University of California
Los Angeles

The Good, the Bad and the Fitting: A Bayesian Hierarchical Model for Patient Preferences Elicited through Discrete Choice Experiments

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Public Health

by

Anna Liza Malazarte Antonio

2017

Abstract of the Dissertation

The Good, the Bad and the Fitting: A Bayesian Hierarchical Model for Patient Preferences Elicited through Discrete Choice Experiments

by

Anna Liza Malazarte Antonio
Doctor of Public Health
University of California, Los Angeles, 2017
Professor Catherine Crespi-Chun, Co-chair
Professor Robert Erin Weiss, Co-chair

In discrete choice experiments, patients are presented with sets of health states described by various attributes and asked to make choices from among them. Discrete choice experiments allow health care researchers to study the preferences of individual patients by eliciting trade-offs between different aspects of health-related quality of life. However, many discrete choice experiments yield data with incomplete ranking information and sparsity due to the limited number of choice sets presented to each patient, making it challenging to estimate patient preferences. Moreover, methods to identify outliers in discrete choice data are lacking. We develop a Bayesian hierarchical random effects rank-ordered multinomial logit model for discrete choice data. Missing ranks are accounted for by marginalizing over all possible permutations of unranked alternatives to estimate individual patient preferences, which are modeled as a function of patient covariates. We provide a Bayesian version of relative attribute importance, and adapt the use of the conditional predictive ordinate to identify outlying choice sets and outlying individuals with unusual preferences compared to the population. The model is applied to data from a study using a discrete choice experiment to estimate individual patient preferences for health states related to prostate cancer treatment.

The dissertation of Anna Liza Malazarte Antonio is approved.

Christopher S Saigal

Catherine Ann Sugar

Robert Erin Weiss, Committee Co-chair

Catherine Crespi-Chun, Committee Co-chair

University of California, Los Angeles

2017

To my parents, Rosendo and Elizabeth Antonio, without whom nothing would have been possible.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my advisors, Dr. Catherine Crespi and Dr. Robert Weiss, and my committee members, Dr. Catherine Sugar, Dr. Christopher Saigal, and the late Dr. Ely Dahan for their guidance and inspiring discussions. I am incredibly lucky to have found mentors who look out for my best interest, who encourage my acheivements, and who have shaped me into the scientist that I am today.

Last, but not least, I would like to thank my family and friends, who have supported me in all my endeavors, who put up with me making myself scarce while writing this dissertation, and who will probably stop reading this about here. You know who you are. Mahal ko kayong lahat mula sa kaibuturan ng aking puso. I love you all from the bottom of my heart.

Vita

2004	B.S. Mathematics. Department of Mathematics, University of Florida, Gainesville, FL.
2004–2006	Teaching Assistant. Courses: Trigonometry, Pre-calulus algebra & Trigonometry, Survey Calculus, Analytic Geometry and Calculus. Department of Mathematics, University of Florida, Gainesville, FL.
2006	M.S. Applied Mathematics. Department of Mathematics, University of Florida, Gainesville, FL.
2006	Independent Instructor. Course: Analytic Geometry and Calculus. Department of Engineering, University of Florida, Gainesville, FL.
2006–2007	Research Assistant. Whitney Laboratory for Marine Bioscience, University of Florida, Gainesville, FL.
2007	Reader. Course: Integration and Infinite Series. Department of Mathematics, University of California, Los Angeles, CA.
2008–2010	Statistical Programmer. Department of Veteran's Affairs, Los Angeles, CA.
2009	M.S. Biostatistics, Department of Biostatistics, University of California, Los Angeles, CA.
2011–2012	Graduate Student Researcher. Center for Population Health and Health Disparities, University of California, Los Angeles, CA.
2011–present	Statistician. Department of Veteran's Affairs, Los Angeles, CA.
2012–2015	Graduate Student Researcher. School of Nursing, University of California, Los Angeles, CA.
2013	Graduate Teaching Assistant. Courses: Basic Biostatistics. School of Nursing, University of California, Los Angeles, CA.
2015–present	Graduate Student Researcher. Faculty Practice Group, UCLA Health System, Los Angeles, CA.

Publications

Li, X., Williams, A., Antonio, A.L.M. (2016) Mental Health and Suicidal Ideation among Chinese Women Who Have Sex With Men Who Have Sex with Men (MSM). *Women and Health.* 56 (8); 940–956.

Mack, J.W., Walling, A., Dy, S., Antonio, A.L., Adams, J., Keating, N.L., Tisnado, D. (2015) Patient beliefs that chemotherapy may be curative and care received at the end of life among patients with metastatic lung and colorectal cancer. *Cancer.* 121(11):1891–1897.

Sarna, L., Bialous, S.A., Nandy, K., Antonio, A.L., Yang, Q. (2014) Changes in smoking prevalences among health care professionals from 2003 to 2010-2011. *JAMA.* 311(2); 197–199.

Ryoo, J.J., Ordin, D.L., Antonio, A.L., Oishi, S.M., Gould, M.K., Asch, S.M., Malin, J.L. (2013) Patient preference and contraindications in measuring quality of care: what do administrative data miss? *Journal of Clinical Oncology.* 31(21); 2716–2723.

Malin, J.L., ONeill, S.M., Asch, S.M., Dy, S.M., Walling, A.M., Tisnado, D., Antonio, A.L., Lorenz, K.A. (2011) Quality of supportive care for patients with advanced cancer in a VA medical center. *Journal of Palliative Medicine.* 14(5); 573–577.

Dy, S.M., Lorenz, K.A., ONeill, S.M., Asch, S.M., Walling, A.M., Tisnado, D., Antonio, A.L., Malin, J.L. (2010) Cancer Quality-ASSIST supportive oncology quality indicator set: feasibility, reliability and validity testing. *Cancer.* 116(13); 3267–3275.

Antonio, A.L., Crespi C.M. (2010) Predictors of interobserver agreement in breast imaging using the Breast Imaging Reporting and Data System. *Breast Cancer Research and Treatment.* 120(3); 539–546.

CHAPTER 1

Introduction

1.1 Motivation

Discrete choice experiments (DCEs) have been increasingly used in health applications to characterize the preferences of individual patients for various health care interventions and services (Lancsar et al., 2013; DeBekker-Grob et al., 2012). In a typical health care DCE, patients are presented with sets of health states described by various attributes and asked to make choices from among them (Ryan et al., 2008). For example, a patient might be asked to choose between a health state with long life expectancy and poor quality of life and a health state with shorter life expectancy and high quality of life. By asking individuals to make choices between health states, they are forced to make trade-offs that reveal information about their preferences for different aspects of health-related quality of life.

Historically, in a DCE, patients provided their most preferred health state or a full ranking of a set of possible health states. However, continued research in discrete choice experiments has led to the development of best-worst designs in which patients indicate their most preferred and least preferred choices from a set (Lancsar and Louviere, 2008; Louviere et al., 2008). While reducing patient burden compared to full rankings, best-worst discrete choice experiments pose new statistical challenges. In such data, incomplete ranking information occurs when choosing best and worst from among four or more health states, and patient-level data are often insufficient to estimate individual-level preferences using maximum likelihood methods.

A number of models have been developed for discrete choice data. The multinomial logit model has been used for best choice data (McFadden, 1974), while the rank-ordered

logit model has been used for full ranking data (Allison and Christakis, 1994). Mixed logit models include random effects that vary across individuals to account for heterogeneity in preferences (Revelt and Train, 1998; McFadden and Train, 2000). More recently, Hernandez-Alava et al. introduced a model for ranked and partially ranked data that includes random effects, and estimated the random effects using Monte Carlo maximum likelihood methods (Hernandez-Alava et al., 2013). Although the model introduced by Hernandez-Alava et al. accommodates partially ranked data, it is not uncommon to obtain coefficient estimates in the wrong direction when using maximum likelihood estimation with sparse data (Rao, 2008). Moreover, their model does not include individual-specific covariates although inference on covariate effects is often of interest and it has been shown that including covariates can improve preference estimates for the mixed logit (Crabbe and Vandebroek, 2011; Orme and Howell, 2009; Greene et al., 2006; Allenby et al., 2005).

In many studies a key purpose of the DCE is to obtain an individual's ranking of various attributes relative to each other. The concept of relative attribute importance is widely used in the marketing research literature to provide rankings of features of consumer products (Paul E. Green, 1978; Halbrendt et al., 1995; Orme, 2010). Recently, this concept has been extended into the health care domain (Dowsey et al., 2016; Kruk et al., 2016; van Dijk et al., 2016). In this context, the purpose of the DCE is to obtain an individual's ranking of various attributes of health care or health-related quality of life, so that this information can be used as part of the health care decision-making process. For example, how a prostate cancer patient values full sexual functioning, long lifespan and no urinary incontinence relative to each other may inform which treatment options are a better match for the patient. While discrete choice data are now routinely analyzed using Bayesian hierarchical models with random effects to accommodate preference heterogeneity (McFadden and Train, 2000; Train, 2001; Allenby et al., 2005; Train, 2009), methods to compute relative attribute importance for such models are not fully developed.

Methods to identify outliers for such models are also lacking. Using the means of the individual-specific parameter distributions, Campbell and Hess (2010) classified individuals in the upper and lower percentiles as outliers. Farrel et al. (2012) proposed a graphical

method to identify outliers by plotting standardized random effects against their expected values for a Bayesian hierarchical logistic regression model. Several approaches for outlier detection in Bayesian models have been explored. For example, using the posterior distribution of the residuals of a regression model, Chaloner and Brant (1988) and Chaloner (1991, 1994) define an outlier as an observation with a large random error and calculate the posterior probabilities that observations are outlying. Other approaches for outlier detection are based on the predictive distribution. The conditional predictive ordinate (CPO), first suggested by Geisser (1980), is a diagnostic measure used to detect observations discrepant with the proposed model (Geisser, 1980, 1987, 1989, 1993; Dey et al., 1997; Pettit, 1990). To our knowledge, CPO has not been used to identify outlying random effects.

1.2 Our Approach & Contributions

We develop a Bayesian hierarchical model for best-worst discrete choice data. Incomplete rankings are handled by marginalizing over all possible permutations of unranked health states in a model that includes random effects to model individual-specific preferences. Bayesian methods are used to overcome the problem of sparse data to obtain estimates of individual preferences. To understand how patient characteristics are related to preferences, we model individual-specific preferences as a function of individual-specific covariates. We also define Bayesian versions of relative attribute importance for individuals and for the population that include random effects and covariates. To identify outliers in DCEs, we adapt the CPO in two ways: we adapt it to include random effects to identify patients who are unusual in their preferences for specific attributes or combinations of attributes, and we adapt it to handle vector outcomes to identify choice sets that are outlying with respect to individual preferences.

The development of best-worst discrete choice designs reduces patient burden compared to full rankings while posing new statistical challenges. By accounting for missing ranking information, patient covariates, and the sparse nature of the individual-level data in a Bayesian framework, our model extends current methods and provides individual-level

preference estimates. Our CPO measures provide some of the first diagnostic techniques for discrete choice models. Our model coupled with our measures of relative importance and outlyingness provide practical methodology for discrete choice modeling applications, in which parameter estimation at the individual-level is desirable, but observed data at the individual-level are limited.

1.3 Overview of the Dissertation

The dissertation is organized as follows. Chapter 2 defines various types of discrete choice experiments. Chapter 3 describes the PROSPECT study. Chapter 4 provides current methods used to analyze discrete choice data. Chapter 5 presents preliminary analyses. Chapter 6 presents the Bayesian hierarchical model for best-worst choice data with random effects and patient covariates, defines measures of relative importance, presents CPO-based measures for outlier detection, and demonstrates application of our methods to data from the PROSPECT study. Chapter 7 presents sensitivity analyses evaluating prior assumptions. Finally, Chapter 8 describes estimation of patient preferences using an adaptive best-worst conjoint method and ordinary least squares regression, and compares the estimated relative attribute importance scores to those obtained using the Bayesian hierarchical model of Chapter 6.

CHAPTER 2

Discrete Choice Experiments

This chapter briefly discusses the history of discrete choice experiments, sets a common nomenclature for discrete choice experiments in this dissertation and describes three common experimental designs for discrete choice experiments.

2.1 Origins

The origin of DCEs can be traced to a family of techniques called conjoint analysis, which grew out of the mathematical area of conjoint measurement (Luce and Tukey, 1964). In the 1930's and 1940's researchers were interested in whether psychological attributes could be quantified, and conjoint measurement provided a means to investigate this. In conjoint measurement, the joint effect of independent variables on a dependent variable is quantified. These variables need not be known quantities and can include psychological attributes, such as attitudes, cognitive abilities, etc. The conjoint measurement model is a deterministic mathematical model and not a statistical one. This distinguishes conjoint measurement from conjoint analysis, where the model is a statistical model with an error term and where the goal is to estimate the parameters of this model. In 1971, Green and Rao introduced conjoint methods to marketing research (Green and Rao, 1971). Its introduction was well received and research in this area continues to this day (Orme, 2010).

2.2 Nomenclature

Carson and Louviere (2011) discuss the need for a common nomenclature for stated preference elicitation approaches. DCEs are a rapidly growing field of research and *conjoint analysis* has become a blanket term for many variations of the orginal methods. Carson and Louviere (2011) argue that the term *conjoint analysis* is vague and should no longer be used because it fails to convey information regarding data collection, the experimental design and what statistical procedures were used for analysis. In their paper, they provide suggestions to improve clarity in the communication of research results. In alignment with this sentiment, this dissertation will use *discrete choice experiments* or *best-worst discrete choice experiments* in lieu of *conjoint analyses*. In addition, we make the following definitions for use throughout this dissertation:

- An *attribute* is a variable that describes a characteristic of a health state. For example, urinary functioning is one attribute of a health state. Sexual functioning and expected lifespan are also attributes of health states.

- An *attribute level* is a category of an attribute. For example, short term urinary issues is an attribute level of the attribute urinary functioning.

- A *profile* is a combination of attribute levels that together describe a health state. For example, a decreased sex life, short term urinary issues, live your expected lifespan, having doctor and family support, taking immediate action, and no surgery describe one health state profile.

- A *choice set* is a set of profiles from which choices are made. For example, if A, B, C and D represent four health state profiles, the set $\{A, B, C, D\}$ is a possible choice set.

- An *alternative* is a profile within a choice set. For example, A, B, C and D are each alternatives in the choice set $\{A, B, C, D\}$.

In the context of this dissertation, the attributes are categorically defined variables describing characteristics of health states which could result from prostate cancer treatment.

A combination of attribute levels constitutes a health state profile. Profiles are assigned to choice sets according to an experimental design. Patients select choices from various choice sets which contain health state profiles as alternatives.

2.3 Traditional Full Profile Conjoint Analysis

Traditional full profile conjoint analysis has been a mainstay of the conjoint community. Individuals view full profiles (all attributes at once) and are asked to either rank order all profiles (if given as a set of profiles) or provide a metric rating, for example, a rating between 0 and 100, of each profile (if viewing a single profile). Based on rankings and ratings, regression coefficients for the attributes (also called partial utilities or partworths) are calculated through regression techniques (for example, ordinary least squares) or through linear programming techniques (for example, Linear Programming Technique for Multidimensional Analysis of Preference (LINMAP)) for each individual or at the population level. This approach is simple in design and execution and all profiles are evaluated in the context of all other alternatives.

2.4 Discrete Choice Experiments

Discrete choice experiments, also known as choice-based conjoint analyses or discrete choice conjoint analyses, have become a popular method for measuring preference (Taneva et al., 2008). In a typical DCE, individuals indicate their best choices from among sets of experimentally designed profiles (Louviere, 1998). In contrast to traditional full profile conjoint analyses, DCEs allow for better representation of actual respondent behavior. Individuals are generally not asked to rate or rank alternatives based on preferences. They simply choose their most preferred alternative from a choice set. The probability of choosing a preferred alternative from among a set of alternatives is generally modeled using a multinomial logit (MNL) model.

Although DCEs reflect actual choice behavior, they are not without disadvantages. DCEs

require a more complicated experimental design in which more choice sets are presented to the respondent. In addition, compared to traditional full profile conjoint analyses, each choice task in DCEs reveals less information since little information is gained about alternatives which were not chosen.

2.5 Best-Worst Discrete Choice Experiments

A special type of DCE is the best-worst discrete choice experiment (BWDCE), in which individuals are shown a set of alternatives and asked to indicate their most preferred alternative (best) and least preferred alternative (worst). Louviere et al. (2008) orginally developed the BWDCE (also referred to as multi-attribute best worst scaling tasks) to elicit additional preference information per choice set. In their version of the BWDCE, individuals are asked to indicate their best choice and their worst choice compared to best choice only, first in the entire choice set and then in each successively smaller subset of unranked profiles until a full ranking is attained. Parameters in models for ranked items have been estimated using the *rank ordered logit* or *exploded logit model* (Chapman and Staelin, 1982) or the sequential best worst multinomial logit model introduced by Lancsar and Louviere (2008).

In the PROSPECT study, patients are presented with choice sets containing four profiles describing treatment related outcomes and asked to indicate their best and worst choices within each set. The two mid-ranked profiles are unranked with respect to each other. To handle ties or incomplete rankings, Allison and Christakis (1994) suggested marginalizing over all possible permutations of unranked items. This idea was implemented by Hernandez-Alava et al. (2013) who introduced a model for partially ranked data that includes random effects estimated using Monte Carlo maximum likelihood methods. We develop a method of implementing this idea and incorporate it into the analysis of the PROSPECT study data using Bayesian methods.

CHAPTER 3

Data Set

This chapter describes the data set, the development of the health state profiles, and the experimental design of the applied study.

3.1 Description of the Data

To understand people's preferences, two types of data reflecting individuals' choices may be elicited: revealed preference data and stated preference data. Revealed preference data reflect actual choices made by individuals in a real-world setting while stated preference data are collected in experimental settings and are elicited as responses to hypothetical, but realistic, choice senarios as presented in a DCE. Stated preference data were collected for the prostate cancer project and this data was analyzed in this dissertation.

The PROSPECT study data originates from a randomized trial designed to compare three methods for assessing preference for health states after prostate cancer treatment: discrete choice experiment, time trade-off, and rating scale. The dataset that we use for this proposal comprises data from 121 men recruited from the West Los Angeles Veteran's Administration Medical Center (WLA VA), the Veteran's Administration Sepulveda Ambulatory Care Center and Olive View-UCLA Medical Center. All men had negative prostate biopsies within one month of enrollment and were randomized into two arms; either the *DCE and rating scale* arm or the *DCE and time trade-off* arm. In both arms, patients were presented with tasks for both preference assessments methods, where the order of preference assessment methods was random. Patients who could not read or speak English were excluded from the study. Table 3.1 describes the sample of 121 men. Approximately half

of the sample were at least 65 years old, and more than half of the sample were non-white, partnered, unemployed, non-smokers or had at least some college education. The majority of the sample made at least $10,000 dollars per year.

Table 3.1: Patient characteristics for 121 men in the PROSPECT study

Patient Characteristic	Category	N (%)
Age		
	GTE 65 years old	59 (0.49)
	LT 65 years old	62 (0.51)
Race		
	White	51 (0.42)
	Black	43 (0.36)
	Other	27 (0.22)
Partnered		
	Yes	81 (0.67)
	No	40 (0.33)
Employed		
	Yes	42 (0.35)
	No	79 (0.65)
Smoker		
	Yes	18 (0.15)
	No	103 (0.85)
Education		
	At least some college	97 (0.80)
	At most high school	24 (0.20)
Income		
	LT 10k USD	18 (0.15)
	10k-30k	53 (0.44)
	GT 30k	50 (0.41)

The discrete choice method for the PROSPECT study is a best-worst discrete choice experiment where a choice task for an individual involves choosing their best choice and their worst choice from a set of four hypothetical health states. In the time trade-off method, patients are presented with a single profile describing a hypothetical current health state and a number line and are asked to indicate the number of years of life in better or perfect health that would be equivalent to the number of years of life in the current hypothetical health state. The rating scale method involves presenting a linear rating scale ranging from 0 (worst possible outcome equivalent to death) to 100 (perfect health) on which individuals indicate their rating for a given health state.

The data used for analysis in this work will pertain only to the discrete choice application and consist of choices from among hypothetical health states which could result from various cancer treatments. Since the study is currently recruiting subjects, the data used in preliminary work for this proposal are only a portion of the data that will be gathered by the conclusion of data collection.

3.2 Development of Health State Profiles

Profiles for health states were developed using the Voice of the Patient Process with a group of seventeen men with localized prostate cancer recruited from the WLA VA for in-person interviews (Saigal and Dahan, 2012). The Voice of the Patient Process is a multistep approach which begins by eliciting important issues from patients regarding their prostate cancer treatments (Dahan and Saigal, 2012). Quotations from patient interviews were transcribed and narrowed by researchers into a smaller set of quotations. Each patient then grouped the quotations into piles that he perceived as similar. For example, "Take charge of your body, take charge of the situation" and "Cancer kills...Do something about it" might be grouped in the same pile. Following the interviews, groupings were evaluated across patients by investigators, who developed the final set of treatment related attributes and levels.

Seven attributes and levels were defined. These are presented in Table 3.2. The attributes include sexual functioning, urinary incontinence, bowel issues, lifespan, others' support, active and cutting. *Active* refers to taking immediate action towards treatment. All attributes were defined by two levels except for sexual functioning and urinary incontinence which were defined by three levels. Hypothetical health state profiles were derived by varying the levels of the seven attributes.

3.3 Experimental Design

The computer-based survey instrument designed for the DCE was developed in Excel by Ely Dahan, PhD, MBA (Dahan and Saigal, 2012). The DCE application was designed to

Table 3.2: Attributes and attribute levels from the PROSPECT Study.

Attributes	Level 1	Level 2	Level 3
Lifespan	Live 5 years fewer than my expected lifespan	Live my expected lifespan	
Bowel Issues	Short term urgent and frequent bowel movements	No bowel issues	
Cutting	Treatment requires surgery with some risks and hospital time	Treatment does not require surgery	
Taking Action	I am not jumping into a radical treatment	I am taking action immediately	
Others' Support	My doctor and family do not favor this treatment	My doctor and family support this treament	
Urinary Incontinence	Long term issues	Short term issues	No urinary issues
Sexual Functioning	Unable to engage in sex	Sex life decreased	Sex life same as before treatment

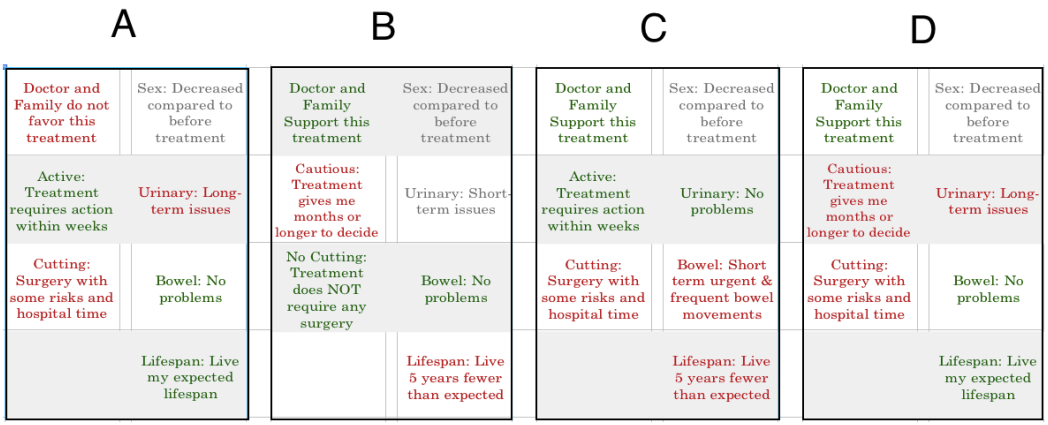

Figure 3.1: Example of a choice set from the PROSPECT study. Patients choose their most and least preferred health state from among the four health states.

present each patient with a series of choice sets each consisting of four health state profiles. Individuals were then asked to identify their best choice and their worst choice in the set presented. No profile was repeated within a choice set and no choice set was repeated in the experiment. Individuals were not given the option to opt out of identifying best and worst choices. A representative screen from the survey is shown in Figure 3.1.

With five 2-level attributes and two 3-level attributes, there are $2^5 * 3^2 = 288$ possible health states in a full factorial design. Sixteen profiles were selected for creation of choice sets because they formed an approximately orthogonal array. These sixteen profiles described by their attribute levels are presented in Table 3.3. In an orthogonal experimental design, the coded attribute levels of the experiment form a set of mutually orthogonal non-zero vectors which are statistically independent. This property is desirable because it allows for an independent determination of each attribute's influence upon the observed choices.

Choice sets of size four were formed using these sixteen identified profiles. The first four choice sets presented were the same for all patients. The selection of subsequent choice sets, determined by the software's algorithm, was dependent on the patient's prior responses. Choice set 5 was constructed by comparing the best choices selected from the first four choice sets while choice set 6 was constructed by comparing the worst choices from the first four choice sets. Choice sets 7 and 8 compared the unranked profiles from the first four choice sets. The remaining choice sets were formed by randomly pairing profiles which had not yet

Table 3.3: Sixteen health state profiles described by attribute levels and utilized in the PROSPECT study where lower attribute levels indicate more side effects, less support, not taking action or surgery.

Profile	Lifespan	Bowel Issues	Cutting	Action	Support	Urinary	Sex
1	2	2	2	2	2	3	3
2	2	2	2	1	2	2	2
3	2	1	2	2	2	1	1
4	2	1	2	1	2	2	2
5	2	1	1	2	1	2	3
6	2	1	1	1	1	3	2
7	2	2	1	2	1	2	1
8	2	2	1	1	1	1	2
9	1	2	2	1	1	1	3
10	1	2	2	2	1	2	2
11	1	1	2	1	1	3	1
12	1	1	2	2	1	2	2
13	1	1	1	1	2	2	3
14	1	1	1	2	2	1	2
15	1	2	1	1	2	2	1
16	1	2	1	2	2	3	2

been compared. Pairs of cards which had been ranked relative to each other or for which a ranking could be inferred were considered a *resolved pair*. In addition to pair resolutions made directly by the individual, pairs are also resolved by an algorithm using transitivity of preference. For example, if $A > B$ and $B > C$ then $A > C$, where $>$ indicates the better choice. The algorithm for choice set creation stopped when all possible paired comparisons were resolved. The men were presented with a total of 10-17 choice sets each consisting of 4 health state profiles. The number of choice sets presented varied across patients because of the adaptive design of the DCE.

CHAPTER 4

Current Models for Discrete Choice Data

This chapter reviews current models that are commonly fit to discrete choice data and describes two methods used to estimate model parameters.

4.1 The Random Utility Model

In economics, discrete choice models are based on the theory of utility maximization, where utility is defined as the total satisfaction received from consuming a good or service. In the context of our project, we assume that patients choose among alternative health states to maximize their utility.

The analytic framework for discrete choice modeling is based on Lancaster's theory of value, where utility is derived from the underlying characteristics or attributes (Ryan et al., 2008), and on the Random Utility Model, where utility has a systematic and a random component (Lancsar and Louviere, 2008). Although utility is not directly observable, it can be estimated from observed choices. The random component may result from unobserved attributes, variations in tastes, or measurement error (Viney et al., 2002).

