1), (s_2, t_2), \ldots, (s_{K_3}, t_{K_3})\}$. The model is then written as,

$$Y(\mathbf{s}, \mathbf{t}) = (\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{t}))'\boldsymbol{\beta} + \epsilon_{\mathbf{s},\mathbf{t}}, \tag{39}$$

where the known regression functions $\mathbf{f} = \{f_1, f_2, \ldots, f_{K_3}\}$, the parameter vector is $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_{K_3}\}$. The error terms, $\epsilon(\mathbf{s}, \mathbf{t})$ are independently and identically distributed from a distribution with a mean of zero. The same identifiability condition on the parameters still holds as in the previous model.

Others have considered methods for optimal designs with partial profiles. Kessels et al. (2011) presents a flexible Bayesian method to determine optimal designs where a couple of attributes are kept constant in each of the profiles and updates for the mixture case is proposed by Ruseckaite (2017). However, Cuervo et al. (2016) cites that the partial profile design by Großman et al. (2009) were of better quality. It is easy to see the application of Großman et al. (2009) in case 3 of the best-worst scaling experiments where best-worst pairs of alternatives are chosen based on their attributes.

3.3 GENERALIZATIONS TO DCEs

The models and associated designs discussed so far in this chapter have applications in DCEs. The work done by Graßhoff et al. (2003) and (2004) and Großman et al. (2009) have dealt with experiments where only two alternatives are present. In Graßhoff et al. (2013), these models were adapted to assess multiple paired comparisons in DCEs where there are more than two alternatives.

According to random utility theory, an alternative is chosen that provides the greatest utility. As mentioned in Train (2009), the differences in utility are what matter in these experiments. The work done by Graßhoff et al. (2013) can be put in terms of utility and the probabilities defined accordingly.

Let there be a set of J alternatives each described by a profile, $\mathbf{x}_j = \{x_{j1}, x_{j2}, \ldots, x_{jK}\}$, for $j = 1, 2, \ldots, J$. The response variable is binary, 1 if the alternative is chosen and 0 otherwise. Graßhoff et al. (2013) write their model in terms of the expected value of Y. With Y being binary, its expected value is the same as the probability of the j^{th} alternative is chosen. The utility corresponding to the j^{th} choice is given by $U_j = V_j + \epsilon_j$, where $V_j = x_j'\beta$ is the systematic component and ϵ_j is the error term for $j = 1, 2, \ldots, J$. Thus we can express the model as in Equation (3) as:

$$\begin{aligned} P(U_j - U_{j'} > 0, \forall j \neq j') &= P(U_j > U_{j'}, \forall j \neq j') \\ &= P(V_j + \epsilon_j > V_{j'} + \epsilon_{j'}, \forall j \neq j') \\ &= P(\epsilon_{j'} - \epsilon_j < V_j - V_{j'}, \forall j \neq j') \qquad (40) \end{aligned}$$

where $j, j' = 1, 2, \ldots, J$.

Referring to Equation (35), the systematic component may be written as

$$V_j = \mathbf{f}_j(\mathbf{x}_j)'\boldsymbol{\beta}.$$

Then,

$$V_j - V_{j'} = (\mathbf{f}_j(\mathbf{x}_j) - \mathbf{f}_{j'}(\mathbf{x}_{j'}))'\boldsymbol{\beta}$$

. If the conditional logit model is used, that is the error terms are assumed to be independent and identically distributed with a type I extreme value distribution, the probability that the j^{th} choice is chosen is then given by:

$$\begin{aligned} P(Y_j = 1) &= P(U_j > U_{j'}, \forall j \neq j') \\ &= P(\epsilon_{j'} - \epsilon_j < V_j - V_{j'}, \forall j \neq j') \\ &= \frac{\exp(\mathbf{f}_j(\mathbf{x}_j)'\boldsymbol{\beta})}{\sum_{j'=1}^{J} \exp(\mathbf{f}_j(\mathbf{x}_{j'})'\boldsymbol{\beta})}, \end{aligned} \tag{41}$$

where $j, j' = 1, 2, \ldots, J$.

As in Großman et al. (2009), $\mathbf{x}_j$ and $\mathbf{x}_{j'}$ may be partial profiles. The model will still be built as given in Equation (41), where the utilities U_j would be built on the partial set of attributes rather than all K attributes. Doing so provides a generalized regression model given in Graßhoff et al. (2013) in terms of utility. However, we can also extend the model to attribute-level best-worst scaling DCEs.

3.4 EXTENSION TO ATTRIBUTE-LEVEL BEST-WORST DCEs

The attribute-level best-worst DCEs are modified traditional DCEs. Models and theory done for traditional DCEs have not been completely evaluated in terms of best-worst scaling experiments. It is of interest to us to extend the model built on a function of the data as presented inGraßhoff et al. (2003), Graßhoff et al. (2004), and Großman et al. (2009) to the attribute-level best-worst DCEs. In extending this work to these experiments, we provide an additional way to define the systematic component that provides flexibility not seen in traditional methods.

Considering functions of the attributes as they enter into the utility function is not a new idea. Van der Pol et al. (2014) presents the systematic components of the utility defined as linear functions, quadratic functions, or as stepwise functions of the attributes. Graßhoff et al. (2013) define the functions as regression functions of the attributes and attribute-levels in the model.

In the attribute-level best-worst DCEs, a set of G profiles, or products, are examined. The profiles are given as $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iK})$, where x_{ij} is the attribute-level in profile $i = 1, 2, \ldots, G$ that corresponds to the j^{th} attribute, where $j = 1, 2, \ldots, K$. The choice task for respondents is to choose the best-worst pair of attribute-levels. In the experiment, respondents make paired comparisons within the profiles instead of between as in traditional DCEs.

In the attribute-level best-worst DCEs, the utility of the pairs is composed of the utility corresponding to the best attribute-level and the worst attribute-level. The regression functions presented in Graßhoff et al. (2003) are applied to the attributes and attribute-levels within the respective systematic components. Let $\mathbf{f}$ be the set of regression functions for the best attribute-levels in the pairs, and $\mathbf{g}$ the set of regression functions for the worst attribute-levels in the pairs. The $p \times 1$ parameter vector $\boldsymbol{\beta}$ still must satisfy the identifiability condition given in Equation (34).

Let us consider the j^{th} choice is given as $(x_{ij}, x_{ij'})$, where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$. The functional form of the data in the systematic component for the best attribute-level, x_{ij} can be found referring back to equations (18) and (35) to be,

$$V_{ij} = \mathbf{f}(\mathbf{x}_{ij})'\boldsymbol{\beta}, \tag{42}$$

where $i = 1, 2, \ldots, G$ and $j = 1, 2, \ldots, K$.

Similarly, the functional form of the data in the systematic component for the worst attribute-level, $x_{ij'}$ is defined as,

$$V_{ij'} = -\mathbf{g}(\mathbf{x}_{ij'})'\boldsymbol{\beta}, \tag{43}$$

where $i = 1, 2, \ldots, G$ and $j = 1, 2, \ldots, K$.

Having defined the systematic components for the best and worst attribute-levels, the systematic component for the pair may be defined. As noted in Marley and Louviere (2005), the inverse random utility model must be used so the properties are met for the conditional logit model for the data. Taking the systematic components defined in equations (42) and (43) together with Equation (19), systematic component for the pair $(x_{ij}, x_{ij'})$ is defined as:

$$V_{ijj'} = V_{ij} - V_{ij'} + \epsilon_{ijj'} = (\mathbf{f}(\mathbf{x}_{ij'}) - \mathbf{g}(\mathbf{x}_{ij'}))'\boldsymbol{\beta}, \tag{44}$$

where $j, j' = 1, 2, \ldots, K$, $j \neq j'$, and $i = 1, 2, \ldots, G$.

The probability an alternative is chosen depends on the definition of the utility and the distribution of the error terms. From equations (42), (43), and (44), we have the systematic components of the best, worst, and best-worst utilities defined. We assume the error terms ϵ_{ij}, $\epsilon_{ij'}$, and $\epsilon_{ijj'}$ are independently and identically distributed as type I extreme value distribution, where $j \neq j'$ and $j, j' = 1, 2, \ldots, K$. For the best attribute-level choice, the probability is given, in reference to Equation (45), under the conditional logit as:

$$\begin{aligned} B_{x_i}(x_{ij}) &= \frac{\exp(V_{ij})}{\sum\limits_{x_{iq} \in \mathbf{x}_i} \exp(V_{iq})} \\ &= \frac{\exp(\mathbf{f}(\mathbf{x}_{ij})'\boldsymbol{\beta})}{\sum\limits_{x_{iq} \in \mathbf{x}_i} \exp(\mathbf{f}(\mathbf{x}_{iq})'\boldsymbol{\beta})}, \end{aligned} \tag{45}$$

where $i = 1, 2, \ldots, G$ and $j = 1, 2, \ldots, K$.

Similarly, the probability associated with the choice in the worst attribute-level is given, in reference to Equation (46), under the conditional logit as:

$$\begin{aligned} W_{x_i}(x_{ij'}) &= \frac{\exp(-V_{ij'})}{\sum\limits_{x_{iq'} \in \mathbf{x}_i} \exp(-V_{iq'})} \\ &= \frac{\exp(-\mathbf{g}(\mathbf{x}_{ij'})'\boldsymbol{\beta})}{\sum\limits_{x_{iq'} \in \mathbf{x}_i} \exp(-\mathbf{g}(\mathbf{x}_{iq'})'\boldsymbol{\beta})}, \end{aligned} \tag{46}$$

where $i = 1, 2, \ldots, G$ and $j' = 1, 2, \ldots, K$.

Finally, the probability for the best and worst attribute-level pair is determined. Referring back to Equation (24) under the conditional logit, the probability is:

$$BW_{x_i}(x_{ij}, x_{ij'}) = \frac{\exp(V_{ijj'})}{\sum\limits_{(x_{iq}, x_{iq'}) \in \mathcal{C}_i} \exp(V_{iqq'})}$$

$$= \frac{\exp(V_{ij} - V_{ij'})}{\sum\limits_{(x_{iq}, x_{iq'}) \in \mathcal{C}_i} \exp(V_{iq} - V_{iq'})}$$

$$= \frac{\exp((\mathbf{f}(\mathbf{x}_{ij}) - \mathbf{g}(\mathbf{x}_{ij'}))'\boldsymbol{\beta})}{\sum\limits_{(x_{iq}, x_{iq'}) \in \mathcal{C}_i} \exp((\mathbf{f}(\mathbf{x}_{iq'}) - \mathbf{g}(\mathbf{x}_{iq'}))'\boldsymbol{\beta})}, \tag{47}$$

where $i = 1, 2, \ldots, G$, $j, j' = 1, 2, \ldots, K$, and $j \neq j'$.

The forms of the systematic components of the utilities as well as their associated probabilities depend on the definition of the regression functions **f** and **g**. We define the regression functions used in the traditional attribute-level best-worst model and extend the definition of the regression functions to a more general form that provides flexibility in the model that is not obtained in the traditional definition in the next section.

3.4.1 REGRESSION FUNCTIONS DEFINITIONS

In Chapter 2, we provided the design, probabilities, and properties associated with attribute-level best-worst DCEs. The data in these experiments are defined as a series of $1's$, $0's$, and $-1's$ corresponding to the best and worst attributes and attribute-levels in a given choice pair. There exist a set of functions **f** and **g** defined on the attribute-level pair that produces traditional methods.

In the attribute-level best-worst DCEs, a set of G profiles, or products, are examined. The profiles are given as $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iK})$, where x_{ij} is the attribute-level in profile $i = 1, 2, \ldots, G$ that corresponds to the j^{th} attribute for $j = 1, 2, \ldots, K$. Let us consider the choice is given as $(x_{ij}, x_{ij'})$, where $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $i = 1, 2, \ldots, G$. Let **f** be the set of regression functions defined on the best attribute-level in a pair and **g** be the set of regression functions defined on the worst attribute-level in a pair.

In the traditional attribute-level best-worst DCE, the regression functions **f** and **g** are defined as indicator functions. The indicator functions are $p \times 1$ vectors. For the attributes, they are defined as,

$$I_{A_k}(x_{ij}) = \begin{cases} 1, & \text{if } x_{ij} \in A_k, \\ 0, & \text{otherwise,} \end{cases} \tag{48}$$

and for the attribute-levels,

$$I_{A_k x_k}(x_{ij}) = \begin{cases} 1, & \text{if } x_{ij} = x_k \text{ for } x_k \in A_k, \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

where $j, k = 1, 2, \ldots, K$ and $i = 1, 2, \ldots, G$.

The systematic component using the regression functions $\mathbf{f}$ for the best attribute-level, x_{ij} is given as,

$$\begin{aligned} V_{ij} &= \mathbf{f}(\mathbf{x}_{ij})'\boldsymbol{\beta} \\ &= \sum_{k=1}^{K} \left[I_{A_k}(x_{ij}) + \sum_{j=1}^{l_k} I_{A_k x_k}(x_{ij}) \right] \\ &= I_{A_j} + I_{A_j x_{ij}}, \end{aligned} \tag{50}$$

where $i = 1, 2, \ldots, G$ and $j = 1, 2, \ldots, K$.

The systematic component using the regression functions $\mathbf{g}$ for the worst attribute-level, $x_{ij'}$ is given as,

$$\begin{aligned} V_{ij'} &= -\mathbf{g}(\mathbf{x}_{ij'})'\boldsymbol{\beta} \\ &= -\sum_{k=1}^{K} \left[I_{A_k}(x_{ij'}) + \sum_{j=1}^{l_k} I_{A_k x_k}(x_{ij'}) \right] \\ &= -I_{A_{j'}} - I_{A_{j'} x_{ij'}}, \end{aligned} \tag{51}$$

where $i = 1, 2, \ldots, G$ and $j' = 1, 2, \ldots, K$.

Using the systematic components for the best and the worst attribute-levels, the best-worst systematic component for the pair $(x_{ij}, x_{ij'})$ is given as,

$$\begin{aligned} V_{ijj'} &= V_{ij} - V_{ij'} \\ &= (\mathbf{f}(\mathbf{x}_{ij}) - \mathbf{g}(\mathbf{x}_{ij'}))'\boldsymbol{\beta} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{K}\left[I_{A_k}(x_{ij}) + \sum_{x_k}^{l_k} I_{A_k x_k}(x_{ij})\right] - \sum_{k=1}^{K}\left[I_{A_k}(x_{ij'}) + \sum_{x_k}^{l_k} I_{A_k x_k}(x_{ij'})\right] \\
&= \sum_{k=1}^{K}\left[I_{A_k}(x_{ij}) - I_{A_k}(x_{ij'}) + \sum_{x_k=1}^{l_k}(I_{A_k x_k}(x_{ij}) - I_{A_k x_k}(x_{ij'}))\right] \\
&= I_{A_j} + I_{A_j x_{ij}} - I_{A_{j'}} - I_{A_{j'} x_{ij'}} \\
&= I_{A_j} - I_{A_{j'}} + I_{A_j x_{ij}} - I_{A_{j'} x_{ij'}}, \qquad (52)
\end{aligned}
$$

where $j, j' = 1, 2, \ldots, K$, $j \neq j'$, and $i = 1, 2, \ldots, G$.

Using the indicator functions of the weights of the A_k and A_{kx_k}, a more general form of the regression functions can be defined. Let b_{A_k} and $b_{A_{kx_k}}$ be constants corresponding to the best attribute and attribute-levels in a pair, and w_{A_k} and $w_{A_{kx_k}}$ be constants corresponding to the worst attribute and attribute-levels in a pair, where $x_k = 1, 2, \ldots, l_k$ and $k = 1, 2, \ldots, K$. The regression functions $\mathbf{f}$ and $\mathbf{g}$ are given as,

$$
\mathbf{f}(x_{ij}) = \sum_{k=1}^{K}\left[b_{A_k} I_{A_k}(x_{ij}) + \sum_{j=1}^{l_k} b_{A_{kx_k}} I_{A_k x_k}(x_{ij})\right] \qquad (53)
$$

and

$$
\mathbf{g}(x_{ij'}) = -\sum_{k=1}^{K}\left[w_{A_k} I_{A_k}(x_{ij'}) + \sum_{j=1}^{l_k} w_{A_{kx_k}} I_{A_k x_k}(x_{ij'})\right] \qquad (54)
$$

where $j, j' = 1, 2, \ldots, K$, $j \neq j'$, and $i = 1, 2, \ldots, G$.

The regression functions defined in this way provide flexibility that the traditional attribute-level best-worst DCEs. Consumer preference in products are constantly changing new information about the product comes to light or as trends come and go. Hence, the data collected today on a product may be obsolete tomorrow. The addition of these constants to the regression functions provides researchers the ability to scale the data to reflect current trends or changes in the products. For example, let us consider the products being modeled are pharmaceuticals such as contraceptives as was done by Knox et al. (2012) and (2013). If new information about a brand of contraceptives posing a health risk was removed from the market, then using regression functions it is possible to update the model to reflect this change. The attribute-level associated with the brand may have $b_{kx_k} = w_{kx_k} = 0$, where $x_k = 1, 2, \ldots, l_k$ and $k = 1, 2, \ldots, K$ to represent its removal from the market. For all the pairs this attribute-level was in, the information the choice pair provides in terms of the other attributes and attribute-levels would remain intact. The model would be estimated again and the parameter vector, $\boldsymbol{\beta}$, would provide the updated impact of the attributes and attribute-levels in the experiment.

We utilize the new definition of the systematic components in the modeling of attribute-level best-worst DCEs across time. In Chapter 4, we extend the work done here to Markov Decision Processes. The generalized form of the systematic components we provided allows for the evaluation of hypothetical future scenarios.

3.4.2 DATA EXAMPLE

In the simulated example an empirical setup is considered. We assume $K = 3$ attributes with $l_1 = 2, l_2 = 3$, and $l_3 = 4$ attribute-levels in an unbalanced design. There are $2 \times 3 \times 4 = 24$ possible profiles, or products, in this experiment. The total number of attribute-levels is $L = \sum_{i=1}^{k} l_i = 9$, and the total number of choice pairs is $J = \sum_{k=1}^{K} l_k(L - l_k) = 52$.

We simulated data for $n = 300$ respondents for 24 profiles. Each choice set has $\tau = K(K-1) = 6$ alternatives to choose from. Using the parameters given in Table 1, we simulated data in R. The data was then exported from R into the SAS environment. Using the SAS procedure called MDC (multinomial discrete choice), the conditional logit model was fitted to the data. The parameter estimates for the generated data are given in Table 1. The parameter estimates are close to the original parameters for this example.

We consider an example where the model is built on the regression functions $\mathbf{f}$ and $\mathbf{g}$ of the data. We define $\mathbf{f}$ and $\mathbf{g}$ as given in equations (53) and (54). The weights used in the regression functions are given as:

$$b_{A_1} = w_{A_1} = -2.$$

$$b_{A_2} = w_{A_2} = 5.$$

$$b_{A_3} = w_{A_3} = 1.$$

$b_{A_1 1} = w_{A_1 1} = b_{A_1 2} = w_{A_1 2} = -2.$

$b_{A_2 1} = w_{A_2 1} = b_{A_2 2} = w_{A_2 2} = b_{A_2 3} = w_{A_2 3} = 5.$

$b_{A_3 1} = w_{A_3 1} = b_{A_3 2} = w_{A_3 2} = b_{A_3 3} = w_{A_3 3} = b_{A_3 4} = w_{A_3 4} = 1.$

The conditional logit model is fit to the data and the resulting parameter estimates are given in Table 1. The parameter estimates provide the adjusted attribute and attribute-level impacts. In Chapter 5, we will evaluate the changes in expected utility for this weighted data in comparison to the original data and model.

Table 1: Parameters and parameter estimates for Simulated Example

		Estimates		Functional Form	
Parameters	β	$\hat{\beta}$	SE	$\hat{\beta}$	SE
β_{A_1}	-2.0000	-2.0711	0.0621	0.9787	0.0289
β_{A_2}	1.5000	1.5248	0.0438	0.3042	0.0082
β_{A_3}	*	*	*	*	*
$\beta_{A_1 1}$	-2.0000	-2.0308	0.0619	0.9838	0.0288
$\beta_{A_1 2}$	2.0000	2.0308	*	-0.9838	*
$\beta_{A_2 1}$	1.9900	2.0970	0.0804	0.3864	0.0148
$\beta_{A_2 2}$	-0.2900	-0.3567	0.0482	-0.0548	0.0092
$\beta_{A_2 3}$	-1.7000	-1.7403	*	-0.3316	*
$\beta_{A_3 1}$	-0.9200	-0.8914	0.0407	-0.8867	0.0410
$\beta_{A_3 2}$	-0.1800	-0.1805	0.0368	-0.1806	0.0368
$\beta_{A_3 3}$	0.5000	0.4911	0.0369	0.4966	0.0366
$\beta_{A_3 4}$	0.6000	0.5808	*	0.5707	*

We can see the impact of weighting as a reciprocal change in the impact of attribute 1 is noticed as its value goes from -2 to 0.9787. We will use these functional forms and included time in them in Chapter 5 in the simulated example with scenarios 5 and 6.

CHAPTER 4

MARKOV DECISION PROCESSES

4.1 INTRODUCTION

Markov decision processes (MDPs) are sequential decision making processes. A decision process is said to be Markovian if the future depends on the present and not the past. In that sense, a Markov process is a memoryless practice. MDPs seek to determine the policy, or set of decision rules, under which maximum reward over time is obtained. According to Puterman (2014), decision processes are defined by the set (S, R, D), where S is the set of states, R is the set of rewards, and D is the set of possible decisions for each time step. Let $s_t \in S$ be the states occupied at time t, $r_t(s_t)$ be the rewards associated with s_t, and $d_t(r_t, s_t)$ is the decision based on possible rewards and states at time t, where $0 \leq t \leq T$. The rewards are defined as the expected gain, or loss, associated with the state. With regards to DCEs, the states are the choices pairs and the rewards are the utility associated with the choice in alternative.

The definition of time is important in the methods for mapping the decision processes. These processes may be discrete or continuous in time with finite or infinite horizon. For the purpose of this dissertation, our interest is in discrete time, finite horizon MDPs, that is $t = 1, 2, \ldots, T$ where $T < \infty$. Numerical methods such as dynamic programming are used to estimate the expected rewards for this type of MDPs.

As the decision process is Markovian, the transition probability to the next state s_{t+1} based solely on the decision made at the current state, s_t, is $p(s_{t+1}|s_t)$, where $t = 1, 2, \ldots, T$ (Puterman, 2014). The transition probabilities are the drivers of this sequential decision making process. The decision process maps the movement from one state to another over time, t, based on rewards received and the optimal decision set. The optimal decision rule is known as the policy, $\delta = (d_1^*, d_2^*, \ldots, d_T^*)$, where

d_t^* is the decision at time $t = 1, 2, \ldots, T$ that yields the maximum expected reward (Puterman, 2014).

MDPs are often applied in inventory monitoring systems. White (1985) provides a literature review of applications for MDPs. Puterman (2014) provides examples of a tire inventory problem and a bus engine replacement problem. For general inventory problems, the decision is made whether or not to order more inventory given the current state of the inventory and transition probabilities related to this state and consumer demand of the products. In such, MDPs have applications in a wide range of problems including econometrics as seen in Diebold et al. (1994). Rust (1994) work focuses on applications in DCEs. It is our interest to extend the work done for DCEs to attribute-level best-worst DCEs.

4.2 APPLICATION TO TRADITIONAL DCEs

The literature on DCEs primarily considers static, or non-time dependent, experiments. The purpose is to apply statistical models to predict consumer behavior under the belief that the choice in alternative is to maximize the utility gained from a product (Marschak, 1960). MDPs follow a similar philosophy with the inclusion of time into the experiments. MDPs are sequential decision making processes where decisions are made under an optimal decision rule δ to maximize future rewards or utilities. The aim is the same at every time point $t = 1, 2, \ldots, T$, to determine the decision/alternative that yields the greatest expected utility.

In DCEs, respondents are asked to choose an alternative based on a set of K attributes describing the product. Let $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tK})$ be the set of K attributes, where x_{tj} is the attribute level of attribute A_j, $j = 1, 2, \ldots, K$ describing the product at time $t = 1, 2, \ldots, T$. The state occupied at time t is defined on $\mathbf{x}_t$ and ϵ_t and is given by $s_t(\mathbf{x}_t, \epsilon_t) \in S$ (Rust, 1994). For DCEs, the reward received in choosing an alternative is its utility, $r(s_t, d_t) = u(s_t, d_t)$, where $t = 1, 2, \ldots, T$.

Decision processes model the sequence of decisions based on expected rewards and transition probabilities. The transition probability from one state to another is defined as,

$$P(s_{t+1}|s_t) = P(s_{t+1} = s'|s_t = s) = P_{ss'}$$

may also be defined in terms of the transition in decisions as $P(d_{t+1}|d_t, x_t)$. The optimal decision made at time t is the d_t that satisfies

$$\max_{d_t \in D} E(U_t(x_t, \epsilon_t)), \tag{55}$$

for $t = 1, 2, \ldots, T$. In this setting, the transition probability can be tied to the decision policy $\delta = (d_1^*, d_2^*, \ldots, d_T^*)$, where d_t^* is the decision at time $t = 1, 2, \ldots, T$ that yields the maximum expected utility as in Equation (55) (Puterman, 2014).

The method for determining δ that produces the maximum expected utility at each time step is done via dynamic programming. Numerical values of the expected utilities are needed and require Monte Carlo estimation. Bellman (1954) utilized dynamic programming to evaluate the value function, also known as Bellman's equation, at each time step. Dynamic programming uses numerical methods to evaluate the value function moving backwards in time. Rust (1994) and Rust (2008) presented the use of dynamic programming for evaluating DCEs as MDPs.

The value function for DCEs defined by Bellman's equation and is given as:

$$\begin{aligned} V_t(x_t, \epsilon_t) &= \max_{d_t \in D} \sum_{t'=t}^{T} P_{ss'}^{t'} \left[\gamma^{t'-t} U(\mathbf{x}_{t'}, d_{t'}) + \epsilon(d_{t'}) | \mathbf{x}_t \right] \\ &= \max_{d_t \in D} E \left(\sum_{t'=t}^{T} \gamma^{t'-t} U(\mathbf{x}_{t'}, d_{t'}) + \epsilon(d_{t'}) | \mathbf{x}_t, \epsilon_t \right), \end{aligned} \tag{56}$$

where $d_t \in D$ is the decision at time t, $U(\mathbf{x}_t, d_t)$ is the derived iterated/expected utility, ϵ_t is the associated error term at time t, where $t = 1, 2, \ldots, T$, and discount utility rate is given by $\gamma \in (0, 1)$ (Bellman, 1954 and 1956). The steps for determining the value function follow (Rust, 1994; Arcidiacono and Miller, 2008; and Ellickson, 2011). There exist J value functions for each of the J alternatives in the experiments evaluated at each time point $t = 1, 2, \ldots, T$. The sum goes from t up to T because it is evaluated using a backwards recursive method, that is we start at the last time point and work our way backwards to earlier time points.