We begin by describing the modeling framework for best choice. For exposition, we assume that each attribute has two levels coded using dummy coding. Suppose that we have N respondents. Further suppose that each respondent i, $i = 1, ..., N$, has to make a choice from a single choice set containing J alternatives. Index the J alternatives by $j = 1, ..., J$. Let Y_i represent individual i's preferred (best) choice among the J alternatives. Then the

basic problem is the estimation of a utility function

$$U_{ij} = f_{ij}(x_{ij1}, ..., x_{ijH}) \tag{4.1}$$

where U_{ij} denotes the utility of alternative j for individual i, $(x_{ij1}, ..., x_{ijH})$ denotes the levels of the H attributes of alternative j presented to individual i and f_{ij} indicates that the utility U_{ij} is a function of the H attribute levels $(x_{ij1}, ..., x_{ijH})$, that in the most general case could be specific to individual i and alternative j. The random utility model assumes that utility can be partitioned into a systematic component and a random component. Thus, the utility of alternative j for individual i can be written as

$$U_{ij} = V_{ij} + \epsilon_{ij} \tag{4.2}$$

where V_{ij} represents the systematic component and ϵ_{ij} represents the random component. If we model the systematic component of utility as a linear function of parameters

$$V_{ij} = \boldsymbol{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} \tag{4.3}$$

where $\boldsymbol{x}_{ij}$ is the $H \times 1$ attribute vector of the jth alternative for individual i and $\boldsymbol{\beta}$ is the $H \times 1$ coefficient vector of the fixed attribute effects in the valuation of alternative j across all individuals, then we can write the utility function of alternative j for individual i as

$$U_{ij} = \boldsymbol{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} + \epsilon_{ij}. \tag{4.4}$$

4.2 Best Choice & Fixed Effects: The Multinomial Logit Model

4.2.1 The Gumbel Distribution

The random component of the random utility model is commonly assumed to be independent and identically distributed (iid) with a Gumbel distribution, also called the Extreme Value Type I distribution. Extreme value distributions arise as the limiting distribution for an

extreme value (maximum or minimum) of a sample of iid random variables. The Gumbel distribution has been used in engineering and hydrology to measure annual flood flows (Kotz and Nadarajah, 2000).

If X has a Gumbel distribution with location parameter μ and scale parameter η, then the probability density function is given by

$$f_X(x|\mu,\eta) = \frac{1}{\eta}\exp\Big(-\frac{x-\mu}{\eta}\Big)\exp\Big(-\exp\Big(-\frac{x-\mu}{\eta}\Big)\Big) \tag{4.5}$$

and the cumulative distribution function is given by

$$F_X(x|\mu,\eta) = \exp\Big(-\exp\Big(-\frac{x-\mu}{\eta}\Big)\Big) \tag{4.6}$$

where x, μ and η are real-valued ($x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\eta \in \mathbb{R}$) and $\eta > 0$. The mean of the Gumbel distribution is

$$E(X) = \mu + \lambda\eta \tag{4.7}$$

where λ is the Euler-Mascheroni constant ≈ 0.5772, the mode is μ and the variance is

$$\mathrm{Var}(X) = \frac{\pi^2\eta^2}{6} \tag{4.8}$$

where $\pi \approx 3.1416$. Using the standard Gumbel distribution with location parameter $\mu = 0$ and scale parameter $\eta = 1$, the probability density function and the cumulative distribution function are given by

$$f_X(x) = \exp(-x)\exp(-\exp(-x)) \tag{4.9}$$

and

$$F_X(x) = \exp(-\exp(-x)), \tag{4.10}$$

respectively, $E(X) = \lambda$ and $\text{Var}(X) = \pi^2/6$. We now derive the logit model choice probability, the probability respondent i chooses alternative j as best from a choice set containing J alternatives.

4.2.2 Derivation of the Logit Model Choice Probability

If we assume that the random components of utility, the ϵ_{ij}'s, are iid with a standard Gumbel distribution, then the probability that alternative j is selected by individual i as the best choice in a choice set is

$$p(Y_i = j) = p(U_{ik} \leqslant U_{ij}, \text{ for all } k \neq j) \tag{4.11}$$

$$= p(V_{ik} + \epsilon_{ik} \leqslant V_{ij} + \epsilon_{ij}, \text{ for all } k \neq j) \tag{4.12}$$

$$= p(\epsilon_{ik} \leqslant V_{ij} - V_{ik} + \epsilon_{ij}, \text{ for all } k \neq j). \tag{4.13}$$

Suppose that ϵ_{ij} is given. Then the conditional probability of alternative j being selected as best by individual i given ϵ_{ij}

$$p(Y_i = j|\epsilon_{ij}) = p(\epsilon_{ik} \leqslant V_{ij} - V_{ik} + \epsilon_{ij}, \text{ for all } k \neq j) \tag{4.14}$$

is the cumulative distribution for each ϵ_{ik} evaluated at $V_{ij} - V_{ik} + \epsilon_{ij}$. Because the ϵ_{ik}'s are assumed to be independent, the cumulative distribution over all $k \neq j$ is the product of the individual cumulative distributions,

$$p(Y_i = j|\epsilon_{ij}) = \prod_{k=1,k\neq j}^{J} p(\epsilon_{ik} \leqslant V_{ij} - V_{ik} + \epsilon_{ij}) \tag{4.15}$$

$$= \prod_{k=1,k\neq j}^{J} \exp\{-\exp[-(V_{ij} - V_{ik} + \epsilon_{ij})]\}. \tag{4.16}$$

Because the ϵ_{ij}'s are not given, the marginal choice probability is the integral of $p(Y_i = j|\epsilon_{ij})$ over all values of ϵ_{ij} weighted by the density, $f(\epsilon_{ij})$. Thus

$$\begin{aligned} p(Y_i = j) &= \int_{-\infty}^{\infty} p(Y_i = j|\epsilon_{ij}) f(\epsilon_{ij}) d\epsilon_{ij} \\ &= \int_{-\infty}^{\infty} \prod_{k=1,k\neq j}^{J} \exp\{-\exp[-(V_{ij} - V_{ik} + \epsilon_{ij})]\} \\ &\quad \cdot \exp(-\epsilon_{ij}) \exp[-\exp(-\epsilon_{ij})] d\epsilon_{ij}. \end{aligned} \tag{4.17}$$

Let $u = \exp(-\epsilon_{ij})$. Then $du = -\exp(-\epsilon_{ij})d\epsilon_{ij}$ and the limits change from $\epsilon_{ij} = -\infty$ to $u = \infty$ and $\epsilon_{ij} = \infty$ to $u = 0$. Now we have that

$$p(Y_i = j) = -\int_{\infty}^{0} \prod_{k=1,k\neq j}^{J} \Big[\exp\Big\{-u \exp(-(V_{ij} - V_{ik}))\Big\}\Big] \exp(-u) du \tag{4.18}$$

$$= \int_{0}^{\infty} \prod_{k=1,k\neq j}^{J} \Big[\exp\Big\{-u \exp(V_{ik} - V_{ij})\Big\}\Big] \exp(-u) du \tag{4.19}$$

$$= \int_{0}^{\infty} \exp\Big[-u\Big\{\sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})\Big\}\Big] \exp(-u) du \tag{4.20}$$

$$= \int_{0}^{\infty} \exp\Big[-u\Big\{1 + \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})\Big\}\Big] du. \tag{4.21}$$

Finally, if we let $w = -u\Big\{1 + \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})\Big\}$, then
$dw = -\Big\{1 + \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})\Big\} du$ and the limits of integration change from $u = \infty$

to $w = -\infty$ and $u = 0$ to $w = 0$. We now have that

$$p(Y_i = j) = \frac{1}{1 + \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})} \int_{-\infty}^{0} \exp(w) dw \tag{4.22}$$

$$= \frac{1}{1 + \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})} \tag{4.23}$$

$$= \frac{1}{1 + \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})} \tag{4.24}$$

$$= \frac{\exp(V_{ij})}{\exp(V_{ij}) + \exp(V_{ij}) \sum_{k=1,k\neq j}^{J} \exp(V_{ik} - V_{ij})} \tag{4.25}$$

$$= \frac{\exp(V_{ij})}{\exp(V_{ij}) + \sum_{k=1,k\neq j}^{J} \exp(V_{ik})} \tag{4.26}$$

$$= \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}. \tag{4.27}$$

Thus, the logit model choice probability, the probability respondent i chooses alternative j as best from a choice set containing J alternatives, is calculable in closed form as

$$p(Y_i = j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}. \tag{4.28}$$

If we model the systematic component of utility as a linear function of parameters

$$V_{ij} = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} \tag{4.29}$$

where $\boldsymbol{x}_{ij}$ is the $H \times 1$ attribute vector of the jth alternative for individual i and $\boldsymbol{\beta}$ is the $H \times 1$ coefficient vector of unknown fixed attribute effects in the valuation of alternative j across all individuals, then

$$p(Y_i = j) = \frac{\exp(\boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta})}{\sum_{k=1}^{J} \exp(\boldsymbol{x}_{ik}^{\mathsf{T}} \boldsymbol{\beta})}. \tag{4.30}$$

From our derivation, we see that the Gumbel distribution leads to the logit choice probability which forms the basis of the multinomial logit model, a model often chosen for its mathematical convenience.

We will model the best choice as the most preferred choice among the four alternatives in a choice set and we will model the worst choice as the least preferred choice among the four alternatives in a choice set.

Define $\delta_{ij} = 1$ if individual i chooses alternative j and $\delta_{ij} = 0$ otherwise and let $p_{ij} = p(Y_i = j)$ be the probability that individual i chooses alternative j as their best choice, then the likelihood contribution for individual i choosing from a single choice set is given by

$$L_i(\boldsymbol{\beta}|y_i) = p_{i1}^{\delta_{i1}} p_{i2}^{\delta_{i2}} ... p_{iJ}^{\delta_{iJ}} \tag{4.31}$$

where J indicates the total number of alternatives and y_i is the observed choice for individual i. Assuming that all individuals are selecting a best choice from a single choice set, the likelihood for a sample of N individuals is

$$L(\boldsymbol{\beta}|y) = \prod_{i=1}^{N} L_i(\boldsymbol{\beta}|y_i) \tag{4.32}$$

where y is the set of observed choices across all patients. The log-likelihood is given by

$$\log L(\boldsymbol{\beta}|y) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \log(p_{ij}) \tag{4.33}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \log\Big(\frac{\exp(\boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta})}{\sum_{k=1}^{J} \exp(\boldsymbol{x}_{ik}^{\mathsf{T}} \boldsymbol{\beta})} \Big). \tag{4.34}$$

4.2.3 The Multinomial Logit Model for Panel Data

In the PROSPECT study, each patient was presented with multiple choice sets from which to make choices. Because each choice task completed by a patient contributes a single observation, the data are a collection of repeated observations for each patient. This type of data is sometimes called panel data (Hsiao, 2003). Because the design of the experiment was adaptive, as described in Chapter 3, the total number of choice sets presented to a patient differed for each respondent. Let T_i be the total number of choice sets presented to individual i in the course of the experiment. With the exception of the first four choice sets, the choice

set presented at a specific time also differed for each respondent. Thus, the J alternatives in a choice set presented at a specific time differed across all patients. If we index choice sets by t, where $t = 1, 2, 3, ..., T_i$, then U_{itj}, the utility of alternative j for individual i in choice set t, is given by

$$U_{itj} = V_{itj} + \epsilon_{itj} \tag{4.35}$$

where V_{itj} is the systematic component of utility and ϵ_{itj} is the random component of utility for alternative j presented to individual i in choice set t. If ϵ_{itj} is iid with a standard Gumble distribution for all i, j and t, then using a similar derivation as in Section 4.2.2, the probability respondent i chooses alternative j as best from a choice set t containing J alternatives is given by

$$p(Y_{it} = j) = \frac{\exp(V_{itj})}{\sum_{k=1}^{J} \exp(V_{itk})}. \tag{4.36}$$

Finally, if we model the systematic component of utility as a linear function of parameters

$$V_{itj} = \boldsymbol{x}_{itj}^{\mathsf{T}} \boldsymbol{\beta} \tag{4.37}$$

where $\boldsymbol{x}_{itj}$ is the $H \times 1$ attribute vector of the jth alternative for individual i in choice set t and $\boldsymbol{\beta}$ is the $H \times 1$ coefficient vector of unknown fixed attribute effects in the valuation of alternative j for all individuals, then the probability of individual i choosing alternative j in choice set t as best choice can be written as

$$p(Y_{it} = j) = \frac{\exp(\boldsymbol{x}_{itj}^{\mathsf{T}} \boldsymbol{\beta})}{\sum_{k=1}^{J} \exp(\boldsymbol{x}_{itk}^{\mathsf{T}} \boldsymbol{\beta})} \tag{4.38}$$

where Y_{it} is the observed best alternative in choice set t for individual i. Define $\delta_{itj} = 1$ if individual i chooses alternative j in choice task t and $\delta_{itj} = 0$ otherwise and let $p_{itj} = p(Y_{it} = j)$ be the probability that individual i chooses alternative j in choice set t as their

best choice. Then the likelihood function for individual i choosing in choice set t is given by

$$L_{it}(\boldsymbol{\beta}|y_{it}) = p_{it1}^{\delta_{it1}} p_{it2}^{\delta_{it2}} ... p_{itJ}^{\delta_{itJ}} \tag{4.39}$$

where J is the total number of alternatives in choice set t and y_{it} is the observed choice for individual i in choice set t. Because each respondent is presented with a varying number of choice sets, the likelihood contribution for individual i making their choices in the course of the experiment is given by

$$L_i(\boldsymbol{\beta}|y_i) = \prod_{t=1}^{T_i} L_{it}(\boldsymbol{\beta}|y_{it}) \tag{4.40}$$

where y_i is the set of observed choices for individual i across all choice sets. The likelihood for the full sample of N individuals is

$$L(\boldsymbol{\beta}|y) = \prod_{i=1}^{N} L_i(\boldsymbol{\beta}|y_i) \tag{4.41}$$

where y is the set of observed choices and the log-likelihood is given by

$$\log L(\boldsymbol{\beta}|y) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} \delta_{itj} \log(p_{itj}) \tag{4.42}$$

$$= \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} \delta_{itj} \log\Big(\frac{\exp(\boldsymbol{x}_{itj}^{\mathsf{T}} \boldsymbol{\beta})}{\sum_{k=1}^{J} \exp(\boldsymbol{x}_{itk}^{\mathsf{T}} \boldsymbol{\beta})} \Big). \tag{4.43}$$

4.2.4 The Property of Independence of Irrelevant Alternatives

A property which results from deriving the multinomial logit choice probabilities as described in (4.38) is the property of independence of irrelevant alternatives (IIA). For any two alternatives j and k, the ratio of the logit probabilities is

$$\frac{p(Y_{it} = j)}{p(Y_{it} = k)} = \frac{\exp(\boldsymbol{x}_{itj}^{\mathsf{T}} \boldsymbol{\beta})}{\exp(\boldsymbol{x}_{itk}^{\mathsf{T}} \boldsymbol{\beta})} = \exp[(\boldsymbol{x}_{itj} - \boldsymbol{x}_{itk})^{\mathsf{T}}] \boldsymbol{\beta}. \tag{4.44}$$

This implies that the choice of one alternative over another does not depend on any remaining alternatives included in the set. In other words, choosing j over k is *independent* of the set that contains j and k and the remaining alternatives are *irrelevant* in the choice between them.

The property of IIA has been identified as a limitation of the MNL model. In cases where choice sets contain similar alternatives or alternatives with a natural order, the property of IIA can be violated (Hernandez-Alava et al., 2013). One well-known example where the property of IIA is inappropriate is called the *red-bus, blue-bus paradox* and involves a choice set that contains two similar alternatives.

Suppose that an individual needs to choose a mode of transportation from a choice set containing two choices:

$$A = \{\text{car}, \text{red bus}\}$$

and that $p_A(Y_i = \text{car}) = 0.7$ and $p_A(Y_i = \text{red bus}) = 0.3$. Now suppose that a blue bus is introduced as a new mode of transportation. Then we now have the choice set:

$$A' = \{\text{car}, \text{red bus}, \text{blue bus}\}$$

Assuming that color does not influence choice of transportation,

$$p_{A'}(Y_i = \text{red bus}) = p_{A'}(Y_i = \text{blue bus}) \tag{4.45}$$

and intuitively we would think that $p_{A'}(Y_i = \text{car}) = 0.7$, $p_{A'}(Y_i = \text{red bus}) = 0.15$ and $p_{A'}(Y_i = \text{blue bus}) = 0.15$. But because of the IIA property, the odds of selecting the car over the red bus does not depend on whether the blue bus is in the choice set or not. According to the IIA property,

$$\frac{p_{A'}(Y_i = \text{car})}{p_{A'}(Y_i = \text{red bus})} = \frac{p_A(Y_i = \text{car})}{p_A(Y_i = \text{red bus})} = \frac{0.7}{0.3} = 2.33. \tag{4.46}$$

Because

$$p_{A'}(Y_i = \text{car}) + p_{A'}(Y_i = \text{red bus}) + p_{A'}(Y_i = \text{blue bus}) = 1 \tag{4.47}$$

equations (4.45) and (4.46) imply that

$$2.33 \cdot p_{A'}(Y_i = \text{red bus}) + 2 \cdot p_{A'}(Y_i = \text{red bus}) = 1, \tag{4.48}$$

and $p_{A'}(Y_i = \text{red bus}) = 0.23 = p_{A'}(Y_i = \text{blue bus})$ and $p_{A'}(Y_i = \text{car}) = 0.54$. We summarize the results in the following table.

Table 4.1: Illustrative results for the red-bus, blue-bus paradox

Mode of transportation	Intuition	IIA
Car	0.7	0.54
Red Bus	0.15	0.23
Blue Bus	0.15	0.23

If one considers the situation where a choice is made between a car and a red bus, and then a blue bus is introduced, because the blue bus is functionally like the red bus, its introduction should draw commuters from primarily the red bus and not from the car. But, as we can see in Table 4.1, assuming that the IIA property holds in the situation where we have similiar alternatives in a choice set we can mis-predict the probability of choice for each of the alternatives. Formal tests have been developed by Hausman and McFadden (1984) to test the IIA property, but Cheng and Long (2007) have shown in a simulation study that the Hausman-McFadden test performs rather poorly, even in large samples and conclude that the test is unsatisfactory for applied work. The IIA property has the potential to be violated in the PROSPECT study, however, in the absence of a well performing test of the IIA property and becasue MNL-based models are commonly fit to discrete choice data, we fit a MNL-based model to the data from the PROSPECT study.

4.3 Best Choice & Random Effects: The Mixed Logit Model

The Mixed Logit (MXL), also sometimes called the Mixed Multinomial Logit Model or the Random-Parameters Logit Model, extends the MNL model by including random effects. It was first described for revealed preference data by Boyd and Mellmand (1980) and Cardell and Dunbar (1980) and later introduced for discrete choice responses by McFadden and Train in 2000 (McFadden and Train, 2000). Increased application of the MXL model occurred with the development of simulation methods which allowed for better estimation of the model. In the mid-1990's, such methods were integrated into software packages and the application of the MXL model increased (Hensher and Greene, 2003). Following its introduction, the MXL model has been applied in a variety of areas, some of which include: transportation (Ben-Akiva and Bolduc, 1996; Brownstone and Train, 1999), willingness to pay (Giergiczny et al., 2012), multiparty elections and food choices (Rigby and Burton, 2006).

Like the MNL model, a MXL model assumes that the error terms are iid according to a Gumbel distribution. However, a MXL model relaxes the restriction that the coefficient vector be fixed for all individuals, which allows one to model heterogeneity or variation in taste by allowing coefficients to vary across individuals (Revelt and Train, 1998). In the MXL model, each individual has their own coefficient vector, $\boldsymbol{\beta}_i$, meaning that each individual has different regression coefficients, also called partial utilities or part-worths, for each attribute. Because the MXL model assumes that differences across patients have some influence on the selection of best choice, the MXL model is a random effects model where $\boldsymbol{\beta}_i$ is the coefficient vector of random attribute effects.

4.3.1 The Mixed Logit Model for Panel Data

The MXL model can be applied to panel data where multiple observations are collected for each individual. In the MXL model, the utility of alternative j for individual i in choice task t is

$$U_{itj} = \boldsymbol{x}_{itj}^{\mathsf{T}} \boldsymbol{\beta}_i + \epsilon_{itj} \tag{4.49}$$

where $\boldsymbol{x}_{itj}$ is the $H \times 1$ attribute vector for alternative j in choice set t, $\boldsymbol{\beta}_i$ is the $H \times 1$ vector of unknown coefficients for respondent i and ϵ_{itj} is a random error term that is iid Gumbel and independent of $\boldsymbol{\beta}_i$. The coefficients vary in the population with density $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the parameters of this population distribution. This model specification is similar to the MNL model for panel data except in this model, the coefficient vector is now allowed to vary over respondents rather than being fixed. Thus, conditional on $\boldsymbol{\beta}_i$, the probability that respondent i chooses alternative j as the best alternative in choice set t is

$$p(Y_{it} = j|\boldsymbol{\beta}_i) = \frac{\exp(\boldsymbol{x}_{itj}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{k=1}^{J}\exp(\boldsymbol{x}_{itk}^{\mathsf{T}}\boldsymbol{\beta}_i)}. \tag{4.50}$$

4.4 Full Ranking & Fixed Effects: Rank Ordered Logit Model

Because of the limited amount of data collected from individuals in a traditional discrete choice experiment, where a single alternative is identified as the preferred choice, individual level models were considered inestimable in the past (Finn and Louviere, 1992). Recently, Louviere et al. (2008) showed that individual models could be estimated using a best-worst discrete choice experiment type method with a more efficient experimental design, instead of relying on only best choices collected from a large number of choice sets (Lancsar and Louviere, 2008). We call the model Louviere et al. (2008) to model full ranking the sequential best-worst (SBW).

In a SBW, the goal is to obtain more information from a single choice set without increasing the total number of choice sets presented to an individual. More information is elicited from the respondent by obtaining a full ranking of the alternatives in each choice set. For a given choice set in Louviere's SBW, the respondent is asked to choose the best and the worst preferred choice. Once identified, these selected alternatives are removed from the choice set and the best-worst task is repeated again on the set of the remaining alternatives. This process continues until a full ranking of alternatives is obtained. An advantage of this method is the increased ability to fit individual-level models. In their paper, Lancsar and Louviere were able to estimate individual level models involving three to five alternatives in

a choice set and six to ten attributes (Louviere et al., 2008). Under the same assumptions as the MNL model, it is possible to define a multinomial logit model on ranked alternatives, also called the rank ordered logit (ROL) model, the sequential multinomial logit or the exploded logit model. Under the ROL model, the probability of observing a ranking is defined as the product of multinomial logit probabilities of selecting a best choice from successively smaller choice sets. We compare the probability of a full ranking under the ROL model and SBW model in Appendix C, Statement 3.

4.4.1 Complete Ranking, 1 Individual, 1 Choice Set

Let $p_C(j)$ denote the probability that an alternative j is chosen as best from a set $C = \{1, 2, 3, ...j..., J\}$ of alternatives and consider the finite set of alternatives. Then the ranked data is a set of permutations, $\pi : C \to C$, mapping alternatives to their ranks (Sun et al., 2012). For permutation π, $\pi(j)$ is the rank assigned to item $j \in C$ and $\pi^{-1}(j)$ is the jth most preferred alternative in C. For example, if $\pi(2) = 1$, then alternative 2 is assigned rank 1. Similarly, if $\pi^{-1}(1) = 2$ then the first ranked alternative is alternative 2. Now, let R_C be the set of all possible permutations of the elements in C and let $\pi \in R_C$ denote the complete ranking $\pi^{-1}(1) > \pi^{-1}(2) > ... > \pi^{-1}(J)$. Here the notation, $i > j$, indicates that alternative i is preferred over j. Finally, let $p_C(\pi)$ equal the probability of the ranking π.

According to Louviere et al. (2008), Bergland (1994) and Chapman and Staelin (1982), the probability of a ranking of alternatives can be written as the probability of a sequence of choices. Applying this approach to the random utility model, for an individual i and ranking

of alternatives $\pi_i = (1 > 2 > ... > j > ... > J)$, the probability of observing the ranking π_i is

$$
\begin{aligned}
p_C(\pi_i) &= p_C(1 > 2 > ... > j > ... > J) && (4.51)\\
&= p_C(\pi^{-1}(1) > \pi^{-1}(2) > ... > \pi^{-1}(J)) && (4.52)\\
&= p_C(U_{i1} > U_{i2} > ... > U_{iJ}) && (4.53)\\
&= \prod_{j=1}^{J} p_C(U_{ij} > U_{ik}, \text{ for all } k > j) && (4.54)\\
&= p_C(U_{i1} > U_{ik}, \text{ for all } k > 1) \cdot p(U_{i2} > U_{ik}, \text{ for all } k > 2) \cdot ... && (4.55)\\
&\cdot p(U_{i(J-1)} > U_{ik}, \text{ for all } k > (J-1))\\
&= p_C(\text{alternative 1 is best}) \cdot p_C(\text{alternative 2 is 2nd best}) \cdot ... && (4.56)\\
&\cdot p_C(\text{alternative } J \text{ is last})
\end{aligned}
$$

where U_{ij} is the utility of alternative j for individual i (Velandia et al., 2011). This equation is derived from the Luce-Suppes Ranking Choice Theorem which decomposes the joint probability $p_C(U_{i1} > U_{i2} > ... > U_{iJ})$ into a series of successive and independent events where U_{ij} represents the utility of the most preferred alternative at each stage of decision (Chapman and Staelin, 1982). According to Velandia et al. (2011), the right hand side of (4.54) can be described as the product of the probability that alternative 1 is preferred over all other choices given the entire choice set, times the probability that alternative 2 is preferred over all other choices given that alternative 1 was already chosen and removed from the choice set, and so on.

In general, let π_i be a permutation in R_C, the set of all permutations of alternatives in the choice set C, for individual i where $\pi_i^{-1}(1) > \pi_i^{-1}(2) > \ldots > \pi_i^{-1}(J)$ is a ranking of the alternatives in the choice set, C. Then

$$
\begin{aligned}
p_C(\pi_i) &= p_C\Big(\pi_i^{-1}(1) > \pi_i^{-1}(2) > ... > \pi_i^{-1}(J)\Big) && (4.57)\\
&= \prod_{j=1}^{J-1} p_{\{\pi_i^{-1}(j),...,\pi_i^{-1}(J)\}}(\pi_i^{-1}(j)) && (4.58)
\end{aligned}
$$

where $\pi_i^{-1}(j)$ represents the jth preferred alternative in the choice set, C. Because the probability of choosing one alternative as the least preferred given that all others were already chosen equals one, the last term, $p_{\{\pi_i^{-1}(J)\}}(\pi_i^{-1}(J))$, equals one and is implicitly included in the equation above.

4.4.1.1 Example: The PROSPECT Study

Consider a ranking, π_i, for individual i of four health states in choice set $C = \{1, 2, 3, 4\}$, such that $4 > 3 > 2 > 1$. Because $p(j)$ equals the multinomial logit probability of choosing j from C, if

$$p_C(j) = \frac{\exp(V_{ij})}{\sum_{k \in C} \exp(V_{ik})}, \tag{4.59}$$

then

$$p(\pi_i) = p_{\{1,2,3,4\}}(4) p_{\{1,2,3\}}(3) p_{\{1,2\}}(2) \tag{4.60}$$

$$= \frac{\exp(V_{i4})}{\sum_{k \in \{1,2,3,4\}} \exp(V_{ik})} \frac{\exp(V_{i3})}{\sum_{k \in \{1,2,3\}} \exp(V_{ik})} \frac{\exp(V_{i2})}{\sum_{k \in \{1,2\}} \exp(V_{ik})} \tag{4.61}$$

where V_{ij} is the linear function from the choice model ($V_{ij} = \boldsymbol{x}_{ij}^{\mathsf{T}} \hat{\boldsymbol{\beta}}$) defined in (4.3).