The value functions are computed recursively via dynamic programming. To determine the value function, backwards recursion must be used. At the last time point, T, the value function is the utilities associated with the different states,

$$V_T(x_T) = U(x_T, d_T). \tag{57}$$

Next we move one time step back and compute,

$$V_{T-1}(x_{T-1}, d_{T-1}) = U(x_{T-1}, d_{T-1}) + \sum_{d_T \in D} \gamma V_T(x_T) P(d_T | d_{T-1}),$$

and another,

$$V_{T-2}(x_{T-2}, d_{T-2}) = U(x_{T-2}, d_{T-2}) + \sum_{d_T \in D} \gamma V_{T-1}(x_{T-1}) P(d_{T-1} | d_{T-2}).$$

Following this pattern, we get

$$V_t(x_t, d_t) = U(x_t, d_t) + \sum_{t'=t+1}^{T} \sum_{d_T \in D} \gamma V_{t'}(x_{t'}) P(d_{t'} | d_t), \tag{58}$$

where $t = 1, 2, \ldots, T$ and V_T is defined as in Equation (57).

The decision rule used by a respondent is the one under which the respondent maximizes utility, but one cannot assume that a person's perceived utility is not impacted by time. To adjust for the impact of time on the expected utility, a discount rate $\gamma \in (0, 1)$ is considered. Frederick et al. (2002) reviewed the work done on the discount utility, where they defined the discount function to be

$$D(k) = \left(\frac{1}{1+\gamma} \right)^k, \tag{59}$$

where γ is a respondent's discount factor after k time steps. The discount utility rate weights the utility a person gains from an option at some time $t + k$ based on their current state at time t and guarantees the convergence in the infinite sum of rewards in an infinite horizon MDP (Rothblum , 1975). We consider a finite horizon MDP in this dissertation. By choosing $\gamma \in (0, 1)$ we follow what has been proposed in the literature until now to give higher rewards to immediate utility and lower rewards to delayed utility (Feinberg and Schwartz, 1994).

4.3 EXTENSIONS TO ATTRIBUTE-LEVEL BEST-WORST DCEs

While there exists some literature on the application of MDPs in traditional DCEs, we have not encountered any work in the literature to extend these methods to best-worst scaling DCEs. In this dissertation, we extend the use of MDPs to Case 2 of best-worst scaling models, the attribute-level best-worst DCEs.

In traditional MDPs, the value functions are computed for each of the J alternatives, or products. At each time point, $t = 1, 2, \ldots, T$, the decision d_t is to choose the alternative that provides the maximum expected utility given information about the state $s_t = (\mathbf{x}_t, \epsilon_t)$, where $\mathbf{x}_t$ is the set of K attributes. The decision made is between alternatives in the traditional DCEs. In attribute-level best-worst DCEs, the experiments model choices within products not between products.

In attribute-level best-worst DCEs, there are K attributes describing a product each with l_k levels, where $k = 1, 2, \ldots, K$. The total number of products in these experiments is $\prod_{k=1}^{K} l_k$. The products are represented in the experiment by a profile. The profile corresponding to the i^{th} product is given as $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})$, where x_{ik} is the attribute level corresponding to the attribute A_k for $k = 1, 2, \ldots, K$ for $i = 1, 2, \ldots, G$. Within each choice set there are $\tau = K(K-1)$ choices. A respondent is asked to evaluate G choice sets in the experiment.

MDPs model the decision process for respondents over multiple time points. For attribute-level best-worst DCEs, the model is built within the choice sets corresponding to each of the G choice sets. In traditional DCEs, there are J alternatives evaluated at each time point producing J value functions at each time point. Attribute-level best-worst DCEs require a respondents to evaluate a series of G choice sets each with τ choices, thus there are τ value functions for each choice set in attribute-level best-worst MDPs. Our interest is to further model the sequence of decisions made by introducing the time element into the experiments. For attribute-level best-worst DCEs, we consider discrete time finite horizon MDPs where:

- G choice sets are modeled across time.
- $\mathbf{x}^t_{ijj'} = (x_{ij}, x_{ij'})$, are the attributes and attribute-levels corresponding to the choices in set $\mathcal{C}_i$, $\quad i = 1, 2, \ldots, G$, $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$.

- The decision set depends on the choice set, called D_i, and we evaluate $d_i^t \in D_i$, where $i = 1, 2, \ldots, G$ and $1 \leq d_{ti} \leq \tau$.

- The set of possible states in the experiment depends on the choice set, called S_i, where $s_i^t = (x_{ij}^t, x_{ij'}^t) \in S_i$, where $j \neq j'$, $j, j' = 1, 2, \ldots, \tau$, $i = 1, 2, \ldots, G$, and $1 \leq s_{ti} \leq \tau$.

- Transition probabilities depend on a set of parameters θ that are assumed known, or estimated from data (Arcidiacono and Ellickson, 2011).

- Transition probability matrices, $P_{s_i s_i'}^t$, are dependent on the choice set being evaluated.

In attribute-level best-worst DCEs, the MDPs model the choices in attribute-level pairs within choice sets over time. Therefore, the transition probabilities and value functions must be defined within the choice sets. Referring to Equation (58) the value function in attribute-level best-worst DCEs is given as,

$$V_i^t(s_i^t, d_i^t) = U(s_i^t, d_i^t) + \sum_{s_i^{'t+1} \in S_i} \gamma V_{t+1}^i(s_i^{'t+1}, d_i^{t+1}) P^t(s_i s_i'), \tag{60}$$

where $U(s_i^t, d_i^t)$ represents the utility associated to the state s_i^t and decision d_i^t and $i = 1, 2, \ldots, G$. The decision $d_i^t = (x_{ij}, x_{ij'})$ is a choice pair within $\mathcal{C}_i$, where $i = 1, 2, \ldots, G$, $j, j' = 1, 2, \ldots, K$, and $j \neq j'$. In the attribute-level best-worst DCEs, there will be $\tau = K(K-1)$ value functions per each of the G choice sets. One of the disadvantages of these experiments is the "curse of dimensionality" (Rust, 2008). As the number of attributes, attribute-levels, and profiles grow in the experiment, the estimation process becomes exponentially more difficult as dynamic programming requires an explicit discretization of the states, decisions, and the value function as seen in Equation (60) depends on the utility and transition probabilities over time. The ability to direct the system, via the transition probabilities, when it is of a higher dimension becomes difficult, if not impossible.

In the following subsections, we provide definitions and insights regarding these components to the value function. In Chapter 3, we provided a functional form of the utility that we can apply in these time dependent processes. Furthermore, we define dynamic transition probabilities that we apply to attribute-level best-worst

DCEs. In Chapter 5, simulations of MDPs for the attribute-level best-worst DCEs are provided.

4.3.1 UTILITY

Marschak (1960) presented random utility theory defining utility to include a systematic component V_{ij} and an unobserved component ϵ_{ij} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$. A consumer chooses the alternative that provides them with the maximum utility. The utility function for traditional DCEs is given as in Equation (2) by:

$$U_{ij} = V_{ij} + \epsilon_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \epsilon_{ij},$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$. For MDPs in the traditional DCEs, the utility is then,

$$U_t(\mathbf{x}_t, d_t) = \mathbf{x}'_t\boldsymbol{\beta} + \epsilon_t, \tag{61}$$

where $t = 1, 2, \ldots, T$. Common models to determine the parameter estimates of $\boldsymbol{\beta}$ are conditional logit, generalized extreme value distributions, and probit models presented in the Introduction. We consider the conditional logit model in this dissertation. Stenberg et al. (2007) provided that the definition of utility/reward in MDPs maybe constant over time or time-dependent/dynamic in nature.

The definition of utility in attribute-level best-worst DCEs that meets the necessary IIA condition for the conditional logit model is given in Equation (19) under the inverse random utility theory presented in Marley and Louviere (2005). For the corresponding choice pair $x^t_{ijj'} = (x^t_{ij}, x^t_{ij'}) \in \mathcal{C}^t_i$ the corresponding utility is given as,

$$U^t_{ijj'} = V^t_{ij} - V^t_{ij'} + \epsilon^t_{ijj'} = V^t_{ijj'} + \epsilon^t_{ijj'},$$

where $i = 1, 2, \ldots, G$, $j, j' = 1, 2, \ldots, K$, and $j \neq j'$.

Referring back to Chapter 3, the systematic component is defined as a model built on functions of the best and worst attribute-levels in the pair, using Equation

(44),

$$V_{ijj'} = (\mathbf{f}_t(\mathbf{x}_{ij}) - \mathbf{g}_t(\mathbf{x}_{ij'}))'\boldsymbol{\beta}, \tag{62}$$

where $\boldsymbol{\beta}$ are the attribute and attribute-level coefficients, $j, j' = 1, 2, \ldots, K$, $j \neq j'$, $i = 1, 2, \ldots, G$, and $\mathbf{f}_t$ and $\mathbf{g}_t$, $t = 1, 2, \ldots, T$, are regression functions defined on the best and worst attributes and attribute-levels, respectively. In the traditional attribute-level best-worst DCEs, the functions $\mathbf{f}_t$ and $\mathbf{g}_t$ are given as indicator functions of the best and worst attributes and attribute-levels as shown in Equation (52).

However, using this functional form of the systematic component, we may consider alternative definitions of the systematic component. In Chapter 3, we provide a weighted function for $\mathbf{f}_t$ and $\mathbf{g}_t$, given in equations (53) and (54). Let $b^t_{A_k}$ and $b^t_{A_k x_k}$ be weights corresponding to the best attribute and attribute-levels in a pair, and $w^t_{A_k}$ and $w^t_{A_k x_k}$ be weights corresponding to the worst attribute and attribute-levels in a pair, where $x_k = 1, 2, \ldots, l_k$, $k = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$. The regression functions $\mathbf{f}$ and $\mathbf{g}$ are given as,

$$\mathbf{f}(x_{ij}) = \sum_{k=1}^{K} \left[b_{A_k} I_{A_k}(x_{ij}) + \sum_{j=1}^{l_k} b_{A_{kx_k}} I_{A_k x_k}(x_{ij}) \right],$$

and

$$\mathbf{g}(x_{ij'}) = -\sum_{k=1}^{K} \left[w_{A_k} I_{A_k}(x_{ij'}) + \sum_{j=1}^{l_k} w_{A_{kx_k}} I_{A_k x_k}(x_{ij'}) \right],$$

where $j, j' = 1, 2, \ldots, K$, $j \neq j'$, and $i = 1, 2, \ldots, G$.

Defining the systematic components according to the weighted function allows the utility to change over time. We considered in Chapter 3 an example where an attribute-level no longer exists in the future. The weighted functions of $\mathbf{f}$ and $\mathbf{g}$ allowed us to update the parameter estimates, thus the utilities, using these weights. It is conceivable in the future that an attribute-level scale may need to be adjusted for possible bettering, worsening, or removal type of conditions for that attribute-level.

4.3.2 TRANSITION PROBABILITIES

MDPs have infinitely many possible futures able to be considered in the simulations. The definition of transition probabilities are the vehicle that drives the processes to these different futures. However, determining transition probabilities for MDPs is a difficult task. One way for estimating the transition probabilities is using maximum likelihood estimates (MLEs). An empirical solution to the transition probabilities may be determined by considering the transition probabilities as a multinomial distribution (Lee et al., 1968).

In the attribute-level best-worst DCEs, there are τ choices within in a choice set $\mathcal{C}_i$, where $i = 1, 2, \ldots, G$. There are τ states, and/or decisions, possible at each of the time points. The transition probabilities are denoted as $P_{ss'} = P(s_{t+1} = s' | s_t = s)$, where $s_t, s_{t+1} \in S$ and $S = \{1, 2, \ldots, \tau\}$. Let N_i be the respondents common to time t and $t+1$ in the experiment and $n_{iss'}$ be the number of respondents who chose s at time t and s' at $t+1$, where $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, G$. The transition choice probability is given by the multinomial distribution as:

$$f(p_{is1}, p_{is2}, \ldots, p_{is\tau}) = \frac{N_i!}{n_{is1}! n_{is2}! \ldots n_{is\tau}!} p_{is1}^{n_{is1}} p_{is2}^{n_{is2}} \ldots p_{is\tau}^{n_{is\tau}}, \tag{63}$$

where $s = 1, 2, \ldots, \tau$, $i = 1, 2, \ldots, G$, $p_{iss'} \geq 0$, and $\sum_{s'=1}^{\tau} p_{iss'} = 1$.

The likelihood function is then given by,

$$L(p_{is1}, p_{is2}, \ldots, p_{is\tau}) = \frac{N_i!}{\prod_{s'=1}^{\tau} n_{iss'}!} \prod_{s'=1}^{\tau} p_{iss'}^{n_{iss'}}$$

and the log likelihood is given as,

$$log(L) = log \left(\frac{N_i!}{\prod_{s'=1}^{\tau} n_{iss'}!} \right) + \sum_{s'=1}^{\tau} n_{iss'} log(p_{iss'}),$$

where $s = 1, 2, \ldots, \tau$, $i = 1, 2, \ldots, G$, $p_{iss'} \geq 0$, and $\sum_{s'=1}^{\tau} p_{iss'} = 1$.

Due to the constraint $\sum_{s'=1}^{\tau} p_{iss'} = 1$, Lagrange multipliers, λ, are used and the

Lagrangian function is given as:

$$G(p_{ss'}) = LL(p_{iss'}) - \lambda(\sum_{s'=1}^{\tau} p_{iss'} - 1),$$

where $s = 1, 2, \ldots, \tau$, $i = 1, 2, \ldots, G$, $p_{iss'} \geq 0$, and $\sum_{s'=1}^{\tau} p_{iss'} = 1$.

We take the partial derivative of the Lagrangian to determine the MLEs:

$$\begin{aligned} \frac{\partial G}{\partial p_{iss'}} &= 0 \\ \frac{n_{iss'}}{p_{iss'}} - \lambda &= 0 \\ \frac{n_{iss'}}{p_{iss'}} &= \lambda \\ n_{iss'} &= \lambda p_{iss'} \\ \frac{n_{iss'}}{\lambda} &= p_{iss'} \end{aligned}$$

,

where $s' = 1, 2, \ldots, \tau$. Under the constraint, $\sum_{s'=1}^{\tau} p_{iss'} = 1$, the value of $\lambda = \sum_{s'=1}^{\tau} n_{iss'} = N_i$. Thus, the MLE for

$$p_{iss'} = \frac{n_{iss'}}{N_i}$$

for $s, s' = 1, 2, \ldots, \tau$ and $i = 1, 2, \ldots, G$.

The MLE of $p_{iss'}$ is computationally simple; however, access to the information needed to compute it may not always be available. To compute the MLE of this nature, we would need to have respondents evaluate the same choice sets at two time periods, which is not necessarily an easy task. Furthermore, this is considering the transition matrix is stationary. It is possible to consider a dynamic transition matrix that changes over time, that is $p^t_{iss'}$ for $t = 1, 2, \ldots, T$. A transition matrix of this nature would need to have multiple time periods of data for the same respondents evaluating the same choice sets to compute the empirical probabilities. Instances where multiple time periods of data for respondents are not possible, one must consider alternative methods for determining the transition probabilities.

There are infinitely many possible transition probabilities in MDPs. Common methods for determining these probabilities is to take a Bayesian approach and the other is a rational observation according to Rust (2008). In the Bayesian approach, a prior distribution is needed. However, as mentioned in Rust (2008), strong assumptions are made when using the prior distributions. According to Ghaoui and Nilim (2005), knowledge about the assumed prior distribution must be complete. Alternatively, the rational observation mentioned by Rust (2008) states that one can rationalize any circumstance for consumer behavior or preferences. Arcidiacono and Ellickson (2011) indicates that the transition probabilities, $P_{ss'} = P(s_{t+1} = s'|s_t = s, \boldsymbol{\theta})$ are a probability function, where the parameters $\boldsymbol{\theta}$ are assumed known. Rust (1994) and Rust (2008) state that discrete decision processes, as we are considering in the attribute-level best-worst models, the transition parameters and probabilities are often times non-parametrically identified. Chades et al. (2014) applied MDPs to solve problems in an ecological setting. As they mentioned, to suggest guidance in transition probabilities would require running several scenarios. To our knowledge, such technique has not yet been applied to consumer choice experiments with attribute and attribute-level best-worst experiments.

We provide a definition of the parameters for the transition probabilities under the rational observation that may be used in stationary or dynamic transition matrices. This method maintains the researcher's ability to guide the MDPs in the direction of their choosing where the transitions occur at a rate determined by the researcher. In such way, the researcher is able to consider a stationary or dynamic transition probabilities to model an evolving MDP over time. The researcher may also determine the amount of time points necessary for the system to converge to the decision they were working towards.

In attribute-level best-worst choice models, a set of G choice sets are considered in the experiment. Within each choice set there are $\tau = K(K-1)$ choices. In MDPs, there exists a set of states $s_t \in S$ and possible decisions in $d_t \in D$ for $t = 1, 2, \ldots, T$. For attribute-level best-worst MDPs, the possible states in each choice set are the alternatives, and the decision made at each time point will also be one of the alternatives. For choice set $\mathcal{C}_i$ the state s_{ti} and decision d_{ti} are such that $1 \leq s_{ti}, d_{ti} \leq \tau$ where $i = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$.

Let $s_{it+1} = s'_i$ and $s_{it} = s_i$, where $s'_i, s_i \in S_i$ for $i = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$.

The transition probability is denoted as $P^t_{iss'} = P^t(s'_i|s_i, \boldsymbol{\theta}_{s_i})$, where

$$\boldsymbol{\theta}^t_{s_i} = (\theta^t_{s_iA_1}, \theta^t_{s_iA_2}, \ldots, \theta^t_{s_iA_K}, \theta^t_{s_iA_11}, \ldots, \theta^t_{s_iA_Kl_k})$$

is the set of parameters guiding the transition from s_i to s'_i, for $i = 1, 2, \ldots, G$. In the attribute-level best-worst models, the parameters would be the relative impact/preference associated with the attributes and attribute-levels corresponding to the different choice pairs, or states, given the current state is s_i, where $i = 1, 2, \ldots, G$. According to Rust (2008) and Arcidiacono and Ellickson (2011), $\boldsymbol{\theta}^t_{s_i}$ is assumed known under some rationale with regards to respondent behavior or preferences.

The parameter estimates determined by fitting the conditional logit model, as described in Chapters 2 and 3, produced $\hat{\boldsymbol{\beta}}$ a $p = K + \sum_{k=1}^{K} l_k$ vector. These parameter estimates measure the relative impact of each attribute and attribute-level in the decisions made by respondents. The parameters $\boldsymbol{\theta}^t_{s_i} = (\theta^t_{s_iA_1}, \theta^t_{s_iA_2}, \ldots, \theta^t_{s_iA_K}, \theta^t_{s_iA_11}, \ldots, \theta^t_{s_iA_Kl_k})$ are the assumed impacts of the attributes and attribute-levels in respondents decisions given they currently occupy state s_i. We define these parameters as functions of the parameter estimates $\hat{\boldsymbol{\beta}}$, where there is a rate of change in the impacts over time. We define $\boldsymbol{\theta}^t_{s_i} = (a_{s_iA_1}(t)\hat{\beta}_{A_1}, a_{s_iA_2}(t)\hat{\beta}_{A_2}, \ldots, a_{s_iA_K}(t)\hat{\beta}_{A_k}, a_{s_iA_11}(t)\hat{\beta}_{A_11}, \ldots, a_{s_iA_Kl_k}(t)\hat{\beta}_{A_Kl_k})$, where a'_is are the time factor change and $\beta's$ are fixed for $i = 1, 2, \ldots, G$, $1 \leq s_i \leq \tau$, and $t = 1, 2, \ldots, T$. The definition of the $\mathbf{a}_{s_i}(t)$ depend on the state s_i and time $t = 1, 2, \ldots, T$. We have considered $a_{s_ij}(t) = a^t_{s_ij}$, where if $|a_{s_ij}| < 1$ the impact of the attribute or attribute-level would be lessening with time, where $i = 1, 2, \ldots, G$ and $j = 1, 2, \ldots, K$. If $a_{s_ij}(t)\hat{\beta}_j = a^t_{s_ij}\hat{\beta}_j > 0$, then the attribute or attribute-level has a positive impact evolving at the rate $a^t_{s_ij}$ over time for $j = 1, 2, \ldots, K$, $i = 1, 2, \ldots, G$, and $t = 1, 2, \ldots, T$. A static, or non-time dependent, system is considered if $a_{s_ij}(t) = 1$, where $i = 1, 2, \ldots, G$, $j = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$.

As mentioned, these $a_{s_ij}(t)$ are rates of change that guide how the dynamic transition of the decision process. We can easily consider them to be non-time dependent, $a_{s_ij}(t) = a_{s_ij}$, defining the transition probabilities as stationary over time. As was mentioned earlier, there are infinitely many possibilities in how we define the transitions. Rust (2008) states that using rational observation to define the transitions any possible choice behavior on the respondents is possible. Chades et al. (2014)

recommends running many scenarios to determine the transition probabilities that will maximize the expected reward. Our definition also offers infinitely many possibilities in terms of the definition; however, we defined a rate of change to consider an evolving system. In this way, the researcher can determine what they consider feasible rates and see if the system eventually evolves to the decision they desire and how long it would take to get there.

Given $\boldsymbol{\theta}^t_{s_i}$, the transition probabilities may be determined using random utility theory, or inverse random utility theory in the case attribute-level best-worst models as shown in Chapter 2. Let $s'_{ijj'} = (x_{ij}, x_{ij'})$, where $j \neq j'$, $j, j' = 1, 2, \ldots, \tau$, and $i = 1, 2, \ldots, G$. The probability that $s'_{ijj'}$ is the chosen state means that given $\boldsymbol{\theta}^t_{s_i}$, the utility for $s'_{ijj'}$ is the maximum utility. The transition probability is given as,

$$\begin{aligned} P^t(s'_{ijj'}|s_i, \boldsymbol{\theta}^t_{s_i}) &= P(U^t_{ijj'} > U^t_{ikk'}, \forall k \neq k' \in \mathcal{C}_i | s_i, \boldsymbol{\theta}^t_{s_i}) \\ &= P^t(V^t_{ijj'} + \epsilon^t_{ijj'} > V^t_{ikk'} + \epsilon^t_{ikk'}, \forall k \neq k' \in \mathcal{C}_i | s_i, \boldsymbol{\theta}_{s_i}) \\ &= P^t(\epsilon^t_{ikk'} < \epsilon^t_{ijj'} + V^t_{ijj'} - V^t_{ikk'}, \forall k \neq k' \in \mathcal{C}_i | s_i, \boldsymbol{\theta}^t_{s_i}) \qquad (64) \end{aligned}$$

where $j \neq j'$, $j, j' = 1, 2, \ldots, \tau$, $i = 1, 2, \ldots, G$, and $t = 1, 2, \ldots, T$. If we assume the random error terms are independently and identically distributed as type I extreme value distribution, the probability would then be found using the conditional logit, and is given as:

$$\begin{aligned} P^t(s'_{ijj'}|s_i, \boldsymbol{\theta}^t_{s_i}) &= P^t(U^t_{ijj'} > U^t_{ikk'}, \forall k \neq k' \in \mathcal{C}_i | s_i, \boldsymbol{\theta}^t_{s_i}) \\ &= \frac{\exp(V^t_{ijj'})}{\sum\limits_{k,k' \in \mathcal{C}_i} \exp(V^t_{ikk'})} \\ &= \frac{\exp((\mathbf{f}_t(\mathbf{x}_{ij}) - \mathbf{g}_t(\mathbf{x}_{ij'}))'\boldsymbol{\theta}^t_{s_i})}{\sum\limits_{k,k' \in \mathcal{C}_i} \exp((\mathbf{f}_t(\mathbf{x}_{ik}) - \mathbf{g}_t(\mathbf{x}_{ik'}))'\boldsymbol{\theta}^t_{s_i})}, \qquad (65) \end{aligned}$$

where $j \neq j'$, $k \neq k'$, $j, j' = 1, 2, \ldots, \tau$, $i = 1, 2, \ldots, G$, and $t = 1, 2, \ldots, T$.

The transition matrix is then a $\tau \times \tau$ matrix of the form,

$$P^t = \begin{pmatrix} P^t_{i11} & P^t_{i12} & . & . & . & P^t_{i1\tau} \\ P^t_{i21} & P^t_{i22} & . & . & . & P^t_{i2\tau} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ P^t_{i\tau 1} & P^t_{i\tau 2} & . & . & . & P^t_{i\tau\tau} \end{pmatrix} = \left(P^t_{iss'}\right)$$

where $i = 1, 2, \ldots, G, \quad s, s' = 1, 2, \ldots, \tau$, and where $\sum_{s'=1}^{\tau} P^t_{iss'} = 1$. The transition matrix may be either stationary or dynamic in nature. In our definition of $\boldsymbol{\theta}^t_{s_i}$, this is determined by the rate $a_{s_ij}(t)$, where $i = 1, 2, \ldots, G$, $1 \leq j \leq p$, and $t = 1, 2, \ldots, T$. In Chapter 5, we provide simulations under stationary and dynamic transition probabilities and make comparisons.

CHAPTER 5

EXAMPLES AND APPLICATIONS

We are now ready to conduct/formalize the attribute-level best-worst DCEs of choosing the pairs and describing the optimal variation over time. Such a process will be done at $T = 5$ time periods under the assumption that the consumer chooses an alternative that provides maximum utility of attributes and attribute-levels. We will use numerical maximization of the expected utility under Bellman's equation of the MDP.

Stationary transition probabilities are considered at first under Scenarios 1 and 3 and then dynamic transition probabilities are presented under Scenarios 2 and 4. Two options will be presented a simulated option and an aggregated real data example from Flynn et al. (2007) of predictive and customer analytic expectation. We consider $T = 5$ time epochs with a discount rate $\gamma = 0.95$.

Our computational process can be divided into 3 steps:

Sample data are generated on the basis of the coefficients.

Estimate of the expected utilities are computed using the transition probabilities.

We repeat the process over time.

Under simulated option, we will also offer Scenarios 5 and 6 using the functional form of the data as discussed in Section 4.2.

5.1 SIMULATED EXAMPLE

In the simulated example, an empirical setup is considered. We assume $K = 3$ attributes with $l_1 = 2, l_2 = 3$, and $l_3 = 4$ attribute-levels in an unbalanced design. There are $2 \times 3 \times 4 = 24$ possible profiles, or products, in this experiment. The total number of attribute-levels is $L = \sum_{i=1}^{k} l_i = 9$, and the total number of choice pairs is $J = \sum_{k=1}^{K} l_k(L - l_k) = 52$.

Louviere and Woodworth (1983), Street and Knox (2012), and Graßhoff et al. (2004) discussed the benefits in using orthogonal arrays. Generally, orthogonal experimental designs are utilized in attribute-level best-worst DCEs due to the large number of profiles in a full factorial design. There is a package in R called DoE.design that creates full factorial and orthogonal designs for a given set of attributes and attribute-levels. To obtain an orthogonal design, the oa.design function is used. For this experiment, the orthogonal design returned the full factorial design, so we used the full set of 24 profiles when simulating this data.

We simulated data for $n = 300$ respondents for 24 profiles. Each choice set has $\tau = K(K-1) = 6$ choices to choose from. Using the parameters given in Table 1, we simulated data in R. The data was then exported from R into the SAS environment. Using the SAS procedure called MDC (multinomial discrete choice), the conditional logit model was fitted to the data. The parameter estimates for the generated data are given in 1. The parameter estimates are close to the original parameters for this example. Using the parameter estimates, the choice utilities were computed and are used to determine the expected utility/value function. The best and worst 3 choice pairs along with their utilities are presented in tables 3 and 4. The opposite of the pairs with the highest utilities have the lowest utilities.