4.4.2 Complete Ranking, N Individuals, T_i Choice Sets for Each Individual

Now suppose that we have N individuals each stating preferences about more than one choice set. Let π_{it} represent the complete ranking of the choice set t, made by individual i, where $\pi_{it}^{-1}(1) > \pi_{it}^{-1}(2) > ... > \pi_{it}^{-1}(J)$, $i = 1, ..., N$, $t = 1, ..., T_i$ and each choice set t contains

exactly J elements. Then the probability of observing the ranking π_{it} is

$$p(\pi_{it}|\boldsymbol{\beta}) = p\Big(\pi_{it}^{-1}(1) > \pi_{it}^{-1}(2) > ... > \pi_{it}^{-1}(J)\Big) \tag{4.62}$$

$$= \prod_{j=1}^{J-1} p_{\{\pi_{it}^{-1}(j),...,\pi_{it}^{-1}(J)\}}(\pi_{it}^{-1}(j)) \tag{4.63}$$

$$= \prod_{j=1}^{J-1} \frac{\exp(V_{i\pi_{it}^{-1}(j)})}{\sum_{J\geq k\geq j}\exp(V_{i\pi_{it}^{-1}(k)})} \tag{4.64}$$

$$= \prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(j)}^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{J\geq k\geq j}\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(k)}^{\mathsf{T}}\boldsymbol{\beta})} \tag{4.65}$$

where $\pi_{it}^{-1}(j)$ represents the jth preferred alternative for individual i in choice set t, $\boldsymbol{x}_{i\pi_{it}^{-1}(j)}$ is the $H \times 1$ attribute vector of the jth ranked alternative for individual i in choice set t and $\boldsymbol{\beta}$ is $H \times 1$ the coefficient vector of the fixed attribute effects.

Now, let $\boldsymbol{\pi_i} = (\pi_{i1}, \pi_{i2}, \pi_{i3}, ..., \pi_{iT_i})$ represent the sequence of rankings made by the respondent over the course of the experiment, where T_i is the total number of choice sets presented to individual i. Then the probability that respondent i makes this sequence of rankings is the product

$$p(\boldsymbol{\pi_i}|\boldsymbol{\beta}) = \prod_{t=1}^{T_i} p(\pi_{it}|\boldsymbol{\beta}) \tag{4.66}$$

$$= \prod_{t=1}^{T_i}\prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(j)}^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{J\geq k\geq j}\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(k)}^{\mathsf{T}}\boldsymbol{\beta})}. \tag{4.67}$$

4.5 Full Ranking & Random Effects: Rank Ordered Mixed Logit Model

A mixed logit model can also be estimated on ranked data. Let $\boldsymbol{\beta}_i$ be the coefficient vector of random attribute effects. If we assume that $\boldsymbol{\beta}_i$ is random and distributed with density $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the parameters of the distribution, then conditional on $\boldsymbol{\beta}_i$, the

probability of observing the ranking of choice set t by individual i, π_{it}, is

$$p(\pi_{it}|\boldsymbol{\beta}_i) = \prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(j)}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{J\geq k\geq j}\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(k)}^{\mathsf{T}}\boldsymbol{\beta}_i)} \tag{4.68}$$

where $\pi_{it}^{-1}(j)$ represents the jth preferred alternative for individual i in choice set t, $\boldsymbol{x}_{i\pi_{it}^{-1}(j)}$ is the $H \times 1$ attribute vector of the jth ranked alternative for individual i in choice set t and $\boldsymbol{\beta}_i$ is the $H \times 1$ coefficient vector of the random attribute effects for individual i.

Let $\boldsymbol{\pi_i} = (\pi_{i1}, \pi_{i2}, \pi_{i3}, ..., \pi_{iT_i})$ represent the sequence of rankings made by the respondent over the course of the experiment, where T_i is the total number of choice sets presented to individual i. Then conditional on $\boldsymbol{\beta}_i$, the probability that respondent i makes this sequence of rankings is the product

$$p(\boldsymbol{\pi_i}|\boldsymbol{\beta}_i) = \prod_{t=1}^{T_i} p(\pi_{it}|\boldsymbol{\beta}_i) \tag{4.69}$$

$$= \prod_{t=1}^{T_i}\prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(j)}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{J\geq k\geq j}\exp(\boldsymbol{x}_{i\pi_{it}^{-1}(k)}^{\mathsf{T}}\boldsymbol{\beta}_i)}. \tag{4.70}$$

For a study where $J = 4$ and conditional on $\boldsymbol{\beta}_i$, the probability of observing a ranking, say $4 > 3 > 2 > 1$, by individual i is given by

$$p(\pi_{it}|\boldsymbol{\beta}_i) = p_{\{1,2,3,4\}}(4)p_{\{1,2,3\}}(3)p_{\{1,2\}}(2) \tag{4.71}$$

$$= \frac{\exp(\boldsymbol{x}_{it4}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{k\in\{1,2,3,4\}}\exp(\boldsymbol{x}_{itk}^{\mathsf{T}}\boldsymbol{\beta}_i)} \frac{\exp(\boldsymbol{x}_{it3}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{k\in\{1,2,3\}}\exp(\boldsymbol{x}_{itk}^{\mathsf{T}}\boldsymbol{\beta}_i)} \frac{\exp(\boldsymbol{x}_{it2}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{k\in\{1,2\}}\exp(\boldsymbol{x}_{itk}^{\mathsf{T}}\boldsymbol{\beta}_i)} \tag{4.72}$$

where X_{itj} is the attribute vector for the jth alternative in choice set t presented to individual i, $\boldsymbol{\beta}_i$ is the coefficient vector of random attribute effects for individual i. Then

$$p(\boldsymbol{\pi_i}|\boldsymbol{\beta}_i) = \prod_{t=1}^{T_i} p(\pi_{it}|\boldsymbol{\beta}_i) \tag{4.73}$$

and the unconditional probability is the integral of this product over the density of $\boldsymbol{\beta}_i$

$$p(\boldsymbol{\pi_i}) = \int p(\boldsymbol{\pi_i}|\boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i|\theta) d\boldsymbol{\beta}_i. \tag{4.74}$$

4.6 Partial Ranking & Fixed Effects

Thus far we have only considered the situation where we elicit a best choice or a full ordering of alternatives. However, it is possible to obtain a partial ordering. The respondent might have difficulty ranking alternatives and thus, leave some alternatives unranked, or the design of the experiment may lend itself to partial rankings. If we consider the data collected in the PROSPECT study, only the best and worst choices from a choice set of four alternative health states are elicited from a respondent. Let $C = \{1, 2, 3, 4\}$ represent a choice set containing four alternative health states presented to a single individual. Suppose that health state 1 is identified as the best choice in the set and health state 4 is identified as the worst choice in the set, which we denote by $1 > \{2, 3\} > 4$. In this case, the middle alternatives, 2 and 3, are unranked relative to each other. The likelihood function for the rank ordered logit model requires data with rankings starting from the most preferred to least preferred choices in a sequential order. Allison and Christakis (1994) proposed an alternative likelihood for ties and incomplete rankings. They assumed that respondents have a preference among the unranked items, e.g., $2 > 3$ or $2 < 3$, and these unobserved events are necessarily mutually exclusive. To handle incomplete or partial rankings, Allison and Christakis (1994) suggest marginalizing over all possible permutations of unranked items. We generalize their discussion below. Hernandez-Alava et al. (2013) published a general version of the model presented below, however did not implement the model on a real dataset in the Bayesian setting.

4.6.1 Partial Ranking, 1 Individual, 1 Choice Set

We begin this section with an example. We first consider the situation where we have a single individual presented with a single choice set. For simplicity, we suppress the indices

representing individuals, i, and choice sets, t. We replace the indices later when generalizing the example to multiple individuals and choice sets.

4.6.1.1 The PROSPECT Study: 4 Alternatives, 1 Best, 1 Worst

Let $\pi^{part} = (1 > \{2, 3\} > 4)$ be a partial ranking of a choice set containing four health state profiles for an individual such that 1 is chosen as best and 4 is chosen as worst. In this example, it is unknown how 2 and 3 rank relative to each other. We have two possible cases: one where $2 > 3$ and one where $3 > 2$. Thus, π^{part} implies two possible rankings: $\pi_1 = 1 > 2 > 3 > 4$ and $\pi_2 = 1 > 3 > 2 > 4$. Because the two cases are mutually exclusive events, the probability of one event or the other occurring is the sum of the probability of each event, i.e., $p(\pi_1 \text{ or } \pi_2) = p(\pi_1) + p(\pi_2)$. Thus

$$
\begin{aligned}
p(\pi^{part}) &= p(\pi_1 \text{ or } \pi_2) && (4.75)\\
&= p(\pi_1) + p(\pi_2) && (4.76)\\
&= p(1 > 2 > 3 > 4) + p(1 > 3 > 2 > 4) && (4.77)\\
&= \frac{\exp(\boldsymbol{x}_1^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{k\in\{1,2,3,4\}} \exp(\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{\beta})} \left(\frac{\exp(\boldsymbol{x}_2^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{k\in\{2,3,4\}} \exp(\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{\beta})} \frac{\exp(\boldsymbol{x}_3^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{k\in\{3,4\}} \exp(\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{\beta})} \right. && (4.78)\\
&\quad \left. + \frac{\exp(\boldsymbol{x}_3^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{k\in\{2,3,4\}} \exp(\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{\beta})} \frac{\exp(\boldsymbol{x}_2^{\mathsf{T}}\boldsymbol{\beta})}{\sum_{k\in\{2,4\}} \exp(\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{\beta})} \right)
\end{aligned}
$$

where $\boldsymbol{x}_j$ is the $H \times 1$ attribute vector of the jth alternative, $j = 1, 2, 3, 4$, for the individual and $\boldsymbol{\beta}$ is the $H \times 1$ coefficient vector of the fixed attribute effects across all individuals.

4.6.1.2 General Discussion: J Alternatives, b Best, w Worst

In the example above, we described the case in which we had an incomplete ranking of the alternatives in the finite choice set, $C = \{1, 2, 3, 4\}$, where the best alternative and the worst alternative were chosen leaving two *mid-ranked* alternatives in the entire choice set unranked

with respect to each other

$$\pi_{part} = \Big(\boxed{\pi^{-1}(1)} > \boxed{\{...\}} > \boxed{\pi^{-1}(4)}\Big). \tag{4.79}$$

Let $\pi : \{1, 2, 3, ...j..., J\} \rightarrow \{1, 2, 3, ...j..., J\}$ be a mapping of the set of alternatives to the set of their ranks. Then $\pi(j)$ is the rank assigned to item j and $\pi^{-1}(j)$ is the jth most preferred alternative in the choice set. Now, suppose an incomplete ranking of the alternatives of the finite choice set, $C = \{1, 2, ..., J\}$ exists, where the b best alternatives and the w worst alternatives are chosen leaving the mid-ranked $J - w - b$ alternatives unranked with respect to each other. We now have the partial ranking

$$\begin{aligned}\pi^{part} = \Big(\boxed{\pi^{-1}(1) > ... > \pi^{-1}(b)} > \boxed{\{\pi^{-1}(b+1)...\pi^{-1}(J-w)\}} \\ > \boxed{\pi^{-1}(J-w+1) > ... > \pi^{-1}(J)}\Big)\end{aligned} \tag{4.80}$$

where the middle $J-w-b$ alternatives, $\{\pi^{-1}(b+1)...\pi^{-1}(J-w)\}$, are unranked with respect to each other. If we assume that a preference order among choice set alternatives exists, then there are $(J - w - b)!$ possible full rankings which are consistent with the partial ranking.

4.6.1.3 General Discussion: Bests, Worsts and Some Middles Ranked

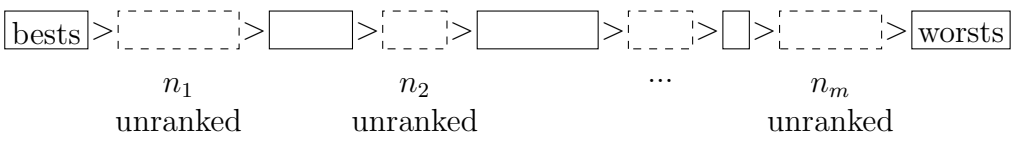

Figure 4.1: A General Partial Ranking

More generally, if we consider the case where we have a more general partial ranking for a single individual and a single choice set (the *bests*, the *worsts* and some of the *middle* alternatives are ranked), then we have the situation illustrated in Figure 4.1, where ranked alternatives are depicted by solid boxes and unranked alternatives are depicted by dashed boxes. If we suppose that there are m sets of unranked *middle* alternatives and if we let n_k be the number of alternatives in unranked set k, $k = 1, 2, 3...m$, then there are $(n_1!)(n_2!)\cdot....\cdot(n_m!)$

possible full rankings which are consistent with the partial ranking described in Figure 4.1

$$\pi_1, \pi_2, ..., \pi_R \tag{4.81}$$

where $R = (n_1!)(n_2!) \cdot ... \cdot (n_m!)$, the total number of possible full rankings which agree with the partial ranking. If we assume a preference ranking exists and that only one can hold, the R full rankings which agree with the partial ranking are mutually exclusive events and the probability of any event occurring is the sum of the probability of the individual events. If we have an incomplete ranking, π^{part}, of the alternatives of the finite choice set, $C = \{1, 2, ..., J\}$, then

$$p(\pi^{part}) = \sum_{r=1}^{R} p(\pi_r) \tag{4.82}$$

$$= \sum_{r=1}^{R} \left(\prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{r\pi_r^{-1}(j)}^{\mathsf{T}} \boldsymbol{\beta})}{\sum_{J \geq k \geq j} \exp(\boldsymbol{x}_{r\pi_r^{-1}(k)}^{\mathsf{T}} \boldsymbol{\beta})} \right) \tag{4.83}$$

where R is the total number of possible full rankings which are consistent with the partial ranking, $\boldsymbol{x}_{r\pi_r^{-1}(j)}$ is the $H \times 1$ attribute vector of the jth ranked alternative in the rth full ranking which is consistent with partial ranking π^{part} and $\boldsymbol{\beta}$ is the $H \times 1$ coefficient vector of fixed attribute effects.

4.6.2 Partial Ranking, N Individuals, T_i Choice Sets for Each Individual

Suppose that we have N individuals each stating preferences about more than one choice set. Let $\boldsymbol{\pi_i^{part}} = (\pi_{i1}^{part}, \pi_{i2}^{part}, \pi_{i3}^{part}, ..., \pi_{iT_i}^{part})$ represent the sequence of incomplete rankings made by individual i over the course of the experiment, where T_i is the total number of choice sets presented to individual i. Then the probability that individual i makes this sequence of incomplete rankings is the product

$$p(\boldsymbol{\pi_i^{part}}) = \prod_{t=1}^{T_i} p(\pi_{it}^{part}) \tag{4.84}$$

$$= \prod_{t=1}^{T_i} \sum_{r=1}^{R} p(\pi_{itr}) \tag{4.85}$$

$$= \prod_{t=1}^{T_i} \sum_{r=1}^{R} \left(\prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{itr\pi_r^{-1}(j)}^{\mathsf{T}} \boldsymbol{\beta})}{\sum_{J \geq k \geq j} \exp(\boldsymbol{x}_{itr\pi_r^{-1}(k)}^{\mathsf{T}} \boldsymbol{\beta})} \right) \tag{4.86}$$

where π_{it}^{part} is the partial ranking of choice set t presented to individual i, π_{itr} is the full ranking r of the alternatives in choice set t presented to individual i which is consistent with π_{it}^{part}, R is the total number of possible full rankings consistent with the partial ranking, π_{it}^{part}, $\boldsymbol{x}_{itr\pi_r^{-1}(j)}$ is the $H \times 1$ attribute vector of the jth ranked alternative in the full ranking r which is consistent with π_{it}^{part} and $\boldsymbol{\beta}$ is the $H \times 1$ coefficient vector of fixed attribute effects.

4.7 Partial Ranking & Random Effects

A model with random effects can also be fit to partially ranked data. Let $\boldsymbol{\beta}_i$ be the coefficient vector of random attribute effects. If we assume that $\boldsymbol{\beta}_i$ is random and distributed with density $f(\boldsymbol{\beta}_i|\theta)$, where θ are the parameters of the distribution, then conditional on $\boldsymbol{\beta}_i$, the probability of an individual's partial ranking of alternatives is given in (4.86) with $\boldsymbol{\beta}_i$ substituted for $\boldsymbol{\beta}$.

Let $\boldsymbol{\pi_i^{part}} = (\pi_{i1}^{part}, \pi_{i2}^{part}, \pi_{i3}^{part}, ..., \pi_{iT_i}^{part})$ represent the sequence of partial rankings made by the respondent over the course of the experiment, where T_i is the total number of choice sets presented to individual i. Then conditional on $\boldsymbol{\beta}_i$, the probability that respondent i

makes this sequence of rankings is the product

$$p(\boldsymbol{\pi_i^{part}}|\boldsymbol{\beta}_i) = \prod_{t=1}^{T_i} p(\pi_{it}^{part}|\boldsymbol{\beta}_i) \tag{4.87}$$

$$= \prod_{t=1}^{T_i} \sum_{r=1}^{R} p(\pi_{itr}|\boldsymbol{\beta}_i) \tag{4.88}$$

$$= \prod_{t=1}^{T_i} \sum_{r=1}^{R} \left(\prod_{j=1}^{J-1} \frac{\exp(\boldsymbol{x}_{itr\pi_r^{-1}(j)}^{\mathsf{T}} \boldsymbol{\beta}_i)}{\sum_{J \geq k \geq j} \exp(\boldsymbol{x}_{itr\pi_r^{-1}(k)}^{\mathsf{T}} \boldsymbol{\beta}_i)} \right) \tag{4.89}$$

where π_{it}^{part} is the partial ranking of choice set t presented to individual i, π_{itr} is the full ranking r of the alternatives in choice set t presented to individual i which agrees with π_{it}^{part}, R is the total number of possible full rankings implied by the partial ranking, π_{it}^{part}, $\boldsymbol{x}_{itr\pi_r^{-1}(j)}$ is the $H \times 1$ attribute vector of the jth ranked alternative in the full ranking r which agrees with π_{it}^{part} and $\boldsymbol{\beta}_i$ is the $H \times 1$ coefficient vector of random attribute effects.

Then the unconditional probability is the integral of this product over the density of $\boldsymbol{\beta}_i$

$$p(\boldsymbol{\pi_i}) = \int p(\boldsymbol{\pi_i}|\boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i|\theta) d\boldsymbol{\beta}_i. \tag{4.90}$$

4.7.1 Special Case: The Rank Ordered Logit Model for Full Rankings

Suppose that for individual i that have partial rankings of a choice sets $t = 1, ..., T_i$ with no unranked middle cards, i.e., a full rankings. Then $R = 0! = 1$ and (4.89) reduces to the probability of observing the full ranking π_{ic} defined in (4.70).

4.7.2 Special Case: The Multinomal Logit Model

Suppose now that for individual i that have partial rankings of a choice sets $t = 1, ..., T_i$ with only best choices ranked. Then (4.89) reduces to the probability of observing the best choices defined in (4.50).

4.8 Methods of Estimation

A challenge to researchers that has arisen when trying to account for heterogeneity in respondents' preferences is the difficulty of obtaining sufficient data to estimate individual-level parameters. With discrete choice experiments, there is difficulty in calculating individual utilities because each respondent provides only a small amount of information. In constructing choice sets for evaluation, a tradeoff is made between the need for a large number of choice sets and the need to minimize respondent fatigue. In typical surveys, respondents make choices from as little as eight to twelve choice sets and it has been documented that increasing the number of choice sets can affect the accuracy in responses (Hauser and Rao, 2002).

While some methods aim to collect more data by making changes to the experimental design to derive individual-level estimates (Louviere et al., 2008; Lancsar and Louviere, 2008), Bayesian methods can also be used to obtain individual-level estimates in the presence of sparse data (Rossi and Allenby, 1993; Allenby, 1995, 1994; Allenby and Rossi, 1999). We first discuss a classical method for estimating individual-level preferences before describing a Bayesian approach. We illustrate the methods using the MXL model to define the choice probability, but we note that the methods are not specific to this model and can be implemented using any of random effects models described in Chapter 4.

4.8.1 Approximate Likelihood Methods

One way to fit mixed logit models is by using maximum simulated likelihood. Maximum simulated likelihood methods involve integration over the distribution of the individual-level preference parameter. If the integral has no closed form solution, one can use simulation and maximize a simulated likelihood. In this method, draws are made from the distribution of the individual-level preference parameter. The likelihood is calculated for each draw and averaged over all draws. This simulated likelihood is then maximized.

In the PROSPECT study, respondents were presented with 10 to 17 choice tasks. Let $\boldsymbol{y_i} = (j_1, j_2, j_3, ..., j_{T_i})$ represent the sequence of preferred choices made by the respondent

over the course of the experiment, where T_i is the total number of choice sets presented to individual i. Conditional on $\boldsymbol{\beta}_i$ and because the ϵ_{itj}'s are independent over time, the conditional likelihood that respondent i makes this sequence of choices is the product

$$L_i(\boldsymbol{\beta}_i|\boldsymbol{y_i}) = P(\boldsymbol{y_i}|\boldsymbol{\beta}_i) = \prod_{t=1}^{T_i}\left[\frac{\exp(\boldsymbol{x}_{itj_t}\boldsymbol{\beta}_i)}{\sum_{k\in C_t}\exp(\boldsymbol{x}_{itk}\boldsymbol{\beta}_i)}\right] \tag{4.91}$$

where $\boldsymbol{y_i}$ is the observed choices for respondent i in the course of the experiment and C_t is the tth choice set presented to individual i, $t = 1, 2, 3, ..., T_i$. A distribution for the coefficients, $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$, is then specified and the parameters, $\boldsymbol{\theta}$, of the coefficient distribution are estimated.

In many studies $f(\boldsymbol{\beta}_i|\theta)$ has been specified to be normal or lognormal (Mehndiratta, 1996; Ben-Akiva and Bolduc, 1996; Revelt and Train, 1998; Johnson, 2000). Triangular and uniform distributions have also been used (Revelt and Train., 2000; Hensher and Greene, 2003). In moving away from finding point estimates of a parameter vector to analyzing the distribution of parameters, the problem of specifying the functional form for the distribution arises. Recent studies have attempted to address this issue (Rigby and Burton, 2006). In this dissertation, we will begin by specifying $\boldsymbol{\beta}_i$ to be normally distributed with parameters $\theta = (\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}})$, which we denote by $\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}})$.

The likelihood function for individual i making their choices in the course of the experiment is the unconditional probability of making that sequence of choices, the integral of $L_i(\boldsymbol{\beta}_i|\boldsymbol{y_i})$ over all values of $\boldsymbol{\beta}_i$,

$$L_i(\boldsymbol{\theta}|\boldsymbol{y_i}) = P(Y_i = \boldsymbol{y_i}|\boldsymbol{\theta}) = \int L_i(\boldsymbol{\beta}_i|\boldsymbol{y_i})f(\boldsymbol{\beta}_i|\boldsymbol{\theta})d\boldsymbol{\beta}_i \tag{4.92}$$

where $\boldsymbol{y_i}$ is the observed choices for respondent i in the course of the experiment. Then the log-likelihood function is

$$l(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{N}\log(L_i(\boldsymbol{\theta}|\boldsymbol{y_i})) \tag{4.93}$$

where $\boldsymbol{y}$ is the observed choices across all choice sets and respondents. Although there is no

closed form solution for the integral in (4.92), $L_i(\boldsymbol{\theta}|\boldsymbol{y_i})$ can be estimated through simulation. A draw of $\boldsymbol{\beta}_i$ for all i is taken from the density $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$ and $L_i(\boldsymbol{\theta}|\boldsymbol{y_i})$ is calculated. This is repeated many times and the results are averaged to approximate the likelihood function for individual i, using

$$SL_i(\boldsymbol{\theta}|\boldsymbol{y_i}) = \frac{1}{R}\sum_{r=1}^{R} P(Y_i = \boldsymbol{y_i}|\boldsymbol{\beta}_i^{(r)}) \tag{4.94}$$

where R is the number of draws, $\boldsymbol{\beta}_i^{(r)}$ is the rth draw from $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$ and $SL_i(\boldsymbol{\theta}|\boldsymbol{y_i})$ is the simulated likelihood function for individual i given their sequence of choices, $\boldsymbol{y_i}$. The simulated log-likelihood function is constructed by summing over the log of the simulated probabilities over all individuals

$$Sl(\boldsymbol{\theta}|\boldsymbol{y_i}) = \sum_{i=1}^{N} \log(SL_i(\boldsymbol{\theta}|\boldsymbol{y_i})). \tag{4.95}$$

The maximum simulated likelihood estimator (MSLE) is the estimate of $\boldsymbol{\theta}$ that maximizes $Sl(\boldsymbol{\theta}|\boldsymbol{y_i})$

$$\hat{\boldsymbol{\theta}}_{MSLE} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Sl(\boldsymbol{\theta}|\boldsymbol{y_i}). \tag{4.96}$$

So, for example, if we assumed that $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}})$, then the MSLEs are the values $\hat{\boldsymbol{\beta}}$ and $\hat{\Sigma}_{\boldsymbol{\beta}}$ that maximize $Sl(\boldsymbol{\theta}|\boldsymbol{y_i})$.

4.8.2 Hierarchical Bayes Approach

In using classic methods of estimation with sparse data, it is not uncommon to obtain estimates of the coefficients in a direction inconsistent with the true values (Rao, 2008). Hierarchical Bayes (HB) provides a method to overcome the problem of sparse information.

Bayesian ideas for MXL models with normally distributed coefficients were introduced by Allenby and Lenk (1994) and Allenby (1994). Allenby and Rossi (1999) showed how Bayesian procedures could be used to obtain estimates for individual parameters within a

random coefficient model and Train (2001) extended the procedures for the MXL model to allow for non-normal distributions (e.g., uniform and lognormal distributions) of the coefficients. For convenience, we will assume that the coefficients are normally distributed.

Consider the utility, U_{itj}, of alternative j for individual i in choice task t

$$U_{itj} = \boldsymbol{x}_{itj}^{\mathsf{T}}\boldsymbol{\beta}_i + \epsilon_{itj} \tag{4.97}$$

where $\boldsymbol{x}_{itj}$ is the attribute vector for alternative j presented to individual i in choice set t, ϵ_{itj} is iid Gumbel and $\boldsymbol{\beta}_i$ is the coefficient vector of random attribute effects for respondent i, and is normally distributed, $\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}})$. Then the posterior distribution of $\boldsymbol{\beta}$ and $\Sigma_{\boldsymbol{\beta}}$ is

$$P(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}}|Y) \propto \prod_{i=1}^{N} P(Y_i = \mathbf{j}|\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}})p(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}}) \tag{4.98}$$

where $p(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}})$ is the prior distribution on $\boldsymbol{\beta}$ and $\Sigma_{\boldsymbol{\beta}}$. If $\boldsymbol{\beta}$ and $\Sigma_{\boldsymbol{\beta}}$ are independent, then $p(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}}) = p(\boldsymbol{\beta})p(\Sigma_{\boldsymbol{\beta}})$. Typically, $p(\boldsymbol{\beta})$ is assumed to be multivariate normal, $\mathcal{N}(\mu, \Sigma)$, and $p(\Sigma_{\boldsymbol{\beta}})$ is assumed to be Inverse Wishart with prior degrees of freedom w and prior precision W. The Inverse Wishart distribution is often used in Bayesian modeling because it is a proper conjugate prior for an unknown covariance matrix in a multivariate normal model (Gelman, 2006).