Table 2: Parameters and parameter estimates for Simulated Example

		Estimates		Functional Form	
Parameters	β	$\hat{\beta}$	SE	$\hat{\beta}$	SE
β_{A_1}	-2.0000	-2.0711	0.0621	0.9787	0.0289
β_{A_2}	1.5000	1.5248	0.0438	0.3042	0.0082
β_{A_3}	*	*	*	*	*
$\beta_{A_1 1}$	-2.0000	-2.0308	0.0619	0.9838	0.0288
$\beta_{A_1 2}$	2.0000	2.0308	*	-0.9838	*
$\beta_{A_2 1}$	1.9900	2.0970	0.0804	0.3864	0.0148
$\beta_{A_2 2}$	-0.2900	-0.3567	0.0482	-0.0548	0.0092
$\beta_{A_2 3}$	-1.7000	-1.7403	*	-0.3316	*
$\beta_{A_3 1}$	-0.9200	-0.8914	0.0407	-0.8867	0.0410
$\beta_{A_3 2}$	-0.1800	-0.1805	0.0368	-0.1806	0.0368
$\beta_{A_3 3}$	0.5000	0.4911	0.0369	0.4966	0.0366
$\beta_{A_3 4}$	0.6000	0.5808	*	0.5707	*

Table 3: Choice pairs with the highest utility in the experiment

Best Attribute	Level	Worst Attribute	Level	Utility
2	1	1	1	12.3633
2	2	1	1	8.8012
3	4	1	1	7.6931

Table 4: Choice pairs with the lowest utility in the experiment

Best Attribute	Level	Worst Attribute	Level	Utility
1	1	2	1	-9.2594
1	1	2	2	-6.5358
1	1	3	4	-5.7929

5.1.1 SCENARIO 1

We ran the simulation under this scenario with an advantageous proposed structure. The intent is to validate/justify our relative performance over time under stationary sparcity.

In this example, respondents are assumed to make similar decisions at each decision epoch that they made at the previous time point. The transition parameters $\boldsymbol{\theta}_{s_i}^t$ where $s_i^t = (x_{ij}, x_{ij'})$ are defined as for the attributes as,

$$\theta_{s_i A_k}^t = \begin{cases} 1.7|\beta_{A_k}|, & \text{if } x_{ij} \in A_k, \\ -1.7|\beta_{A_k}|, & \text{if } x_{ij'} \in A_k, \\ \beta_{A_k}, & \text{otherwise,} \end{cases} \tag{66}$$

and for the attribute-levels,

$$\theta_{s_i A_k x_{ik}}^t = \begin{cases} 1.7|\beta_{A_k x_{ik}}|, & \text{if } x_{ij} = x_{ik} \text{ where } x_{ik} \in A_k, \\ -1.7|\beta_{A_k x_{ik}}|, & \text{if } x_{ij'} = x_{ik} \text{ where } x_{ik} \in A_k, \\ \beta_{A_k x_{ik}}, & \text{otherwise,} \end{cases} \tag{67}$$

where $j \neq j'$, $j, j', k = 1, 2, \ldots, K$, $1 \leq x_k \leq l_k$, and $i = 1, 2, \ldots, G$. The transition parameters do not change with time, so the transition matrix is stationary. The goal of this scenario was to design the transition probabilities in a way that the choice made at t is most likely to be made at $t + 1$. If we considered $a_{s_i m}(t) = \beta_m$ for $i = 1, 2, \ldots, G$, and $m = 1, 2, \ldots, p$, then the system would remain static and every row of the transition matrix would be the same. Recall that $p = K + \sum_{k=1}^{K} l_k = 12$ is the number of parameters. We consider $1.7|\beta_m|$ when a state or choice pair at time $t + 1$ has the same best attribute and attribute-level as the state occupied at time t, and $-1.7|\beta_m|$ when a state or choice pair at time $t + 1$ has the same worst attribute and attribute-level as the state occupied at time t. We consider $|\beta_m|$ to control the direction of the impact making sure it is positive for the best attribute and attribute-level of s_i and use $-|\beta_m|$ to make sure its negative for the worst attribute and attribute-level of s_i. We use 1.7 to increase the impact of the best and worst attributes and attribute-levels of s_i. The definition of $a_{s_i m}(t)$ in this way insures that states with common best and worst attributes and attribute-levels as the present state occupied, $s_i^t = (x_{ij}, x_{ij'})$, have a greater probability of being transitioned to, where $i = 1, 2, \ldots, G$, $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$.

The value function/expected utilities for Profile 1 are displayed in Figure 3 along with the difference in value function. Choice pair (x_{22}, x_{12}), where x_{22} is the 2^{nd} level of attribute 2 is the best and x_{12} is the 2^{nd} level of attribute 1 is the worst, corresponds to the highest expected utility. The opposite pair (x_{12}, x_{22}) is the worst choice pair.

Table 5: Stationary transition matrix in Scenario 1 for Profile 1

(x_{12}, x_{22})	0.9837	0.0000	0.0136	0.0000	0.0000	0.0027
(x_{22}, x_{12})	0.0000	0.8924	0.0000	0.1074	0.0003	0.0000
(x_{12}, x_{34})	0.0038	0.0000	0.9932	0.0000	0.0030	0.0000
(x_{34}, x_{12})	0.0000	0.0038	0.0000	0.9613	0.0000	0.0003
(x_{22}, x_{34})	0.0000	0.4289	0.0004	0.0002	0.5705	0.0000
(x_{34}, x_{22})	0.0001	0.0004	0.0000	0.7113	0.0000	0.2882

The model applied here views the attribute-level best-worst DCEs as sequential leading to a partial separation best-worst choices over time. Validity is guided by the transition probabilities under Scenario 1, the participants follow the same choice preferences. In Table 5, the transition probabilities are generally highest on the diagonal and the same at each time period as we would expect in this setup. As expected the trend in the utility is kept.

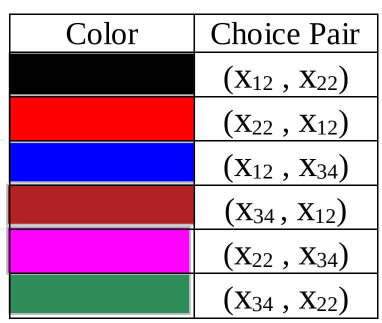
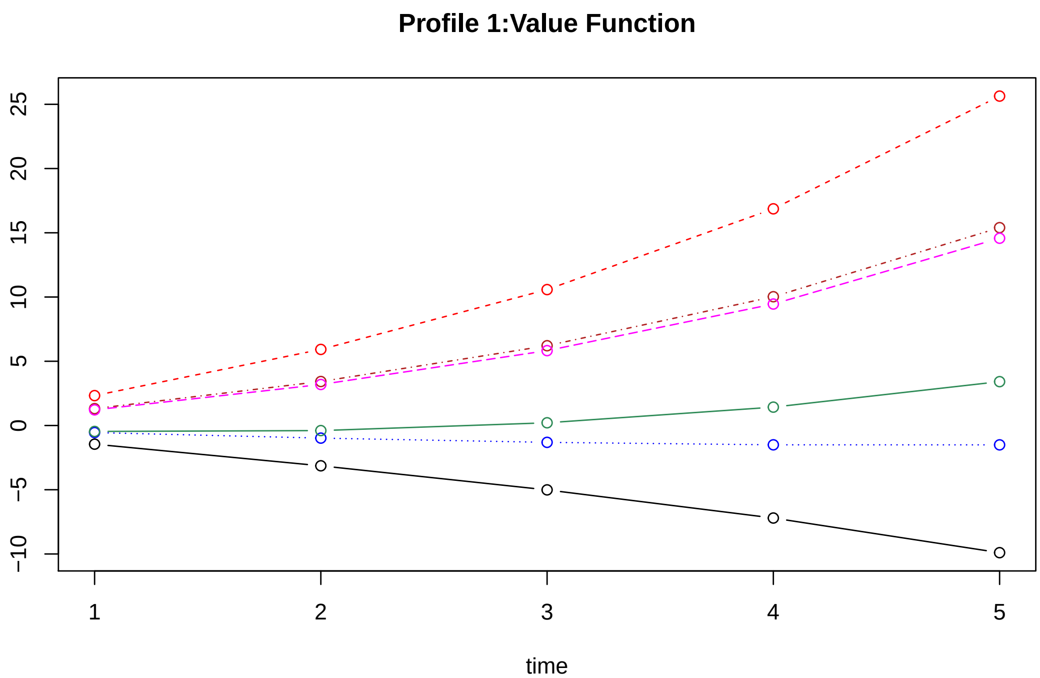


Figure 2: Legend corresponding to Figure 3



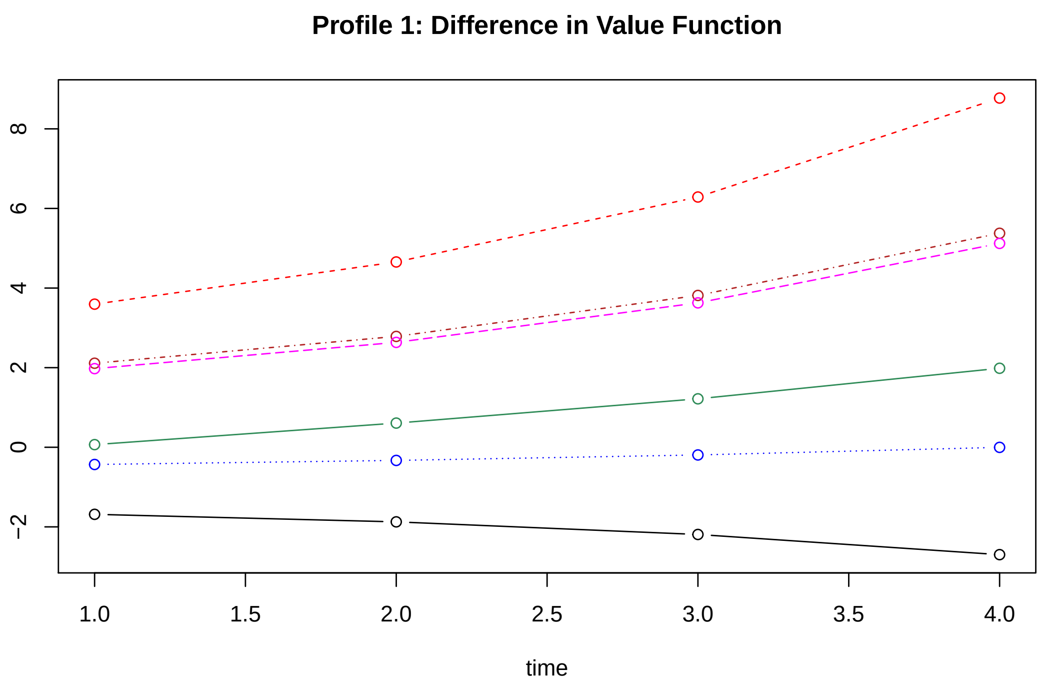


Figure 3: Expected discounted utility and their differences over time for the Profile 1

5.1.2 SCENARIO 2

In Scenario 2, respondents are allowed to make similar decisions at each time epoch with a different rate of change, making the transition probabilities dynamic. The transition parameters $\boldsymbol{\theta}_{s_i}^t$ where $s_i^t = (x_{ij}, x_{ij'})$ are defined as for the attributes as,

$$\theta_{s_i A_k}^t = \begin{cases} 1.7^t|\beta_{A_k}|, & \text{if } x_{ij} \in A_k, \\ -1.7^t|\beta_{A_k}|, & \text{if } x_{ij'} \in A_k, \\ \beta_{A_k}, & \text{otherwise}, \end{cases} \tag{68}$$

and for the attribute-levels,

$$\theta_{s_i A_k x_{ik}}^t = \begin{cases} 1.7^t|\beta_{A_k x_{ik}}|, & \text{if } x_{ij} = x_{ik} \text{ where } x_{ik} \in A_k, \\ -1.7^t|\beta_{A_k x_{ik}}|, & \text{if } x_{ij'} = x_{ik} \text{ where } x_{ik} \in A_k, \\ \beta_{A_k x_{ik}}, & \text{otherwise}, \end{cases} \tag{69}$$

where $j \neq j'$, $j, j', k = 1, 2, \ldots, K$, $1 \leq x_k \leq l_k$, and $i = 1, 2, \ldots, G$.

The transition matrix at time $t = 1$ is kept the same as it was Scenario 1 in Table 5, and subsequent transition probabilities at time $t = 2, 3$, and 4 are given in tables 6, 7, and 8, respectively. The transition probabilities are highest on the diagonal verifying the direction we wanted in the transitions. The value function/expected utilities for Profile 1 are displayed in Figure 5 along with the difference in value function. The same best-worst pair (x_{22}, x_{12}) is the optimal choice over time for Profile 1 as it was in Scenario 1 with a slight difference in the expected utilities. This is because the transition matrices are reinforcing those choices over time. This explains why the transition matrix under Table 8 is the identity matrix as expected under the trend in the utility.

Table 6: Dynamic transition matrix in Scenario 2 at time $t = 2$ for Profile 1

(x_{12}, x_{22})	0.9985	0.0000	0.0015	0.0000	0.0000	0.0000
(x_{22}, x_{12})	0.0000	0.9873	0.0000	0.0127	0.0000	0.0000
(x_{12}, x_{34})	0.0002	0.0000	0.9998	0.0000	0.0000	0.0000
(x_{34}, x_{12})	0.0000	0.0019	0.0000	0.9981	0.0000	0.0000
(x_{22}, x_{34})	0.0000	0.0337	0.0001	0.0000	0.9663	0.0000
(x_{34}, x_{22})	0.0000	0.0000	0.0000	0.2082	0.0000	0.7918

Table 7: Dynamic transition matrix in Scenario 2 at time $t = 3$ for Profile 1

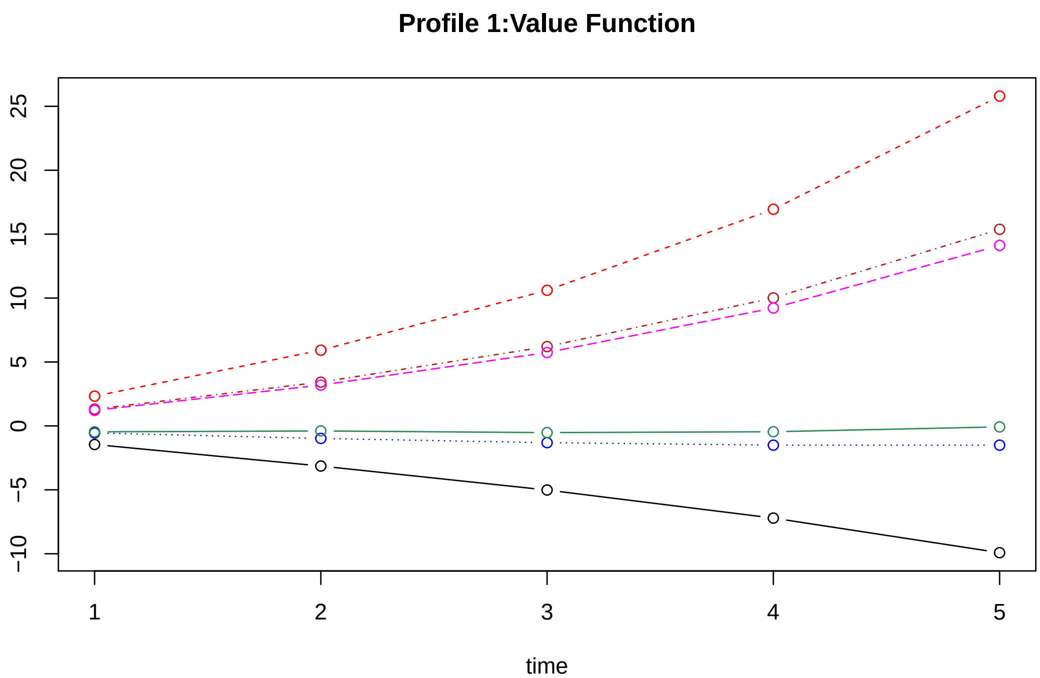
(x_{12}, x_{22})	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
(x_{22}, x_{12})	0.0000	0.9997	0.0000	0.0003	0.0000	0.0000
(x_{12}, x_{34})	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
(x_{34}, x_{12})	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
(x_{22}, x_{34})	0.0000	0.0002	0.0000	0.0000	0.9998	0.0000
(x_{34}, x_{22})	0.0000	0.0000	0.0000	0.0058	0.0000	0.9942

Table 8: Dynamic transition matrix in Scenario 2 at time $t = 4$ for Profile 1

(x_{12}, x_{22})	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
(x_{22}, x_{12})	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
(x_{12}, x_{34})	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
(x_{34}, x_{12})	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
(x_{22}, x_{34})	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
(x_{34}, x_{22})	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Color	Choice Pair
	(x_{12}, x_{22})
	(x_{22}, x_{12})
	(x_{12}, x_{34})
	(x_{34}, x_{12})
	(x_{22}, x_{34})
	(x_{34}, x_{22})

Figure 4: Legend corresponding to Figure 5



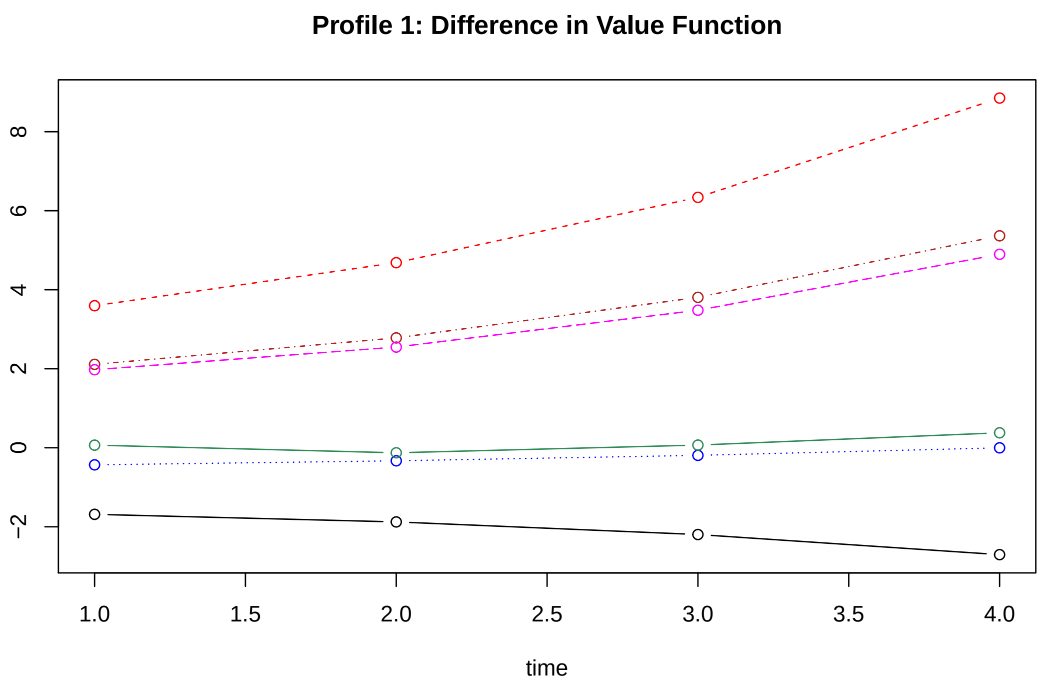


Figure 5: Expected discounted utility and their differences over time for the Profile 1

5.1.3 SCENARIO 3

In this scenario, respondents are allowed to make different decisions under a stationary transition probability. Table 9 provide the transition probabilities for choices with some of the most noticeable shifts. Respondents who used to consider attribute 2 with positive utility are now considering it with negative utility because $-0.7^{T-1} = -0.24$ has been attached to the transition probability of attribute 2. Similar changes can be seen in 9. We provide the transition matrix for Profile 1 built under this consideration in Table 10.

Table 9: Parameters and parameter estimates from simulations

Parameters	(x_{21}, x_{31})	(x_{31}, x_{21})	(x_{21}, x_{34})	(x_{34}, x_{21})	(x_{11}, x_{33})	(x_{33}, x_{11})
θ_1	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$	$0.9\hat{\beta}_1$	$1.7\hat{\beta}_1$
θ_2	$-0.7^{T-1}\hat{\beta}_2$	$1.1\hat{\beta}_2$	$-0.7^{T-1}\hat{\beta}_2$	$1.1\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$
θ_3	$0.9\hat{\beta}_3$	$1.7\hat{\beta}_3$	$0.9\hat{\beta}_3$	$1.7\hat{\beta}_3$	$0.9\hat{\beta}_3$	$1.7\hat{\beta}_3$
θ_{11}	$\hat{\beta}_{11}$	$\hat{\beta}_{11}$	$\hat{\beta}_{11}$	$\hat{\beta}_{11}$	$0.9\hat{\beta}_{11}$	$1.1\hat{\beta}_{11}$
θ_{12}	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$
θ_{21}	$0.5\hat{\beta}_{21}$	$1.1\hat{\beta}_{21}$	$0.5\hat{\beta}_{21}$	$1.1\hat{\beta}_{21}$	$\hat{\beta}_{21}$	$\hat{\beta}_{21}$
θ_{22}	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$
θ_{23}	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$
θ_{31}	$1.15\hat{\beta}_{31}$	$0.9\hat{\beta}_{31}$	$\hat{\beta}_{31}$	$\hat{\beta}_{31}$	$\hat{\beta}_{31}$	$\hat{\beta}_{31}$
θ_{32}	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$
θ_{33}	$\hat{\beta}_{33}$	$\hat{\beta}_{33}$	$\hat{\beta}_{33}$	$\hat{\beta}_{33}$	$0.95\hat{\beta}_{22}$	$1.3\hat{\beta}_{33}$
θ_{34}	$\hat{\beta}_{34}$	$\hat{\beta}_{34}$	$0.9\hat{\beta}_{34}$	$1.3\hat{\beta}_{34}$	$\hat{\beta}_{34}$	$\hat{\beta}_{34}$

Note: A_3 has no parameter estimate as it was the baseline category in the conditional logit model. For the purpose of the transition matrix, we take $\hat{\beta}_3$ to be 0.5.

Table 10: Stationary transition matrix in Scenario 3 for Profile 1

(x_{12}, x_{22})	0.3060	0.0023	0.0131	0.0543	0.0011	0.6231
(x_{22}, x_{12})	0.1475	0.0284	0.0172	0.2439	0.0076	0.5554
(x_{12}, x_{34})	0.0522	0.2578	0.1806	0.0745	0.4013	0.0335
(x_{34}, x_{12})	0.0250	0.4221	0.0731	0.1443	0.3004	0.0351
(x_{22}, x_{34})	0.3396	0.0435	0.1255	0.1176	0.0449	0.3288
(x_{34}, x_{22})	0.4094	0.0071	0.0470	0.0162	0.0062	0.4687

The value function/expected utilities for Profile 1 are displayed in Figure 7 along with the difference in the value function over time. The choice pair that started as the best (x_{22}, x_{12}), where x_{22} is the 2^{nd} level of attribute 2 is the best and x_{12} is the 2^{nd} level of attribute 1 is the worst, has now been changed to (x_{34}, x_{12}) starting at time $t = 3$. The intent was to validate the transition probability moving from attribute 2 to attribute 3 as the best and from attribute-level x_{22} to x_{34} as the best. The opposite pair (x_{12}, x_{22}) stayed the worst choice pair throughout the experiment. The worst pair has not changed because the focus was on altering the best pair.

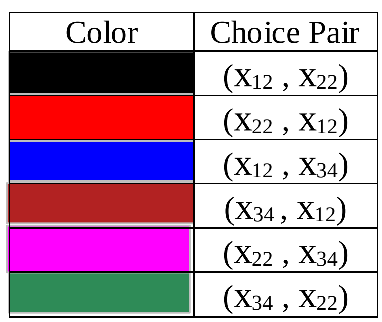
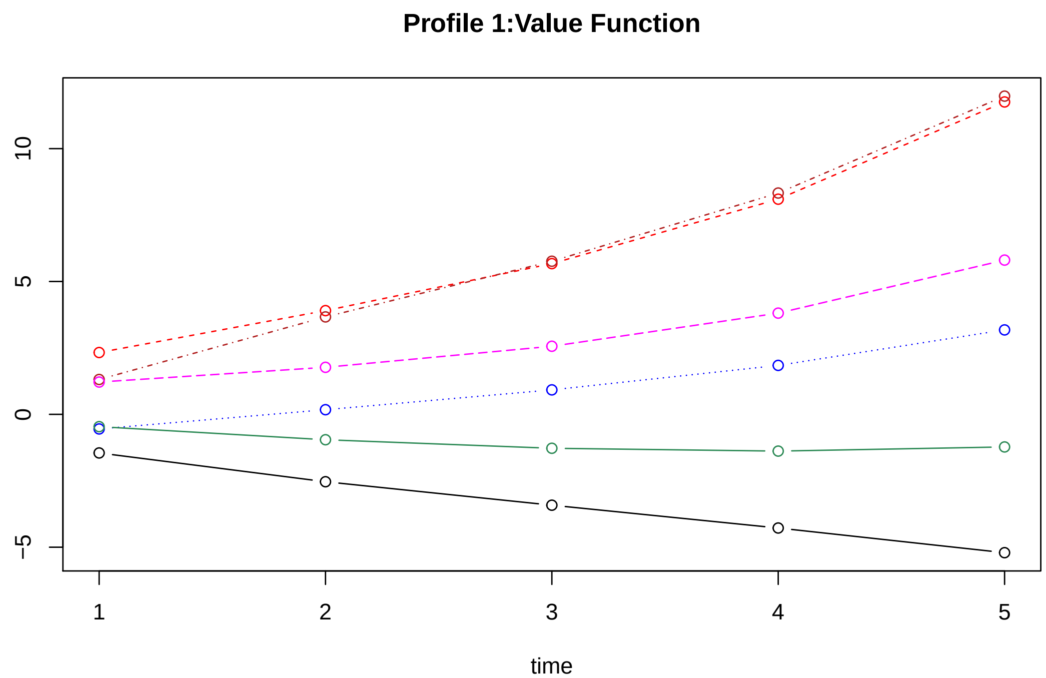


Figure 6: Legend corresponding to Figure 7



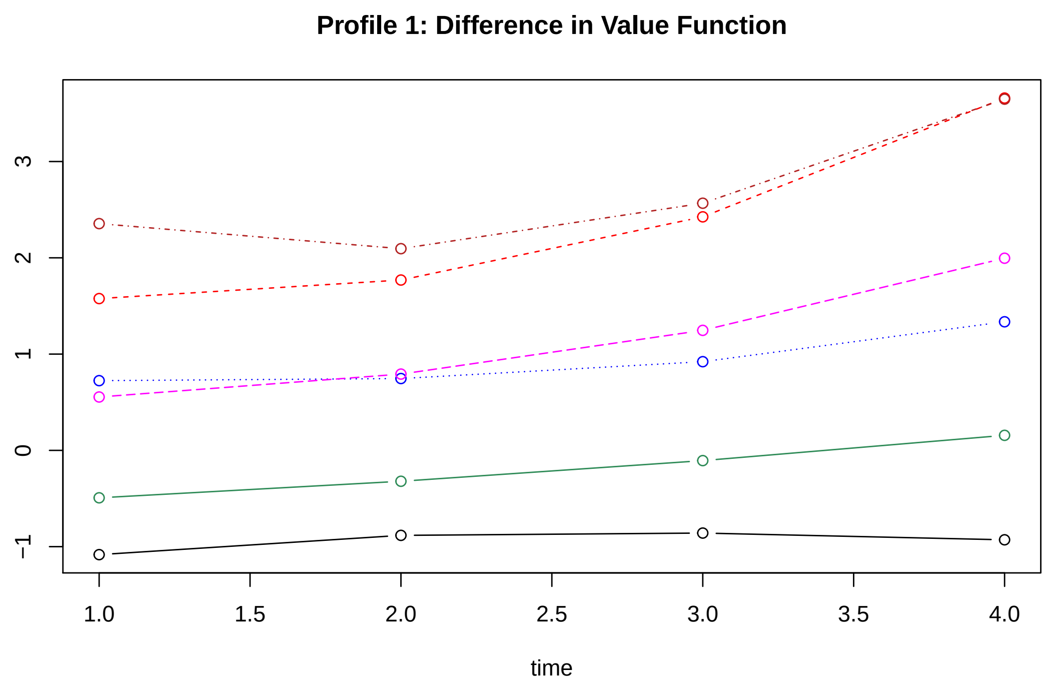


Figure 7: Expected discounted utility and their differences over time for Profile 1

5.1.4 SCENARIO 4

Scenario 4 is an extension of Scenario 3 in the sense that the coefficient associated with the transition probabilities are dynamic in the fact that we exponentiated the coefficient to the t^{th} power. So, respondents are allowed to make different decisions under a dynamic transition probabilities. Among the probabilities that have the noticeable changes in the transition probabilities are given in Table 11.