If X has an Inverse Wishart distribution with scale matrix W and degrees of freedom parameter w (denote this by $X \sim \text{inverseWishart}(w, W)$), then the probability density function is given by

$$f_X(x|\mu, \eta) = \frac{|W|^{\frac{w}{2}}}{2^{\frac{wp}{2}}\Gamma_p(\frac{w}{2})}|X|^{-\frac{w+p+1}{2}} \exp^{-\frac{1}{2}\text{tr}(WX^{-1})} \tag{4.99}$$

where X and W are $p \times p$ positive definite matrices, Γ_p is the multivariate gamma function, tr is the trace function, w is real-valued ($w \in \mathbb{R}$) and $w > p + 1$. The mean of the Inverse

Wishart distribution is

$$E(X) = \frac{W}{w - p - 1}. \tag{4.100}$$

where $w > p + 1$ and $w \in \mathbb{R}$.

The conditional posteriors can be shown to be

$$p(\boldsymbol{\beta}|\Sigma_{\boldsymbol{\beta}}, \boldsymbol{\beta}_i \text{ for all i}) \propto \mathcal{N}\Big(\sum_{i=1}^{N} \boldsymbol{\beta}_i/N, \Sigma_{\boldsymbol{\beta}}/N\Big) \tag{4.101}$$

and

$$p(\Sigma_{\boldsymbol{\beta}}|\boldsymbol{\beta}, \boldsymbol{\beta}_i \text{ for all i}) \propto \text{inverseWishart}\Big(w + N, \frac{wW + NG}{w + N}\Big) \tag{4.102}$$

where N is the number of individuals and $G = \sum_{i=1}^{N}(\boldsymbol{\beta}_i - \boldsymbol{\beta})(\boldsymbol{\beta}_i - \boldsymbol{\beta})'/N$ (Train, 2001).

Using Gibbs sampling we can draw from $p(\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}}|Y)$ in three steps (Train, 2001):

1. Take a draw of $\boldsymbol{\beta}$ conditional on $\Sigma_{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_i$ for all i.
2. Take a draw of $\Sigma_{\boldsymbol{\beta}}$ conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_i$ for all i.
3. Take a draw of $\boldsymbol{\beta}_i$ for all i conditional on values of $\boldsymbol{\beta}$ and $\Sigma_{\boldsymbol{\beta}}$.

The posterior for each person's coefficient vector, $\boldsymbol{\beta}_i$, conditional on their choices and the population mean, $\boldsymbol{\beta}$, and variance, $\Sigma_{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}_i$, is

$$p(\boldsymbol{\beta}_i|\boldsymbol{\beta}, \Sigma_{\boldsymbol{\beta}}, Y_i = \mathbf{j}) \propto P(Y_i = \mathbf{j}|\boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i|\theta) \tag{4.103}$$

where we have assumed that $\boldsymbol{\beta}_i$ is normally distributed with mean $\boldsymbol{\beta}$ and variance $\Sigma_{\boldsymbol{\beta}}$. The three steps above are repeated for many iterations and the resulting values converge to draws from the joint posterior of $\boldsymbol{\beta}$, $\Sigma_{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_i$ for all i individuals. The mean and standard deviation of these draws can then be calculated to obtain estimates and standard errors of the parameters.

CHAPTER 5

Preliminary Analyses

This chapter presents the results of early analyses conducted using data from the first 44 patients while patient recruitment was ongoing. First, using only best choices, a multinomial logit model using maximum likelihood and a mixed logit model using simulated maximum likelihood were fit to the data and compared. Then two hierarchical Bayes models (one using best choices only and another using best and worst choices) were fit to the data and compared. Relative attribute importance was then calculated using Bayesian methods and finally, principal components analysis was used to identify attribute groupings which may inform us about possible underlying choice processes.

5.1 Best Choice: Fitting the Multinomial Logit and Mixed Logit Models Using Maximum Likelihood and Simulated Maximum Likelihood in R

We fit three models for best choice to the PROSPECT study data. For the analyses in this section, we ignore the additional information gained by asking for the worst alternatives. Attributes were included as dummy variables in the models. The models include the multinomial logit (MNL) using maximum likelihood estimation and two mixed logit models based on different specifications for the covariance matrix of the random coefficients using simulated maximum likelihood (MXL1 and MXL2). In the MNL model, the attribute coefficients are considered fixed for all patients. The coefficients for the mixed logit, on the other hand, are considered to vary randomly in the population. In the first model MXL1, it is assumed that each coefficient is independently distributed according to a normal distribution. In

model MXL2, all coefficients are normally distributed and allowed to be correlated. Monte Carlo simulation for the likelihood of the mixed logit models was performed using 500 draws for each participant. Analysis was performed using the mlogit package in R.

Table 5.1 presents the estimated population parameters for the three models. For the MNL model, all coefficients are significantly different from zero except for *Taking Action* (taking action immediately vs not jumping into a radical treatment). Similar results are found for the two mixed logit models MXL1 and MXL2. Except for *Taking Action*, the signs and significance of the coefficients are consistent with a priori expectations.

For MXL1, the estimated standard deviations of the random effects for all attributes, except for active, decreased urinary function and decreased sexual function, are significant. The estimated standard deviations vary in the population, which implies that there is considerable heterogeneity in patients' preference for a full lifespan, bowel issues, surgery, others' support, same urinary functioning and same sexual functioning. By allowing the parameters to vary, the log likelihood increases. From Table 5.3, the likelihood ratio test comparing the MNL model to the MXL1 is significant, indicating that the mixed logit model provides a better fit for the data.

The random coefficients are specified to be independently distributed in MXL1, but it is possible that the coefficients are correlated. For example, patients concerned about urinary functioning might also be concerned about sexual functioning. Thus, for MXL2, the coefficients are specified to be normally distributed with a covariance matrix with possibly non-zero off-diagonal entries. From Table 5.1, we see that the MXL2 yields similar estimates compared to MXL1. However, from Table 5.3, we see that the likelihood ratio test comparing the MXL1 to MXL2 is significant indicating that the mixed logit model where the coefficients are allowed to be correlated provides a better fit. Table 5.2 presents the estimated covariance matrix and estimates for the correlation matrix. Seventeen covariances were found to be significantly different from zero.

The coefficient for full lifespan is negatively correlated with the coefficient of no bowel problems and positively correlated with the coefficient of no cutting. This implies that pa-

tients who value a full lifespan tend to value no cutting and tend not to value no bowel issues. The coefficient of no cutting is negatively associated with the coefficients of active, others support and sexual functioning. This implies that the patients who value no cutting also tend to not value taking an active role in treatment, others support and sexual functioning. The coefficient of others' support is negatively associated with the coefficients for urinary functioning and sexual functioning, and the coefficients for urinary functioning are positively correlated with the coefficients for sexual functioning.

5.2 Best Choice and Best-Worst Choices: Fitting the Multinomial Logit and the Rank Ordered Logit Model for Best and Worst Choices Using Bayesian Methods

We fit a MNL model with random effects for best choice and a rank-ordered logit (ROL) model with random effects for best-worst choices to the PROSPECT study data using Bayesian methods. The models are defined in more detail in Section 6.3. The estimates were obtained using Gibbs sampling using JAGS software. The JAGS model specifications can be found in Appendices A.1 and A.2.

Table 5.4 presents the results for our best and best-worst models. For these patients, the most valued attributes (and most variable) in both models were full lifespan, urinary functioning and sexual functioning. For these analyses, we considered those coefficients with high posterior probabilities of being non-zero to be those whose 95% posterior credible interval did not contain zero. All estimated coefficients for the population means and standard deviations are considered significantly different from zero under both models except for the posterior mean estimates of taking action. In addition, because the 95% credible intervals for these posterior mean estimates overlapped we find that there is no significant difference between these mean estimates.

Compared to the MNL for best choice with random effects, the estimated coefficients for the model incorporating worst choices are slightly higher across nearly all attributes.

Table 5.1: Results for the Multinomial Logit and Mixed Logit Models.

	Multinomial Logit (MNL) Fixed Effects			Mixed Logit (MXL1) Random Effects, Independent			Mixed Logit (MXL2) Random Effects, Correlated		
Attribute	Est	SE	P value	Est	SE	P value	Est	SE	P value
Full Lifespan	1.30	0.12	<0.0001	2.24	0.22	<0.0001	2.82	0.31	<0.0001
No Bowel Issues	0.77	0.12	<0.0001	1.13	0.16	<0.0001	1.82	0.22	<0.0001
No Cutting	0.92	0.13	<0.0001	1.36	0.20	<0.0001	1.87	0.29	<0.0001
Taking Action	-0.02	0.11	0.84	0.11	0.15	0.45	0.17	0.23	0.45
Others' Support	0.58	0.12	<0.0001	0.90	0.18	<0.0001	1.35	0.27	<0.0001
Short Term Urinary Issues	0.74	0.15	<0.0001	1.00	0.19	<0.0001	1.50	0.34	<0.0001
Full Urinary Functioning	1.15	0.18	<0.0001	1.65	0.23	<0.0001	2.25	0.41	<0.0001
Short Term Sexual Issues	0.90	0.16	<0.0001	1.39	0.21	<0.0001	2.42	0.42	<0.0001
Full Sexual Functioning	1.79	0.18	<0.0001	2.66	0.28	<0.0001	3.62	0.54	<0.0001
Standard Deviations of Random Effects									
Full Lifespan				1.57	0.23	<0.0001	2.38	0.34	<0.0001
No Bowel Issues				0.98	0.26	0.0002	1.33	0.3	0.01
No Cutting				1.52	0.29	<0.0001	3.71	0.31	0.07
Taking Action				0.46	0.28	0.10	0.7	0.28	0.94
Others' Support				1.09	0.24	<0.0001	2.17	0.31	<0.0001
Short Term Urinary Issues				0.43	0.32	0.18	2.36	0.43	0.0004
Full Urinary Functioning				0.60	0.27	0.02	3.15	0.32	0.11
Short Term Sexual Issues				0.37	0.32	0.24	4.28	0.77	<0.0001
Full Sexual Functioning				1.26	0.25	<0.0001	5.93	0.28	0.01
Log-Likelihood	-616.69			-558.95			-528.79		
Number of draws	500			500			500		

Table 5.2: Covariances and correlations among the coefficients in the Mixed Logit Model.

	Full Lifespan	No Bowel Issues	No Cutting	Active	Others' Support	Urinary Issues	Urinary Functioning	Short Term Issues	Full Sexual Functioning
Estimated Covariance Matrix									
Full Lifespan	5.66	-2.6	2.72	-0.65	-0.93	0.53	1.19	-1.18	-0.98
No Bowel Issues		1.77	-3.9	0.22	1.05	0.08	-0.89	1.08	0.12
No Cutting			13.77	-0.3	-3.89	-1.22	1.96	-3.61	-0.27
Taking Action				0.49	0.7	-0.13	0.09	0.67	1.56
Others' Support					4.69	-2.21	-2.87	-0.15	-0.71
Short Term Urinary Issues						5.56	6.62	3.79	2.99
Full Urinary Functioning							9.93	5.78	6.99
Short Term Sexual Issues								18.32	23.71
Full Sexual Functioning									35.21
Correlation Matrix									
Full Lifespan	1	-0.82	0.31	-0.39	-0.18	0.09	0.16	-0.12	-0.07
No Bowel Issues		1	-0.79	0.23	0.36	0.03	-0.21	0.19	0.02
No Cutting			1	-0.12	-0.48	-0.14	0.17	-0.23	-0.01
Taking Action				1	0.46	-0.08	0.04	0.22	0.38
Others' Support					1	-0.43	-0.42	-0.02	-0.06
Short Term Urinary Issues						1	0.89	0.37	0.21
Full Urinary Functioning							1	0.43	0.37
Short Term Sexual Issues								1	0.93
Full Sexual Functioning									1

* Red highlights indicate significant at 0.05

Table 5.3: Likelihood Ratio Tests

Model	df	LogLik	Chisq	Chisq df	P-value
Multinomial Logit	9	-616.69			
Mixed Logit with independent coefficients	18	-558.95	115.49	9	<0.0001
Mixed Logit with independent coefficients	18	-558.95			
Mixed Logit with correlated coefficients	54	-528.79	60.31	36	0.01

In addition, by incorporating worst-choices, the standard deviations of the mean estimates decrease. We see a similar results for the estimated standard deviations of the random effects.

Table 5.5 presents the correlation matrix for the random effects. Under the MNL model for best choice, the estimated mean correlation between short term urinary issues and full lifespan is positive and the 95% credible interval does not contain zero (highlighted), which implies that an average patient who values full lifespan may be willing endure short term urinary issues. Under the ROL model for best and worst choices, patients who tend to value urinary functioning also tend to value sexual functioning. The negative correlations in the table describe the trade-offs in attribute preferences. For example, patients who tend to value full lifespan may be willing to undergo surgery.

5.3 Classification of Health State Attributes Using Principal Components Analysis

The principal aim of this analysis was to ascertain if an underlying structure could be identified to assist with classification of the health state attributes in the population. Classification of the health state attributes may assist the clinician and the patient in an initial discussion regarding treatment, which may prompt the patient to begin exploring their own personal preferences.

Correlations among the nine attributes were examined using principal components analysis (PCA). PCA examines the correlations among measured variables to determine if there are groups of variables that are correlated. It is a tool used to extract components that simplify the data, while retaining as much information as possible. PCA initially generates

Table 5.4: Results using the HB Rank-Ordered Logit Random Effects Model.

	MNL Model (Best Choice)				ROL Model (Best-Worst Choices)			
Attribute	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
Full Lifespan	1.55	0.25	1.06	2.01	1.77	0.22	1.35	2.20
No Bowel Issues	0.86	0.15	0.61	1.17	1.03	0.15	0.75	1.32
Short Term Urinary Issues	0.72	0.14	0.43	0.99	0.91	0.11	0.70	1.12
Full Urinary Functioning	1.15	0.20	0.72	1.47	1.28	0.14	1.02	1.55
Short Term Sexual Issues	1.00	0.26	0.48	1.44	1.34	0.17	1.02	1.66
Full Sexual Functioning	1.88	0.34	1.20	2.51	2.07	0.25	1.59	2.56
No Cutting	1.05	0.23	0.55	1.45	1.10	0.16	0.81	1.42
Taking Action	0.04	0.16	-0.28	0.32	0.11	0.08	-0.05	0.28
Others' Support	0.76	0.17	0.45	1.10	0.80	0.13	0.54	1.06
Standard Deviations of Random Effects								
Full Lifespan	0.99	0.25	0.49	1.47	1.31	0.20	0.96	1.77
No Bowel Prob	0.46	0.16	0.20	0.78	0.71	0.14	0.45	0.99
Short Term Urinary Issues	0.40	0.11	0.19	0.62	0.37	0.11	0.17	0.59
Full Urinary Functioning	0.37	0.14	0.16	0.71	0.49	0.16	0.23	0.84
Short Term Sexual Issues	0.52	0.22	0.16	1.01	0.94	0.17	0.65	1.3
Full Sexual Functioning	1.00	0.44	0.37	1.98	1.57	0.24	1.17	2.13
No Cutting	0.70	0.23	0.30	1.12	0.84	0.16	0.57	1.20
Taking Action	0.26	0.10	0.12	0.49	0.29	0.08	0.16	0.46
Others' Support	0.48	0.19	0.20	0.93	0.67	0.12	0.46	0.92

Table 5.5: Correlations among the random effects for the Best and Best-Worst Models.

Best Choice									
	Full Lifespan	No Bowel Issues	Short Term Urinary Issues	Full Urinary Functioning	Short Term Sexual Issues	Full Sexual Functioning	No Cutting	Taking Action	Others' Support
Full Lifespan	1	-0.34	0.65	0.2	0.57	0.27	0.78	-0.04	-0.08
No Bowel Issues		1	0.03	-0.08	-0.58	-0.6	0.44	-0.1	-0.44
Short Term Urinary Issues			1	0.28	0.23	-0.13	-0.55	-0.04	-0.25
Full Urinary Functioning				1	0.17	-0.21	-0.27	0.02	0.21
Short Term Sexual Issues					1	0.57	-0.61	0.07	0.28
Full Sexual Functioning						1	-0.29	0.11	0.16
No Cutting							1	0.08	-0.17
Taking Action								1	0.06
Others' Support									1

Best-Worst Choices									
	Full Lifespan	No Bowel Issues	Short Term Urinary Issues	Full Urinary Functioning	Short Term Sexual Issues	Full Sexual Functioning	No Cutting	Taking Action	Others' Support
Full Lifespan	1	-0.17	0.27	0.09	0.36	0.34	-0.46	0.11	-0.18
No Bowel Issues		1	-0.34	-0.38	-0.3	-0.26	0.06	0.36	-0.2
Short Term Urinary Issues			1	0.73	0.65	0.63	-0.43	0.05	-0.07
Full Urinary Functioning				1	0.69	0.68	-0.35	0.03	-0.03
Short Term Sexual Issues					1	0.96	-0.42	0.01	-0.38
Full Sexual Functioning						1	-0.32	-0.04	-0.48
No Cutting							1	-0.48	-0.36
Taking Action								1	0.24
Others' Support									1

the same number of components as the number of variables, but most of the components will explain very little variance. Components that explain much of the variance in the data are identified on a scree plot as those *above the elbow* and/or those with eigenvalues greater than one. The eigenvalues measure the amount of variance accounted for by each principal component. The sum of the eigenvalues equals the total number of principal components and the proportion of variance accounted for by each principal component is calculated by dividing the eigenvalue corresponding to the principal component by the total number of principal components.

Interpretation of the principal components relies on the factor loadings, which are the correlations between the original variable values and the extracted components. Components are considered to represent those variables with which they have moderate to high correlations. For this analysis, the variables were the estimate individual-level random effects. A cutoff of 0.4 was used and attributes with factor loadings of 0.4 or greater were considered to load highly on the component. The value of using PCA lies in the fact that attribute groupings may inform us about possible underlying choice processes. For example, if lifespan, bowel issues, cutting, sexual and urinary functioning were grouped together (loaded highly on the same component), then these might be interpreted as issues that patients tend to be considered together.

The estimated individual-level random effects from the Best-Worst model, which has results in Table 5.4, are presented in Table 5.6 and a scatterplot matrix of the estimated random effects is depicted in Figure 5.1. From Figure 5.1, we observe that nearly all pairs of attributes appear to have some degree of correlation.

Principal components analysis was conducted using the posterior mean estimates for each of the components of the correlation matrix presented in Table 5.5. The PCA results are presented in Table 5.7. The eigenvector with the highest eigenvalue is the first principal component of the data. In Table 5.7, the highest eigenvalue is 3.78. If the data are projected onto the line defined by the eigenvector, this line describes the direction where the variability is maximized. Comp.1 accounts for 42% (3.78/9) of the total variation in the data and Comp.2 accounts for 20%.

To reduce dimensionality in the data, we could use the scree plot in Figure 5.2 to identify components with variances above the *elbow*. Following this rule of thumb, we may decide to keep the first four components. Another rule of thumb recommends that we keep components whose eigenvalues are > 1 (Afifi et al., 2003). Since the eigenvalue corresponding to the fourth component is approximately equal to one, using this rule we would keep the first four components. Together, these first four components explain 88% of the total variance of the data.

To interpret the components we look at the attributes corresponding to the highest vector components (loadings) in absolute value. Factor loadings with absolute value greater than or equal to 0.4 are highlighted in red. We could characterize the first principal component as a vector describing *urinary and sexual functioning*. The second principal component focuses on *treatment issues: surgery, support & action*, while the third principal component reflects *bowel functioning*. The fourth component reflects *expected lifespan*. PCA clarified choice processes in the population by identifying meaningful groups of health state attributes. The results suggest that health state attributes related to prostate cancer treatment can be summarized by four components related to urinary and sexual functioning, treatment issues, bowel functioning, and expected lifespan. These components are useful for future research and my have implications for treating patients.

5.4 Remarks

These early analyses helped with the iterative process of model development. We moved forward with the mixed logit where all coefficients are assumed to be normally distributed and allowed to be correlated. We did not directly use the PCA approach in subsequent analyses; this could be an area for future research.

Table 5.6: Individual-Level Random Effects

Patient ID	Full Lifespan	No Bowel Issues	Short Term Urinary Issues	Full Urinary Functioning	Short Term Sexual Issues	Full Sexual Functioning	No Cutting	Taking Action	Others' Support
1	1.35	1.49	1.16	1.73	2.31	3.69	0.38	0.34	0.35
2	0.79	0.61	0.61	0.99	0.72	1.24	2.47	-0.11	0.67
3	0.67	0.88	1.08	1.64	1.30	1.89	1.11	0.17	1.07
4	0.93	1.89	0.15	0.30	-0.04	-0.09	2.01	-0.07	0.73
5	2.50	0.77	0.78	1.00	1.41	2.31	1.53	-0.11	0.29
6	0.86	1.14	0.87	1.25	1.09	1.71	1.05	0.28	1.29
7	1.15	0.39	0.91	1.25	1.20	2.00	1.12	-0.05	1.20
8	1.91	1.90	0.36	0.45	0.61	1.17	1.78	0.03	0.15
9	1.40	1.21	0.73	0.96	0.66	0.93	0.80	0.05	1.38
10	2.69	0.08	1.25	1.78	2.58	4.25	0.58	0.05	0.66
11	2.36	0.02	0.74	1.15	1.19	1.82	1.68	-0.20	0.84
12	-0.40	1.76	0.51	0.87	0.45	0.95	2.62	-0.07	0.29
13	0.86	0.93	1.16	1.75	1.95	2.96	0.97	0.22	0.59
14	3.21	0.90	0.77	0.85	1.09	1.39	0.64	0.10	0.93
15	0.09	0.78	0.92	1.53	0.38	0.30	0.60	0.27	2.45
16	0.74	0.98	0.78	1.16	0.34	0.18	0.56	0.35	2.26
17	1.09	0.97	0.46	0.82	0.67	1.11	2.66	-0.14	0.19
18	3.22	0.57	0.93	1.10	1.51	2.14	0.47	0.07	1.03
19	1.86	0.36	1.33	2.17	3.45	5.99	0.91	0.06	-0.21
20	2.06	1.27	1.07	1.41	2.31	3.71	0.53	0.15	0.12
21	1.63	0.86	1.15	1.64	1.93	3.04	0.53	0.25	0.85
22	0.08	1.21	0.33	0.71	0.15	0.28	2.55	-0.12	0.91
23	1.59	0.45	0.83	1.19	1.28	2.13	1.61	-0.12	0.75
24	4.76	0.32	1.06	1.34	2.47	3.90	0.14	0.11	0.45
25	4.20	0.70	0.98	1.16	1.91	2.92	0.01	0.15	0.78
26	2.06	0.92	0.98	1.64	2.44	3.97	0.53	0.16	0.41
27	3.21	0.39	0.75	0.99	0.74	0.76	0.45	0.09	1.78
28	1.39	1.27	0.47	0.74	0.86	1.48	2.26	-0.02	0.16
29	0.54	1.42	0.53	0.89	0.15	0.02	1.65	0.16	1.35
30	0.77	2.07	0.56	0.87	0.77	1.22	1.18	0.20	0.77
31	0.34	0.50	0.94	1.60	0.98	1.44	1.04	0.17	1.85
32	2.44	0.16	0.84	1.30	1.92	3.24	1.49	-0.08	0.44
33	1.49	2.88	0.50	0.61	0.72	1.01	0.72	0.36	0.37
34	2.00	1.39	0.66	0.80	0.44	0.43	0.69	0.25	1.31
35	3.05	0.52	0.64	0.94	1.67	2.75	1.50	-0.06	0.29
36	3.27	0.95	0.62	0.86	1.15	1.52	0.62	0.10	0.99
37	0.34	1.25	0.76	1.27	2.00	3.58	2.12	-0.07	-0.20
38	2.48	0.82	1.14	1.59	2.38	3.62	0.44	0.21	0.54
39	3.66	1.61	0.69	0.62	1.03	1.47	0.47	0.19	0.50
40	2.05	1.45	0.86	1.12	1.39	1.97	0.51	0.26	0.87
41	1.71	0.91	1.22	1.77	2.53	4.25	0.65	0.15	0.23
42	1.82	1.96	0.81	1.10	1.69	2.58	0.64	0.28	0.28
43	2.70	0.99	1.10	1.38	1.60	2.29	0.21	0.19	1.00
44	1.14	0.55	0.88	1.42	1.46	2.05	1.03	-0.03	1.09