Respondents who used to consider attribute 2 with positive utility are now considering it with negative utility because -0.7^{T-t} has been attached to the transition probability of attribute 2 changing over time. As for attribute 3, the coefficients change at the rate 1.3^t at the attribute-level x_{34}. Similar changes can be seen in the transition parameters given in Table 11. We provide the transition matrix for Profile 1 at time $t = 2, 3, 4$ built under this consideration in tables 12, 13, and 14, respectively.

Table 11: Parameters and parameter estimates from simulations

Parameters	(x_{21}, x_{31})	(x_{31}, x_{21})	(x_{21}, x_{34})	(x_{34}, x_{21})	(x_{11}, x_{33})	(x_{33}, x_{11})
θ_1	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$	$0.9^t\hat{\beta}_1$	$1.7^t\hat{\beta}_1$
θ_2	$-0.7^{T-t}\hat{\beta}_2$	$1.1^t\hat{\beta}_2$	$-0.7^{T-t}\hat{\beta}_2$	$1.1^t\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$
θ_3	$0.9^t\hat{\beta}_3$	$1.7^t\hat{\beta}_3$	$0.9^t\hat{\beta}_3$	$1.7^t\hat{\beta}_3$	$0.9^t\hat{\beta}_3$	$1.7^t\hat{\beta}_3$
θ_{11}	$\hat{\beta}_{11}$	$\hat{\beta}_{11}$	$\hat{\beta}_{11}$	$\hat{\beta}_{11}$	$0.9^t\hat{\beta}_{11}$	$1.1^t\hat{\beta}_{11}$
θ_{12}	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$	$\hat{\beta}_{12}$
θ_{21}	$0.5^t\hat{\beta}_{21}$	$1.1^t\hat{\beta}_{21}$	$0.5^t\hat{\beta}_{21}$	$1.1^t\hat{\beta}_{21}$	$\hat{\beta}_{21}$	$\hat{\beta}_{21}$
θ_{22}	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$	$\hat{\beta}_{22}$
θ_{23}	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$	$\hat{\beta}_{23}$
θ_{31}	$1.15^t\hat{\beta}_{31}$	$0.9^t\hat{\beta}_{31}$	$\hat{\beta}_{31}$	$\hat{\beta}_{31}$	$\hat{\beta}_{31}$	$\hat{\beta}_{31}$
θ_{32}	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$	$\hat{\beta}_{32}$
θ_{33}	$\hat{\beta}_{33}$	$\hat{\beta}_{33}$	$\hat{\beta}_{33}$	$\hat{\beta}_{33}$	$0.95^t\hat{\beta}_{22}$	$1.3^t\hat{\beta}_{33}$
θ_{34}	$\hat{\beta}_{34}$	$\hat{\beta}_{34}$	$0.9^t\hat{\beta}_{34}$	$1.3^t\hat{\beta}_{34}$	$\hat{\beta}_{34}$	$\hat{\beta}_{34}$

Note: A_3 has no parameter estimate as it was the baseline category in the conditional logit model. For the purpose of the transition matrix, we take $\hat{\beta}_3$ to be 0.5.

Table 12: Dynamic transition matrix in Scenario 4 at time $t = 2$ for Profile 1

(x_{12}, x_{22})	0.4076	0.0009	0.0142	0.0259	0.0007	0.5508
(x_{22}, x_{12})	0.1043	0.0506	0.0157	0.3361	0.0109	0.4823
(x_{12}, x_{34})	0.0785	0.1708	0.2696	0.0497	0.3976	0.0337
(x_{34}, x_{12})	0.0186	0.4735	0.0376	0.2341	0.1898	0.0464
(x_{22}, x_{34})	0.3000	0.0641	0.1422	0.1353	0.0657	0.2926
(x_{34}, x_{22})	0.3431	0.0039	0.0222	0.0607	0.0024	0.5678

Table 13: Dynamic transition matrix in Scenario 4 at time $t = 3$ for Profile 1

(x_{12}, x_{22})	0.5142	0.0003	0.0143	0.0114	0.0004	0.4595
(x_{22}, x_{12})	0.1149	0.0120	0.0065	0.2113	0.0021	0.6531
(x_{12}, x_{34})	0.1081	0.1028	0.3686	0.0301	0.3595	0.0309
(x_{34}, x_{12})	0.0109	0.4190	0.0108	0.4242	0.0667	0.0684
(x_{22}, x_{34})	0.4179	0.0128	0.0744	0.0717	0.0130	0.4102
(x_{34}, x_{22})	0.2122	0.0015	0.0053	0.0612	0.0005	0.7193

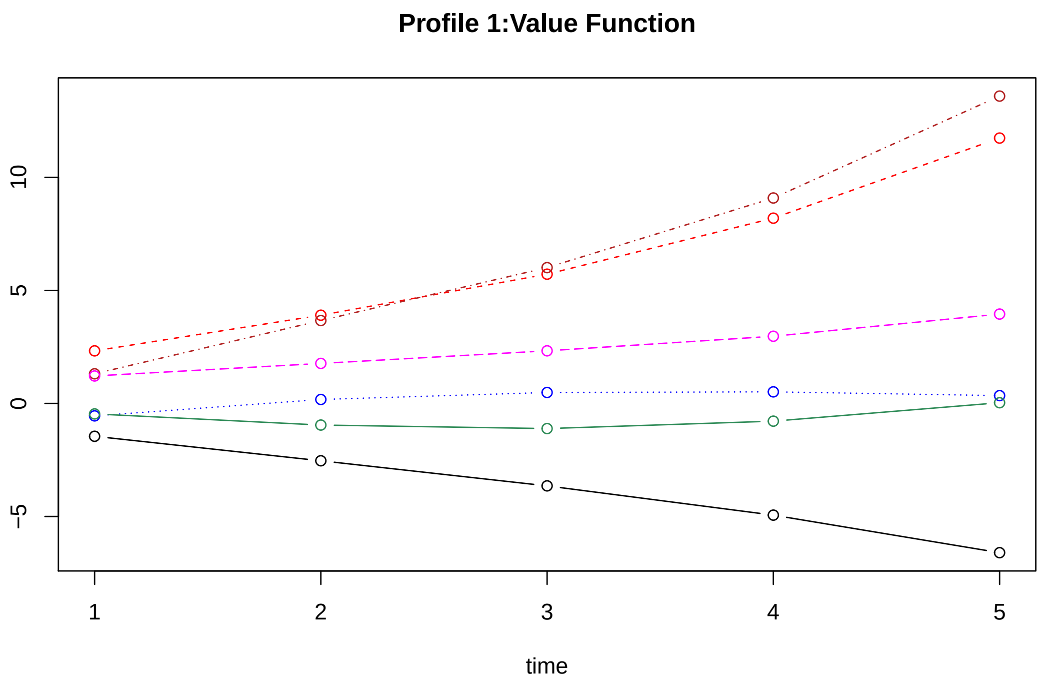
Table 14: Dynamic transition matrix in Scenario 4 at time $t = 4$ for Profile 1

(x_{12}, x_{22})	0.6188	0.0001	0.0134	0.0046	0.0002	0.3629
(x_{22}, x_{12})	0.1047	0.0047	0.0032	0.1554	0.0007	0.7314
(x_{12}, x_{34})	0.1374	0.0563	0.4658	0.0166	0.2981	0.0259
(x_{34}, x_{12})	0.0032	0.1847	0.0008	0.7117	0.0063	0.0933
(x_{22}, x_{34})	0.4563	0.0040	0.0431	0.0420	0.0040	0.4505
(x_{34}, x_{22})	0.0677	0.0003	0.0003	0.0579	0.0000	0.8737

The value function/expected utilities for Profile 1 are displayed in Figure 9 along with the difference in the value function over time. The choice pair that started as the best (x_{22}, x_{12}), where x_{22} is the 2^{nd} level of attribute 2 is the best and x_{12} is the 2^{nd} level of attribute 1 is the worst, has now been changed to (x_{34}, x_{12}) between $t = 2$ and $t = 3$ in contrast with Scenario 3. The intent to validate the transition probability moving from attribute 2 to attribute 3 as the best as in Scenario 3.

Color	Choice Pair
	(x_{12}, x_{22})
	(x_{22}, x_{12})
	(x_{12}, x_{34})
	(x_{34}, x_{12})
	(x_{22}, x_{34})
	(x_{34}, x_{22})

Figure 8: Legend corresponding to Figure 9



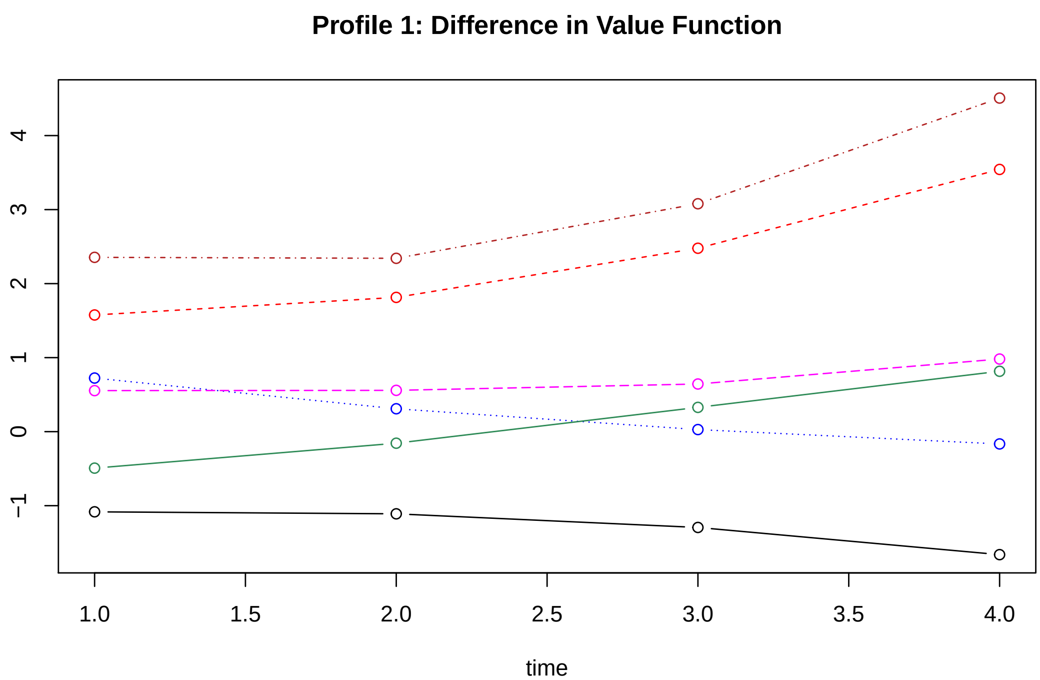


Figure 9: Expected discounted utility and their differences over time for Profile 1

5.1.5 SCENARIO 5

We ran the simulation under this scenario with advantageous proposed hybrid structure of Scenario 1 using the functional form of the utility as described in Chapter 4. The transition matrix is stationary and defined as in Scenario 1. The weights associated to the attributes and attribute levels are selected as:

$b_{A_1} = w_{A_1} = -2.$

$b_{A_2} = w_{A_2} = 5.$

$b_{A_3} = w_{A_3} = 1.$

$b_{A_1 1} = w_{A_1 1} = b_{A_1 2} = w_{A_1 2} = -2.$

$b_{A_2 1} = w_{A_2 1} = b_{A_2 2} = w_{A_2 2} = b_{A_2 3} = w_{A_2 3} = 5.$

$b_{A_3 1} = w_{A_3 1} = b_{A_3 2} = w_{A_3 2} = b_{A_3 3} = w_{A_3 3} = b_{A_3 4} = w_{A_3 4} = 1.$

Referring back to Chapter 3, the systematic component as a function of the best and worst attribute-level in the pair, is as in Equation (44),

$$V_{ijj'} = (\mathbf{f}_t(\mathbf{x}_{ij'}) - \mathbf{g}_t(\mathbf{x}_{ij'}))'\boldsymbol{\beta},$$

where **f** and **g**, as in Equation (53) and (54) are defined as,

$$\mathbf{f}(x_{ij}) = \sum_{k=1}^{K} \left[b_{A_k} I_{A_k}(x_{ij}) + \sum_{j=1}^{l_k} b_{A_{k x_k}} I_{A_k x_k}(x_{ij}) \right]$$

and

$$\mathbf{g}(x_{ij'}) = -\sum_{k=1}^{K} \left[w_{A_k} I_{A_k}(x_{ij'}) + \sum_{j=1}^{l_k} w_{A_{k x_k}} I_{A_k x_k}(x_{ij'}) \right].$$

The value function/expected utilities for Profile 1 are displayed in Figure 11 along with the difference in the value functions over time. Choice pair (x_{22}, x_{12}), where x_{22} is the 2^{nd} level of attribute 2 is the best and x_{12} is the 2^{nd} level of attribute 1 is the worst, still remains the choice with the highest expected utility as in Scenario 1. The opposite pair (x_{12}, x_{22}) is the worst choice pair. The pair (x_{34}, x_{22}) has a sharp drop between time $t = 3$ and $t = 4$ because of the change in the weights applied to the attributes and attribute-levels from Equation (44).

Color	Choice Pair
	(x_{12}, x_{22})
	(x_{22}, x_{12})
	(x_{12}, x_{34})
	(x_{34}, x_{12})
	(x_{22}, x_{34})
	(x_{34}, x_{22})

Figure 10: Legend corresponding to Figure 11

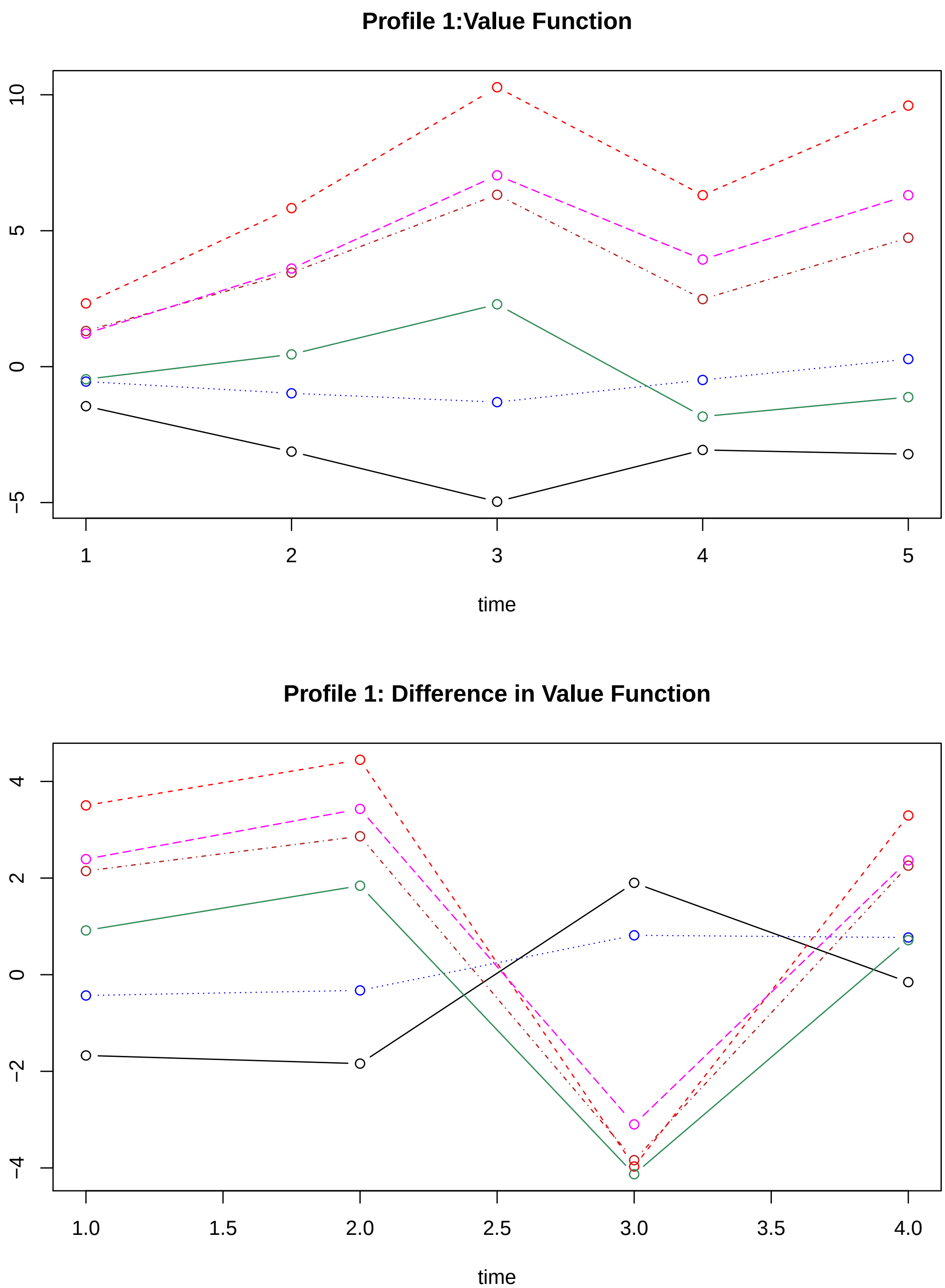


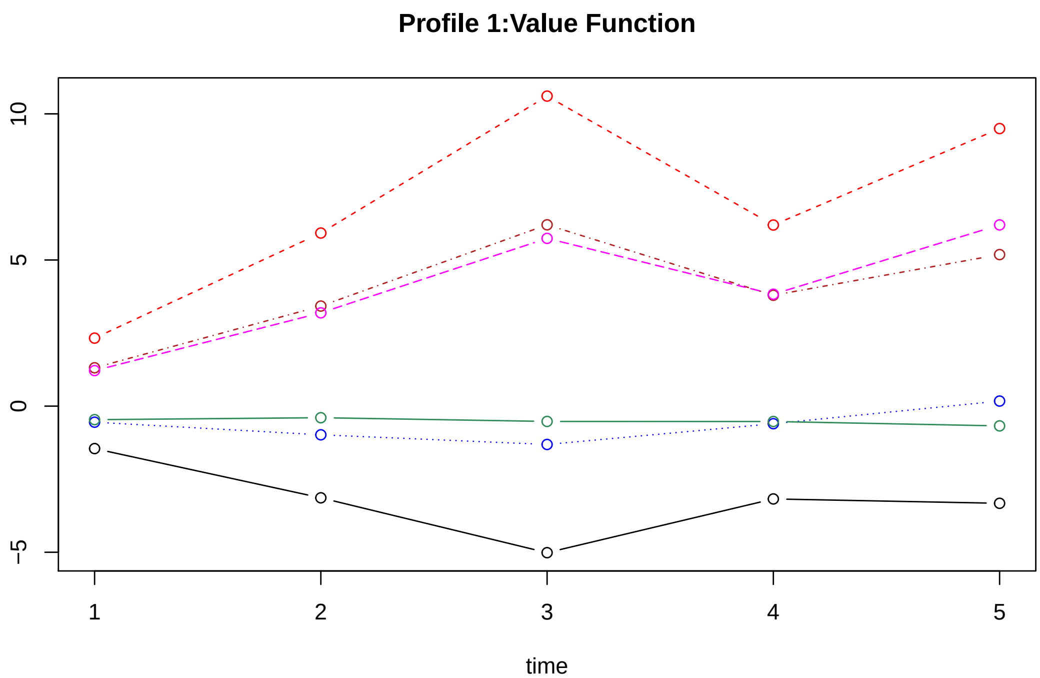
Figure 11: Expected discounted utility and their differences over time for Profile 1

5.1.6 SCENARIO 6

We ran the simulation under this scenario with advantageous proposed hybrid structure of Scenario 2 using the functional form as described in Scenario 5. The value function/expected utilities for Profile 1 are displayed in Figure 13 along with the difference in value functions. Choice pair (x_{22}, x_{12}), where x_{22} is the 2^{nd} level of attribute 2 is the best and x_{12} is the 2^{nd} level of attribute 1 is the worst, still remains the choice with the highest expected utility as in Scenario 5. The opposite pair (x_{12}, x_{22}) is the worst choice pair. We also notice more shifts in expected utility than in previous scenarios for Profile 1. Scaling the data makes the utilities shift in much more extreme values.

Color	Choice Pair
	(x_{12}, x_{22})
	(x_{22}, x_{12})
	(x_{12}, x_{34})
	(x_{34}, x_{12})
	(x_{22}, x_{34})
	(x_{34}, x_{22})

Figure 12: Legend corresponding to Figure 13



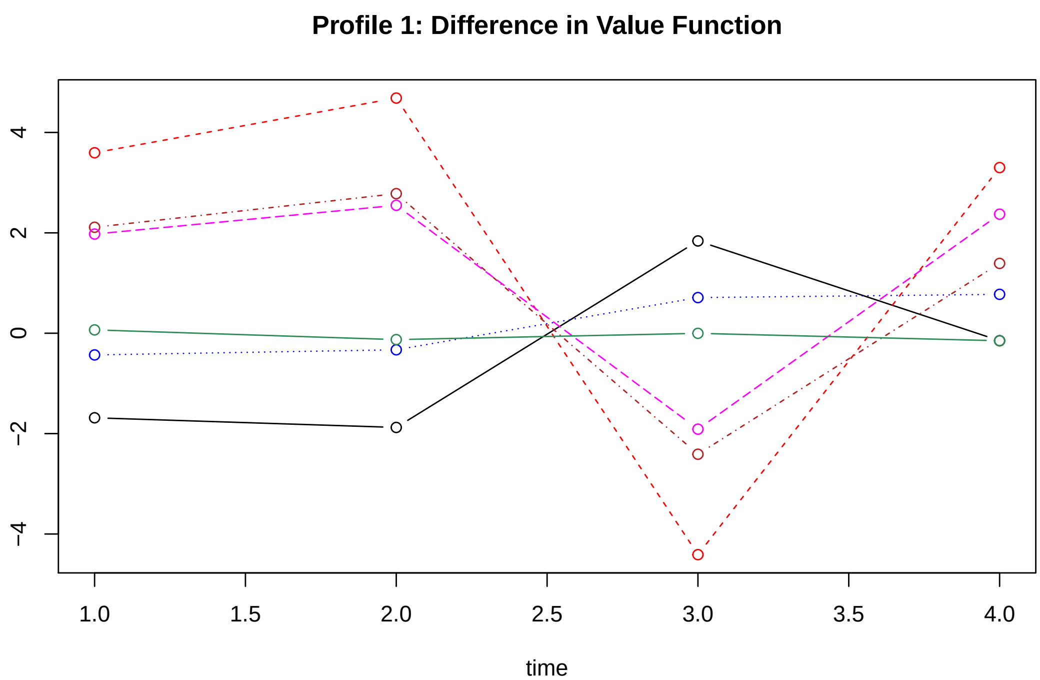


Figure 13: Expected discounted utility and their differences over time for Profile 1

5.2 AGGREGATED DATA EXAMPLE

We adapt our simulations of experiments to Flynn et al. (2007). The latter conducted an attribute-level best-worst scaling type of study to examine the quality of life of seniors. They considered a balanced design with five attributes (attachment, security, role, enjoyment, and control) each with four attribute levels (none, little, lot, and all) for attachment, security, and enjoyment and (none, few, many, all) for role and control. The attribute-levels are about the hypothetical quality of life states of 30 people of age 65 or more studied at one time. In their paper, they provide a partial look at their data and include the parameter estimates along with their standard errors. Since the data is not entirely published/available, using their parameter estimates, the data was generated under above rationale and simulations conducted. There are $K = 5$ attributes, and each attribute with $l_k = 4, \quad k = 1, 2, \ldots, 5,$ attribute-levels.

We use that study done by Flynn et al. (2007) as the basis for our simulations, considering a full factorial study, with a total of 1024 profiles. As mentioned in Street and Knox (2012), a full factorial design is often times costly and places an overwhelming choice task on the shoulders of the respondents. Therefore, an optimal partial factorial design, OMEP as in Chapter 3 was considered. In doing the computations in R, we utilized a package DoE.design. Using the oa.design option in that R package, a subset of 32 profiles was produced and used in the simulations based on a sample of $n = 100$ respondents.

Using the parameter estimates from Flynn et al. (2007), we generated data from that model. We first compare the model parameters with those obtained in Flynn et al. (2007). Using the model parameters, the utilities are estimated.

Attribute and attribute-level data in the experiments are a series of $1's$ and $0's$, indicating the attributes and attribute-levels in the choice pair. Looking Flynn et al. (2007), the attribute-level data when the l_i^{th} attribute-level is chosen as best the data is coded as -1 for the attribute-levels $1, 2, \ldots, l_i - 1$ and the data is coded as 1 for the attribute-levels $1, 2, \ldots, l_i - 1$ when it is the worst to satisfy the identifiability conditions. In fact Street and Burgess (2007) explained that for attribute-level point estimates, they satisfy the following identifiability condition:

$$\Sigma_{i=1}^{l_k} \beta_i = 0$$

or

$$\beta_{l_k} = -\Sigma_{j=1}^{l_i-1}\beta_j$$

for all $k = 1, 2, \ldots, K$.

The data was then exported from R into the SAS environment. Using the SAS procedure called MDC (multinomial discrete choice), the conditional logit model was fitted to the data. The parameter estimates from the Flynn et al. (2007) paper and from our simulated data are given in Table 15. The simulated data appears to be similar to their data based on how close the parameter estimates are. Our Akaike Information Criterion (AIC) statistic for the model is $16,384$; however, this provides little to no insight with regards to the fit of the model as mentioned in Flynn et al. (2010). The constant in the choice probability equation cancels out and has no impact on the simulation of the choices.