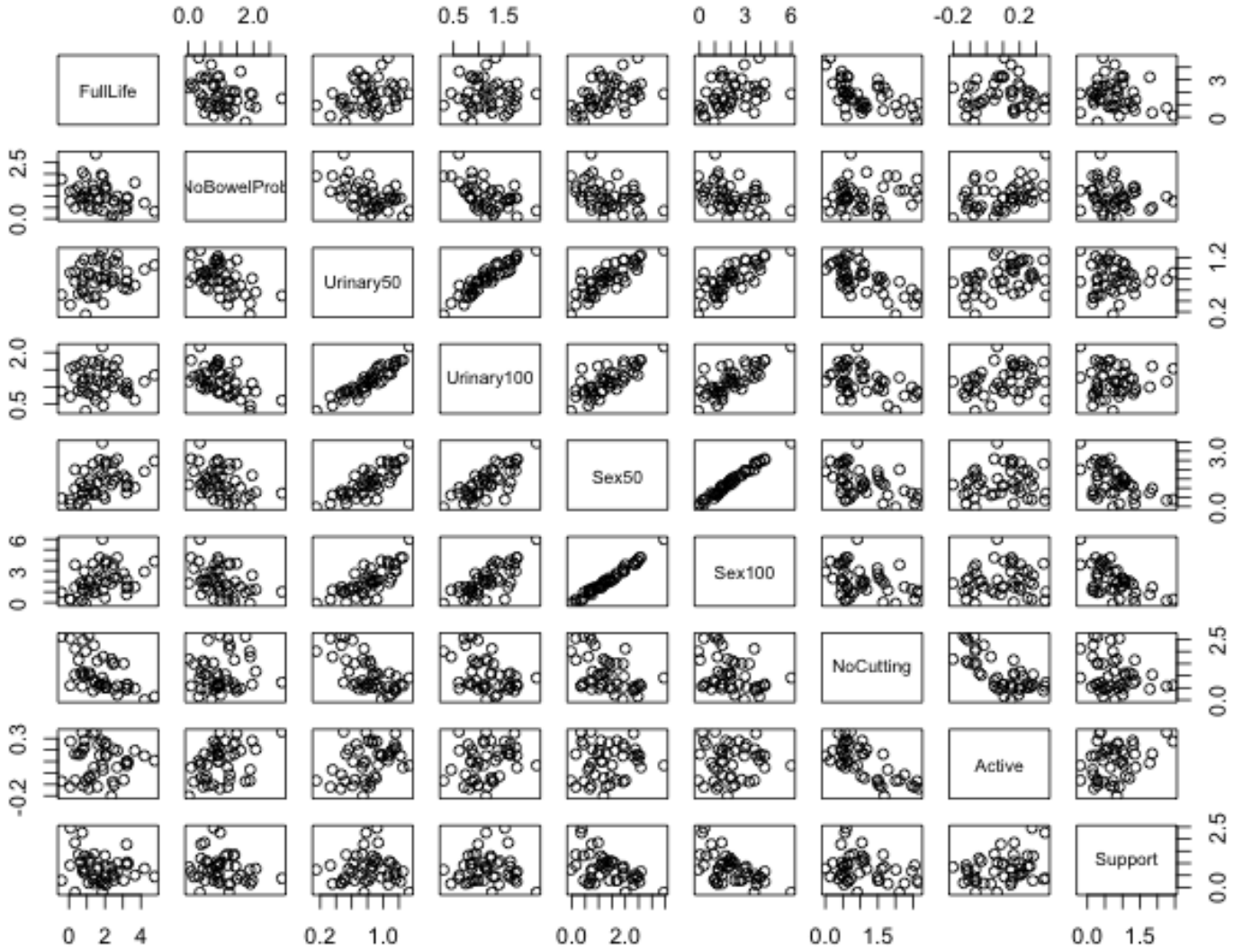

Figure 5.1: Scatterplot of Random Effects

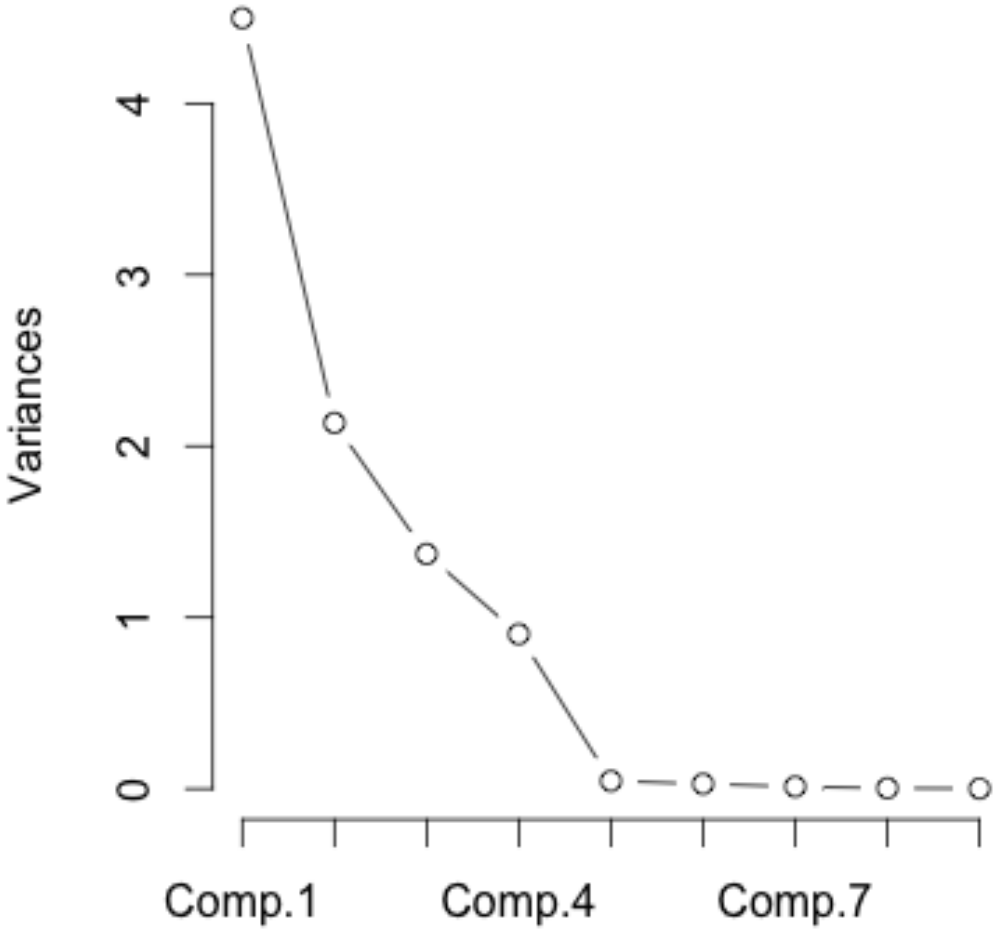

Figure 5.2: Scree plot

Table 5.7: Results of the principal components analysis using Bayesian estimates for correlation matrix

Eigenvalues	3.71	1.72	1.31	1.16	0.47	0.30	0.17	0.12	0.03
Eigenvectors	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
FullLife	0.227	-0.022	-0.387	0.687	-0.075	-0.296	-0.408	0.256	-0.018
Bowel100	-0.238	0.043	-0.569	-0.440	0.493	0.013	-0.201	0.375	0.008
Urinary50	0.418	0.168	0.186	-0.151	0.358	-0.723	0.291	0.063	-0.020
Urinary100	0.404	0.139	0.327	-0.309	0.041	0.119	-0.768	-0.094	-0.002
Sex50	0.482	-0.185	-0.104	-0.067	-0.065	0.277	0.208	0.263	0.724
Sex100	0.464	-0.254	-0.092	-0.111	-0.104	0.286	0.206	0.306	-0.686
NoCutting	-0.288	-0.463	0.306	-0.223	-0.384	-0.359	-0.154	0.506	0.060
Active	0.081	0.510	-0.369	-0.327	-0.674	-0.178	0.050	-0.031	-0.002
Support	-0.129	0.614	0.368	0.205	0.064	0.239	0.084	0.601	-0.013

CHAPTER 6

A Hierarchical Bayes Model for Discrete Choice Data in Health Care

This chapter presents the Bayesian hierarchical model for best-worst choice data with random effects and patient covariates, defines measures of relative importance, presents CPO-based measures for outlier detection, and demonstrates application of our methods to data from the PROSPECT study.

6.1 Introduction

Discrete choice experiments (DCEs) have been increasingly used in health applications to characterize the preferences of individual patients for various health care interventions and services (Lancsar et al., 2013; DeBekker-Grob et al., 2012). In a typical health care DCE, patients are presented with sets of health states described by various attributes and asked to make choices from among them (Ryan et al., 2008). For example, a patient might be asked to choose between a health state with long life expectancy and poor quality of life and a health state with shorter life expectancy and high quality of life. By asking individuals to make choices between health states, they are forced to make trade-offs that reveal information about their preferences for different aspects of health-related quality of life.

Historically, in a DCE, patients provided their most preferred health state or a full ranking of a set of possible health states. However, continued research in discrete choice experiments has led to the development of best-worst designs in which patients provide only their most preferred and least preferred choices (Lancsar and Louviere, 2008; Louviere et al., 2008). While reducing patient burden compared to full rankings, best-worst discrete choice

experiments pose new statistical challenges. In such data, incomplete ranking information occurs when choosing best and worst from among four or more health states, and patient-level data are often insufficient to estimate individual-level preferences using maximum likelihood methods as it is not uncommon to obtain estimates of the coefficients in the wrong direction with sparse data (Allenby et al., 2005; Rao, 2008).

A number of models have been developed for discrete choice data. The multinomial logit models the probability of observing best choices (McFadden, 1974), while the rank-ordered logit models the probability of full rankings (Allison and Christakis, 1994). Mixed logit models include random effects that vary across individuals to account for heterogeneity in preferences (Revelt and Train, 1998; McFadden and Train, 2000). More recently, Allenby et al. (2005) developed a Bayesian hierarchical model for best choices with random effects and individual-level covariates and Hernandez-Alava et al. (2013) introduced a model for ranked and partially ranked data that includes random effects, and estimated the random effects using Monte Carlo maximum likelihood methods. Although the model introduced by Hernandez-Alava et al. accommodates partially ranked data, it is not uncommon to obtain coefficient estimates in the wrong direction when using maximum likelihood estimation with sparse data (Rao, 2008). Moreover, their model does not include individual-specific covariates although inference on covariate effects is often of interest and it has been shown that including covariates can improve preference estimates for the mixed logit (Crabbe and Vandebroek, 2011; Orme and Howell, 2009; Greene et al., 2006; Allenby et al., 2005).

In many studies a key purpose of the DCE is to obtain an individual's ranking of various attributes relative to each other. The concept of relative attribute importance is widely used in the marketing research literature to provide rankings of features of consumer products (Paul E. Green, 1978; Halbrendt et al., 1995; Orme, 2010). Recently, this concept has been extended into the health care domain (Dowsey et al., 2016; Kruk et al., 2016; van Dijk et al., 2016). In this context, the purpose of the DCE is to obtain an individual's ranking of various attributes of health care or health-related quality of life, so that this information can be used as part of the health care decision-making process. For example, how a prostate cancer patient values full sexual functioning, long lifespan and no urinary incontinence relative to

each other may inform which treatment options are a better match for the patient. While discrete choice data are now routinely analyzed using Bayesian hierarchical models with random effects to accommodate preference heterogeneity (McFadden and Train, 2000; Train, 2001; Allenby et al., 2005; Train, 2009), methods to compute relative attribute importance for such models are not fully developed.

Methods to identify outliers for such models are also lacking. Using the means of the individual-specific parameter distributions, Campbell and Hess (2010) classified individuals in the upper and lower percentiles as outliers. Farrel et al. (2012) proposed a graphical method to identify outliers by plotting standardized random effects against their expected values for a Bayesian hierarchical logistic regression model. Several approaches for outlier detection in Bayesian models have been explored. For example, using the posterior distribution of the residuals of a regression model, Chaloner and Brant (1988) and Chaloner (1991, 1994) define an outlier as an observation with a large random error and calculate the posterior probabilities that observations are outlying. Other approaches for outlier detection are based on the predictive distribution. The conditional predictive ordinate (CPO), first suggested by Geisser (1980), is a diagnostic measure used to detect observations discrepant with the proposed model (Geisser, 1980, 1987, 1989, 1993; Dey et al., 1997; Pettit, 1990). To our knowledge, CPO has not been used to identify outlying random effects.

In this paper, we develop a Bayesian hierarchical model for best-worst discrete choice data. Our model extends previous approaches. Incomplete rankings are handled by marginalizing over all possible permutations of unranked health states in a model that includes random effects to model individual-specific preferences. Bayesian methods are used to overcome the problem of sparse data to obtain estimates of individual preferences. To enable analysis of how patient characteristics are related to preferences, we model individual-specific preferences as a function of individual-specific covariates. We also define Bayesian versions of relative attribute importance for individuals and for the population that handle random effects and covariates. To identify outliers in DCEs, we adapt the CPO in two ways: we adapt it to include random effects to identify patients who are unusual in their preferences for specific attributes or combinations of attributes, and we adapt it to handle vector outcomes

to identify choice sets that are outlying with respect to individual preferences.

The paper is organized as follows. Section 6.2 describes the motivating dataset and defines terms used throughout the remainder of the paper. Section 6.3 presents the Bayesian hierarchical model for best-worst choice data with random effects and patient covariates. Section 6.4 defines measures of relative importance, while Section 6.5 presents CPO-based measures for outlier detection. Section 6.6 demonstrates application of our methods to data from the PROSPECT study. This is followed by a discussion in Section 6.7.

6.2 Motivating Example: The PROSPECT Study

The methods are motivated by the PROSPECT (PROState cancer PrEferenCes for Treatment) study, which used a DCE to understand patient preferences for aspects of health-related quality of life associated with prostate cancer treatment outcomes (Saigal and Dahan, 2012). The 121 patients were men with negative prostate biopsies.

We make the following definitions. An *attribute* is a characteristic of a treatment or a health state resulting from a treatment. For simplicity of discussion, we define an *attribute* as a characteristic of a health state. Attributes are defined by at least two attribute levels. For example, sexual functioning is an attribute with three attribute levels, no sexual functioning, decreased sexual functioning and full sexual functioning. Investigators identified seven attributes important for prostate cancer treatment decision making using a *Voice of the Patient* process (Saigal and Dahan, 2012). In addition to sexual functioning, these were urinary incontinence, bowel issues, expected lifespan, others' support for the proposed treatment, cutting and taking immediate action towards treatment. Table 3.2 presents the seven attributes with their attribute levels. *Health state attribute variables* are dummy variables for health state attributes with the lowest attribute level as the reference group. A *health state* is defined by specifying attribute levels for each of the seven attributes. Sets of health states from which patients make choices are called *choice sets* and a health state contained in a choice set is called an *alternative*. An example of a choice set is shown in Figure 3.1.

In the PROSPECT study, Patients were presented with choice sets comprised of four

hypothetical health states that could result from various cancer treatments, and asked to choose their most and least preferred health state from each set, leaving two health states unranked. Sixteen health states were selected by investigators for creation of choice sets. These sixteen health states described by their attribute levels are presented in Table 3.3. The first four choice sets were the same for all patients and consisted of health states {1,3,9,15}, {2,4,10,14}, {5,6,11,12} and {7,8,13,16}. An algorithm was used to create the remaining choice sets for each patient. The algorithm composed subsequent choice sets in a manner that achieved an implicit ranking of the sixteen states using the minimum of choice sets. As a result, the number of choice sets as well as the choice sets presented to each patient differed. The number of choice sets per patient ranged from 10 to 17.

6.3 Bayesian Hierarchical model for Best-Worst Choice Data

Our model includes a probability model for best-worst choice data with incomplete rankings, a hierarchical prior distribution, and individual-specific covariates predicting an individual's preference scores for attributes.

6.3.1 Probability Model

Let $i = 1, ..., N$ index patients, $t = 1, ..., T_i$ index choice sets within patient i, and $j = 1, ..., J_{it}$ index the health states within choice set t presented to patient i, where N is the total number of patients, T_i is the total number of choice sets presented to patient i, and J_{it} is the total number of health states in choice set t presented to patient i.

Let $\boldsymbol{Y}_{it}$ be a $J_{it} \times 1$ vector describing an observed full ranking of a choice set, where element y_{itj} of $\boldsymbol{Y}_{it}$ is the observed jth ranked health state in choice set t presented to individual i. For example, suppose patient i gives a full ranking $D > A > C > B$ of choice set $t = \{A, B, C, D\}$ where D is most preferred and B is least preferred. Then $\boldsymbol{Y}_{it} = (y_{it1}, y_{it2}, y_{it3}, y_{it4})^{\mathsf{T}} = (D, A, C, B)^{\mathsf{T}}$.

We use a linear predictor to relate choices to the attribute levels of health states (Hauber

et al., 2016). Let $\boldsymbol{x}_{itj}$ be an $H \times 1$ vector encoding the attribute levels of the jth ranked health state in choice set t presented to individual i, where H is the total number of health state attribute variables. Let $\boldsymbol{\beta}_i$ be an $H \times 1$ unknown vector of preference scores for individual i.

Suppose that individual i provides only a most preferred health state for choice set t. Then the probability that individual i chooses the jth health state as the best state in choice set t is

$$p(y_{itj}|\boldsymbol{\beta}_i) = \frac{\exp(\boldsymbol{x}_{itj}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{k=1}^{J_{it}} \exp(\boldsymbol{x}_{itk}^{\mathsf{T}}\boldsymbol{\beta}_i)}, \tag{6.1}$$

where the summation is over the health states in the choice set (McFadden, 1974).

Patient i is presented with T_i choice sets, each of size $J_{it} = 4$. Eliciting best and worst choices from a choice set of size four yields a partial ranking of the choice set. Two possible full rankings are consistent with each partial ranking. For example, the full rankings $A > B > C > D$ and $A > C > B > D$ are consistent with the partial ranking $A > \{B, C\} > D$, where A is most preferred and D is least preferred. We can model the probability of observing a full ranking of health states as a product of probabilities, where each factor in the product is the probability of observing a best choice from a subsequently smaller choice set. For example, the probability of observing the full ranking $A > B > C > D$ is the product of the probability of choosing A as best from the choice set $\{A, B, C, D\}$ times the probability of choosing B as best from the choice set $\{B, C, D\}$ times the probability of choosing C as best from the choice set $\{C, D\}$. The probability of choosing D from $\{D\}$ is one.

Let $r = 1, ..., R_{it}$ index the full rankings consistent with an elicited partial ranking of choice set t for patient i, where R_{it} is the total number of possible full rankings consistent with the partial ranking. Then the probability of observing $\boldsymbol{Y}_{rit}$, a full ranking consistent with the partial ranking $\boldsymbol{Y}_{it}$, is written as the probability of a sequence of choices (Louviere et al., 2008; Bergland, 1994; Chapman and Staelin, 1982),

$$p(\boldsymbol{Y}_{rit}|\boldsymbol{\beta}_i) = \prod_{j=1}^{J_{it}-1} \frac{\exp(\boldsymbol{x}_{ritj}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{J_{it} \geq k \geq j} \exp(\boldsymbol{x}_{ritk}^{\mathsf{T}}\boldsymbol{\beta}_i)}. \tag{6.2}$$

Because the set of R_{it} full rankings consistent with $\boldsymbol{Y}_{it}$ is a set of mutually exclusive events, marginalizing over all possible permutations of unranked health states amounts to summing over all possible full rankings, and the probability of observing a partial ranking $\boldsymbol{Y}_{it}$ is

$$p(\boldsymbol{Y}_{it}|\boldsymbol{\beta_i}) = \sum_{r=1}^{R_{it}} \left(\prod_{j=1}^{J_{it}-1} \frac{\exp(\boldsymbol{x}_{ritj}^{\mathsf{T}}\boldsymbol{\beta}_i)}{\sum_{J_{it}\geq k\geq j} \exp(\boldsymbol{x}_{ritk}^{\mathsf{T}}\boldsymbol{\beta}_i)} \right). \tag{6.3}$$

If a patient is asked to provide their most preferred and least preferred health states from a choice set t containing fewer than four alternatives or if choice set t is fully ranked, then Equation (6.3) simplifies to Equation (6.2). Moreover, if we observe only best choices, then Equation (6.3) simplifies to Equation (6.1).

Let $\boldsymbol{Y}_i = \{\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{iT_i}\}$ represent the set of partial rankings made by patient i over the course of the experiment. Then assuming that each set of rankings $\boldsymbol{Y}_{it}$ is conditionally independent given $\boldsymbol{\beta}_i$, the likelihood contribution for individual i is given by

$$p(\boldsymbol{Y}_i|\boldsymbol{\beta}_i) = \prod_{t=1}^{T_i} p(\boldsymbol{Y}_{it}|\boldsymbol{\beta_i}). \tag{6.4}$$

6.3.2 Hierarchical Prior Distributions

Let $\boldsymbol{z_i}$ be a $Q \times 1$ vector of patient covariates for individual i including an intercept. For example, suppose we want to include an indicator for patient age greater than 65 years in the model. Then we could let $\boldsymbol{z_i} = (1, z_{i1})^{\mathsf{T}}$ where $z_{i1} = 1$ when patient age is greater than 65 and $z_{i1} = 0$ otherwise. To model patient preferences as a function of patient covariates, we model random effect $\boldsymbol{\beta}_i$ as a linear function of $\boldsymbol{z_i}$ plus error as

$$\boldsymbol{\beta}_i = \boldsymbol{\Gamma}\boldsymbol{z_i} + \boldsymbol{\epsilon_i}, \tag{6.5}$$

where $\boldsymbol{\Gamma}$ is an unknown $H \times Q$ matrix of fixed regression coefficients and $\boldsymbol{\epsilon_i}$ is an $H \times 1$ mean zero random effect vector that allows patients with the same covariates to have different

values for $\boldsymbol{\beta}_i$. We model $\boldsymbol{\epsilon_i}$ as

$$\boldsymbol{\epsilon_i}|\boldsymbol{\Sigma} \sim \text{Normal}_H(\mathbf{0}, \boldsymbol{\Sigma}), \tag{6.6}$$

a multivariate normal distribution with mean vector $\mathbf{0}$ and $H \times H$ covariance matrix $\boldsymbol{\Sigma}$. Let $h = 1, ..., H$ index health state attribute variables, and $q = 1, ..., Q$ index patient covariates. Then each element γ_{hq} of $\boldsymbol{\Gamma}$ describes the effect of covariate q on patient preference for attribute variable h. We set the prior for the γ_{hq} as

$$\gamma_{hq} \sim \text{Normal}(0, 1), \tag{6.7}$$

and the prior for $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} \sim \text{inverseWishart}(w, \boldsymbol{W}), \tag{6.8}$$

an inverse Wishart distribution with w degrees of freedom and scale matrix $\boldsymbol{W}$. We set the prior mean of the inverse Wishart distribution equal to the identity matrix. If no covariates are included, then Equation (6.5) reduces to $\boldsymbol{\beta}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon_i}$, where $\boldsymbol{\mu} = (\mu_h)$ is the $H \times 1$ unknown population mean vector of the distribution of $\boldsymbol{\beta}_i$. In this case, we set a prior for $\boldsymbol{\mu}$ as $\boldsymbol{\mu} \sim \text{Normal}_H(\mathbf{0}, \mathbb{I}_H)$, where $\mathbf{0}$ is the $H \times 1$ zero vector and $\mathbb{I}_H$ is the $H \times H$ identity matrix.

6.4 Relative Attribute Importance

An attribute may be represented using two, three or more levels. When using dummy variable coding, this yields one, two or more coefficients where the coefficient for the reference level is defined to be zero. In market research, the difference between the estimated maximum and minimum attribute-level coefficients has been used as a measure of attribute importance (Paul E. Green, 1978; Halbrendt et al., 1995; Orme, 2010). Relative attribute importance is calculated by normalizing attribute importance measures to sum to one, so that the relative

importance of an attribute is a proportional contribution to the importance of all attributes jointly (Soofi et al., 2000). Although model coefficients can be estimated using maximum likelihood or Bayesian methods (Orme, 2010), current methods only provide point estimates of relative importance. We extend current measures by defining relative attribute importance as a function of the random-effects $\boldsymbol{\beta}_i$, and describe Bayesian versions of relative atribute importance.

Let $a = 1, ..., A$ index health state attributes, where A is total number of health state attributes and let $k = 1, ..., K_a$ index the attribute levels of attribute a, where K_a is the total number of attribute levels for attribute a. In the PROSPECT study, we consider seven health state attributes. Urinary functioning and sexual functioning each have three attribute levels, while the other attributes have two levels. The importance of attribute a for individual i is defined as

$$\max_k \beta_{iak} - \min_k \beta_{iak},$$

where β_{iak} is an unknown preference score for the k$^{\text{th}}$ attribute level within attribute a for patient i. Using Equation (6.5), we define the relative importance (RI) of attribute variable a for individual i as the proportional contribution of attribute variable A to the sum of all attributes' importance,

$$\text{RI}_{ia} = \frac{\max_k(\boldsymbol{\gamma}_{ak}^{\mathsf{T}}\boldsymbol{z}_i + \epsilon_{iak}) - \min_k(\boldsymbol{\gamma}_{ak}^{\mathsf{T}}\boldsymbol{z}_i + \epsilon_{iak})}{\sum_{f=1}^{A} \max_k(\boldsymbol{\gamma}_{fk}^{\mathsf{T}}\boldsymbol{z}_i + \epsilon_{ifk}) - \min_k(\boldsymbol{\gamma}_{fk}^{\mathsf{T}}\boldsymbol{z}_i + \epsilon_{ifk})}, \tag{6.9}$$

where $\boldsymbol{\gamma}_{ak}^{\mathsf{T}}$ is the row of $\boldsymbol{\Gamma}$ corresponding to attribute level k within attribute a, and ϵ_{iak} is the random effect for the k$^{\text{th}}$ attribute level within attribute a for individual i. If no patient covariates are included in the model, then $\boldsymbol{\gamma}_{ak}^{\mathsf{T}}\boldsymbol{z}_i$ reduces to μ_{ak}, the k$^{\text{th}}$ attribute level population preference score within attribute a for attribute a.

We can define the average relative importance (ARI) of attribute a for the population as the arithmetic average of Equation (6.9) over all patients,

$$\text{ARI}_a = \frac{1}{N}\sum_{i=1}^{N} \text{RI}_{ia}. \tag{6.10}$$

For a specific set of patient covariates $\boldsymbol{z}$, we define the relative importance of attribute a for the population as

$$\mathrm{RI}_{az} = \frac{\max_k(\boldsymbol{\gamma}_{ak}^{\mathsf{T}}\boldsymbol{z}) - \min_k(\boldsymbol{\gamma}_{ak}^{\mathsf{T}}\boldsymbol{z})}{\sum_{f=1}^{A} \max_k(\boldsymbol{\gamma}_{fk}^{\mathsf{T}}\boldsymbol{z}) - \min_k(\boldsymbol{\gamma}_{fk}^{\mathsf{T}}\boldsymbol{z})}, \tag{6.11}$$

where the summation is over all attributes. This formulation can be used to, for example, compute marginal predictions at specific patient covariate values. If no patient covariates are included in the model, then $\boldsymbol{\gamma}_{ak}^{\mathsf{T}}\boldsymbol{z}$ reduces to μ_{ak} and we get estimates of relative importance at the population level.

Equation (6.11) differs from Equation (6.10) in that relative importance is calculated from population parameters, rather than as an average of the individual preference scores.

The posterior means and standard deviations of Equations (6.9), (6.10), and (6.11) are estimated as the means and standard deviations of the MCMC samples of relative importance scores, calculated using randomly sampled draws from the posterior distributions of the relevant parameters.

6.5 Outlier Statistics for Choice Sets and Preferences

We use the conditional predictive ordinate (CPO) (Geisser, 1980, 1987, 1989, 1993; Dey et al., 1997; Pettit, 1990) to identify outliers in discrete choice data. In general, suppose we have a set of observations $\boldsymbol{Y} = (Y_1, \ldots, Y_S)$ which we model using parameters $\boldsymbol{\theta}$. Let $\boldsymbol{Y}_{(s)}$ be the vector $\boldsymbol{Y}$ after omitting Y_s. The CPO for observation Y_s is the predictive density of Y_s conditional upon the model and all other observations $\boldsymbol{Y}_{(s)}$ (Geisser, 1980)

$$\begin{aligned} \mathrm{CPO}_s &= p(Y_s|\boldsymbol{Y}_{(s)}) && (6.12)\\ &= \int p(Y_s|\boldsymbol{\theta}, \boldsymbol{Y}_{(s)}) p(\boldsymbol{\theta}|\boldsymbol{Y}_{(s)}) d\boldsymbol{\theta}, && (6.13) \end{aligned}$$

where $p(Y_s|\boldsymbol{\theta}, \boldsymbol{Y}_{(s)})$ is the distribution of Y_s given $\boldsymbol{\theta}$ and $\boldsymbol{Y}_{(s)}$. Small values of CPO indicate that observation Y_s is a poor fit to a given model.

We can use CPO to identify outlying choice sets as follows. If we let $\boldsymbol{Y} = (\boldsymbol{Y}_1^{\mathsf{T}}, \ldots, \boldsymbol{Y}_S^{\mathsf{T}})^{\mathsf{T}}$ be the vector of $S = \sum_{i=1}^{N} T_i$ observed choice set rankings across all N patients and let $\boldsymbol{Y}_{(s)}$ be the vector after omitting choice set s, we can use Equation (6.13) to calculate CPO for the observed ranking of choice set s, $\boldsymbol{Y}_s$. To find outlying choice sets inconsistent with a patient's preferences, we can calculate and compare the CPOs for each of their choice sets, CPO-SET$_{i1}, \ldots,$ CPO-SET$_{iT_i}$.

Gelfand et al. (1992), Dey et al. (1997), Gelfand (1996), Pettit (1990), and Weiss (1994, 1996) observed that

$$\text{CPO}_s = \left\{ \text{E}_{\boldsymbol{\theta}|\boldsymbol{Y}} \left[\frac{1}{p(Y_s|\boldsymbol{\theta}, \boldsymbol{Y}_{(s)})} \right] \right\}^{-1}, \tag{6.14}$$

and showed that Monte Carlo integration can be used to estimate CPO (Gelfand et al., 1992; Gelfand, 1996) using a posterior sample from $p(\boldsymbol{\theta}|\boldsymbol{Y})$. Drawing an MCMC sample $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^G$ of size G, where $g = 1, ..., G$ indexes iterations of the Gibbs sampler, from the full posterior density after the burn-in period allows us to obtain a Monte Carlo approximation of CPO for choice set s as

$$\text{CPO-SET}_s \approx \left\{ \frac{1}{G} \sum_{g=1}^{G} \frac{1}{p(\boldsymbol{Y}_s|\boldsymbol{\theta}^g, \boldsymbol{Y}^{(s)})} \right\}^{-1}. \tag{6.15}$$

We also use CPO to identify patients with outlying preferences with respect to the population. To do so, we define several varieties of the conditional predictive ordinate for preferences. Suppose we want to identify patients with outlying preferences on a single attribute variable h. Let $\text{L}_h = (0, ..., 0, 1, 0, ..., 0)^{\mathsf{T}}$ be an $H \times 1$ indicator vector for attribute variable h, where the single 1 in L_h^{T} corresponds to the h^{th} component of L_h^{T} and all other components are zero. Then $\text{L}_h^{\mathsf{T}}\boldsymbol{\beta}_i = \beta_{ih}$, where β_{ih} the unknown preference score for individual i and attribute variable h. Let $\boldsymbol{\theta}_{(ih)}$ be the vector of model parameters $\boldsymbol{\theta}$ after omitting $\text{L}_h^{\mathsf{T}}\boldsymbol{\beta}_i$. Then the CPO for individual i and attribute variable h, which we denote CPO-UVP (univariate preference), is defined as the inverse of the posterior mean of the inverse prior

density of $\mathrm{L}_h^\mathsf{T}\boldsymbol{\beta}_i$ from equation (6.6),

$$\text{CPO-UVP}_i^h = p(\mathrm{L}_h^\mathsf{T}\boldsymbol{\beta}_i|\boldsymbol{\theta}_{(ih)}) \tag{6.16}$$

$$= \Big\{ \int \frac{1}{p(\mathrm{L}_h^\mathsf{T}\boldsymbol{\beta}_i|\boldsymbol{\theta}_{(ih)}, \boldsymbol{Y})} p(\boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\theta} \Big\}^{-1} \tag{6.17}$$

$$= \Big\{ \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}} \Big[\frac{1}{p(\mathrm{L}_h^\mathsf{T}\boldsymbol{\beta}_i|\boldsymbol{\theta}_{(ih)}, \boldsymbol{Y})} \Big] \Big\}^{-1}, \tag{6.18}$$

where $p(\mathrm{L}_h^\mathsf{T}\boldsymbol{\beta}_i|\boldsymbol{\theta}_{(ih)}, \boldsymbol{Y})$ is the distribution of $\mathrm{L}_h^\mathsf{T}\boldsymbol{\beta}_i$ given $\boldsymbol{\theta}_{(ih)}$ and $\boldsymbol{Y}$.

More generally, suppose we want to identify patients with outlying preferences on a combination of attribute variables. For example, in our application, urinary functioning and sexual functioning are represented by two attribute variables and thus two component $\boldsymbol{\beta}_i$. To do so, we can define an appropriate $H \times M$ indicator matrix L_c in which each row selects one of the desired attribute variables. For example, to select the 8^{th} and 9^{th} elements of the attribute vector corresponding to short term sexual issues and full sexual functioning, we can use

$$\mathrm{L}_c^\mathsf{T} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then the CPO for individual i and combination of attribute preferences c, which we could here denote as CPO-BVP (bivariate preference), is defined using Equation (6.18)

$$\text{CPO-BVP}_i^c = \Big\{ \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Y}} \Big[\frac{1}{p(\mathrm{L}_c^\mathsf{T}\boldsymbol{\beta}_i|\boldsymbol{\theta}_{(ic)}, \boldsymbol{Y})} \Big] \Big\}^{-1}, \tag{6.19}$$

where $\boldsymbol{\theta}_{(ic)}$ is the vector of model parameters $\boldsymbol{\theta}$ minus $\mathrm{L}_c^\mathsf{T}\boldsymbol{\beta}_i$. CPO-BVP can also be computed for other combinations of attributes, for example, full lifespan and others' support. Drawing an MCMC sample of size G, $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^G$, from the full posterior density after the burn-in period allows us to obtain the following Monte Carlo approximations of CPO-BVP for individual i and list of attributes c

$$\text{CPO-BVP}_i^c \approx \Big\{ \frac{1}{G} \sum_{g=1}^{G} \frac{1}{p(\mathrm{L}_c^\mathsf{T}\boldsymbol{\beta}_i|\boldsymbol{\theta}_{(ic)}^g, \boldsymbol{Y})} \Big\}^{-1}. \tag{6.20}$$

We can also compute a more global outlier statistic for preferences. Identifying patients with outlying preferences on all attributes is a special case in which L_c is the identity matrix of size H. We call this statistic CPO-MVP$_i$.

6.6 Results

We fit the Bayesian hierarchical model of Section 6.3 to data from the 121 patients in the PROSPECT study. Dummy variables for three patient covariates were included in the model. These were: age ($\geq$ 65 years vs. age $<$ 65 years), race (black vs. white, other race vs. white), and partnered (vs. unpartnered). We chose a proper prior distribution (Gelman, 2006) for $\boldsymbol{\Sigma}^{-1}$ as Wishart$(9, \frac{1}{9}\mathbb{I}_9)$, where $\mathbb{I}_9$ is the 9×9 identity matrix, and we used Gibbs sampling implemented in JAGS (Plummer, 2003) to obtain posterior samples. Three Markov chains were run, each with a burn-in of 20,000 iterations, followed by 100,000 iterations keeping every 10^{th} draw of the chain. The final posterior sample consisted of 30,000 iterations (3 chains $\times$ 10,000 iterations).

The last two columns of Table 6.1 present the posterior means and standard deviations of the population mean preferences $\boldsymbol{\mu}$ for the model without patient covariates. Attribute variables were considered significant if the posterior probability that the parameter is greater than zero was at least 95% or at most 5%. Preferences for all attributes were nonzero except for taking action. Sexual functioning appeared to be the most important attribute affecting health state preference followed by full lifespan, urinary functioning, no bowel issues, no cutting, and others's support. For comparison, we also fit the model for best choices to our data (first and second columns of results in Table 6.1). The comparison shows that, by using all available information (best and worst choices), we gain precision in our estimates (smaller posterior standard deviations). From Table 6.1, we see that our model for best-worst choices consistently provides more precise estimates than does the model for best choices, while providing similar results. Table 6.1 also presents the standard deviations of the attribute-specific random effects, which describe the between-subject variation. For both models, we can see a relatively high standard deviation of the random effect for full life, indicating

Table 6.1: Posterior means (standard deviations) of the components of the vector of population mean preferences $\boldsymbol{\mu}$ and the standard deviations of the random effects.

Attribute Variable	Best Choice		Best-Worst Choice	
	Population Mean μ_h	SD of random effect ϵ_{ih}	Population Mean μ_h	SD of random effect ϵ_{ih}
Full Life	**2.67 (0.21)**	**1.79 (0.20)**	**2.20 (0.17)**	**1.68 (0.16)**
No Bowel Issues	**1.54 (0.15)**	**1.21 (0.15)**	**1.47 (0.13)**	**1.25 (0.12)**
No Cutting	**0.88 (0.17)**	**1.33 (0.16)**	**0.79 (0.11)**	**0.97 (0.10)**
Action	**0.27 (0.12)**	**0.82 (0.10)**	0.08 (0.08)	**0.61 (0.06)**
Others' Support	**0.94 (0.14)**	**1.06 (0.13)**	**0.69 (0.10)**	**0.83 (0.08)**
Short Term Urinary Issues	**1.21 (0.17)**	**1.27 (0.17)**	**1.22 (0.13)**	**1.21 (0.12)**
Full Urinary Functioning	**1.98 (0.20)**	**1.48 (0.21)**	**1.83 (0.16)**	**1.53 (0.15)**
Short Term Sexual Issues	**1.99 (0.23)**	**1.97 (0.25)**	**1.55 (0.15)**	**1.47 (0.14)**
Full Sexual Functioning	**3.16 (0.29)**	**2.70 (0.31)**	**2.48 (0.21)**	**2.03 (0.19)**

Boldface denotes that the posterior probability that the parameter is greater than zero is greater than 95% or less than 5%

substantial heterogeneity in preference between patients. Full sexual functioning also had high variance. In contrast, the random effects for taking action and others' support have relatively low standard deviations indicating less heterogeneity.

Table 6.2 presents the correlation matrix of the random effects. The correlation between short term urinary functioning and full urinary functioning is 0.84, as might be expected, since they measure the same attribute. We find the same relationship between short term sexual functioning and full sexual functioning. No cutting is negatively correlated with each of the other attributes implying that patients who prefer no cutting place less value on all of the other attributes.

Table 6.3 presents the posterior means and standard deviations of the regression coefficients $\mathbf{\Gamma}$ for the model including patient covariates and thus shows how preferences vary with age, race, and partnership status. The column labeled *Intercept* contains the posterior means and standard deviations corresponding to younger (<65 years old), white, and unpartnered patients. For this particular group, preferences for all attributes except taking action were nonzero. Older men (>= 65 years old) appeared to favor full lifespan and urinary functioning more than younger men. For older men, each of these attributes were associated with approximately a 0.7 and 0.6 (respectively) higher estimated patient preference score than younger men. Differences in preferences were also found by partnership status. Partnered men favored full lifespan more than unpartnered men by 0.98 points.

Figure 6.1 presents the posterior mean average relative importance scores for each health state attribute for the population and the posterior mean relative importance scores for fourteen sample patients. To select the fourteen patients in Figure 6.1, patients were sorted by decreasing relative importance score on full lifespan and every 10^{th} ranked patient was selected. This figure shows the heterogeneity of preferences for health state attributes in the sample. Greater heterogeneity in preference for full lifespan and lower heterogeneity in preference for taking action were apparent.

Table 6.4 presents the posterior mean relative importance scores for each health state attribute for the population and for three sample patients. In general, the standard devia-

Table 6.2: Posterior means (standard deviations) of the elements of the correlation matrix of the residual effect $\boldsymbol{\epsilon}_i$ for the model without patient covariates.

	Full Life	No Bowel Issues	Short Term Urinary Issues	Full Urinary Functioning	Short Term Sexual Issues	Full Sexual Functioning	No Cutting	Taking Action	Others' Support
Full Life	1								
No Bowel Issues	0.15 (0.11)	1							
Short Term Urinary Issues	**0.28 (0.11)**	**0.29 (0.11)**	1						
Full Urinary Functioning	**0.26 (0.11)**	**0.35 (0.11)**	**0.84 (0.04)**	1					
Short Term Sexual Issues	0.16 (0.11)	0.02 (0.12)	**0.26 (0.11)**	**0.29 (0.11)**	1				
Full Sexual Functioning	**0.19 (0.11)**	-0.04 (0.11)	**0.24 (0.11)**	**0.26 (0.11)**	**0.88 (0.03)**	1			
No Cutting	**-0.22 (0.12)**	-0.09 (0.12)	**-0.22 (0.12)**	**-0.24 (0.12)**	**-0.28 (0.11)**	**-0.24 (0.12)**	1		
Taking Action	0.005 (0.13)	0.18 (0.13)	0.12 (0.13)	0.14 (0.13)	0.05 (0.13)	0.02 (0.14)	-0.11 (0.13)	1	
Others Support	-0.03 (0.13)	-0.09 (0.13)	-0.18 (0.12)	-0.15 (0.13)	-0.14 (0.12)	**-0.21 (0.12)**	-0.06 (0.13)	0.08 (0.13)	1

Boldface denotes that the posterior probability that the parameter is greater than zero is greater than 95% or less than 5%

Table 6.3: Posterior means (standard deviations) of the elements of the matrix of regression coefficients $\mathbf{\Gamma}$ and the standard deviations of the random effects.

Attribute Variable	Patient-Specific Covariate					
	Intercept	$\geq$ 65 years (vs. $<$ 65 years)	Black (vs. White)	Other Race (vs. White)	Has Partner (vs. Does not)	SD of residual effect ϵ_{ih}
Full Life	**1.35 (0.35)**	**0.72 (0.32)**	-0.14 (0.34)	-0.12 (0.39)	**0.98 (0.33)**	**1.70 (0.17)**
No Bowel Issues	**1.36 (0.29)**	0.11 (0.26)	0.38 (0.29)	0.23 (0.32)	-0.12 (0.27)	**1.31 (0.12)**
No Cutting	**0.66 (0.25)**	0.06 (0.22)	0.03 (0.24)	0.21 (0.28)	0.10 (0.23)	**1.01 (0.10)**
Taking Action	0.26 (0.18)	-0.12 (0.15)	-0.01 (0.17)	-0.17 (0.20)	-0.12 (0.16)	**0.63 (0.06)**
Others Support	**0.76 (0.22)**	-0.35 (0.19)	-0.12 (0.21)	-0.16 (0.24)	0.28 (0.20)	**0.83 (0.09)**
Short Term Urinary Issues	**1.14 (0.27)**	**0.54 (0.24)**	0.03 (0.27)	-0.39 (0.30)	-0.11 (0.25)	**1.22 (0.12)**
Full Urinary Functioning	**1.67 (0.34)**	**0.60 (0.31)**	0.14 (0.33)	-0.29 (0.37)	-0.09 (0.32)	**1.55 (0.16)**
Short Term Sexual Issues	**1.64 (0.30)**	-0.01 (0.28)	0.004 (0.30)	-0.72 (0.33)	0.18 (0.28)	**1.48 (0.15)**
Full Sexual Functioning	**2.27 (0.40)**	-0.05 (0.38)	0.13 (0.40)	-0.24 (0.45)	0.45 (0.39)	**2.12 (0.20)**

Boldface denotes that the posterior probability that the parameter is greater than zero is greater than 95% or less than 5%

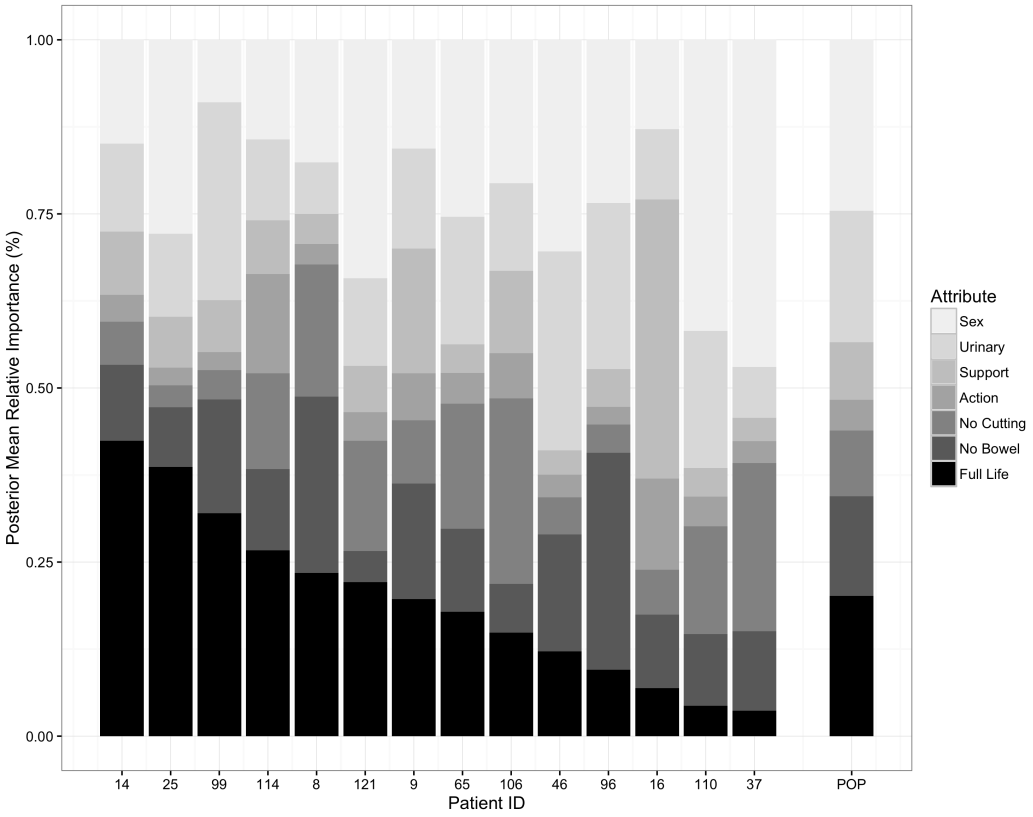

Figure 6.1: Posterior mean relative attribute importance scores for each health state attribute for fourteen men and for the population.

Table 6.4: Posterior mean (standard deviation) of relative attribute importance scores for three men and for the population.

Attribute	Patient 115	Patient 13	Patient 108	Population
Full Life	0.41 (0.05)	0.07 (0.04)	0.22 (0.05)	0.201 (0.005)
No Bowel Issues	0.22 (0.03)	0.09 (0.04)	0.11 (0.05)	0.143 (0.004)
No Cutting	0.07 (0.03)	0.07 (0.04)	0.07 (0.04)	0.094 (0.004)
Taking Action	0.02 (0.02)	0.07 (0.03)	0.04 (0.03)	0.044 (0.004)
Others Support	0.07 (0.03)	0.04 (0.03)	0.07 (0.05)	0.082 (0.004)
Urinary Functioning	0.10 (0.04)	0.31 (0.05)	0.16 (0.06)	0.189 (0.006)
Sexual Functioning	0.12 (0.04)	0.35 (0.06)	0.34 (0.07)	0.245 (0.006)

tions of the relative importance scores are small relative to the posterior means, suggesting that the posterior means provide a reliable ranking of attributes by relative importance. At a population level, sexual functioning, urinary functioning and full lifespan appear to be the three most important attribute variables, whereas taking action appears to be the least important. Patient 115 clearly placed highest importance on full lifespan, moderate importance on bowel issues, and low importance on all other attributes. Patient 13 placed highest importance on urinary functioning and sexual functioning. Patient 108 has posterior mean estimates similar to those of the population.

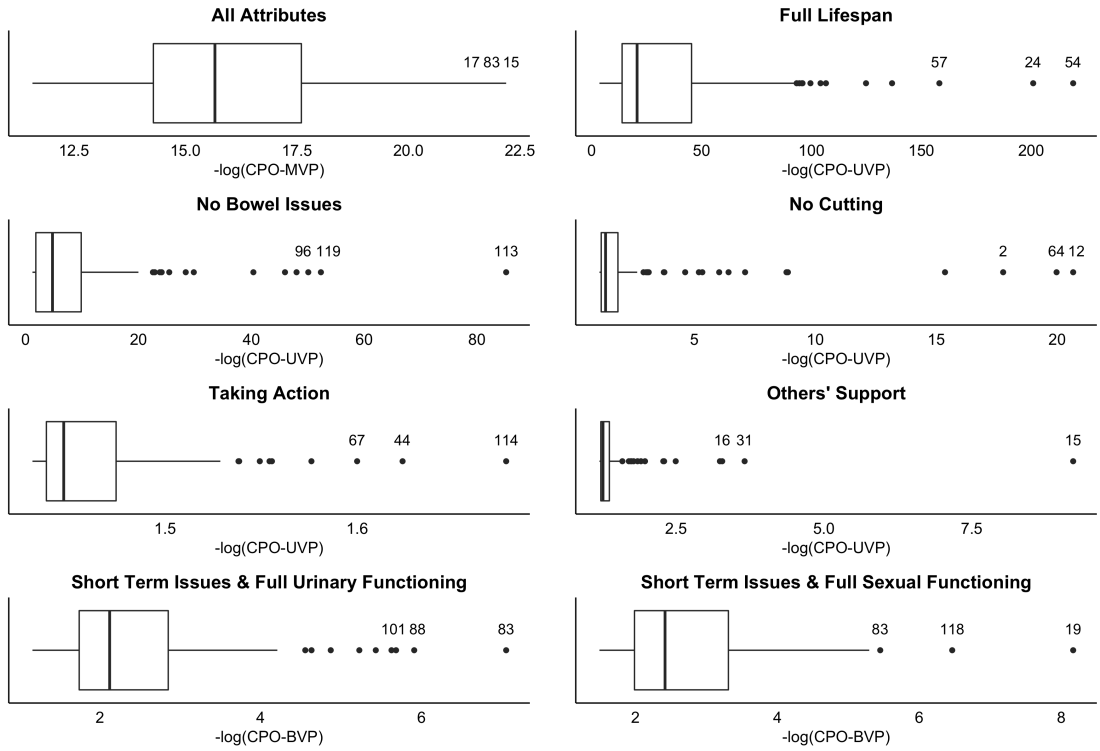

Figure 6.2: Plot of the -log(CPO-MVP)s on all health state attributes, the -log(CPO-UVP)s for specific attributes, and the -log(CPO-BVP)s for the bivariate combinations of attributes for urinary and sexual functioning for 121 patients. Patients with values of the outlier statistic in the upper 2.5^{th} percentile are labeled with ID numbers.

Figure 6.2 presents boxplots of CPO-MVP values for the set of all attributes, CPO-UVP values for specific attributes, and CPO-BVP values for the two bivariate combinations of attributes for urinary and sexual functioning for the 121 patients. A negative log-transformation was applied to the CPOs to better visualize small values. High values of

negative log-transformed CPOs indicate possible outliers (low CPO). Patients 15, 83, and 17 are multivariate outliers on the set of all attribute variables by CPO-MVP. Patient 83 is also outlying on the bivariate CPO for urinary functioning and the bivariate CPO for sexual functioning. Patient 15 had highest negative log-transformed CPO-UVP values on others' support. Patient 17 is an example of a multivariate outlier that cannot be detected by looking at outliers on specific health state attributes, while patient 54 is an example of a patient with outlying preferences on a single attribute who is not a multivariate outlier.

Figure 6.3 presents time series of the negative log CPO values for choice sets presented to eight patients. DCEs require patients to evaluate a number of different choice sets and some patients may undergo a learning effect where accuracy in responses improves with time. Conversely, some patients may become fatigued and accuracy of their responses may degrade as the number of questions increases (Bradlow et al., 1998; Hauser and Rao, 2002). By examining these time series, we can gain insight as to an individual's performance on discrete choice tasks, and observe possible learning effects or fatigue effects, and whether they made choices on specific sets that were inconsistent with their preferences. High values of negative log-transformed CPO indicate possible outlying choice sets. Patient 115 is an example of a patient with consistent responses and no outliers. In contrast, patient 52 shows highly variable responses, which might indicate more difficulty with the choice tasks. Patient 10 has an outlier on the first choice set, which may indicate a cognitive error early in the exercise. Patient 109 shows an upward trend suggesting a possible fatigue effect and an especially inconsistent choice on the second to last choice set. For patient 74, we observe a downward trend suggesting a learning effect where patient performance on choice tasks improves over time.

We conducted sensitivity analyses on the prior assumptions Wishart$(w, \boldsymbol{W})$ for the random effects precision matrix $\boldsymbol{\Sigma}^{-1}$ by comparing the posterior results over variations of the prior. With degrees of freedom parameter w and scale matrix $\boldsymbol{W}$ we explored the following Wishart specifications: $w = 9$ and $\boldsymbol{W} = \frac{1}{9}I$, $w = 9$ and $\boldsymbol{W} = \frac{1}{18}I$, $w = 18$ and $\boldsymbol{W} = \frac{1}{9}I$, $w = 18$ and $\boldsymbol{W} = \frac{1}{18}I$. There was little change in the posterior estimates for the elements of the correlation matrix or the population preference parameters with different specifications,

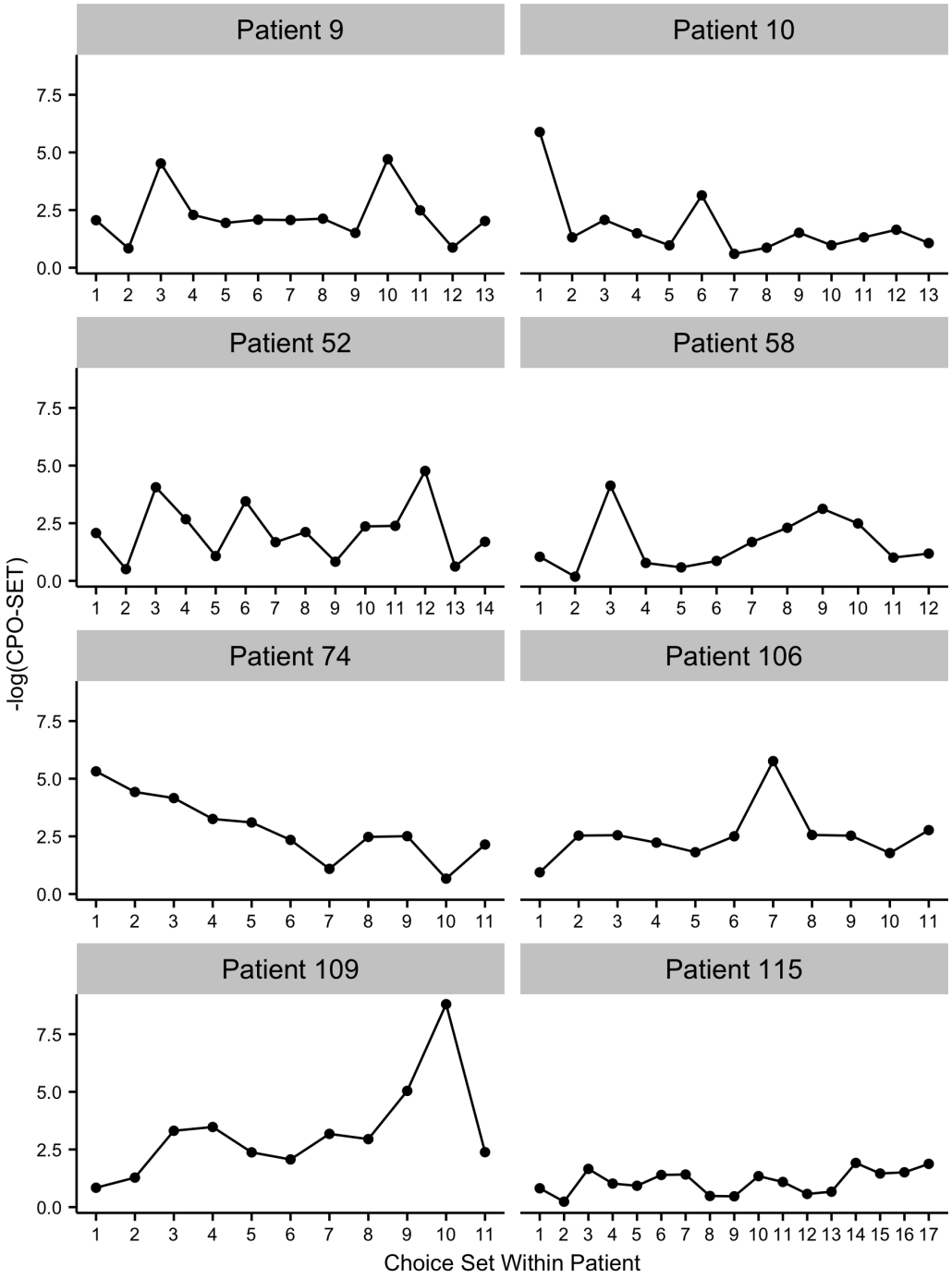

Figure 6.3: Plot of the -log(CPO-SET)s calculated for each choice set presented to eight patients.

indicating that the results were fairly robust to changes in the hyperprior specifications. The details of this analysis are given in Chapter 7.