The probabilities to simulate choice behavior were computed using Equation (23). Using the original parameter estimates provided in Flynn et al. (2007), in Table 15, the values of $V_{ijj'}$ were computed. Lets consider the choice pair (Attachment None, Enjoyment Lot). From Table 15, the associated parameter estimates with this pair are given as,

$\hat{\beta}_{A_1} = 0.8142$ for Attachment,

$\hat{\beta}_{A_3} = 0.2842$ for Enjoyment,

$\hat{\beta}_{A_1x_1} = -1.8535$ for Attachment None,

and $\hat{\beta}_{A_3x_3} = 0.6844$ for Enjoyment Lot.

Referring to Equation (27) the value of $V_{ijj'}$ for this pair would be,

$$\begin{aligned}\hat{V}_{ijj'} &= exp(\hat{V}_{ij} - \hat{V}_{ij'}) = exp((\hat{\beta}_{A_j} + \hat{\beta}_{A_jx_j}) - (\hat{\beta}_{A_{j'}} + \hat{\beta}_{A_{j'}x_{j'}})\\ &= exp((0.8142 + 0.2842) - (-1.8535 + 0.6844)).\end{aligned}$$

Obtaining these values for all choice pairs the probabilities of choice selection were determined per profile and consumer choices were simulated.

Table 15: Parameter estimates from Flynn et al. (2007) paper and our simulation of their data

	Flynn et al.		*Simulated*	
Parameters	Estimates	SE	Estimates	SE
Constant	-0.3067	0.0750	0.0500	*
Attachment	0.8105	0.0803	0.8142	*
Security	*	*	*	*
Enjoyment	0.2632	0.1010	0.2842	0.0394
Role	0.1908	0.0974	0.1611	0.0400
Control	0.1076	0.0971	0.1148	0.0402
Attachment None	-1.9678	0.1129	-1.8535	0.0548
Attachment Little	0.1694	0.1012	0.1389	0.0532
Attachment Lot	0.9053	0.0905	0.9210	0.0561
Attachment All	0.8932	*	0.7936	*
Security None	-0.6123	0.1180	-0.6262	0.0541
Security Little	-0.3761	0.1302	-0.4077	0.0547
Security Lot	0.0373	0.1153	0.1027	0.0543
Security All	0.9511	*	0.9312	*
Enjoyment None	-0.8888	0.1286	-0.8166	0.0542
Enjoyment Little	-0.3367	0.1632	-0.3814	0.0544
Enjoyment Lot	0.6561	0.1493	0.6844	0.0548
Enjoyment All	0.5695	*	0.5136	*
Role None	-0.8956	0.1239	-0.8903	0.0546
Role Few	-0.0277	0.1532	-0.0079	0.0546
Role Many	0.4435	0.1363	0.4007	0.0546
Role All	0.4798	*	0.4975	*
Control None	-0.8085	0.1122	-0.7254	0.0546
Control Few	0.0835	0.1596	0.0755	0.0552
Control Many	0.2780	0.1376	0.2592	0.0543
Control All	0.4471	*	0.3907	*

From the parameter estimates provided in Table 15, we can determine the choice pairs with the highest and lowest utilities for the experiment. The choice pairs with the highest utilities are given in Table 16, and the pairs with the lowest utilities are given in Table 17. The pair (x_{13}, x_{51}) provides the greatest utility of any pair in the experiment as seen in Table 16. The pair (x_{13}, x_{51}) has the attribute Attachment and attribute-level Lot as the best and attribute Control and level None as the worst. Looking at the attribute and attribute-level impacts, or the parameter estimates, given in Table 15, this choice pair having the highest utility makes sense. We see that the attribute attachment has the largest impact in comparison to security, which was also noted in Flynn et al. (2007). The attribute with the smallest impact in comparison to security was control. Looking at the attribute-levels for these attributes, we see that the level lot for attachment has the largest positive impact, and attribute-level none for control has the largest negative impact.

Table 16: Choice pairs with the highest utility in the experiment

Best Attribute	Level	Worst Attribute	Level	Utility
1	3	5	1	8.9107
1	3	4	1	7.7977
1	4	5	1	7.2599
1	3	3	1	6.9108
1	4	4	1	6.6562
1	3	2	1	6.4402

Table 17: Choice pairs with the lowest utility in the experiment

Best Attribute	Level	Worst Attribute	Level	Utility
5	1	1	3	-4.3159
4	1	1	3	-4.1167
5	1	1	4	-3.9912
3	1	1	3	-3.9082
4	1	1	4	-3.8493
2	1	1	3	-3.7974

5.2.1 SCENARIO 1

For the simulated data of Flynn et al. (2007), we consider MDPs where the respondents are more likely to choose the same alternative at each time point. The transition parameters $\boldsymbol{\theta}^t_{s_i}$, where $s^t_i = (x_{ij}, x_{ij'})$ are defined as for the attributes as,

$$\theta^t_{s_iA_k} = \begin{cases} 3|\beta_{A_k}|, & \text{if } x_{ij} \in A_k, \\ -3|\beta_{A_k}|, & \text{if } x_{ij'} \in A_k, \\ \beta_{A_k}, & \text{otherwise}, \end{cases} \tag{70}$$

and for the attribute-levels,

$$\theta^t_{s_iA_kx_{ik}} = \begin{cases} 3|\beta_{A_kx_{ik}}|, & \text{if } x_{ij} = x_{ik} \text{ where } x_{ik} \in A_k, \\ -3|\beta_{A_kx_{ik}}|, & \text{if } x_{ij'} = x_{ik} \text{ where } x_{ik} \in A_k, \\ \beta_{A_kx_{ik}}, & \text{otherwise}, \end{cases} \tag{71}$$

where $j \neq j'$, $j, j', k = 1, 2, \ldots, K$, $1 \leq x_k \leq l_k$, and $i = 1, 2, \ldots, G$. The goal of this scenario was to design the transition probabilities in a way that the choice made at t is most likely to be made at $t+1$. If we considered $a_{s_im}(t) = \beta_m$ for $i = 1, 2, \ldots, G$, and $m = 1, 2, \ldots, p$, then the system would remain static and every row of the transition matrix would be the same. Recall that $p = K + \sum_{k=1}^{K} l_k = 25$ is the number of parameters. We consider $3|\beta_m|$ when a state or choice pair at time $t+1$ has the same best attribute and attribute-level as the state occupied at time t, and $-3|\beta_m|$ when a state or choice pair at time $t+1$ has the same worst attribute and attribute-level as the state occupied at time t. We consider $|\beta_m|$ to control the direction of the impact making sure it is positive for the best attribute and attribute-level of s_i and use $-|\beta_m|$ to make sure its negative for the worst attribute and attribute-level of s_i. We use 3 to increase the impact of the best and worst attributes and attribute-levels of s_i. The definition of $a_{s_im}(t)$ in this way insures that states with common best and worst attributes and attribute-levels as the present state occupied, $s^t_i = (x_{ij}, x_{ij'})$, have a greater probability of being transitioned to, where $i = 1, 2, \ldots, G$, $j \neq j'$, $j, j' = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$.

Table 16 reveals that attachment one of the most important preference for the DCEs. In fact it confirms the result obtained in Flynn et al. (2010). Since only one level of attachment is represented in each profile, we do not compare the attribute-levels with those found in Flynn et al. (2010). Flynn et al. (2010) mention that the opposite relative preference are for attachment and control. We see the same pattern in Tabel 16.

The value function/expected utilities for Profile 1 are displayed in Figure 15 along with the difference in value function. Choice pair (x_{12}, x_{51}), where x_{12} is the attribute-level little of attribute attachment is the best and x_{51} is the attribute-level none of attribute control is the worst, corresponds to the highest expected utility. This makes the simulation study a solution as mentioned in Flynn et al. (2007) in the best-worst studies of expected heterogeneity in respondents preferences.

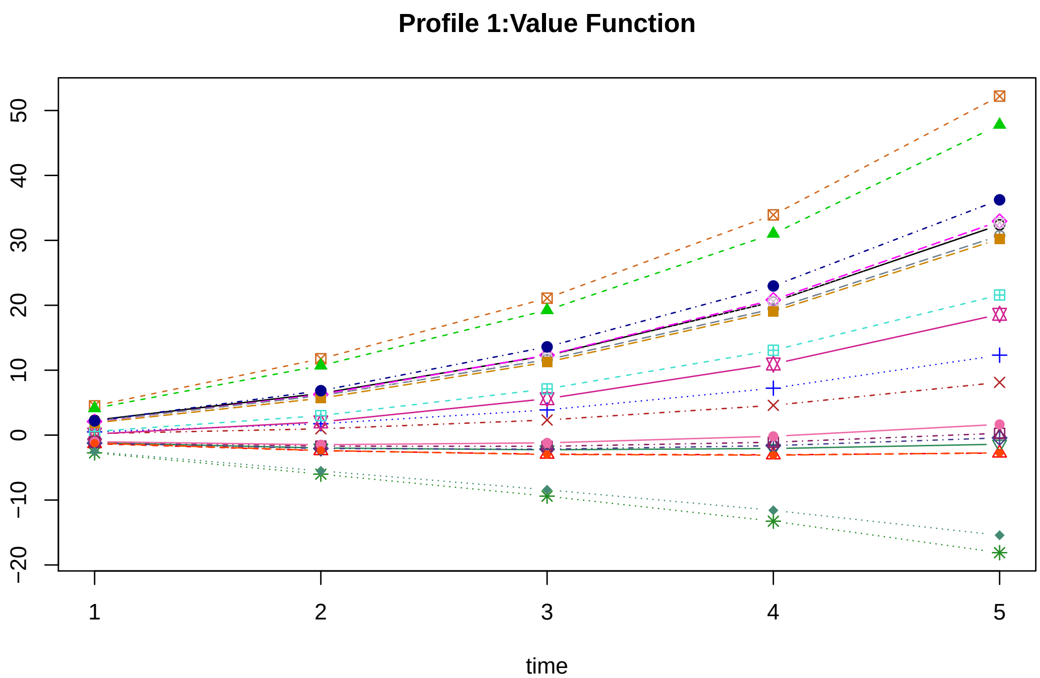
The value function/expected utilities for Profile 9 are displayed in Figure 17 along with the difference in value function. Choice pair (x_{34}, x_{11}), where x_{34} is the attribute-level all of attribute enjoyment is the best and x_{11} is the attribute-level none of attribute attachment is the worst, corresponds to the highest expected utility. In Profile 1, the attribute attachment as the best, and in Profile 9, attachment is the worst. As mentioned in Flynn et al. (2007), the levels for attachment have a greater range than the other attributes. This range of levels is how attachment is the best attribute in one profile and the worst in another.

Table 18: Stationary transition matrix in Scenario 1 for Profile 1

(x_{12}, x_{23})	0.518	0.000	0.038	0.001	0.081	0.000	0.192	0.000	0.000	0.069	0.001	0.032	0.002	0.014	0.011	0.002	0.026	0.001	0.012	0.002
(x_{23}, x_{12})	0.000	0.487	0.000	0.226	0.000	0.105	0.001	0.044	0.010	0.002	0.022	0.001	0.052	0.000	0.010	0.002	0.024	0.001	0.011	0.002
(x_{12}, x_{33})	0.025	0.000	0.703	0.000	0.037	0.000	0.088	0.000	0.065	0.000	0.003	0.002	0.008	0.001	0.000	0.043	0.000	0.018	0.005	0.001
(x_{33}, x_{12})	0.000	0.076	0.000	0.703	0.000	0.051	0.000	0.021	0.000	0.021	0.003	0.002	0.008	0.001	0.032	0.000	0.075	0.000	0.005	0.001
(x_{12}, x_{42})	0.091	0.001	0.063	0.001	0.331	0.000	0.032	0.000	0.006	0.012	0.031	0.002	0.030	0.002	0.044	0.002	0.043	0.002	0.008	0.009
(x_{42}, x_{12})	0.000	0.223	0.000	0.320	0.000	0.262	0.001	0.063	0.005	0.010	0.006	0.008	0.024	0.002	0.008	0.005	0.034	0.001	0.028	0.002
(x_{12}, x_{51})	0.025	0.000	0.018	0.000	0.038	0.000	0.701	0.000	0.002	0.003	0.004	0.002	0.065	0.000	0.005	0.001	0.093	0.000	0.043	0.000
(x_{51}, x_{12})	0.000	0.076	0.000	0.110	0.000	0.051	0.000	0.681	0.002	0.003	0.003	0.002	0.000	0.020	0.005	0.001	0.000	0.014	0.000	0.031
(x_{23}, x_{33})	0.003	0.011	0.229	0.000	0.012	0.002	0.029	0.001	0.469	0.000	0.025	0.001	0.058	0.001	0.000	0.101	0.001	0.043	0.013	0.002
(x_{33}, x_{23})	0.082	0.000	0.001	0.034	0.013	0.003	0.031	0.001	0.000	0.501	0.001	0.036	0.002	0.015	0.078	0.000	0.185	0.000	0.013	0.002
(x_{23}, x_{42})	0.009	0.038	0.020	0.018	0.103	0.003	0.100	0.003	0.040	0.009	0.211	0.002	0.206	0.002	0.098	0.004	0.095	0.004	0.018	0.019
(x_{42}, x_{23})	0.226	0.001	0.016	0.015	0.020	0.012	0.084	0.003	0.001	0.215	0.001	0.176	0.006	0.042	0.019	0.013	0.079	0.003	0.065	0.004
(x_{23}, x_{51})	0.002	0.010	0.005	0.005	0.011	0.002	0.206	0.000	0.011	0.002	0.023	0.001	0.422	0.000	0.011	0.002	0.196	0.000	0.091	0.000
(x_{51}, x_{23})	0.095	0.000	0.007	0.006	0.015	0.003	0.001	0.038	0.001	0.090	0.001	0.042	0.000	0.557	0.014	0.003	0.001	0.040	0.001	0.087
(x_{33}, x_{42})	0.014	0.006	0.002	0.058	0.053	0.002	0.051	0.002	0.001	0.088	0.035	0.003	0.034	0.003	0.319	0.000	0.312	0.000	0.009	0.010
(x_{42}, x_{33})	0.011	0.005	0.324	0.000	0.010	0.006	0.040	0.001	0.214	0.000	0.006	0.009	0.027	0.002	0.000	0.251	0.001	0.060	0.031	0.002
(x_{33}, x_{51})	0.004	0.002	0.000	0.016	0.006	0.001	0.110	0.000	0.000	0.024	0.004	0.002	0.073	0.000	0.036	0.000	0.672	0.000	0.049	0.000
(x_{51}, x_{33})	0.004	0.002	0.113	0.000	0.006	0.001	0.000	0.015	0.075	0.000	0.004	0.002	0.000	0.023	0.000	0.050	0.000	0.668	0.000	0.035
(x_{42}, x_{51})	0.010	0.004	0.007	0.006	0.008	0.005	0.272	0.000	0.00	0.009	0.00	0.008	0.179	0.000	0.008	0.005	0.258	0.000	0.211	0.000
(x_{51}, x_{42})	0.020	0.009	0.014	0.013	0.073	0.002	0.002	0.078	0.009	0.019	0.048	0.004	0.002	0.118	0.070	0.003	0.002	0.082	0.000	0.432

Color	Choice Pair	Color	Choice Pair
	(x_{12}, x_{23})		(x_{23}, x_{42})
	(x_{23}, x_{12})		(x_{42}, x_{23})
	(x_{12}, x_{33})		(x_{23}, x_{51})
	(x_{33}, x_{12})		(x_{51}, x_{23})
	(x_{12}, x_{42})		(x_{33}, x_{42})
	(x_{42}, x_{12})		(x_{42}, x_{33})
	(x_{12}, x_{51})		(x_{33}, x_{51})
	(x_{51}, x_{12})		(x_{51}, x_{33})
	(x_{23}, x_{33})		(x_{42}, x_{51})
	(x_{33}, x_{23})		(x_{51}, x_{42})

Figure 14: Legend corresponding to Figure 15.



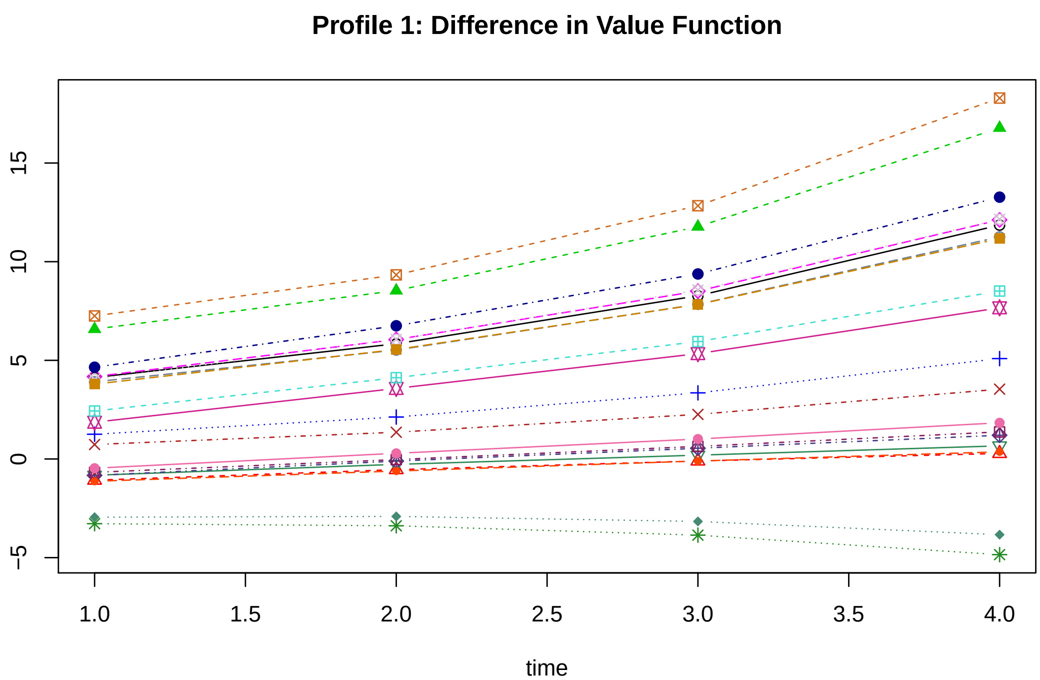


Figure 15: Expected discounted utility and their differences over time for the Profile 1

Color	Choice Pair	Color	Choice Pair
	(x_{14}, x_{24})		(x_{24}, x_{44})
	(x_{24}, x_{14})		(x_{44}, x_{24})
	(x_{14}, x_{34})		(x_{24}, x_{51})
	(x_{34}, x_{14})		(x_{51}, x_{24})
	(x_{14}, x_{44})		(x_{34}, x_{44})
	(x_{44}, x_{14})		(x_{44}, x_{34})
	(x_{14}, x_{51})		(x_{34}, x_{51})
	(x_{51}, x_{14})		(x_{51}, x_{34})
	(x_{24}, x_{34})		(x_{44}, x_{51})
	(x_{34}, x_{24})		(x_{51}, x_{44})

Figure 16: Legend corresponding to Figure 17

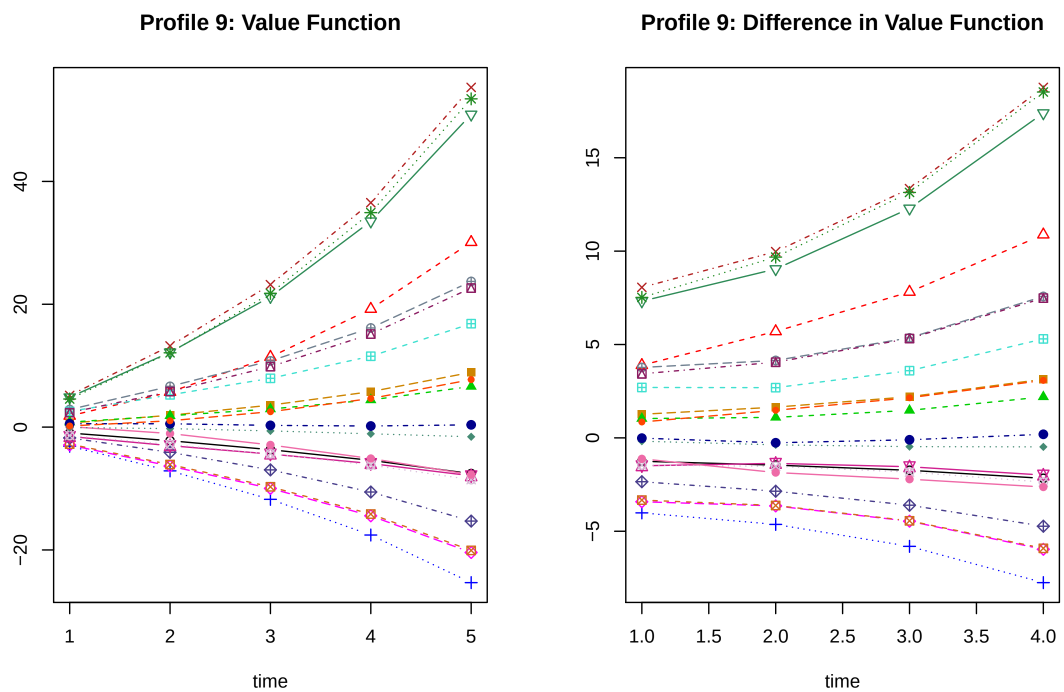


Figure 17: Expected discounted utility and their differences over time for the Profile 1

5.2.2 SCENARIO 2

In Scenario 2, respondents are allowed to make similar decisions at each time epoch with a different rate, making the transition probabilities dynamic. For the simulated data as in Flynn et al. (2007), we consider MDPs where in the respondents are more likely to choose the same alternative at each time point. The transition parameters $\boldsymbol{\theta}^t_{s_i}$ where $s^t_i = (x_{ij}, x_{ij'})$ are defined as for the attributes as,

$$\theta^t_{s_i A_k} = \begin{cases} 3^t|\beta_{A_k}|, & \text{if } x_{ij} \in A_k, \\ -3^t|\beta_{A_k}|, & \text{if } x_{ij'} \in A_k, \\ \beta_{A_k}, & \text{otherwise,} \end{cases} \tag{72}$$

and for the attribute-levels,

$$\theta^t_{s_i A_k x_{ik}} = \begin{cases} 3^t|\beta_{A_k x_{ik}}|, & \text{if } x_{ij} = x_{ik} \text{ where } x_{ik} \in A_k, \\ -3^t|\beta_{A_k x_{ik}}|, & \text{if } x_{ij'} = x_{ik} \text{ where } x_{ik} \in A_k, \\ \beta_{A_k x_{ik}}, & \text{otherwise,} \end{cases} \tag{73}$$

where $j \neq j'$, $j, j', k = 1, 2, \ldots, K$, $1 \leq x_k \leq l_k$, and $i = 1, 2, \ldots, G$.

The transition matrix at time $t = 1$ is kept the same as it was Scenario 1 in Table 18, and subsequent transition probabilities at time $t = 2, 3, 4$ are given in tables 19, 20, and 21, respectively. The same best-worst pair (x_{12}, x_{51}), where x_{12} is the attribute-level little of attribute attachment is the best and x_{51} is the attribute-level none of attribute control is the worst, corresponds to the highest expected utility as it was in Scenario 1 with a slight difference in the expected utilities. This is because the transition matrices are reinforcing those choices over time. This explains why the transition matrix under Table 21 is the identity matrix as expected under the trend in the utility. We also note a clustering in Figure 19 of the expected utilities into 5 groups. This membership in the utilities seems to better capture the estimates from the DCEs as was suggested in Flynn et al. (2010) when they considered gender in the quality of life study.