6.7 Discussion

We developed a Bayesian hierarchical model for best-worst discrete choice data that accounts for incomplete rankings and includes patient covariates. The model can handle sparse data and is particularly useful when discrete choice experiments involve relatively few choice sets per patient. Although our application had choice sets of size four, the model can be applied to studies with larger choice sets.

The main goal of our discrete choice experiment was to identify health state attributes that are most important to individual patients to guide that individual's treatment; thus, we presented Bayesian versions of a commonly used measure of relative attribute importance. The estimates of relative attribute importance include posterior standard deviations that reflect uncertainty; in the literature, many studies only provide point estimates which may give false confidence about how the patient ranks the attributes. Our method for computing relative attribute importance is not specific to best-worst DCE and can be applied to other DCE designs. The concept of relative attribute importance is akin to the concept of variable importance in regression and prediction modeling. We have not explored other possible measures of variable importance that might be applied to DCE. The measurement of relative variable importance is an active area of research (Kruskal and Majors, 1989; Retzer et al., 2009; Johnson and Lebreton, 2004; Bi, 2012; Grömping, 2015; Harris and Burch, 2005).

We have shown how the conditional predictive ordinate can be adapted to identify outlying choice sets and outlying patients with unusual preferences in discrete choice data. Our CPO for identifying preference outliers finds outliers in the random effects. Random effects are a common feature of Bayesian models, and this new application of the CPO could have broader application in Bayesian modeling. The method is quite flexible and general, and can even identify outliers on sets of multiple random effects. We have shown how the method can be applied to identify outliers on categorical attributes modeled using two coefficients. The

CPO for identifying outlying choice sets utilizes a vector outcome and is also an important extension of the CPO that could be used in other applications.

Our application includes two attributes, sexual functioning and urinary functioning, whose attribute levels are naturally ordered; the levels of sexual functioning are none, decreased and full, and the levels of urinary functioning are long term issues, short term issues and full functioning. One approach to estimating the corresponding coefficients would be to impose order constraints, such that the coefficient for decreased functioning must be less than or equal to the coefficient for full functioning. This could be accomplished by specifying a truncated multivariate prior density on the vector of random effects and the vector of population effects (Gelfand et al., 1992). However, we obtained satisfactory results without imposing such constraints.

Experimental design for DCEs is an area of active research (Johnson et al., 2013; Jaynes et al., 2016); however, there is little consensus on the optimal design of choice experiments, including how to generate choice sets (Lusk and Norwood, 2005; Louviere et al., 2011). A recent report described alternative approaches to experimental design for DCEs (Johnson et al., 2013), but did not recommend any specific approach as best practice. The choice of alternatives for each choice set and the choice sets presented to each patient are important with regard to statistical efficiency. Random selection of profiles to choice sets may result in choice sets for which little information is gained on relative preferences because the attributes are not varied sufficiently. In addition, increasing the number of choice sets presented to patients can increase cognitive burden, jeopardizing the quality of patient responses. When creating a DCE, a trade-off is made between maximizing statistical efficiency and maximizing respondent efficiency (measurement error related to the quality of responses). A direction for future research would be to formally evaluate the impact of the experimental design on estimation of preferences.

Our DCE uses factors with different numbers of levels. Studies have shown that there is a positive association between the number of attribute levels and attribute importance scores (Wittink et al., 1982, 1990). Designing a study with the same number of attribute levels for each attribute may not be acceptable for some applications. In our study, all of

our attributes have either two or three levels. We think it reasonable that a priori important variables, such as urinary and sexual functioning, would be modeled using more levels. We fit the model after collapsing the two highest categories of urinary functioning and sexual functioning into a single category and obtained similar posterior means. Hence we surmise that the different numbers of levels did not appreciably affect our results.

The development of best-worst discrete choice designs reduces patient burden compared to full rankings while posing new statistical challenges. By accounting for missing ranking information, patient covariates, and the sparse nature of the individual-level data in a Bayesian framework, our model extends current methods and provides individual-level preference estimates. Our CPO measures provide some of the first diagnostic techniques for discrete choice models. Our model coupled with our measures of relative importance and outlyingness, provide practical methodology for discrete choice modeling applications, in which parameter estimation at the individual-level is desirable, but observed data at the individual-level are limited.

CHAPTER 7

Sensitivity Analyses

This chapter presents sensitivity analyses to evaluate the prior assumptions of the Bayesian hiearchical model developed in Chapter 6.

7.1 The Wishart Distribution for the Between-Attribute Precision Matrix

For the multivariate normal distribution of the random attribute effects vector of Chapter 6, the conjugate prior distribution for the between-attribute precision matrix is a Wishart distribution,

$$\boldsymbol{\Sigma^{-1}} \sim \text{Wishart}(w, \boldsymbol{W}), \tag{7.1}$$

with inverse-scale matrix $\boldsymbol{W}$ and w degrees of freedom where w is at least the length of the random attribute effects vector. The prior mean of the precision matrix is $w * \boldsymbol{W}^{-1}$, and smaller values of w imply a less informative distribution. The least informative, proper Wishart prior is obtained by setting w equal to the length of the random attribute effects vector.

We conducted sensitivity analyses for the specification of the Wishart$(w, \boldsymbol{W})$ prior for the random effects precision matrix $\boldsymbol{\Sigma}^{-1}$ by comparing the posterior results for different specifications of the prior. With degrees of freedom parameter w and scale matrix $\boldsymbol{W}$ with prior mean $w * \boldsymbol{W}^{-1}$, we explored the following Wishart specifications: $w = 9$ and $\boldsymbol{W} = (1/9) * \mathbb{I}_9$, $w = 9$ and $\boldsymbol{W} = (1/18) * \mathbb{I}_9$, $w = 18$ and $\boldsymbol{W} = (1/9) * \mathbb{I}_9$, $w = 18$ and

$\boldsymbol{W} = (1/18) * \mathbb{I}_9$.

Tables 7.1 and 7.2 summarize the posterior results for important parameters. Resulting changes in the posterior estimates for the elements of the correlation matrix (Table 7.2) as well as for the population preference parameters (Table 7.1) were relatively small, indicating that the posterior results were fairly robust to changes in the hyperprior specification.

Table 7.1: Sensitivity analyses for the specification of the Wishart prior for the random effects precision matrix $\boldsymbol{\Sigma}^{-1}$. The table provides posterior means and standard deviations of the components of the vector of population mean preferences $\boldsymbol{\mu}$ and the standard deviations of the random effect ϵ_{ih} for the model without patient covariates. The posterior probability that the parameters are greater than zero is also provided for each parameter.

	Wishart$(9, (1/9)\mathbb{I}_9)$			Wishart$(9, (1/18)\mathbb{I}_9)$			Wishart$(18, (1/9)\mathbb{I}_9)$			Wishart$(18, (1/18)\mathbb{I}_9)$		
Parameter	Mean	SD	Pr(>0)	Mean	SD	Pr(>0)	Mean	SD	Pr(>0)	Mean	SD	Pr(>0)
μ_1	2.20	0.17	1.00	2.32	0.18	1.00	2.09	0.15	1.00	2.22	0.17	1.00
μ_2	1.47	0.13	1.00	1.55	0.14	1.00	1.39	0.12	1.00	1.48	0.13	1.00
μ_3	1.22	0.13	1.00	1.28	0.14	1.00	1.16	0.12	1.00	1.23	0.13	1.00
μ_4	1.83	0.16	1.00	1.93	0.18	1.00	1.75	0.15	1.00	1.85	0.16	1.00
μ_5	1.55	0.15	1.00	1.63	0.16	1.00	1.48	0.14	1.00	1.57	0.15	1.00
μ_6	2.47	0.21	1.00	2.60	0.22	1.00	2.36	0.19	1.00	2.50	0.20	1.00
μ_7	0.79	0.11	1.00	0.85	0.12	1.00	0.75	0.10	1.00	0.80	0.11	1.00
μ_8	0.08	0.08	0.85	0.08	0.09	0.84	0.07	0.07	0.85	0.08	0.08	0.84
μ_9	0.69	0.10	1.00	0.73	0.11	1.00	0.66	0.09	1.00	0.70	0.10	1.00
SD of ϵ_{i1}	1.68	0.16		1.82	0.17		1.48	0.14		1.63	0.15	
SD of ϵ_{i2}	1.25	0.12		1.38	0.12		1.11	0.10		1.24	0.11	
SD of ϵ_{i3}	1.21	0.12		1.34	0.13		1.06	0.11		1.19	0.11	
SD of ϵ_{i4}	1.53	0.15		1.67	0.16		1.34	0.14		1.48	0.14	
SD of ϵ_{i5}	1.47	0.14		1.61	0.15		1.29	0.12		1.43	0.13	
SD of ϵ_{i6}	2.03	0.19		2.19	0.20		1.79	0.17		1.95	0.18	
SD of ϵ_{i7}	0.97	0.10		1.11	0.10		0.85	0.08		0.99	0.09	
SD of ϵ_{i8}	0.61	0.06		0.74	0.06		0.55	0.05		0.68	0.06	
SD of ϵ_{i9}	0.82	0.08		0.96	0.09		0.73	0.07		0.87	0.08	

Table 7.2: Sensitivity analyses for the specification of the Wishart prior for the random effects precision matrix $\mathbf{\Sigma}^{-1}$. The table provides posterior means and standard deviations of the elements of the correlation matrix of the random effect $\boldsymbol{\epsilon}_i$ for the model without patient covariates. The posterior probability that the parameters are greater than zero is also provided for each parameter.

	Wishart(9,(1/9)I9)			Wishart(9,(1/18)I9)			Wishart(18,(1/9)I9)			Wishart(18,(1/18)I9)		
Parameter	Mean	SD	Pr(¿0)	Mean	SD	Pr(¿0)	Mean	SD	Pr(¿0)	Mean	SD	Pr(¿0)
ρ_{12}	0.15	0.11	0.91	0.15	0.11	0.92	0.15	0.11	0.91	0.14	0.14	0.91
ρ_{13}	0.28	0.11	0.99	0.27	0.11	0.99	0.28	0.11	0.99	0.27	0.11	0.99
ρ_{14}	0.26	0.11	0.99	0.25	0.11	0.99	0.25	0.11	0.98	0.24	0.11	0.98
ρ_{15}	0.16	0.11	0.92	0.15	0.11	0.92	0.15	0.11	0.91	0.15	0.15	0.91
ρ_{16}	0.18	0.11	0.95	0.19	0.11	0.96	0.18	0.11	0.95	0.17	0.17	0.95
ρ_{17}	-0.22	0.12	0.03	-0.19	0.11	0.05	-0.24	0.12	0.02	-0.20	0.11	0.04
ρ_{18}	0.004	0.13	0.51	0.01	0.12	0.52	0.01	0.13	0.51	0.01	0.01	0.51
ρ_{19}	-0.02	0.13	0.42	-0.01	0.12	0.45	-0.03	0.13	0.40	-0.02	0.12	0.43
ρ_{23}	0.29	0.11	0.99	0.27	0.11	0.99	0.30	0.11	0.99	0.27	0.11	0.99
ρ_{24}	0.36	0.11	1.00	0.34	0.11	1.00	0.37	0.11	1.00	0.34	0.10	1.00
ρ_{25}	0.02	0.12	0.56	0.02	0.11	0.56	0.01	0.12	0.52	0.01	0.01	0.54
ρ_{26}	-0.04	0.12	0.37	-0.03	0.11	0.39	-0.05	0.12	0.32	-0.04	0.11	0.36
ρ_{27}	-0.09	0.12	0.24	-0.06	0.12	0.30	-0.11	0.12	0.20	-0.08	0.11	0.25
ρ_{28}	0.18	0.13	0.91	0.15	0.12	0.89	0.19	0.13	0.93	0.15	0.15	0.90
ρ_{29}	-0.09	0.13	0.24	-0.07	0.12	0.27	-0.09	0.13	0.23	-0.07	0.12	0.26
ρ_{34}	0.84	0.04	1.00	0.79	0.05	1.00	0.82	0.04	1.00	0.78	0.05	1.00
ρ_{35}	0.26	0.11	0.99	0.25	0.11	0.98	0.27	0.12	0.99	0.25	0.11	0.98
ρ_{36}	0.24	0.24	0.98	0.22	0.22	0.97	0.24	0.24	0.98	0.23	0.23	0.97
ρ_{37}	-0.22	0.12	0.04	-0.19	0.11	0.05	-0.24	0.12	0.03	-0.20	0.11	0.04
ρ_{38}	0.12	0.13	0.82	0.09	0.12	0.78	0.14	0.13	0.85	0.10	0.10	0.81
ρ_{39}	-0.18	0.12	0.08	-0.15	0.12	0.10	-0.19	0.12	0.06	-0.15	0.12	0.10
ρ_{45}	0.29	0.11	0.99	0.27	0.11	0.99	0.29	0.11	0.99	0.27	0.11	0.99
ρ_{46}	0.26	0.11	0.99	0.25	0.11	0.99	0.26	0.11	0.99	0.25	0.11	0.98
ρ_{47}	-0.24	0.12	0.03	-0.21	0.12	0.04	-0.27	0.12	0.02	-0.22	0.11	0.03
ρ_{48}	0.15	0.13	0.86	0.12	0.12	0.83	0.16	0.13	0.89	0.13	0.13	0.85
ρ_{49}	-0.15	0.13	0.11	-0.12	0.12	0.15	-0.16	0.13	0.10	-0.13	0.12	0.14
ρ_{56}	0.88	0.03	1.00	0.85	0.03	1.00	0.87	0.03	1.00	0.84	0.04	1.00
ρ_{57}	-0.28	0.11	0.01	-0.24	0.11	0.02	-0.29	0.11	0.01	-0.25	0.11	0.01
ρ_{58}	0.05	0.13	0.66	0.04	0.12	0.64	0.05	0.13	0.66	0.04	0.04	0.64
ρ_{59}	-0.15	0.12	0.12	-0.11	0.12	0.16	-0.16	0.12	0.11	-0.12	0.12	0.14
ρ_{67}	-0.24	0.12	0.02	-0.21	0.11	0.03	-0.26	0.12	0.02	-0.22	0.11	0.02
ρ_{68}	0.03	0.13	0.58	0.02	0.12	0.57	0.02	0.13	0.58	0.02	0.02	0.58
ρ_{69}	-0.21	0.12	0.04	-0.17	0.11	0.07	-0.22	0.12	0.04	-0.18	0.11	0.06
ρ_{78}	-0.11	0.13	0.21	-0.09	0.12	0.24	-0.11	0.13	0.20	-0.09	0.12	0.22
ρ_{79}	-0.06	0.13	0.31	-0.04	0.12	0.36	-0.06	0.13	0.30	-0.04	0.12	0.35
ρ_{89}	0.07	0.13	0.71	0.07	0.12	0.72	0.06	0.13	0.69	0.06	0.06	0.70

CHAPTER 8

Comparing models of patient preference: the PROSPECT study

This chapter describes estimation of patient preferences using an adaptive best-worst conjoint method and ordinary least squares regression. We present and compare estimated relative attribute importance scores for patients obtained using adaptive best-worst conjoint and the Bayesian hierachical model of Chapter 6.

8.1 Dahan's Adaptive Best-worst Conjoint method: Using Ordinary Least Squares Regression

Dahan's Adaptive Best-worst Conjoint method is a method for discrete choice experiments developed by Ely Dahan that elicits discrete choice data using the experimental design described in Chapter 3 and uses ordinary least squares regression to estimate patient preferences. The development of the method was motivated by the need for preferences to be quickly estimated immediately after a patient completed the DCE module and then used during a discussion regarding a patient's treatment plan with the patient's physician.

In the PROSPECT study, the DCE module used to collect best-worst choices monitors which pairs of health states have been *resolved*. Resolved health states are pairs of health states that have been ranked relative to each other or for which a ranking can be inferred. For example, if $A > B$ and $B > C$, then it is inferred that $A > C$. Sixteen health states were used in the PROSPECT study. In Dahan's method, an ordinary least squares regression was perfomed for each patient using only that patient's data. The outcome for each individual-

level regression is a vector of length 16 in which each component c corresponds to one of the 16 health states and indicates the number of the remaining health states that ranked lower than c.

For a specific patient i, let $\boldsymbol{y_i} = (y_{i1}, ..., y_{i16})^{\mathsf{T}}$ be a 16×1 vector where each component y_{ic} represents the number of the remaining health states ranked lower than health state c for individual i. Let $\boldsymbol{X}$ denote the 16×9 covariate matrix of attribute information for the 16 health states where each row $\boldsymbol{x_c^{\mathsf{T}}}$ corresponds to the covariate vector of attribute information for health state c. The regression model is

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i16} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \dots & x_{1,9} \\ 1 & x_{2,1} & x_{2,2} \dots & x_{2,9} \\ \vdots & \vdots & & \vdots \\ 1 & x_{16,1} & x_{16,2} \dots & x_{16,9} \end{bmatrix} \times \begin{bmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{i9} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i16} \end{bmatrix}.$$

Using matrix notation, this is represented by

$$\underset{16\times1}{\boldsymbol{y_i}} = \underset{16\times9}{\boldsymbol{X}} \times \underset{9\times1}{\boldsymbol{\beta}_i} + \underset{16\times1}{\boldsymbol{\epsilon}_i}\ . \tag{8.1}$$

To estimate relative attribute importance for each patient, Dahan's used ordinary least squares regression to estimate the elements of the parameter vector $\boldsymbol{\beta}_i$ of unknown patient preference scores. We use the abbreviation LinEST to refer to this method and we use HB to refer to our Bayesian method and we compare the relative importance scores estimated using the LinEST method with the relative importance scores estimated using our method. Point estimates of relative importance for the both methods are calculated using the definition in Chapter 6.

8.2 Comparing Methods: Relative Importance Scores

Figure 8.1 presents the posterior mean estimates of relative attribute importance ± 1 SD for the Bayesian hiearchical model without patient covariates and the point estimates of relative attribute importance using the LinEST estimates for 10 randomly selected men. For

many men, it appears that the methods provide similar results; using the posterior mean relative importance scores provides a ranking of attributes for each patient similar to that obtained using the LinEST point estimates of relative importance. Using the LinEST point estimates for patient 3, urinary functioning followed by sexual functioning appear to be most important and second most important attributes affecting health state preference followed by cutting, bowel issues, lifespan, others' support and, lastly, taking action. Using the HB point estimates for patient 3, we obtain a similar ranking where bowel issues appears to be ranked more favorably over cutting.

Although the point estimates give a ranking of attributes by importance, the Bayesian results suggest that some attributes maybe similarly ranked. From Figure 8.1, we see that for patient 3 the error bars for lifespan, bowel issues, cutting, active role and others support appear to overlap indicating that the posterior mean estimates of relative importance for these attributes may not be significantly different, and thus maybe similarly ranked. Moreover, the methods can provide wildy different results for some patients. Patient 121 is an example of a patient who did not have similar rankings according to the HB and LinEST methods. For patient 121, the HB method ranked sexual functioning as the most important attribute, while the LinEST method ranked sexual functioning as the least important.

Table 8.1 summarizes the LinEST point estimates, the HB point estimates, and the differences between the LinEST and HB point estimates across all patients. While the mean differences are close to zero, the minimum and maximum differences show that the methods can disagree by over 0.3 points and are frequently over 0.1 points discrepant.

8.3 Mean Squared Difference

We use the root mean squared difference (RMSD) to measure the degree of dissimilarity between our HB model estimates of relative importance and the LinEST estimates. The RMSD measures the average absolute difference in relative attribute importance between the LinEST and HB estimates. Lower values of RMSD indicate less discrepancy and the RMSD has the same units as the estimates of relative importance (%).

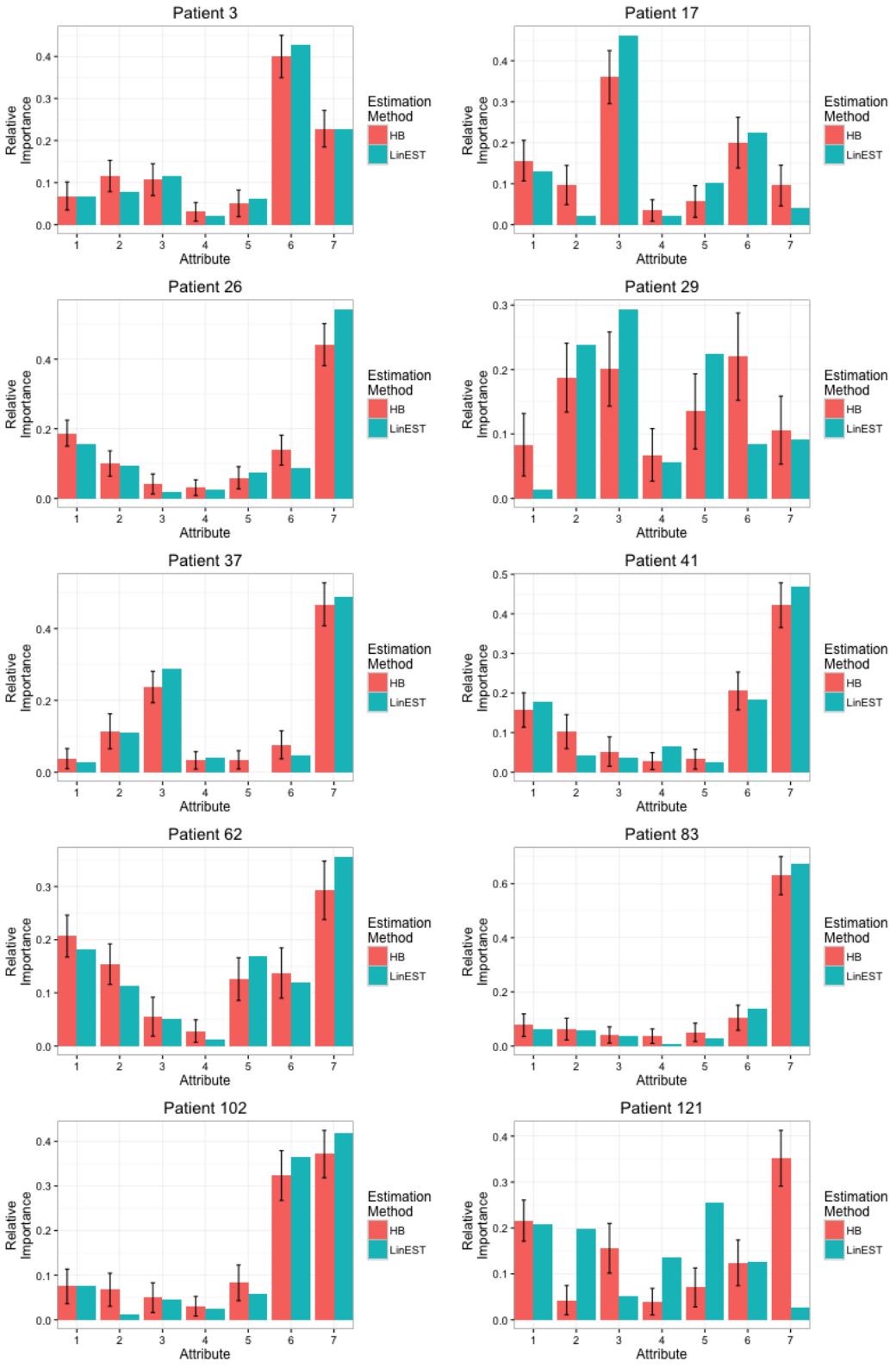

Figure 8.1: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

Table 8.1: Descriptive statistics of the LinEST and HB point estimates by health state attribute. The LinEST point estimates are individual-level ordinary least squares estimate and the HB point estimates are the individual-level posterior means.

	LinEST			HB			Difference (LinEST-HB)		
Attribute	Mean	SD	(Min, Max)	Mean	SD	(Min, Max)	Mean	SD	(Min, Max)
Lifespan	0.207	0.126	(0.126,0.577)	0.202	0.103	(0.038,0.434)	0.006	0.048	(-0.132,0.168)
Bowel Issues	0.135	0.099	(0.099,0.500)	0.143	0.079	(0.034,0.389)	-0.008	0.050	(-0.144,0.164)
Cutting	0.090	0.101	(0.101,0.462)	0.095	0.075	(0.027,0.362)	-0.004	0.050	(-0.138,0.253)
Action	0.056	0.046	(0.046,0.249)	0.043	0.020	(0.020,0.139)	0.012	0.038	(-0.050,0.124)
Others' Support	0.084	0.091	(0.091,0.482)	0.083	0.064	(0.026,0.389)	0.002	0.048	(-0.091,0.184)
Urinary Functioning	0.198	0.105	(0.105,0.597)	0.189	0.075	(0.065,0.436)	0.008	0.058	(-0.136,0.192)
Sexual Functioning	0.229	0.142	(0.142,0.674)	0.245	0.121	(0.066,0.629)	-0.016	0.063	(-0.326,0.137)

We define the root mean squared difference as

$$\text{RMSD} = \Big[\frac{1}{N}\sum_{i=1}^{N}(R_{ia}^{HB} - R_{ia}^{LinEST})^2\Big]^{1/2} \tag{8.2}$$

where N is the number of patients in the sample and R_{ia}^{HB} is the relative attribute importance of attribute a for individual i for the HB method and R_{ia}^{LinEST} is the relative attribute importance of attribute a for individual i for the LinEST method. The RMSD was calculated for each of the seven health state attributes.

Table 8.2: Root mean square difference in relative importance scores between the HB method and the LinEST method by attribute

Attribute	RMSD
Lifespan	0.0477
Bowel Issues	0.0508
Cutting	0.0497
Taking Action	0.0397
Others' Support	0.0480
Urinary Functioning	0.0580
Sexual Functioning	0.0650

Table 8.2 presents the RMSDs for each attribute. The RMSD ranged from a minimun of 0.0397 for taking action to 0.0650 for sexual functioning.

8.4 Remarks

Dissimilarity in estimated relative importance scores was greatest for lifespan, bowel issues, urinary functioning and sexual functioning, which are attributes determined to be most important in the population under the HB model. One potential problem in fitting a multiple linear regression model to observations which are essentially vectors of rankings is that for pairwise samples of $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{N}$ are not independent, thus violating the assumption that the errors $\boldsymbol{\epsilon}_i$ are independent. Comparison of the LinEST and HB methods should be more formally investigated via a simulation study. This is a direction for future research. The code provided in Appendix B can be used to simulate the observed data from our specific

discrete choice experiment.