Table 19: Dynamic transition matrix in Scenario 2 for Profile 1 at time $t = 2$

(x_{12}, x_{23})	0.980	0.000	0.002	0.000	0.005	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{23}, x_{12})	0.000	0.974	0.000	0.015	0.000	0.007	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{12}, x_{33})	0.000	0.000	0.999	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{33}, x_{12})	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{12}, x_{42})	0.048	0.000	0.033	0.000	0.748	0.000	0.170	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{42}, x_{12})	0.000	0.129	0.000	0.185	0.000	0.650	0.000	0.036	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{12}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{51}, x_{12})	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.998	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{23}, x_{33})	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.973	0.000	0.000	0.000	0.001	0.000	0.000	0.007	0.000	0.003	0.000	0.000
(x_{33}, x_{23})	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.980	0.000	0.000	0.000	0.000	0.005	0.000	0.012	0.000	0.000	0.000
(x_{23}, x_{42})	0.000	0.031	0.001	0.001	0.012	0.000	0.003	0.000	0.033	0.000	0.737	0.000	0.167	0.000	0.011	0.000	0.003	0.000	0.000	0.002
(x_{42}, x_{23})	0.180	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.171	0.000	0.600	0.000	0.033	0.000	0.002	0.002	0.000	0.008	0.000
(x_{23}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.962	0.000	0.000	0.000	0.015	0.000	0.007	0.000
(x_{51}, x_{23})	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.000	0.000	0.989	0.000	0.000	0.000	0.002	0.000	0.005
(x_{33}, x_{42})	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.749	0.000	0.170	0.000	0.000	0.000
(x_{42}, x_{33})	0.000	0.000	0.193	0.000	0.000	0.000	0.000	0.000	0.127	0.000	0.000	0.000	0.000	0.000	0.000	0.643	0.000	0.036	0.000	0.000
(x_{33}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.998	0.000	0.000	0.000
(x_{51}, x_{33})	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.998	0.000	0.000
(x_{42}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.168	0.000	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.000	0.160	0.000	0.561	0.000
(x_{51}, x_{42})	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.037	0.000	0.000	0.000	0.000	0.000	0.055	0.000	0.000	0.000	0.039	0.000	0.868

Table 20: Dynamic transition matrix in Scenario 2 for Profile 1 at time $t = 3$

(x_{12}, x_{23})	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{23}, x_{12})	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{12}, x_{33})	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{33}, x_{12})	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{12}, x_{42})	0.001	0.000	0.000	0.000	0.996	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{42}, x_{12})	0.000	0.002	0.000	0.004	0.000	0.993	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{12}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{51}, x_{12})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{23}, x_{33})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{33}, x_{23})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{23}, x_{42})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.996	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{42}, x_{23})	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.992	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{23}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{51}, x_{23})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
(x_{33}, x_{42})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.996	0.000	0.003	0.000	0.000	0.000
(x_{42}, x_{33})	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.993	0.000	0.001	0.000	0.000
(x_{33}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
(x_{51}, x_{33})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
(x_{42}, x_{51})	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.004	0.000	0.990	0.000
(x_{51}, x_{42})	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.998

Table 21: Dynamic transition matrix in Scenario 2 for Profile 1 at time $t = 4$

(x_{12}, x_{23})	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{23}, x_{12})	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{12}, x_{33})	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{33}, x_{12})	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{12}, x_{42})	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{42}, x_{12})	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{12}, x_{51})	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
(x_{51}, x_{12})	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
(x_{23}, x_{33})	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
(x_{33}, x_{23})	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
(x_{23}, x_{42})	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
(x_{42}, x_{23})	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
(x_{23}, x_{51})	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
(x_{51}, x_{23})	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
(x_{33}, x_{42})	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
(x_{42}, x_{33})	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
(x_{33}, x_{51})	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
(x_{51}, x_{33})	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
(x_{42}, x_{51})	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
(x_{51}, x_{42})	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Color	Choice Pair	Color	Choice Pair
	(x_{12}, x_{23})		(x_{23}, x_{42})
	(x_{23}, x_{12})		(x_{42}, x_{23})
	(x_{12}, x_{33})		(x_{23}, x_{51})
	(x_{33}, x_{12})		(x_{51}, x_{23})
	(x_{12}, x_{42})		(x_{33}, x_{42})
	(x_{42}, x_{12})		(x_{42}, x_{33})
	(x_{12}, x_{51})		(x_{33}, x_{51})
	(x_{51}, x_{12})		(x_{51}, x_{33})
	(x_{23}, x_{33})		(x_{42}, x_{51})
	(x_{33}, x_{23})		(x_{51}, x_{42})

Figure 18: Legend corresponding to Figure 19

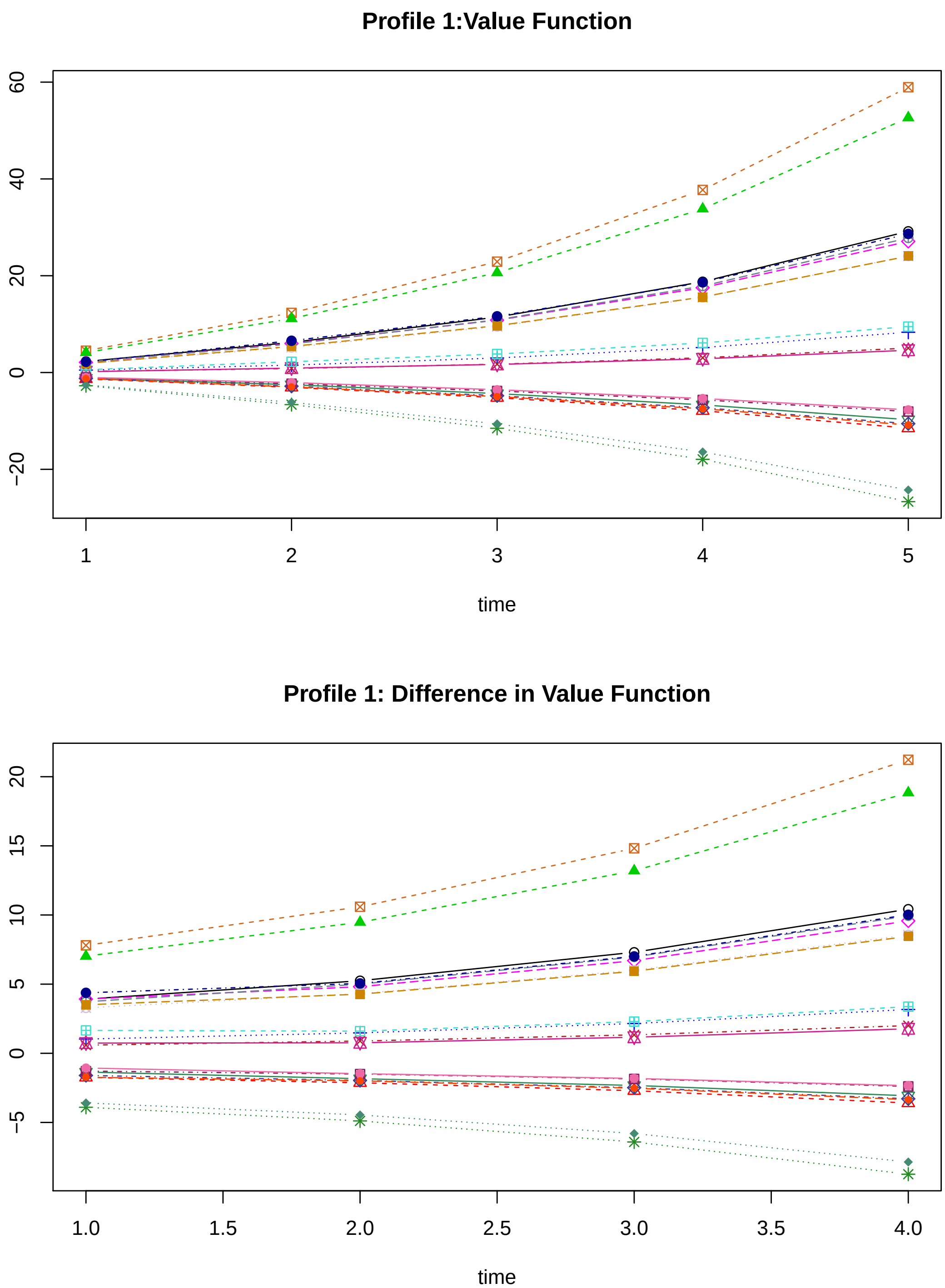


Figure 19: Expected discounted utility and their differences over time for the Profile 1

5.2.3 SCENARIO 3

In this scenario, respondents are allowed to make different decisions under a stationary transition probability. Among the probabilities that have noticeable changes in transition are displayed in 22. Respondents who would consider attachment the best attribute are now deflecting it with the rate of 0.3 and those that used to consider the attribute-level control all as the best is inflated at a rate of 2.1. Similar changes can be seen in 22. We provide the transition matrix for Profile 1 built under this consideration in Table 23.

The value function/expected utilities for Profile 1 are displayed in Figure 21 along with the difference in value function. The same best-worst pair (x_{12}, x_{51}), where x_{12} is the attribute-level little of attribute attachment is the best and x_{51} is the attribute-level none of attribute control is the worst, corresponds to the highest expected utility as it was in Scenario 2 with a slight difference in the expected utilities. The clustering is much more perceptible.

Table 22: Transition parameters for Scenario 3

Parameters	(x_{12}, x_{33})	(x_{12}, x_{31})	(x_{12}, x_{41})	(x_{12}, x_{51})	(x_{33}, x_{51})	(x_{43}, x_{11})	(x_{54}, x_{11})
θ_1	$0.3\hat{\beta}_1$	$0.3\hat{\beta}_1$	$0.3\hat{\beta}_1$	$0.3\hat{\beta}_1$	$0.85\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$
θ_2	$0.85\hat{\beta}_2$	$0.85\hat{\beta}_2$	$0.85\hat{\beta}_2$	$0.85\hat{\beta}_2$	$0.85\hat{\beta}_2$	$0.85\hat{\beta}_2$	$0.85\hat{\beta}_2$
θ_3	$\hat{\beta}_3$	$\hat{\beta}_3$	$0.85\hat{\beta}_3$	$0.85\hat{\beta}_3$	$0.95\hat{\beta}_3$	$0.85\hat{\beta}_3$	$0.85\hat{\beta}_3$
θ_4	$0.85\hat{\beta}_4$	$0.85\hat{\beta}_4$	$\hat{\beta}_4$	$0.85\hat{\beta}_4$	$0.85\hat{\beta}_4$	$0.95\hat{\beta}_4$	$0.85\hat{\beta}_4$
θ_5	$0.85\hat{\beta}_5$	$0.85\hat{\beta}_5$	$0.85\hat{\beta}_5$	$\hat{\beta}_5$	$\hat{\beta}_5$	$0.85\hat{\beta}_5$	$0.95\hat{\beta}_5$
θ_{11}	$0.85\hat{\beta}_{11}$	$0.85\hat{\beta}_{11}$	$0.85\hat{\beta}_{11}$	$0.85\hat{\beta}_{11}$	$0.85\hat{\beta}_{11}$	$0.9\hat{\beta}_{11}$	$0.9\hat{\beta}_{11}$
θ_{12}	$0.7\hat{\beta}_{12}$	$0.7\hat{\beta}_{12}$	$0.7\hat{\beta}_{12}$	$0.7\hat{\beta}_{12}$	$0.85\hat{\beta}_{12}$	$0.85\hat{\beta}_{12}$	$0.85\hat{\beta}_{12}$
θ_{13}	$0.85\hat{\beta}_{13}$	$0.85\hat{\beta}_{13}$	$0.85\hat{\beta}_{13}$	$0.85\hat{\beta}_{13}$	$0.85\hat{\beta}_{13}$	$0.85\hat{\beta}_{13}$	$0.85\hat{\beta}_{13}$
θ_{14}	$0.85\hat{\beta}_{14}$	$0.85\hat{\beta}_{14}$	$0.85\hat{\beta}_{14}$	$0.85\hat{\beta}_{14}$	$0.85\hat{\beta}_{14}$	$0.85\hat{\beta}_{14}$	$0.85\hat{\beta}_{14}$
θ_{21}	$0.85\hat{\beta}_{21}$	$0.85\hat{\beta}_{21}$	$0.85\hat{\beta}_{21}$	$0.85\hat{\beta}_{21}$	$0.85\hat{\beta}_{21}$	$0.85\hat{\beta}_{21}$	$0.85\hat{\beta}_{21}$
θ_{22}	$0.85\hat{\beta}_{22}$	$0.85\hat{\beta}_{22}$	$0.85\hat{\beta}_{22}$	$0.85\hat{\beta}_{22}$	$0.85\hat{\beta}_{22}$	$0.85\hat{\beta}_{22}$	$0.85\hat{\beta}_{22}$
θ_{23}	$0.85\hat{\beta}_{23}$	$0.85\hat{\beta}_{23}$	$0.85\hat{\beta}_{23}$	$0.85\hat{\beta}_{23}$	$0.85\hat{\beta}_{23}$	$0.85\hat{\beta}_{23}$	$0.85\hat{\beta}_{23}$
θ_{24}	$-0.80\hat{\beta}_{24}$	$-0.80\hat{\beta}_{24}$	$-0.80\hat{\beta}_{24}$	$-0.80\hat{\beta}_{24}$	$-0.80\hat{\beta}_{24}$	$-0.80\hat{\beta}_{24}$	$-0.80\hat{\beta}_{24}$
θ_{31}	$0.85\hat{\beta}_{31}$	$0.50\hat{\beta}_{31}$	$0.85\hat{\beta}_{31}$	$0.85\hat{\beta}_{31}$	$0.85\hat{\beta}_{31}$	$0.85\hat{\beta}_{31}$	$0.85\hat{\beta}_{31}$
θ_{32}	$0.85\hat{\beta}_{32}$	$0.85\hat{\beta}_{32}$	$0.85\hat{\beta}_{32}$	$0.85\hat{\beta}_{32}$	$0.85\hat{\beta}_{32}$	$0.85\hat{\beta}_{32}$	$0.85\hat{\beta}_{32}$
θ_{33}	$0.9\hat{\beta}_{33}$	$0.85\hat{\beta}_{33}$	$0.85\hat{\beta}_{33}$	$0.85\hat{\beta}_{33}$	$1.50\hat{\beta}_{33}$	$0.85\hat{\beta}_{33}$	$0.85\hat{\beta}_{33}$
θ_{34}	$0.85\hat{\beta}_{34}$	$0.85\hat{\beta}_{34}$	$0.85\hat{\beta}_{34}$	$0.85\hat{\beta}_{34}$	$0.85\hat{\beta}_{34}$	$0.85\hat{\beta}_{34}$	$0.85\hat{\beta}_{34}$
θ_{41}	$0.85\hat{\beta}_{41}$	$0.85\hat{\beta}_{41}$	$1.70\hat{\beta}_{41}$	$0.85\hat{\beta}_{41}$	$0.85\hat{\beta}_{41}$	$0.85\hat{\beta}_{41}$	$0.85\hat{\beta}_{41}$
θ_{42}	$0.85\hat{\beta}_{42}$	$0.85\hat{\beta}_{42}$	$0.85\hat{\beta}_{42}$	$0.85\hat{\beta}_{42}$	$0.85\hat{\beta}_{42}$	$0.85\hat{\beta}_{42}$	$0.85\hat{\beta}_{42}$
θ_{43}	$0.85\hat{\beta}_{43}$	$0.85\hat{\beta}_{43}$	$0.85\hat{\beta}_{43}$	$0.85\hat{\beta}_{43}$	$0.85\hat{\beta}_{43}$	$0.70\hat{\beta}_{43}$	$0.85\hat{\beta}_{43}$
θ_{44}	$0.85\hat{\beta}_{44}$	$0.85\hat{\beta}_{44}$	$0.85\hat{\beta}_{44}$	$0.85\hat{\beta}_{44}$	$0.85\hat{\beta}_{44}$	$0.85\hat{\beta}_{44}$	$0.85\hat{\beta}_{44}$
θ_{51}	$0.85\hat{\beta}_{51}$	$0.85\hat{\beta}_{51}$	$0.85\hat{\beta}_{51}$	$0.70\hat{\beta}_{51}$	$0.70\hat{\beta}_{51}$	$0.85\hat{\beta}_{51}$	$0.85\hat{\beta}_{51}$
θ_{52}	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$
θ_{52}	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$	$0.85\hat{\beta}_{52}$
θ_{53}	$0.85\hat{\beta}_{53}$	$0.85\hat{\beta}_{53}$	$0.85\hat{\beta}_{53}$	$0.85\hat{\beta}_{53}$	$0.85\hat{\beta}_{53}$	$0.85\hat{\beta}_{53}$	$0.85\hat{\beta}_{53}$
θ_{54}	$0.85\hat{\beta}_{54}$	$0.85\hat{\beta}_{54}$	$0.85\hat{\beta}_{54}$	$0.85\hat{\beta}_{54}$	$0.85\hat{\beta}_{54}$	$0.85\hat{\beta}_{54}$	$2.1\hat{\beta}_{54}$

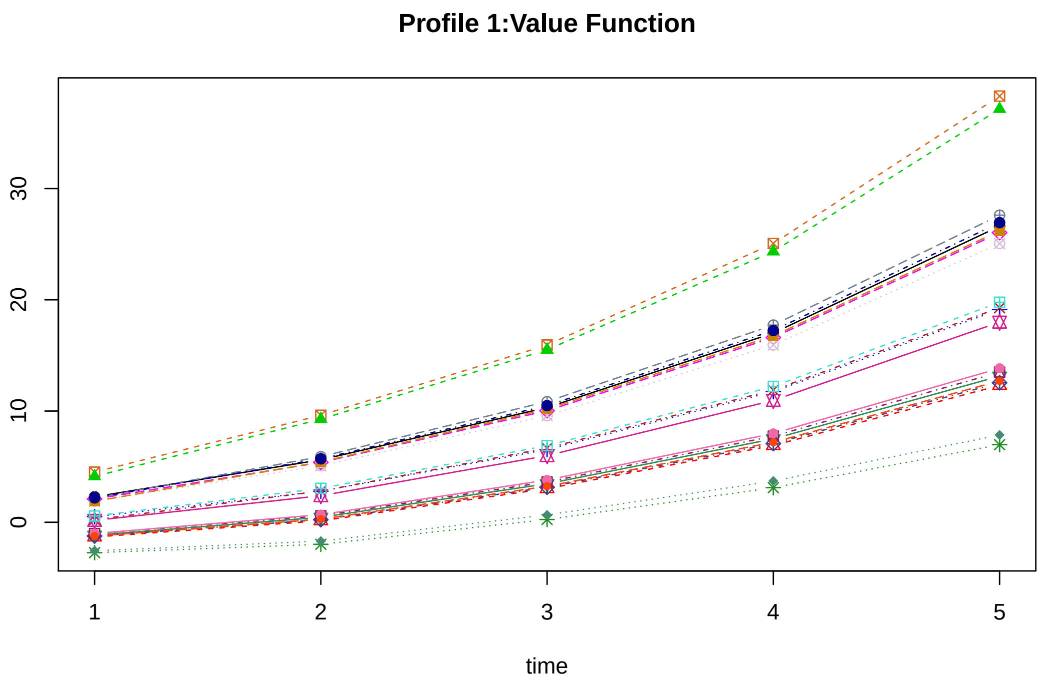
Note: Security has no parameter estimate as it was the baseline category in the conditional logit model. For the purpose of the transition matrix, we take $\hat{\beta}_2$ to be 0.5.

Table 23: Stationary transition matrix in Scenario 3 for Profile 1

(x_{12}, x_{23})	0.032	0.046	0.025	0.059	0.048	0.031	0.099	0.015	0.030	0.049	0.057	0.026	0.119	0.012	0.073	0.020	0.153	0.010	0.080	0.018
(x_{23}, x_{12})	0.052	0.023	0.041	0.029	0.079	0.015	0.165	0.007	0.028	0.043	0.053	0.023	0.110	0.011	0.067	0.018	0.139	0.009	0.072	0.017
(x_{12}, x_{33})	0.052	0.028	0.023	0.063	0.048	0.031	0.099	0.015	0.017	0.086	0.035	0.041	0.074	0.020	0.079	0.019	0.165	0.009	0.080	0.018
(x_{33}, x_{12})	0.075	0.012	0.023	0.040	0.069	0.013	0.144	0.006	0.009	0.099	0.028	0.033	0.058	0.016	0.092	0.010	0.192	0.005	0.063	0.014
(x_{12}, x_{42})	0.053	0.029	0.025	0.060	0.050	0.031	0.102	0.015	0.019	0.081	0.037	0.041	0.075	0.020	0.077	0.019	0.156	0.009	0.079	0.019
(x_{42}, x_{12})	0.086	0.014	0.041	0.029	0.078	0.015	0.166	0.007	0.017	0.072	0.031	0.038	0.067	0.018	0.065	0.018	0.139	0.009	0.074	0.016
(x_{12}, x_{51})	0.055	0.030	0.027	0.063	0.051	0.033	0.093	0.018	0.020	0.085	0.038	0.044	0.069	0.024	0.079	0.021	0.143	0.012	0.074	0.023
(x_{51}, x_{12})	0.085	0.014	0.041	0.029	0.078	0.015	0.017	0.007	0.016	0.071	0.032	0.037	0.068	0.017	0.066	0.018	0.141	0.008	0.073	0.016
(x_{23}, x_{33})	0.046	0.027	0.034	0.036	0.071	0.018	0.147	0.008	0.026	0.047	0.054	0.023	0.112	0.011	0.073	0.017	0.151	0.008	0.073	0.017
(x_{33}, x_{23})	0.042	0.023	0.021	0.048	0.063	0.016	0.132	0.008	0.016	0.064	0.042	0.021	0.098	0.010	0.096	0.010	0.200	0.005	0.066	0.015
(x_{23}, x_{42})	0.047	0.027	0.037	0.034	0.073	0.017	0.149	0.009	0.028	0.045	0.056	0.023	0.114	0.011	0.070	0.018	0.143	0.009	0.072	0.018
(x_{42}, x_{23})	0.048	0.027	0.038	0.034	0.070	0.018	0.15	0.009	0.028	0.046	0.053	0.024	0.112	0.012	0.067	0.019	0.144	0.009	0.076	0.017
(x_{23}, x_{51})	0.050	0.029	0.039	0.036	0.076	0.019	0.138	0.010	0.030	0.047	0.058	0.025	0.105	0.014	0.073	0.020	0.132	0.011	0.069	0.021
(x_{51}, x_{23})	0.047	0.026	0.037	0.034	0.071	0.018	0.153	0.008	0.028	0.045	0.053	0.024	0.114	0.011	0.068	0.018	0.146	0.009	0.076	0.016
(x_{33}, x_{42})	0.067	0.014	0.020	0.047	0.064	0.015	0.130	0.007	0.009	0.102	0.029	0.033	0.060	0.016	0.097	0.010	0.197	0.005	0.063	0.015
(x_{42}, x_{33})	0.077	0.016	0.034	0.036	0.069	0.018	0.148	0.008	0.016	0.079	0.032	0.039	0.068	0.018	0.071	0.018	0.015	0.008	0.075	0.017
(x_{33}, x_{51})	0.071	0.015	0.022	0.050	0.066	0.016	0.119	0.009	0.010	0.108	0.030	0.035	0.055	0.020	0.100	0.011	0.181	0.006	0.059	0.018
(x_{51}, x_{33})	0.076	0.016	0.034	0.036	0.070	0.017	0.151	0.008	0.016	0.078	0.032	0.038	0.069	0.018	0.072	0.017	0.155	0.008	0.075	0.016
(x_{42}, x_{51})	0.082	0.017	0.040	0.036	0.074	0.019	0.138	0.010	0.018	0.079	0.034	0.041	0.063	0.023	0.071	0.020	0.132	0.011	0.070	0.020
(x_{51}, x_{42})	0.077	0.016	0.037	0.034	0.073	0.017	0.153	0.008	0.017	0.074	0.034	0.037	0.070	0.018	0.070	0.018	0.146	0.009	0.074	0.017

Color	Choice Pair	Color	Choice Pair
	(x_{12}, x_{23})		(x_{23}, x_{42})
	(x_{23}, x_{12})		(x_{42}, x_{23})
	(x_{12}, x_{33})		(x_{23}, x_{51})
	(x_{33}, x_{12})		(x_{51}, x_{23})
	(x_{12}, x_{42})		(x_{33}, x_{42})
	(x_{42}, x_{12})		(x_{42}, x_{33})
	(x_{12}, x_{51})		(x_{33}, x_{51})
	(x_{51}, x_{12})		(x_{51}, x_{33})
	(x_{23}, x_{33})		(x_{42}, x_{51})
	(x_{33}, x_{23})		(x_{51}, x_{42})

Figure 20: Legend corresponding to Figure 21



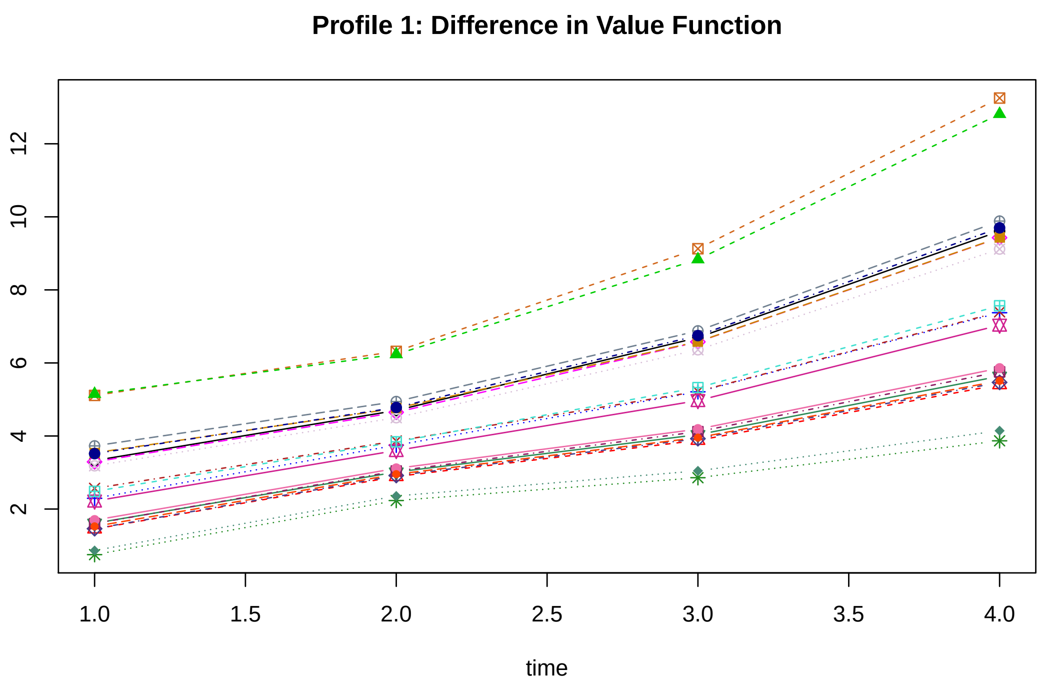


Figure 21: Expected discounted utility and their differences over time for the Profile 1

5.2.4 SCENARIO 4

Scenario 4 is an extension of Scenario 3 in the sense that the coefficient associated with the transition probabilities are dynamic. In fact, we exponentiate the coefficient to the t^{th} power. So, respondents are allowed to make different decisions under a dynamic transition probabilities. Among the probabilities that have the noticeable changes in the transition probabilities are given in Table 24.

In this scenario, respondents are allowed to make different decisions under a dynamic transition probability. Among the probabilities that have noticeable changes in transition are displayed in 24. Respondents who would consider attachment the best attribute are now deflecting it with the rate of 0.3^t and those that used to consider the attribute-level control all as the best is inflated at a rate of 2.1^t. Similar changes can be seen in 24. We provide the transition matrix for Profile 1 at time $t = 2, 3, 4$ built under this consideration in tables 25, 26, and 27, respectively.

The value function/expected utilities for Profile 1 are displayed in Figure 23 along with the difference in the value function over time. The pair that started as the best (x_{12}, x_{51}) , where x_{12} is the attribute-level little of attribute attachment is the best and x_{51} is the attribute-level none of attribute control is the worst, has converted to (x_{33}, x_{51}), where x_{33} is the attribute level lot of attribute enjoyment. The clustering in expected utilities is much more perceptible.

Table 24: Transition parameters for Scenario 4

Parameters	(x_{12}, x_{33})	(x_{12}, x_{31})	(x_{12}, x_{41})	(x_{12}, x_{51})	(x_{33}, x_{51})	(x_{43}, x_{11})	(x_{54}, x_{11})
θ_1	$0.3^t\hat{\beta}_1$	$0.3^t\hat{\beta}_1$	$0.3^t\hat{\beta}_1$	$0.3^t\hat{\beta}_1$	$0.85^t\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_1$
θ_2	$0.85^t\hat{\beta}_2$	$0.85^t\hat{\beta}_2$	$0.85^t\hat{\beta}_2$	$0.85^t\hat{\beta}_2$	$0.85^t\hat{\beta}_2$	$0.85^t\hat{\beta}_2$	$0.85^t\hat{\beta}_2$
θ_3	$\hat{\beta}_3$	$\hat{\beta}_3$	$0.85^t\hat{\beta}_3$	$0.85^t\hat{\beta}_3$	$0.95^t\hat{\beta}_3$	$0.85^t\hat{\beta}_3$	$0.85^t\hat{\beta}_3$
θ_4	$0.85^t\hat{\beta}_4$	$0.85^t\hat{\beta}_4$	$\hat{\beta}_4$	$0.85^t\hat{\beta}_4$	$0.85^t\hat{\beta}_4$	$0.95^t\hat{\beta}_4$	$0.85^t\hat{\beta}_4$
θ_5	$0.85^t\hat{\beta}_5$	$0.85^t\hat{\beta}_5$	$0.85^t\hat{\beta}_5$	$\hat{\beta}_5$	$\hat{\beta}_5$	$0.85^t\hat{\beta}_5$	$0.95^t\hat{\beta}_5$
θ_{11}	$0.85^t\hat{\beta}_{11}$	$0.85^t\hat{\beta}_{11}$	$0.85^t\hat{\beta}_{11}$	$0.85^t\hat{\beta}_{11}$	$0.85^t\hat{\beta}_{11}$	$0.9^t\hat{\beta}_{11}$	$0.9^t\hat{\beta}_{11}$
θ_{12}	$0.7^t\hat{\beta}_{12}$	$0.7^t\hat{\beta}_{12}$	$0.7^t\hat{\beta}_{12}$	$0.7^t\hat{\beta}_{12}$	$0.85^t\hat{\beta}_{12}$	$0.85^t\hat{\beta}_{12}$	$0.85^t\hat{\beta}_{12}$
θ_{13}	$0.85^t\hat{\beta}_{13}$	$0.85^t\hat{\beta}_{13}$	$0.85^t\hat{\beta}_{13}$	$0.85^t\hat{\beta}_{13}$	$0.85^t\hat{\beta}_{13}$	$0.85^t\hat{\beta}_{13}$	$0.85^t\hat{\beta}_{13}$
θ_{14}	$0.85^t\hat{\beta}_{14}$	$0.85^t\hat{\beta}_{14}$	$0.85^t\hat{\beta}_{14}$	$0.85^t\hat{\beta}_{14}$	$0.85^t\hat{\beta}_{14}$	$0.85^t\hat{\beta}_{14}$	$0.85^t\hat{\beta}_{14}$
θ_{21}	$0.85^t\hat{\beta}_{21}$	$0.85^t\hat{\beta}_{21}$	$0.85^t\hat{\beta}_{21}$	$0.85^t\hat{\beta}_{21}$	$0.85^t\hat{\beta}_{21}$	$0.85^t\hat{\beta}_{21}$	$0.85^t\hat{\beta}_{21}$
θ_{22}	$0.85^t\hat{\beta}_{22}$	$0.85^t\hat{\beta}_{22}$	$0.85^t\hat{\beta}_{22}$	$0.85^t\hat{\beta}_{22}$	$0.85^t\hat{\beta}_{22}$	$0.85^t\hat{\beta}_{22}$	$0.85^t\hat{\beta}_{22}$
θ_{23}	$0.85^t\hat{\beta}_{23}$	$0.85^t\hat{\beta}_{23}$	$0.85^t\hat{\beta}_{23}$	$0.85^t\hat{\beta}_{23}$	$0.85^t\hat{\beta}_{23}$	$0.85^t\hat{\beta}_{23}$	$0.85^t\hat{\beta}_{23}$
θ_{24}	$-0.80^t\hat{\beta}_{24}$	$-0.80^t\hat{\beta}_{24}$	$-0.80^t\hat{\beta}_{24}$	$-0.80^t\hat{\beta}_{24}$	$-0.80^t\hat{\beta}_{24}$	$-0.80^t\hat{\beta}_{24}$	$-0.80^t\hat{\beta}_{24}$
θ_{31}	$0.85^t\hat{\beta}_{31}$	$0.50^t\hat{\beta}_{31}$	$0.85^t\hat{\beta}_{31}$	$0.85^t\hat{\beta}_{31}$	$0.85^t\hat{\beta}_{31}$	$0.85^t\hat{\beta}_{31}$	$0.85^t\hat{\beta}_{31}$
θ_{32}	$0.85^t\hat{\beta}_{32}$	$0.85^t\hat{\beta}_{32}$	$0.85^t\hat{\beta}_{32}$	$0.85^t\hat{\beta}_{32}$	$0.85^t\hat{\beta}_{32}$	$0.85^t\hat{\beta}_{32}$	$0.85^t\hat{\beta}_{32}$
θ_{33}	$0.9^t\hat{\beta}_{33}$	$0.85^t\hat{\beta}_{33}$	$0.85^t\hat{\beta}_{33}$	$0.85^t\hat{\beta}_{33}$	$1.50^t\hat{\beta}_{33}$	$0.85^t\hat{\beta}_{33}$	$0.85^t\hat{\beta}_{33}$
θ_{34}	$0.85^t\hat{\beta}_{34}$	$0.85^t\hat{\beta}_{34}$	$0.85^t\hat{\beta}_{34}$	$0.85^t\hat{\beta}_{34}$	$0.85^t\hat{\beta}_{34}$	$0.85^t\hat{\beta}_{34}$	$0.85^t\hat{\beta}_{34}$
θ_{41}	$0.85^t\hat{\beta}_{41}$	$0.85^t\hat{\beta}_{41}$	$1.70^t\hat{\beta}_{41}$	$0.85^t\hat{\beta}_{41}$	$0.85^t\hat{\beta}_{41}$	$0.85^t\hat{\beta}_{41}$	$0.85^t\hat{\beta}_{41}$
θ_{42}	$0.85^t\hat{\beta}_{42}$	$0.85^t\hat{\beta}_{42}$	$0.85^t\hat{\beta}_{42}$	$0.85^t\hat{\beta}_{42}$	$0.85^t\hat{\beta}_{42}$	$0.85^t\hat{\beta}_{42}$	$0.85^t\hat{\beta}_{42}$
θ_{43}	$0.85^t\hat{\beta}_{43}$	$0.85^t\hat{\beta}_{43}$	$0.85^t\hat{\beta}_{43}$	$0.85^t\hat{\beta}_{43}$	$0.85^t\hat{\beta}_{43}$	$0.70^t\hat{\beta}_{43}$	$0.85^t\hat{\beta}_{43}$
θ_{44}	$0.85^t\hat{\beta}_{44}$	$0.85^t\hat{\beta}_{44}$	$0.85^t\hat{\beta}_{44}$	$0.85^t\hat{\beta}_{44}$	$0.85^t\hat{\beta}_{44}$	$0.85^t\hat{\beta}_{44}$	$0.85^t\hat{\beta}_{44}$
θ_{51}	$0.85^t\hat{\beta}_{51}$	$0.85^t\hat{\beta}_{51}$	$0.85^t\hat{\beta}_{51}$	$0.70^t\hat{\beta}_{51}$	$0.70^t\hat{\beta}_{51}$	$0.85^t\hat{\beta}_{51}$	$0.85^t\hat{\beta}_{51}$
θ_{52}	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$
θ_{52}	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$	$0.85^t\hat{\beta}_{52}$
θ_{53}	$0.85^t\hat{\beta}_{53}$	$0.85^t\hat{\beta}_{53}$	$0.85^t\hat{\beta}_{53}$	$0.85^t\hat{\beta}_{53}$	$0.85^t\hat{\beta}_{53}$	$0.85^t\hat{\beta}_{53}$	$0.85^t\hat{\beta}_{53}$
θ_{54}	$0.85^t\hat{\beta}_{54}$	$0.85^t\hat{\beta}_{54}$	$0.85^t\hat{\beta}_{54}$	$0.85^t\hat{\beta}_{54}$	$0.85^t\hat{\beta}_{54}$	$0.85^t\hat{\beta}_{54}$	$2.1^t\hat{\beta}_{53}$

Note: Security has no parameter estimate as it was the baseline category in the conditional logit model. For the purpose of the transition matrix, we take $\hat{\beta}_2$ to be 0.5.

Table 25: Dynamic transition matrix in Scenario 4 at $t = 2$ for Profile 1

(x_{12}, x_{23})	0.028	0.059	0.024	0.069	0.042	0.039	0.079	0.021	0.035	0.048	0.061	0.027	0.114	0.015	0.071	0.023	0.133	0.013	0.076	0.022
(x_{23}, x_{12})	0.054	0.025	0.049	0.028	0.085	0.016	0.158	0.009	0.033	0.041	0.057	0.024	0.107	0.013	0.064	0.021	0.119	0.011	0.069	0.020
(x_{12}, x_{33})	0.045	0.037	0.021	0.078	0.042	0.039	0.078	0.021	0.019	0.086	0.038	0.043	0.071	0.023	0.080	0.020	0.150	0.011	0.076	0.022
(x_{33}, x_{12})	0.062	0.010	0.011	0.055	0.058	0.011	0.108	0.006	0.005	0.136	0.023	0.027	0.044	0.014	0.126	0.005	0.237	0.003	0.047	0.013
(x_{12}, x_{42})	0.047	0.038	0.025	0.071	0.049	0.036	0.081	0.022	0.023	0.079	0.044	0.040	0.073	0.024	0.082	0.022	0.137	0.013	0.070	0.025
(x_{42}, x_{12})	0.091	0.015	0.049	0.028	0.082	0.017	0.159	0.009	0.020	0.069	0.033	0.041	0.064	0.021	0.062	0.022	0.120	0.011	0.072	0.019
(x_{12}, x_{51})	0.049	0.040	0.026	0.075	0.046	0.043	0.069	0.029	0.024	0.083	0.042	0.048	0.062	0.032	0.078	0.026	0.116	0.017	0.067	0.030
(x_{51}, x_{12})	0.089	0.015	0.048	0.027	0.083	0.016	0.164	0.008	0.019	0.067	0.034	0.039	0.066	0.020	0.063	0.021	0.124	0.011	0.071	0.018
(x_{23}, x_{33})	0.044	0.033	0.035	0.042	0.068	0.021	0.128	0.011	0.030	0.048	0.059	0.024	0.110	0.013	0.075	0.019	0.140	0.010	0.071	0.020
(x_{33}, x_{23})	0.032	0.022	0.010	0.075	0.048	0.015	0.090	0.008	0.008	0.091	0.040	0.018	0.074	0.010	0.135	0.005	0.253	0.003	0.050	0.014
(x_{23}, x_{42})	0.045	0.033	0.040	0.037	0.078	0.019	0.130	0.011	0.034	0.043	0.067	0.022	0.112	0.013	0.075	0.020	0.125	0.012	0.064	0.023
(x_{42}, x_{23})	0.047	0.032	0.041	0.038	0.068	0.022	0.132	0.012	0.033	0.046	0.056	0.027	0.109	0.014	0.065	0.023	0.127	0.012	0.076	0.020
(x_{23}, x_{51})	0.049	0.036	0.043	0.040	0.076	0.023	0.114	0.015	0.037	0.047	0.065	0.027	0.098	0.018	0.073	0.024	0.109	0.016	0.063	0.028
(x_{51}, x_{23})	0.046	0.032	0.040	0.037	0.069	0.021	0.136	0.011	0.033	0.045	0.057	0.026	0.112	0.013	0.067	0.022	0.131	0.011	0.075	0.019
(x_{33}, x_{42})	0.049	0.013	0.009	0.071	0.051	0.013	0.086	0.008	0.005	0.138	0.027	0.024	0.044	0.015	0.144	0.004	0.240	0.003	0.042	0.015
(x_{42}, x_{33})	0.074	0.020	0.035	0.042	0.066	0.022	0.129	0.011	0.018	0.081	0.034	0.042	0.067	0.022	0.073	0.020	0.141	0.010	0.074	0.020
(x_{33}, x_{51})	0.054	0.014	0.010	0.078	0.050	0.015	0.076	0.010	0.005	0.151	0.026	0.030	0.039	0.020	0.141	0.006	0.212	0.004	0.042	0.019
(x_{51}, x_{33})	0.072	0.019	0.034	0.041	0.068	0.021	0.133	0.011	0.018	0.079	0.040	0.040	0.069	0.020	0.074	0.019	0.146	0.010	0.074	0.019
(x_{42}, x_{51})	0.081	0.022	0.044	0.040	0.073	0.024	0.114	0.015	0.023	0.078	0.038	0.047	0.059	0.030	0.070	0.025	0.110	0.016	0.066	0.027
(x_{51}, x_{42})	0.074	0.020	0.040	0.037	0.078	0.019	0.136	0.011	0.021	0.071	0.040	0.037	0.070	0.021	0.075	0.020	0.131	0.011	0.067	0.022

Table 26: Dynamic transition matrix in Scenario 4 at $t = 3$ for Profile 1

(x_{12}, x_{23})	0.028	0.065	0.026	0.070	0.042	0.044	0.071	0.026	0.040	0.046	0.064	0.029	0.108	0.017	0.069	0.027	0.117	0.016	0.073	0.025
(x_{23}, x_{12})	0.056	0.026	0.055	0.027	0.089	0.017	0.151	0.010	0.038	0.039	0.061	0.024	0.103	0.014	0.061	0.024	0.104	0.014	0.065	0.023
(x_{12}, x_{33})	0.044	0.041	0.022	0.082	0.041	0.043	0.070	0.025	0.021	0.085	0.040	0.045	0.068	0.026	0.081	0.022	0.137	0.013	0.072	0.025
(x_{33}, x_{12})	0.039	0.007	0.003	0.075	0.037	0.007	0.063	0.004	0.001	0.185	0.015	0.017	0.026	0.010	0.017	0.002	0.030	0.001	0.027	0.009
(x_{12}, x_{42})	0.045	0.042	0.027	0.071	0.059	0.032	0.072	0.026	0.026	0.074	0.057	0.033	0.070	0.027	0.098	0.019	0.118	0.016	0.053	0.036
(x_{42}, x_{12})	0.095	0.016	0.056	0.027	0.085	0.018	0.152	0.010	0.023	0.065	0.034	0.043	0.062	0.024	0.059	0.025	0.105	0.014	0.069	0.021
(x_{12}, x_{51})	0.048	0.045	0.028	0.076	0.045	0.048	0.060	0.036	0.027	0.079	0.044	0.049	0.057	0.038	0.075	0.029	0.098	0.022	0.061	0.036
(x_{51}, x_{12})	0.092	0.015	0.054	0.026	0.087	0.016	0.158	0.009	0.022	0.064	0.035	0.040	0.064	0.022	0.060	0.023	0.109	0.013	0.068	0.021
(x_{23}, x_{33})	0.042	0.038	0.035	0.046	0.066	0.024	0.112	0.014	0.033	0.048	0.063	0.025	0.107	0.015	0.076	0.021	0.130	0.012	0.068	0.024
(x_{33}, x_{23})	0.019	0.015	0.003	0.112	0.028	0.010	0.048	0.006	0.002	0.013	0.025	0.012	0.043	0.007	0.189	0.002	0.316	0.001	0.029	0.010
(x_{23}, x_{42})	0.041	0.038	0.041	0.038	0.091	0.017	0.111	0.014	0.039	0.040	0.087	0.018	0.106	0.015	0.088	0.018	0.107	0.014	0.048	0.032
(x_{42}, x_{23})	0.046	0.037	0.043	0.040	0.065	0.027	0.117	0.015	0.039	0.045	0.059	0.030	0.105	0.016	0.063	0.027	0.113	0.015	0.075	0.023
(x_{23}, x_{51})	0.047	0.042	0.046	0.043	0.074	0.027	0.096	0.021	0.043	0.045	0.070	0.028	0.092	0.021	0.071	0.028	0.093	0.021	0.058	0.034
(x_{51}, x_{23})	0.045	0.036	0.042	0.039	0.067	0.024	0.122	0.013	0.038	0.044	0.060	0.027	0.109	0.015	0.065	0.025	0.118	0.014	0.074	0.022
(x_{33}, x_{42})	0.026	0.009	0.002	0.098	0.035	0.007	0.042	0.005	0.001	0.173	0.020	0.011	0.024	0.009	0.228	0.001	0.277	0.001	0.018	0.012
(x_{42}, x_{33})	0.071	0.023	0.035	0.047	0.063	0.026	0.114	0.014	0.020	0.082	0.036	0.045	0.065	0.025	0.073	0.022	0.131	0.012	0.072	0.023
(x_{33}, x_{51})	0.031	0.010	0.003	0.115	0.029	0.011	0.038	0.008	0.002	0.202	0.017	0.019	0.022	0.014	0.191	0.002	0.250	0.001	0.023	0.013
(x_{51}, x_{33})	0.069	0.023	0.034	0.046	0.065	0.024	0.119	0.013	0.020	0.080	0.037	0.042	0.068	0.023	0.076	0.021	0.137	0.011	0.072	0.022
(x_{42}, x_{51})	0.079	0.026	0.046	0.043	0.070	0.029	0.097	0.021	0.026	0.076	0.040	0.050	0.055	0.036	0.068	0.029	0.094	0.021	0.062	0.032
(x_{51}, x_{42})	0.070	0.023	0.041	0.039	0.092	0.017	0.120	0.013	0.023	0.068	0.053	0.030	0.068	0.023	0.089	0.018	0.116	0.014	0.052	0.031

Table 27: Dynamic transition matrix in Scenario 4 at $t = 4$ for Profile 1

(x_{12}, x_{23})	0.029	0.068	0.029	0.069	0.043	0.046	0.067	0.029	0.044	0.045	0.065	0.0300	0.103	0.019	0.066	0.030	0.104	0.019	0.070	0.028
(x_{23}, x_{12})	0.058	0.027	0.062	0.025	0.092	0.017	0.144	0.011	0.042	0.037	0.063	0.025	0.099	0.016	0.059	0.027	0.093	0.017	0.062	0.025
(x_{12}, x_{33})	0.044	0.043	0.023	0.083	0.042	0.045	0.065	0.029	0.022	0.084	0.041	0.046	0.065	0.029	0.080	0.023	0.126	0.015	0.068	0.028
(x_{33}, x_{12})	0.016	0.003	0.000	0.095	0.015	0.003	0.024	0.002	0.000	0.232	0.006	0.007	0.010	0.004	0.221	0.000	0.347	0.000	0.010	0.004
(x_{12}, x_{42})	0.041	0.040	0.026	0.063	0.091	0.018	0.061	0.027	0.026	0.063	0.090	0.018	0.060	0.027	0.141	0.012	0.095	0.017	0.027	0.060
(x_{42}, x_{12})	0.097	0.016	0.062	0.026	0.087	0.018	0.145	0.011	0.025	0.062	0.036	0.045	0.059	0.027	0.056	0.028	0.093	0.017	0.067	0.024
(x_{12}, x_{51})	0.048	0.047	0.031	0.074	0.046	0.050	0.054	0.042	0.030	0.075	0.045	0.050	0.054	0.042	0.071	0.032	0.085	0.027	0.057	0.040
(x_{51}, x_{12})	0.095	0.016	0.060	0.025	0.090	0.017	0.153	0.010	0.025	0.061	0.037	0.041	0.063	0.024	0.058	0.026	0.098	0.015	0.066	0.023
(x_{23}, x_{33})	0.040	0.043	0.035	0.050	0.064	0.030	0.100	0.017	0.036	0.048	0.067	0.026	0.105	0.017	0.077	0.023	0.121	0.014	0.065	0.027
(x_{33}, x_{23})	0.007	0.007	0.000	0.151	0.011	0.005	0.017	0.003	0.000	0.157	0.010	0.005	0.016	0.003	0.231	0.000	0.363	0.000	0.011	0.004
(x_{23}, x_{42})	0.034	0.037	0.036	0.034	0.127	0.010	0.085	0.015	0.038	0.033	0.132	0.009	0.089	0.014	0.123	0.010	0.083	0.015	0.024	0.053
(x_{42}, x_{23})	0.045	0.042	0.045	0.042	0.063	0.030	0.105	0.018	0.043	0.044	0.060	0.031	0.101	0.019	0.061	0.031	0.102	0.019	0.073	0.026
(x_{23}, x_{51})	0.044	0.048	0.047	0.045	0.071	0.030	0.084	0.025	0.049	0.043	0.074	0.029	0.087	0.024	0.069	0.031	0.082	0.026	0.055	0.039
(x_{51}, x_{23})	0.044	0.041	0.044	0.041	0.065	0.028	0.110	0.016	0.042	0.043	0.063	0.029	0.106	0.017	0.063	0.028	0.107	0.017	0.072	0.025
(x_{33}, x_{42})	0.008	0.003	0.000	0.107	0.018	0.001	0.012	0.002	0.000	0.173	0.011	0.002	0.007	0.003	0.384	0.000	0.258	0.000	0.003	0.007
(x_{42}, x_{33})	0.068	0.026	0.035	0.051	0.061	0.029	0.102	0.018	0.022	0.082	0.038	0.047	0.063	0.028	0.073	0.024	0.123	0.015	0.071	0.025
(x_{33}, x_{51})	0.011	0.004	0.000	0.152	0.011	0.005	0.013	0.004	0.000	0.246	0.007	0.007	0.008	0.006	0.234	0.000	0.278	0.000	0.008	0.006
(x_{51}, x_{33})	0.067	0.026	0.034	0.050	0.063	0.027	0.107	0.016	0.021	0.080	0.039	0.043	0.067	0.026	0.076	0.022	0.129	0.013	0.070	0.024
(x_{42}, x_{51})	0.075	0.029	0.048	0.045	0.067	0.032	0.085	0.026	0.029	0.073	0.042	0.052	0.053	0.041	0.065	0.033	0.083	0.026	0.059	0.037
(x_{51}, x_{42})	0.060	0.023	0.038	0.036	0.133	0.010	0.096	0.014	0.024	0.058	0.082	0.017	0.060	0.023	0.129	0.011	0.094	0.015	0.027	0.051

Color	Choice Pair	Color	Choice Pair
	(x_{12}, x_{23})		(x_{23}, x_{42})
	(x_{23}, x_{12})		(x_{42}, x_{23})
	(x_{12}, x_{33})		(x_{23}, x_{51})
	(x_{33}, x_{12})		(x_{51}, x_{23})
	(x_{12}, x_{42})		(x_{33}, x_{42})
	(x_{42}, x_{12})		(x_{42}, x_{33})
	(x_{12}, x_{51})		(x_{33}, x_{51})
	(x_{51}, x_{12})		(x_{51}, x_{33})
	(x_{23}, x_{33})		(x_{42}, x_{51})
	(x_{33}, x_{23})		(x_{51}, x_{42})

Figure 22: Legend corresponding to Figure 23

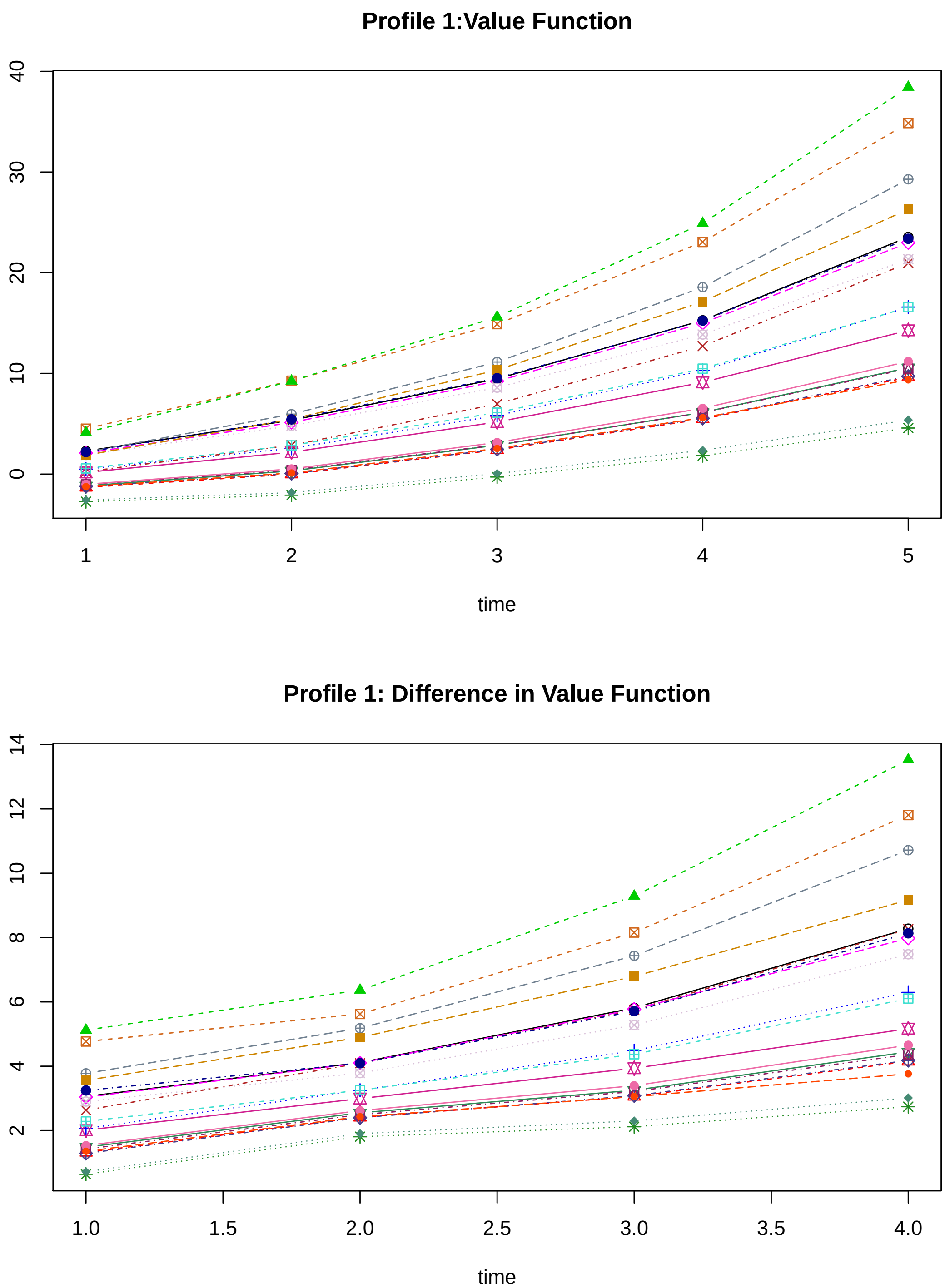


Figure 23: Expected discounted utility and their differences over time for the Profile 1

CHAPTER 6

CONCLUSION

Modeling DCEs has applications in many areas. However, it is challenging because of the large number of covariates, issues with reliability, and the condition that consumer behaviors is a forwards evolving activity/practice. By extending the idea of stationary process, we present a dynamic model with evaluation under random utility analysis. Generic preferences are presented and best-worst choices are identified.

Such results could guide researchers/practitioners in areas of health systems research, public policy, transportation research, and economics, and clinical decision. Steps that are the most appropriate for primary objective goals are provided. Different choice pairs were prioritized/arranged under selected transition probabilities. We initialize the process under an MDP algorithm for the proposed data generations and model building over time. We show relevant utility changes/classification in the applications/examples.

The attribute-level best-worst DCMs considered in this dissertation provide a general insight for modeling complex dependencies of time evolving decisions. The decisions for the large data set are guided by a functional form of the expected utility under identifiability constraints. Profile specific trends are displayed and pattern behaviors are exhibited. We highlighted compelling situations that allow shrinkage towards referenced choices and show efficacy in both low and high dimensional data examples to make inferences on the best-worst decisions of interest. Our simulated and aggregated data examples show the flexibility and wide applications of our proposed techniques. Our methodology is easily reproducible. The functional dependency and time evolving structure may accommodate additional arrangements/setups.

A potential area of concern in the application of MDPs for attribute-level best-worst DCEs is the "curse of dimensionality" as mentioned in Rust (2008). As the number of attributes, attribute-levels, and profiles grow in the experiment, the estimation process becomes exponentially more difficult. DCEs with larger number of attributes and attribute-levels have more choice sets and pairs to model across time. For discrete processes as is considered in the attribut-level best-worst DCEs, the amount of information that needs to be stored becomes overwhelming. The ability to guide the system becomes difficult due to the increased number of states and choice sets considered. These issues should be considered when using MDPs.

Extensions of this work may include interactions of choice pairs under different correlation structures. The first order Markov dependency structure presented here may be extended to higher order decision processes under stationary and dynamic transition probabilities. Extensions to the continuous time scale case are being explored.

REFERENCES

Adamsen, J.M., Rungle-Thiele, S., and Whitty, J.A. (2013), "Best-worst scaling . . . reflections on presentation, analysis, and lessons learnt from case 3 BWS experiments," *Market and Social Research*, 21(1), 9-27.

Agresti, A. (2007), *An Introduction To Categorical Data Analysis* (2nd ed.). New York: Wiley.

Agresti, A. (1990), *Categorical Data Analysis* (2nd ed.). New York: Wiley.

Anas, A. (1983), "Discrete choice theory, information theory and the multinomial logit and gravity models." *Transportation Research Part B: Methodological*, 17(1), 13-23.

Anderson, S.P., De Palma, A. and Thisse, J.F. (1992), *Discrete choice theory of product differentiation*. MIT press.

Arcidiacono, P., and Ellickson, P.B. (2011), "Practical methods for estimation of dynamic discrete choice models," *Annual Review Econonmics*, 3(1), 363-394.

Auger, P., Devinney, T.M., and Louviere, J.J., (2007), "Using Best-Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries," *Journal of Business Ethics*, 70(3), 299-326.

Bellman, R. (1954), *The theory of dynamic programming* (No. RAND-P-550), RAND CORP SANTA MONICA CA.

Bellman, R. (1956), "Dynamic programming" (Vol. 295). RAND CORP SANTA MONICA CA.

Ben-Akiva, M. and Bierlaire, M. (1999), "Discrete Choice Methods and their Application to Short Term Decisions," *Handbook of Transportation Science*, 23, 5-33.

Chades, I., Chapron, G., Cros, M.J., Garcia, F. and Sabbadin, R. (2014), "MDPtoolbox: a multiplatform toolbox to solve stochastic dynamic programming problems," *Ecography*, 37(9), 916-920.

Coast, J., Salisbury, C., De Berker, D., Noble, A., Horrocks, S., Peters, T.J. and Flynn, T.N. (2006), "Preferences for aspects of a dermatology consultation," *British Journal of Dermatology*, 155(2), 387-392.

Cuervo, D.P., Kessels, R., Goos, P. and Srensen, K. (2016), "An integrated algorithm for the optimal design of stated choice experiments with partial profiles," *Transportation Research Part B: Methodological*, 93, 648-669.

Dean, A., Morris, M., Stufken, J. and Bingham, D. eds. (2015), *Handbook of design and analysis of experiments* (Vol. 7). CRC Press.

DeShazo, J.R. and Fermo, G. (2002), "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency," *Journal of Environmental Economics and Management*, 44(1), 123-43.

Diebold, F.X., Lee, J.H. and Weinbach, G.C. (1994), "Regime switching with time-varying transition probabilities," *Business Cycles: Durations, Dynamics, and Forecasting*, 144-165.

Fedorov, V.V. (1972), *Theory of optimal experiments*, Elsevier.

Feinberg, E. and Schwartz, A. (1994), "Markov Decision Models with Weighted Discounted Criteria," *Mathematics of Operations Research*, 19(1), 152-168.

Finn, A. and Louviere, J.J. (1992), "Determining the appropriate response to evidence of public concern: the case of food safety," *Journal of Public Policy and Marketing*, 12-25.

Flynn, T.N.,Louviere, J.J.,Peters, T.J., and Coast, J. (2007), "Best-worst scaling: what it can do for health care research and how to do it," *Journal of Health Economics*, 26(1), 171-189.

Flynn, T.N.,Louviere, J.J.,Peters, T.J., and Coast, J. (2008), "Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis," *BMC Medical Research Methodology*, 8(1), 76.

Flynn, T.N.,Louviere, J.J.,Peters, T.J., and Coast, J. (2010), "Valuing citizen and patient preferences in health: recent developments in three types of bestworst scaling," *Expert review of pharmacoeconomics and outcomes research*, 10(3), 259-267.

Frederick, S., Loewenstein, G. and O'Donoghue, T. (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40(2), 351-401.

Goodman, S., Lockshin, L., and Cohen, E. (2006), "Using the Best-Worst method to examine market segment and identify different influences of consumer choice," *International Wine Business and Marketing Conference*.

Graßhoff, U., Großmann, H., Holling, H. and Schwabe, R. (2003), "Optimal paired comparison designs for first order interactions," *Statistics*, 37(5), 373-386.

Graßhoff, U., Großmann, H., Holling, H. and Schwabe, R. (2004), "Optimal designs for main effects in linear paired comparison models," *Journal of Statistical Planning and Inference*, 126(1), 361-376.

Graßhoff, U., Großmann, H., Holling, H. and Schwabe, R. (2013), "Optimal designs for discrete choice experiments," *Journal of Statistical Planning and Inference*, 143(1), 167-175.

Green, P. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," *Journal of Consumer Research*, 1(2), 61-68.

Großmann, H., Graßhoff, U., and Schwabe, R. (2009), "Approximate and exact optimal designs for paired comparisons of partial profiles when there are two groups of factors," *Journal of Statistical Planning and Inference*, 139, 1171-1179.

Hausman, J.A. and Wise, D.A. (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences,"*Econometric*, 146(2), 403-426.

Hess, S. and Hensher, D.A. (2010), "Using conditioning on observed choices to retrieve individual-specific attribute process startegies," *Transportation Research*, 44B, 781-790.

Hingkelmann, K., ed., (2011), *Design and analysis of experiments*, 3, John Wiley and Sons.

Hoffman, S.D. and Duncan, G.J. (1988), "Multinomial and Conditional Logit Discrete-Choice Models in Demography,"*Demography*, 25(3), 415-417.

Hole, A.R. (2011), “A discrete choice model with endogeneous attribute attendance,” *Economics Letters*, 110, 203-205.

Hurtubia, R., Nguyen, M.H., Glerum, A., and Bierlaire, M. (2014), “Integrating psychometric indicators in latent class choice,” *Transportation Research Part A*, 64, 135-146.

Kamakura, W.A. and Russell, G. (1989), “A probabilistic choice model for market segmentation and elasticity structure.”

Kessels, R., Jones, B. and Goos, P. (2011), “Bayesian optimal designs for discrete choice experiments with partial profiles,” *Journal of Choice Modelling*, 4(3), 52-74.

Knox, S.A., Viney, R.C., Street, D.J., Haas, M.R., Fiebig, D.G., Weisberg, E., and Bateson, D. (2012), “What’s Good and Bad About Contraceptive Products?,” *Pharmaeconomics*, 30(12), 1-16.

Knox, S.A., Vineya, R.C., Gu, Y., Hole, A.R., Fiebig, D.G., Street, D.J., Haas, M.R., Weisberg, E., Bateson, D. (2013), ”The effect of adverse information and positive promotion on women’s preferences for prescribed contraceptive products,” *Social Science and Medicine*, 83, 70-80.

Lancsar, E., Louviere, J. and Flynn, T. (2007), “Several methods to investigate relative attribute impact in stated preference experiments,” *Social Science and Medicine*, 64(8), 1738-1753.

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., and Burgess, L. (2013),“Best worst discrete choice experiments in health: methods and an application,” *Social Science and Medicine*, 76, 74-82.

Lee, T.C., Judge, G.G. and Zellner, A. (1968), “Maximum likelihood and Bayesian estimation of transition probabilities”, *Journal of the American Statistical Association*, 63(324), 162-1179.

Loose, S.M. and Lockshin, L. (2013), “Tesing the robustness of best worst scaling for cross-national segmentation with different numbers of choice sets,” *Food Quality and Preference*, 27(2), 230-242.

Louviere, J.J., Street, D., Burgess, L., Wasi, N., Islam, T. and Marley, A.A. (2008), "Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information." *Journal of choice modelling*, 1(1), 128-164.

Louviere, J. and Timmermans, H. (1990), "Hierarchical information integration applied to residential choice behavior" *Geographical Analysis*, 22(2), 127-144.

Louviere, J.J., Islam, T., Wasi, N., Street, D, and Burgess, L. (2008), "Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price?," *Journal of Consumer Research*, 35(2), 360-375.

Louviere, J.J. and Woodworth, G. (1983), "Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data," *Journal of Marketing Research*, 350-367.

Louviere, J.J. and Woodworth, G.G. (1991), "Best-worst scaling: A model for the largest difference judgement," *University of Alberta: Working Paper.*

Louviere, J.J., Street, D., Burgess, L., Wasi, N., Islam, T., and Marley, A.A. (2008), "Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information," *Journal of Choice Modeling*, 1(1), 128-164.

Louviere, J., Lings, I., Islam, T., Gudergan, S., and Flynn, T. (2013), "An introduction to the application of (case 1) best-worst scaling in marketing research", *International Journal of Research in Marketing*, 30(3), 292-303.

Luce, D. (1959), *Individual Choice Behavior*, New York: John Wiley and Sons.

Marschak, J. (1960), "Binary Choice Constraints on Random Utility Indications," *Stanford Symposium on Mathematical Methods in the Social Sciences*, ed. K. Arrow, Stanford, CA: Stanford University Press, 312-329.

Marley, A.A.J. and Louviere, J.J. (2005), "Some probabilistic models of best, worst, and best-worst choices," *Journal of Mathematical Psychology*, 49, 464-480.

Marley, A.A.J., Flynn, T.N., and Louviere, J.J. (2008), "Probabilistic models of set-dependent and attribute-level best-worst choice," *Journal of Mathematical Psychology*, 52, 281-296.

Marley, A.A.J. and Pihlens, D. (2012), "Models of best-worst choice and ranking among multiattribute options (profiles)," *Journal of Mathematical Psychology*, 56, 24-34.

Massey, G.R., Wang, P.Z., Waller, D.S., and Lanasier, E.V. (2015), "Best-worst scaling: a new method for advertisement evaluation," *Journal of Marketing Communications*, 21(6), 425-449.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, ed. P. Zambreka , New York: Academic Press, 105-142.

McFadden, D. (1978), "Modeling the Choice of Residential Location," *Spatial Interaction Theory and Planning Models*, eds. A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, North-Holland, Amsterdam, pp. 75-96.

McFadden, D. and Train, K. (2000), "Mixed MNL Models for Discrete Repsonse," *Journal of Applied Econometrics*, 15, 447-470.

Mueller, S. and Rungie, C. (2009), "Is there more information in best-worst choice data? Using the attitude heterogeneity structure to identify consumer segments," *International Journal of Wine Business Research*, 21(1), 24-40.

Ghaoui, L.E. and Nilim, A. (2005), "Robust solutions to markov decision problems with uncertain transition matrices," *Operations Research*, 53(5).

Parvin, S., Wang, P., and Uddin, J. (2016), "Using best-worst scaling method to examine consumers' value preferences: A multidimensional perspective," *Cogent Business and Management*, 3(1),1199110.

Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., and Brazier, J.E. (2011), "Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data", *Social Science and Medicine*, 72(10), 1717-1727.

Puterman, M.L. (2014), *Markov decision processes: discrete stochastic dynamic programming*, Wiley and Sons.

Rothblum, U. (1975), "Normalized Markov Decision Chains I; Sensitive Discount Optimality," *Operations Research*, 23(4), 785-795.

Ruseckaite, A., Goos, P. and Fok, D. (2017), "Bayesian Doptimal choice designs for mixtures," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(2), 363-386.

Rust, J. (1994), "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4, 3081-3143.

Rust, J. (2008), Dynamic programming. *The New Palgrave Dictionary of Economics.*

Small, K. (1987), "A Discrete Choice Model for Ordered Alternatives," *Econometrica*, 55, 409-424.

Stenberg, F., Manca, R. and Silvestrov, D. (2007), "An algorithmic approach to discrete time non-homogeneous backward semi-Markov reward processes with an application to disability insurance", *Methodology and Computing in Applied Probability*, 9(4), 497-519.

Street, D.J. and Burgess, L. (2007), "The construction of optimal stated choice experiments: Theory and methods," (Vol. 647). John Wiley and Sons.

Street, D.J. and Knox, S.A. (2012), "Designing for Attribute-Level Best-Worst Choice Experiments," *Journal of Statistical Theory and Practice*, 6(2), 363-375.

Swait, J. (2001),"Choice set generation within the generalized extreme value family of discrete choice models," *Transportation Research Part B: Methodological*, 35(7), 643-666.

Thisse, J. and Norman, G. (1994), *The Economics of product differentiation.* Edward Elgar Publishing.

Thurstone, L. L. (1927), "A Law of Comparative Judgment," *Psychological Review*, 34(4), 273-286.

Train, K. (2009), *Discrete Choice Models with Simulations* (2nd Ed.), Cambridge University Press.

van der Pol, M., Currie, G., Kromm, S., and Ryan, M. (2014), "Specification of the Utility Function in Discrete Choice Experiments," *Value in Health*, 17, 297-301.

White, D.J. (1985), "Real Applications of Markov Decision Processes," *Interfaces*, 15(6), 73-83.

Whitty, J.A., Ratcliffe, J., Chen, G., and Scuffman, P.A. (2014), "Australian public preferences for the funding of new health technologies: a comparison of discrete choice and profile case best-worst scaling methods", *Medical Decision Making*, 34(5), 638-654.

Yoo, H.I. and Doiron, D. (2013), "The use of alternative preference elicitation methods in complex discrete choice experiments," *Journal of Health Economics*, 32(6), 1166-1179.

APPENDIX A

SELECTED R CODE

```
########################################################
#MDC_Data to simulate the data
MDC_Data<-function(n,k,m,est,nc,nCH,profiles,alt2_sort)
{
  #alt2_sort are the choices within each profile
  new<-1
  pairs<-k*(k-1)
  g<-n*m*pairs
  subj<-vector(,g)
  prof<-vector(,g)
  Ch.opt<-vector(,g)
  Y<-vector(,g)
  Cov<-matrix(nrow=g,ncol=nc,byrow=TRUE)
  pi<-vector(,nCH)

  for(k in 1:nCH){
    pi[k]<-sum(est*X[k,])
  }
  for(i in 1:m){
  q<-Set.pairs[profiles[i],]
    p<-exp(pi[q])/sum(exp(pi[q]))
    Ch<-rmultinom(n,1,p)

    for(j in 1:n){
      new1<-new+pairs-1
      Y[new:new1]<-Ch[,j]
      Ch.opt[new:new1]<-q
      subj[new:new1]<-j
      prof[new:new1]<-profiles[i]
     choices<-c(alt2_sort[i,1],alt2_sort[i,2],alt2_sort[i,3],
                alt2_sort[i,4],alt2_sort[i,5],alt2_sort[i,6])
      R1<-new;R2<-new+1;R3<-new+2;R4<-new+3;R5<-new+4;R6<-new+5;
      #rows for each of the k(k-1)=6 choices
      c1<-choices[1];c2<-choices[2];c3<-choices[3];c4<-choices[4];
      c5<-choices[5];c6<-choices[6]
      Cov[R1,]<-X[c1,]
      Cov[R2,]<-X[c2,]
```

```
        Cov[R3,]<-X[c3,]
        Cov[R4,]<-X[c4,]
        Cov[R5,]<-X[c5,]
        Cov[R6,]<-X[c6,]
        new<-new+pairs
} }

  return(df3<-data.frame(subject=subj,profile=prof,mode=Ch.opt,y=Y,
          A1=Cov[,1],A2=Cov[,2],A3=Cov[,3],A111=Cov[,4],A112=Cov[,5],
          A121=Cov[,6],A122=Cov[,7],A123=Cov[,8],A131=Cov[,9],
          A132=Cov[,10],A133=Cov[,11],A134=Cov[,12]))
}
########################################################################
T.Prob<-function(T,t,beta.est,X,k,levels,m,profiles){
  k<-num.att
  tau<-k*(k-1)
  num.cov<-k+sum(levels)
    Trans.par<-matrix(nrow=num.choices, ncol=num.cov)
  beta<-vector(,num.cov)
  t1<-T-t

for(i in 1:num.choices){
Trans.par[i,]<-beta.est
if (X[i,1]==1){Trans.par[i,1]<-(1.7^t)*abs(beta.est[1])} else
  if (X[i,1]==-1){Trans.par[i,1]<--(1.7^t)*abs(beta.est[1])}
   else {Trans.par[i,1]<-beta.est[1]}
if (X[i,2]==1){Trans.par[i,2]<-(1.7^t)*abs(beta.est[2])} else
  if (X[i,2]==-1){Trans.par[i,2]<--(1.7^t)*abs(beta.est[2])}
    else {Trans.par[i,2]<-beta.est[2]}
if (X[i,3]==1){Trans.par[i,3]<-(1.7^t)*abs(2)} else
  if (X[i,3]==-1){Trans.par[i,3]<--(1.7^t)*abs(2)}
   else {Trans.par[i,3]<-0.5}
if (X[i,4]==1){Trans.par[i,4]<-(1.7^t)*abs(beta.est[4])} else
  if (X[i,4]==-1){Trans.par[i,4]<--(1.7^t)*abs(beta.est[4])}
     else {Trans.par[i,4]<-beta.est[4]}
if (X[i,5]==1){Trans.par[i,5]<-(1.7^t)*abs(beta.est[4])} else
  if (X[i,5]==-1){Trans.par[i,5]<--(1.7^t)*abs(beta.est[4])}
     else {Trans.par[i,5]<-beta.est[4]}
if (X[i,6]==1){Trans.par[i,6]<-(1.7^t)*abs(beta.est[6])} else
  if (X[i,6]==-1){Trans.par[i,6]<--(1.7^t)*abs(beta.est[6])}
     else {Trans.par[i,6]<-beta.est[6]}
if (X[i,7]==1){Trans.par[i,7]<-(1.7^t)*abs(beta.est[7])} else
  if (X[i,7]==-1){Trans.par[i,7]<--(1.7^t)*abs(beta.est[7])}
     else {Trans.par[i,7]<-beta.est[7]}
if (X[i,8]==1){Trans.par[i,8]<-(1.7^t)*abs(beta.est[6]+beta.est[7])}
```

```
else if (X[i,8]==-1){Trans.par[i,8]<--(1.7^t)*abs(beta.est[6]+beta.est[7])}
     else {Trans.par[i,8]<--(beta.est[6]+beta.est[7])}
if (X[i,9]==1){Trans.par[i,9]<-(1.7^t)*abs(beta.est[9])} else
   if (X[i,9]==-1){Trans.par[i,9]<--(1.7^t)*abs(beta.est[9])}
     else {Trans.par[i,9]<-beta.est[9]}
if (X[i,10]==1){Trans.par[i,10]<-(1.7^t)*abs(beta.est[10])} else
   if (X[i,10]==-1){Trans.par[i,10]<--(1.7^t)*abs(beta.est[9])}
     else {Trans.par[i,10]<-beta.est[10]}
if (X[i,11]==1){Trans.par[i,11]<-(1.7^t)*abs(beta.est[11])} else
   if (X[i,11]==-1){Trans.par[i,11]<--(1.7^t)*abs(beta.est[11])}
     else {Trans.par[i,11]<-beta.est[11]}
if (X[i,12]==1){Trans.par[i,12]<-(1.7^t)*abs(beta.est[9]+
                                        beta.est[10]+beta.est[11])}
else if (X[i,12]==-1){Trans.par[i,12]<--(1.7^t)*abs(beta.est[9]+
                              beta.est[10]+beta.est[11])}
else {Trans.par[i,12]<--(beta.est[9]+beta.est[10]+beta.est[11])}
   }
  Trans.prob<-matrix(nrow=num.choices,ncol=num.choices)
for(i in 1:num.choices){
   for(j in 1:num.choices){
     Trans.prob[i,j]<-sum(Trans.par[i,]*X[j,])
   }  }
Trans.prob<-exp(Trans.prob)
t.prob<-matrix(nrow=m*tau,ncol=tau)
NEW<-1
for(i in 1:m){
   set<-Choices[i,]
   Parameters<-Trans.prob
   for(j in 1:tau){
     for(k in 1:tau){
        t.prob[NEW,k]<-Parameters[set[j],set[k]]/sum(Parameters[set[j],set])
     }
     NEW<-NEW+1
   }   }
return(t.prob)
}
##############################################################################
#Creating all of the choice pairs
num.att <- 3
levels <- c(2,3,4)
choice.pairs <- data.frame(bestAtt=numeric(0),bestLevel=numeric(0),
                          worstAtt=numeric(0),worstLevel=numeric(0))
#we create all possible pairs
num.choices<-nrow(choice.pairs)
nCol<-sum(num.att,levels)
```

```
##########################################################################
# Choice Specific Explanatory Variables
X<-matrix(nrow=num.choices,ncol=nCol,byrow=TRUE)
for(i in 1:num.choices){
  new<-num.att
  for(k in 1:num.att){
if (choice.pairs$bestAtt[i]==k){X[i,k]<-1} else
  if (choice.pairs$worstAtt[i]==k){X[i,k]<--1}
    else {X[i,k]<-0}
for(j in 1:levels[k]){
  new<-new+1
  h<-levels[k]
if (X[i,k]==1){if (j==h)
{if(choice.pairs$bestLevel[i]==j){X[i,(new-(h-1)):(new-1)]<--1;
                                    X[i,new]<-1}
else X[i,new]<-0} else
  {if (choice.pairs$bestLevel[i]==j){X[i,new]<-1}
      else X[i,new]<-0}}
else if (X[i,k]==-1){if (j==h){if (choice.pairs$worstLevel[i]==j)
  {X[i,(new-(h-1)):(new-1)]<-1; X[i,new]<--1}
  else X[i,new]<-0} else {if (choice.pairs$worstLevel[i]==j)
    {X[i,new]<--1} else X[i,new]<-0}}
else {X[i,new]<-0}
    } } }
Choice<-seq(1:num.choices)
####################################################################
Xtrans<-matrix(nrow=num.choices,ncol=nCol,byrow=TRUE)
for(i in 1:num.choices){
new<-num.att
for(j in 1:num.att){
if (choice.pairs$bestAtt[i]==j) {Xtrans[i,j]<-1} else
if (choice.pairs$worstAtt[i]==j) {Xtrans[i,j]<--1}
else {Xtrans[i,j]<-0}
for(k in 1:levels[j]){
new<-new+1
if(Xtrans[i,j]==1){if (choice.pairs$bestLevel[i]==k){Xtrans[i,new]<-1}
else {Xtrans[i,new]<-0}}
else if (Xtrans[i,j]==-1){if(choice.pairs$worstLevel[i]==k)
                              {Xtrans[i,new]<--1}
                             else {Xtrans[i,new]<-0}}
else {Xtrans[i,new]<-0}
    }    }    }
##########################################################
library(DoE.base) #Design of Experiments package
```

```
OMEP<-oa.design(nlevels=levels)
#used to determine OMEP, in this case there is no subset
#of profiles that work so all 24 are used

choice.profiles<-data.frame(profile=numeric(0),
                    bestAtt=numeric(0),bestLevel=numeric(0),
                    worstAtt=numeric(0),worstLevel=numeric(0))
row.number <- 0
#Determines which profile each choice is in
nr<-dim(OMEP)[1]
nc<-dim(OMEP)[2]
#pairs corresponding to profiles
for(i in 1:nr){
  for(j in 1:(nc-1)){
    for(k in (j+1):nc){
row.number <- row.number + 1
profile<-i
bestAtt<-j
bestLevel<-OMEP[i,j]
worstAtt<-k
worstLevel<-OMEP[i,k]
choice.profiles[row.number,]<-c(profile,bestAtt,bestLevel,
                                worstAtt,worstLevel)
row.number<-row.number+1
choice.profiles[row.number,]<-c(profile,worstAtt,worstLevel,
                                bestAtt,bestLevel)
    } } }
nr2<-dim(choice.profiles)[1]
nr3<-dim(choice.pairs)[1]
set<-vector(,nr2)
#matching the pairs in the profiles to choice pairs
for(i in 1:nr2){
  for(j in 1:nr3){
if (choice.profiles[i,2]==choice.pairs[j,1]
    && choice.profiles[i,3]==choice.pairs[j,2]
&& choice.profiles[i,4]==choice.pairs[j,3]
&& choice.profiles[i,5]==choice.pairs[j,4]) {set[i]<-j}
} }
cbind(choice.profiles,set)
prof.pairs<-num.att*(num.att-1)
Set.pairs<-matrix(nrow=nr,ncol=prof.pairs,byrow=TRUE, set)
df2<-data.frame(Ch1=Set.pairs[,1],Ch2=Set.pairs[,2],Ch3=Set.pairs[,3],
                Ch4=Set.pairs[,4],Ch5=Set.pairs[,5],Ch6=Set.pairs[,6])
profiles1<-seq(1:24)
beta.est<-c(-2,1.5,0,-2,0,1.99,-0.29,0,-.92,-0.18,0.5,0)
```

```
n<-300 #number of subjects
nc<-length(beta.est)
m<-length(profiles1)
datK3<-MDC_Data(n,num.att,m,beta.est,nc,num.choices,profiles1,Set.pairs)
##################################################################
#Utility Computations
V<-pi<-vector(,num.choices)
beta.mod<-c(-2.0711,1.5248,0,-2.0308,0,2.0970,-0.3567,
            0,-0.8914,-0.1805,0.4911,0)
for(i in 1:num.choices){
  V[i]<-sum(beta.mod*X[i,])
  pi[i]<-exp(V[i])
}
Prob<-pi/sum(pi)

library(QRM) #Needed for the Type I Extreme Value Distribution
Quantiles<-vector(,num.choices)
for(i in 1:num.choices){
  Quantiles[i]<-sum(V < V[i])
}
Quantiles<-(Quantiles+0.5)/num.choices
Error<-qGumbel(Quantiles)

Utility1<-Utility2<-vector(,num.choices)
for(i in 1:num.choices){
  Utility1[i]<-V[i]+Error[i]
  Utility2[i]<-V[i]
}
########################################################
T<-5
discount=0.95
r<-m*tau
Exp.Value1<-matrix(nrow=r,ncol=T)
k<-num.att
tau<-k*(k-1) #number of choices per profile
Choices<-choice.prob<-matrix(nrow=m,ncol=tau)
for(i in 1:tau){
  for(j in 1:m){
    Choices[j,]<-Set.pairs[profiles[j],]
    q<-Choices[j,]
    choice.prob[j,]<-pi[q]/sum(pi[q])
}  }
new<-1
for(i in 1:m){
```

```
  new2<-new+tau-1
  for(t in 2:T){
    for(j in 1:tau){
new3<-new+j-1
Exp.Value1[new:new2,1]<-Utility1[Choices[i,]]
trans<-T.Prob(T,,beta.mod,Xtrans,k,levels,m,profiles1)
Trans<-trans[new:new2,]
Valuepre<-Exp.Value1[new:new2,t-1]*trans[j,]
library(QRM) #Needed for the Type I Extreme Value Distribution
UtilityCh<-Utility2[Choices[i,j]]+
           qGumbel(Quantiles[Choices[i,j]],0,1.5^t)
Exp.Value1[new3,t]<-UtilityCh + discount*sum(Valuepre)
}  }
  new<-new2+1
}
Decisions1<-matrix(nrow=m,ncol=T-1)
new<-1
for(i in 1:m){
  new2<-new+tau-1
  for(t in 2:T){
    MAX<-max(Exp.Value1[new:new2,t])
     Decisions1[i,t-1]<-which(Exp.Value1[new:new2,t] %in% MAX)
  }
  new<-new2+1
}
```

VITA

Amanda Faye Working
Department of Computational and Applied Mathematics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D	Old Dominion University, Norfolk, VA. (December 2017) Major: Computational and Applied Mathematics. (Statistics)
MS	Old Dominion University, Norfolk, VA. (May 2012) Major: Computational and Applied Mathematics. (Statistics)
B.S.	Old Dominion University, Norfolk, VA. (December 2009) Major: Computational and Applied Mathematics.
B.A.	Virginia Wesleyan College, Norfolk, VA. (May 2008) Major: German Minors: Mathematics and History.

Experience

Data Analyst, Positive Behavioral Interventions and Supports (PBIS) a component of the Virginia Tiered Systems of Supports (VTSS), Old Dominion University, Norfolk, VA, (05/2016 - Present).

Graduate Teaching Assistant, Old Dominion University, Norfolk, VA, (01/2010 - 12/2017).

Typeset using LaTeX.