Appendix A

JAGS Code

A.1 Bayesian Hierarchical Model for Best Choices With No Patient Covariates

```
# Purpose: To fit the Bayesian hierarchical model for best choices with no
#          patient covariates
# Author: Anna Liza Malazarte Antonio
# Developed using JAGS version 4.2.0

# Assumptions:
#          1) Each choice set contains 4 health states
#          2) Nine health state attribute variables

# User input:
#          N (total number of choice sets)
#          nsubj (total number of patients
#          mean.mu.b (mean of the hyperprior distribution for the population
#               mean vector of preference scores mu.b)
#          prec.mu.b (precision matrix of the hyperprior distribution for the
#               population mean vector of preference scores mu.b)
#          df (degrees of freedom parameter of the hyperprior distribution for
#               the precision matrix of the prior distribution of the vector of
#               patient preference scores prec.b)
#          Omega (scale matrix of the hyperprior distribution for the precision
#               matrix of the prior distribution of the vector of patient
#               preference scores prec.b)

# Data description:
#          X (an N by 36 matrix where each row corresponds to a unique
#               patient-choice set combination and contains the attribute
#               information for each of the four health states in that specific
#               choice set)
#          X[,1:9] (attribute information for the best health state)
#          X[,10:18] (attribute information for one of the mid-ranked health
```

```
#               states)
#          X[,19:27] (attribute information for the remaining mid-ranked
#               health state)
#          X[,28:36] (attribute information for the worst health state)

data {
       for (j in 1:N) {
       ones[j] <- 1
       }
}

model {
       for (i in 1:N){                      # Loop over choice sets
       for (j in 1:4){                      # Linear predictor
       mu[i,j]<-inprod(beta[id[i],1:9],X[i,(9*(j-1)+1):(9*j)])
       expmu[i,j]<-exp(mu[i,j])
       }

       L[i] <- expmu[i,1]/sum(expmu[i,])    # Best choice probability

       # Ones trick
       phi[i] <-  L[i]
       ones[i] ~ dbern(phi[i])
       }

       # Prior for the random effects
       for(n in 1:nsubj){
       beta[n,1:9] ~ dmnorm(mu.b,prec.b)
       }

        # Hyperpriors
        mu.b[1:9] ~ dmnorm(mean.mu.b, prec.mu.b)
        prec.b[1:9,1:9] ~ dwish(Omega,df)

        # Convert precision to covariance matrix
        sigma.b[1:9,1:9] <- inverse(prec.b[,])
            # Standard deviations
            for(k in 1:9){
            sd.b[k] <- sqrt(sigma.b[k,k])
            }
            # Correlations
            for(p in 1:9){
            for(q in 1:9){
            corr.b[p,q] <- sigma.b[p,q]/sqrt(sigma.b[p,p]*sigma.b[q,q])
            }
```

```
                }
}
```

A.2 Bayesian Hierarchical Model for Best-Worst Choices With No Patient Covariates

```
# Purpose: To fit the Bayesian hierarchical model for best-worst choices with no
#          patient covariates
# Author: Anna Liza Malazarte Antonio
# Developed using JAGS version 4.2.0

# Assumptions:
#          1) Each choice set contains 4 health states
#          2) Nine health state attribute variables

# User input:
#          N (total number of choice sets)
#          nsubj (total number of patients
#          mean.mu.b (mean of the hyperprior distribution for the population
#               mean vector of preference scores mu.b)
#          prec.mu.b (precision matrix of the hyperprior distribution for the
#               population mean vector of preference scores mu.b)
#          df (degrees of freedom parameter of the hyperprior distribution for
#               the precision matrix of the prior distribution of the vector of
#               patient preference scores prec.b)
#          Omega (scale matrix of the hyperprior distribution for the precision
#               matrix of the prior distribution of the vector of patient
#               preference scores prec.b)

# Data description:
#          X (an N by 36 matrix where each row corresponds to a unique
#               patient-choice set combination and contains the attribute
#               information for each of the four health states in that specific
#               choice set)
#          X[,1:9] (attribute information for the best health state)
#          X[,10:18] (attribute information for one of the mid-ranked health
#               states)
#          X[,19:27] (attribute information for the remaining mid-ranked
#               health state)
#          X[,28:36] (attribute information for the worst health state)

data {
       for (a in 1:N) {
       ones[a] <- 1
       }
}
```

```
model {
       for (i in 1:N){                        # Loop over choice sets
       for (j in 1:4){                        # Linear predictor
       mu[i,j]<-inprod(beta[id[i],1:9],X[i,(9*(j-1)+1):(9*j)])
       expmu[i,j]<-exp(mu[i,j])
       }

       p[i,1]<-expmu[i,1]/sum(expmu[i,])          # Best choice
       p[i,2]<-expmu[i,2]/sum(expmu[i,2:4])       # 1st factor in first summand
       p[i,3]<-expmu[i,3]/sum(expmu[i,3:4])       # 2nd factor in first summand
       p[i,4]<-expmu[i,3]/sum(expmu[i,2:4])       # 1st factor in second summand
       summand[i,1]<-expmu[i,2]
       summand[i,2]<-expmu[i,4]
       p[i,5]<-expmu[i,2]/sum(summand[i,])        # 2nd factor in second summand

       # Best-worst choice probability
       L[i] <- p[i,1]*p[i,2]*p[i,3] + p[i,1]*p[i,4]*p[i,5]

       # Ones trick
       phi[i] <-  L[i]
       ones[i] ~ dbern(phi[i])
       }

       # Prior for the random effects
       for(n in 1:nsubj){
       beta[n,1:9] ~ dmnorm(mu.b,prec.b)
       }

       # Hyperpriors
       mu.b[1:9] ~ dmnorm(mean.mu.b, prec.mu.b)
       prec.b[1:9,1:9] ~ dwish(Omega,df)

       # Convert precision to covariance matrix
       sigma.b[1:9,1:9] <- inverse(prec.b[,])
            # Standard deviations
            for(k in 1:9){
            sd.b[k] <- sqrt(sigma.b[k,k])
            }
            # Correlations
            for(p in 1:9){
            for(q in 1:9){
            corr.b[p,q] <- sigma.b[p,q]/sqrt(sigma.b[p,p]*sigma.b[q,q])
            }
            }
}
```

A.3 Bayesian Hierarchical Model for Best-Worst Choices With Patient Covariates

```
# Purpose: 1) To fit the Bayesian hierarchical model for best-worst choices
#          with patient covariates
#          2) To calculate the inverse CPO-SET for each unique patient-choice
#          set combination, and the inverse CPO-MVP and inverse CPO-MVP for
#          each patient

# Author: Anna Liza Malazarte Antonio
# Developed using JAGS version 4.2.0

# Assumptions:
#          1) Each choice set contains 4 health states
#          2) Nine health state attribute variables

# User input:
#          N (total number of choice sets)
#          nsubj (total number of patients
#          ncov (total number of patient covariate variables)
#          pie (number pi)
#          mu.b (mean of the hyperprior distribution for the residual effect
#               epsilon[n,1:9] of the vector of preference scores)
#          df (degrees of freedom parameter of the hyperprior distribution for
#               the precision matrix of the prior distribution of the vector of
#               patient preference scores prec.b)
#          Omega (scale matrix of the hyperprior distribution for the precision
#               matrix of the prior distribution of the vector of patient
#               preference scores prec.b)

# Data description:
#          X (an N by 36 matrix where each row corresponds to a unique
#               patient-choice set combination and contains the attribute
#               information for each of the four health states in that specific
#               choice set)
#          X[,1:9] (attribute information for the best health state)
#          X[,10:18] (attribute information for one of the mid-ranked health
#               states)
#          X[,19:27] (attribute information for the remaining mid-ranked
#               health state)
#          X[,28:36] (attribute information for the worst health state)

#          zeta (an nsubj by ncov matrix where each row corresponds to a unique
```

```
#               patient and contains the patient's covariate information)

data {
       for (a in 1:N) {
       ones[a] <- 1
        }
}

model {
       for (i in 1:N){                              # Loop over choice sets
       for (j in 1:4){                              # Linear predictor
       mu[i,j]<-inprod(beta[id[i],1:9],X[i,(9*(j-1)+1):(9*j)])
       expmu[i,j]<-exp(mu[i,j])
       }

       p[i,1]<-expmu[i,1]/sum(expmu[i,])          # Best choice
       p[i,2]<-expmu[i,2]/sum(expmu[i,2:4])       # 1st factor in first summand
       p[i,3]<-expmu[i,3]/sum(expmu[i,3:4])       # 2nd factor in first summand
       p[i,4]<-expmu[i,3]/sum(expmu[i,2:4])       # 1st factor in second summand
       summand[i,1]<-expmu[i,2]
       summand[i,2]<-expmu[i,4]
       p[i,5]<-expmu[i,2]/sum(summand[i,])        # 2nd factor in second summand

       # Best-worst choice probability
       L[i] <- p[i,1]*p[i,2]*p[i,3] + p[i,1]*p[i,4]*p[i,5]

       # Ones trick
       phi[i] <-  L[i]
       ones[i] ~ dbern(phi[i])
       }

       # Calculate 1/CPO-SET to identify choice sets which are outlying
       #      with respect to patient prefereces
       for (b in 1:N){   # Loop over choice sets
       invcpo.set[b] <- 1/L[b]
       }

       for(n in 1:nsubj){
       # Prior for the random effects
       mu.beta[n,1:9] <- zeta[n,1:ncov]%*%Gamma[1:ncov,1:9]
       beta[n,1:9] <- mu.beta[n,1:9] + epsilon[n,1:9]
       epsilon[n,1:9] ~ dmnorm(mu.b,prec.b)

       # Calculate 1/CPO-MVP to identify patients with outlying preferences
       #      on all attributes
```

```
        ppo.beta[n] <-sqrt(exp(logdet(prec.b[1:9,1:9])))/sqrt(pow((2*pie),9))
        *exp(-0.5%*%t(beta[n,1:9]-mu.b[1:9])
        %*%prec.b[1:9,1:9]%*%(beta[n,1:9]-mu.b[1:9]))
        invcpo.beta[n] <- 1/ppo.beta[n]

        # Calculate 1/CPO-UVP to identify patients with outlying preferences
        #      on individual attributes
        for (r in 1:9){ # Loop over attributes
        ppo.beta.marg[n,r] <-dnorm(beta[n,r], mu.b[r], sigma.b[r,r])
        invcpo.beta.marg[n,r] <- 1/ppo.beta.marg[n,r]
        }

        }

        for(l in 1:ncov){  # Loop over columns of Gamma
        for (m in 1:9){  # Loop over rows of Gamma
        Gamma[l,m] ~ dnorm(0,1)
        }
        }

        # Hyperpriors
        prec.b[1:9,1:9] ~ dwish(Omega,df)

        # Convert precision to covariance matrix
        sigma.b[1:9,1:9] <- inverse(prec.b[,])
             # Standard deviations
             for(k in 1:9){
             sd.b[k] <- sqrt(sigma.b[k,k])
             }
             # Correlations
             for(p in 1:9){
             for(q in 1:9){
             corr.b[p,q] <- sigma.b[p,q]/sqrt(sigma.b[p,p]*sigma.b[q,q])
             }
             }
}
```

Appendix B

R Code

B.1 Code to Generate Discrete Choice Data

B.1.1 (001) Simulate Data - Functions.R

```
# Author: Anna Liza Malazarte Antonio
# Developed using R version 3.3.1
# Contingencies: None

# Program Description ---------------------------------------------------------
# The purpose of this code is to:
# Define functions to be used in the script "(003) Simulate Data.R"

# prob.best -------------------------------------------------------------------

# Purpose: To calculate the probabilities of best choice in a choice set
# Arguments:
# M = covariate matrix
# b = beta vector
# nalts = size of the choice set
# nsets = number of choice sets

# Output: Matrix of probabilities

prob.best<-function(M,b,nsets,nalts){
  # Numerators
  Xbeta=M%*%b
  expXbeta=matrix(exp(Xbeta),byrow=TRUE,ncol=nalts)

  # Denominators
  denom=as.matrix(rowSums(expXbeta))
  denominvmat.b=matrix(rep(denom^-1,each=nalts), byrow=T, nrow=nsets, ncol=nalts)

  # Probabilities: Best Choice
```

```
  # element-wise multiplication; by row, each element/rowsum
  bprob=expXbeta*denominvmat.b
}

# draw.y ------------------------------------------------------------------

# Purpose: Draw Y as a multinomial response by applying the random multinomial
#         function to each row of the probability matrix (Will identify best
#         in choice set)
# Arguments:
# M = covariate matrix
# p = probability matrix
# nalts = size of the choice set
# nsets = number of choice sets

draw.y <-function(M,p,nsets,nalts){
  y = matrix(data=NA,nrow=nsets,ncol=nalts)         # define a null matrix
  y = apply(p,1,function(M){rmultinom(1,1,M)})      # draw multinomial response
  y =t(y)                                           # transpose to row vectors
}


# id.best -----------------------------------------------------------------

# Purpose: To identify the cards (by label) chosen as best
# Arguments:
# S = matrix of cards (labels); each row represents a choice set
# y = observation matrix
# nalts = size of the choice set
# nsets = number of choice sets

id.best <-function(S,y,nsets,nalts){
  bchoice = matrix(data=NA, nrow=1, ncol=nsets)
  for (i in 1:nsets) {
    for (j in 1:nalts) {
      if (y[i,j]==1){bchoice[i]=S[i,j]}
    }
  }
  bchoice
}

# remove.choicesX ---------------------------------------------------------

# Purpose: To remove covariate data for past choices within a choice set
# Arguments:
```

```
# y = matrix of observations
# M = covariate matrix

remove.choicesX <-function(y,M){
  y.vec = as.vector(t(y))
  subset = subset(cbind(M,y.vec), y.vec==0,-c(y.vec)); subset
}

# remove.choicesS ------------------------------------------------------------

# Purpose: To remove cards(labels) for past choices within a choice set
# Arguments:
# y = matrix of observations
# S = card (labels) matrix

remove.choicesS <-function(y,S){
  y.vec = as.vector(t(y))
  cards.sets.sub = subset(cbind(as.vector(t(S)),y.vec), y.vec==0,-c(y.vec))
  cards.sets.sub = matrix(cards.sets.sub, nrow=nrow(S), ncol=ncol(S)-1, byrow=T)
}

# id.bestworst ---------------------------------------------------------------

# Purpose: To identify the cards (by label) chosen as best and worst
# Arguments:
# Yb = observed matrix of best choices
# S = card (labels) matrix by choice set
# w = vector of worst choice cards (labels)

id.bestworst <- function(Yb,S,w){
  Y.bw = Yb
  for (i in 1:nrow(S)) {
    for (j in 1:ncol(S)) {
      for(k in 1:ncol(w)){
        if (S[i,j]==w[1,k]){Y.bw[i,j] = -1}
      }
    }
  }
  Y.bw
}

# resolvedpairs.counts.setup -------------------------------------------------

# Purpose: To count the number of times of a card beats (count>0) another card
#          Winners along x-axis and losers along y-axis
```

```
#          Example: if card j beats card j n times,
#                   then count[i,j] = n, n an integer, n>0
# Arguments:
# ncards = number of unique cards presented to subjects

resolvedpairs.counts.setup <-function(ncards){
  counts = matrix(data=NA, nrow = ncards, ncol = ncards)
  for (i in 1:ncards){
    for (j in 1:ncards) {
      if (j>i | i>j){counts[i,j] = 0}

    }
  }
  counts
}

# resolvedpairs.setup -------------------------------------------------------

# Purpose: To identify which card pairs have been resolved
#          Winners along x-axis and losers along y-axis
#          Example: if the card pair (i,j) has been resolved , then ind[i,j] = 1
# Arguments:
# ncards = number of unique cards presented to subjects

resolvedpairs.setup <-function(ncards){
  ind = matrix(data=NA, nrow = ncards, ncol = ncards)
  for (i in 1:ncards){
    for (j in 1:ncards) {
      if (j>i){ind[i,j] = 0}

    }
  }
  ind
}

# resolvedpairs.counts.update -----------------------------------------------

# Purpose: To update counts which describe the number of times card pairs have
#          been resolved
#          Winners along x-axis and losers along y-axis
#          Example: if the card pair (i,j) has been resolved , then ind[i,j] = 1
#          Accounts for 5 comparisons for each choice set
# Note: Loop operates on each choice set
# Arguments:
# R = matrix of ranked cards (row is a choice set)
```

```
resolvedpairs.counts.update <-function(R){

for (i in 1:nrow(R)){
  print(R[i,])
  # work with best
  for (j in 2:4){
    resolvedpairs.counts[R[i,j],R[i,1]] = resolvedpairs.counts[R[i,j],R[i,1]] + 1
  }
  # work with worst
  for (k in 2:3){
    resolvedpairs.counts[R[i,4],R[i,k]] = resolvedpairs.counts[R[i,4],R[i,k]] + 1
  }
}
resolvedpairs.counts

}

# resolvedpairs.update ------------------------------------------------------

# Purpose: To update the indicator matrix which describes which card pairs have
#          been resolved
#          Winners along x-axis and losers along y-axis
#          Example: if the card pair (i,j) has been resolved , then ind[i,j] = 1
#          Accounts for 5 comparisons for each choice set
# Arguments:
# Mc = matrix which describes the number of times card pairs have been resolved
# Mi = matrix which describes which card pairs have been resolved

resolvedpairs.update <-function(Mc,Mi){
  # Keep track of what has been compared
  for (i in 1:16){
    for (j in 1:16){
      # in lower triange of paircounts
      if (Mc[i,j] > 0 & i>j){Mi[j,i]=1}
      # in upper triange of paircounts
      if (Mc[i,j] > 0 & j>i){Mi[i,j]=1}
    }
  }
  Mi
}

# listunresolvedpairs -------------------------------------------------------

# Purpose: To output a list of unresolved pairs
```

```
# Arguments:
# M = indicator matrix describing which card pairs (described by row and col
#     indices) have been resolved

listunresolvedpairs <- function(M){
  n = 16
  index = which(M==0)       # gives indices of matrix (by column) where == 0
  columnlist = ceiling(index/n)
  rowlist = index%%n; rowlist                  # vector index mod 16
  pairs = as.list(data.frame(t(cbind(rowlist, columnlist)))); pairs
}

# newchoiceset ------------------------------------------------------------

# Purpose: To form a new choice set using a list of unresolved pairs
# Arguments:
# l = list of unresolved pairs

newchoiceset <- function(l){

  # Permute list of unresolved pairs
  perm = sample(l)

  # Select first two pairs to create a new choice set

  # If there is a duplicate card in the set, resample a new pair
  #   duplicated(v) outputs a indicator vector for input vector v,
  #   where a component = 1 identifies a duplicate
  # Example:
  # if pairs 1 and 2 have duplicates then pick pairs 1 and 3
  # if pairs 1 and 3 have duplicates then pick pairs 1 and 4...
  # repeat until a set of four unique cards are found

  for (i in 2:length(perm)){
    cset = c(rbind(as.vector(unlist(perm[1])),as.vector(unlist(perm[i]))))
    if(sum(duplicated(cset))==0){break}
  }
cset
}

# transitivity ------------------------------------------------------------
# Purpose: To account for card 'wins' derived from observed 'wins' using
#     the property of transitivity, e.g., if 1>2 two times and 2>3 one time
#     then 1>3 three times
# Arguments:
```

```
# R = 16x16 matrix of resolved pair counts

transitivity <- function(R){

  # Remove "NAs" from count matrix diagonal for arithmetic
  for (j in 1:16){
    R[j,j] = 0
  }

  # Update resolved pair counts
  for (i in 1:16){
    for (j in 1:16){
      for (k in 1:16){

        if(i!=k & R[i,j] != 0 & R[j,k] != 0){
          R[i,k] = R[i,k] + R[i,j] + R[j,k]
        }

      }
    }
  }

  # Replace "NAs" in count matrix
  for (m in 1:16){
    R[m,m] = NA
  }

  return(R)
}


# create data ------------------------------------------------------------

# Purpose: To generate data which replicates the adaptive choice elicitation
#          process created by Ely Dahan
#          Function creates data using the random effects for each patient

# Arguments:
# betavec = random effects vector for one patient

# Output:
# Let Nchoicesets be total number choice sets presented to a single patient
#         in a DCE
# Y = NchoicesetsX4 matrix indicating best and worst choices; each row
#       corresponds to a choice set; within a row, 1 = best choice in choice
```

```
#       set, -1 = worst choice in choice set and 0 = mid-ranked choices in
#       choice set
# cardsets = NchoicesetsX4 matrix which describes the cards presented in
#      a choice set by label; each row corresponds to a choice set
# X = (NchoicesetsX4)X9 matrix that describes the attribute information for
#      each card in the DCE; each row corresponds to a card; row order is
#      determined by the choice set and the order in which it is presented in
#      the choice set, e.g., the first row corresponds to the first card
#      displayed in the first choice set
# beta = random effects vector for one patient
# ranked = NchoicesetsX4 matrix which describes the cards ranked in a choice
#      set by label; each row corresponds to a choice set
# Ycounts = 1x16 vector where Ycounts[i] is the number of times card i 'beat'
#      the other 15 cards by majority vote

createdata <- function(betavec) {

  # Count all the paired comparisons (1v2,...1v16,2v3,...,2v16,...,15v16)
  # Winners along x-axis; Losers along y-axis
  resolvedpairs.counts = resolvedpairs.counts.setup(16)

  # Keep track of resolved pairs
  resolvedpairs <- resolvedpairs.setup(16)

  # ------------------------------------------------------------------------
  # Choices sets 1-4 : Present all 16 cards ---------------------------------
  # ------------------------------------------------------------------------

  # Subset the data ---------------------------------------------------------

  cards.sets1_4.1 = matrix(data.matrix(attribdata[,3], rownames.force = NA)
                           ,byrow=TRUE,ncol=4)
  X.1_4.1 = data.matrix(attribdata[,4:12], rownames.force = NA)
  beta = data.matrix(betavec, rownames.force = NA)
  # beta dataset above loads with an extra column ("X" = obs num)
  # Code below excludes the first variable/column
  beta = beta[c(-1)]

  # Best Choices ------------------------------------------------------------

  # Construct Linear Predictor and Probabilities
  prob.1_4.1 = prob.best(X.1_4.1,beta,4,4)

  # Draw Y as a Multinomial Response
  # Apply random multinomial function to each row of prob matrix
```

```
Y.1_4.1 = draw.y(X.1_4.1,prob.1_4.1,4,4)

# Identify Best Choices
bchoice.1_4.1 = id.best(cards.sets1_4.1,Y.1_4.1,4,4)

# Remove best choices (Y.b_vec==0 for other than best choices)
# Convert obs matrix to a vector by row (use t())
X.1_4.2 = remove.choicesX(Y.1_4.1,X.1_4.1)
cards.sets1_4.2 = remove.choicesS(Y.1_4.1,cards.sets1_4.1)

# 2nd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.1_4.2 = prob.best(X.1_4.2,beta,4,3)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.1_4.2 = draw.y(X.1_4.2,prob.1_4.2,4,3)

# Identify Best Choices
bchoice.1_4.2 = id.best(cards.sets1_4.2,Y.1_4.2,4,3)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.1_4.3 = remove.choicesX(Y.1_4.2,X.1_4.2)
cards.sets1_4.3 = remove.choicesS(Y.1_4.2,cards.sets1_4.2)

# 3rd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.1_4.3 = prob.best(X.1_4.3,beta,4,2)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.1_4.3 = draw.y(X.1_4.3,prob.1_4.3,4,2)

# Identify Best Choices
bchoice.1_4.3 = id.best(cards.sets1_4.3,Y.1_4.3,4,2)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.1_4.4 = remove.choicesX(Y.1_4.3,X.1_4.3)
cards.sets1_4.4 = remove.choicesS(Y.1_4.3,cards.sets1_4.3)

# Worst Choices ---------------------------------------------------------
```

```
bchoice.1_4.4=t(cards.sets1_4.4)

# RESULTS ------------------------------------------------------------------

# Display cards in order presented by choice set ---------------------------
cards.sets1_4.1

# Identify cards in order presented by choice set --------------------------
Y.bw.1_4 = id.bestworst(Y.1_4.1,cards.sets1_4.1,bchoice.1_4.4)

# Display cards in ranked order by choice set ------------------------------
cards.ranked.1_4 = t(rbind(bchoice.1_4.1,bchoice.1_4.2,bchoice.1_4.3
                   ,bchoice.1_4.4))

# ACCOUNTING: Resolved Pairs -----------------------------------------------

# Update resolved pair counts
resolvedpairs.counts <- resolvedpairs.counts.update(cards.ranked.1_4)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

# Count number of unresolved pairs
# which(): gives indices of matrix (by column) where == 0
unresolvedpairs_num = length(which(resolvedpairs==0))

# ------------------------------------------------------------------------------
# Choices set 5 : Compare "best" choices from choice sets 1-4 ----------------
# ------------------------------------------------------------------------------

# Setup --------------------------------------------------------------------

# "Best" choices from choice sets 1-4
cards.sets5.1 = bchoice.1_4.1

# Attributes of best cards
X.5.1 = data.matrix(subset(data.frame(attribdata[,3:12])
      , (Card == bchoice.1_4.1[1,1])|(Card == bchoice.1_4.1[1,2])
      |(Card == bchoice.1_4.1[1,3])|(Card == bchoice.1_4.1[1,4])
      ,-c(Card)),rownames.force = NA)

# Best Choices -------------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.5.1 = prob.best(X.5.1,beta,1,4)
```

```
# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.5.1 = draw.y(X.5.1,prob.5.1,1,4)

# Identify Best Choices
bchoice.5.1 = id.best(cards.sets5.1,Y.5.1,1,4)

# Remove best choices (Y.b_vec==0 for other than best choices)
# Convert obs matrix to a vector by row (use t())
X.5.2 = remove.choicesX(Y.5.1,X.5.1)
cards.sets5.2 = remove.choicesS(Y.5.1,cards.sets5.1)

# 2nd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.5.2 = prob.best(X.5.2,beta,1,3)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.5.2 = draw.y(X.5.2,prob.5.2,1,3)

# Identify Best Choices
bchoice.5.2 = id.best(cards.sets5.2,Y.5.2,1,3)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.5.3 = remove.choicesX(Y.5.2,X.5.2)
cards.sets5.3 = remove.choicesS(Y.5.2,cards.sets5.2)

# 3rd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.5.3 = prob.best(X.5.3,beta,1,2)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.5.3 = draw.y(X.5.3,prob.5.3,1,2)

# Identify Best Choices
bchoice.5.3 = id.best(cards.sets5.3,Y.5.3,1,2)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.5.4 = remove.choicesX(Y.5.3,X.5.3)
```

```
cards.sets5.4 = remove.choicesS(Y.5.3,cards.sets5.3)

# Worst Choices ------------------------------------------------------------
bchoice.5.4=t(cards.sets5.4)

# RESULTS ------------------------------------------------------------------

# Display cards in order presented by choice set ---------------------------
cards.sets5.1

# Identify cards in order presented by choice set --------------------------
Y.bw.5 = id.bestworst(Y.5.1,cards.sets5.1,bchoice.5.4)

# Display cards in ranked order by choice set ------------------------------
cards.ranked.5 = t(rbind(bchoice.5.1,bchoice.5.2,bchoice.5.3,bchoice.5.4))


# ACCOUNTING: Resolved Pairs -----------------------------------------------

# Update resolved pair counts
resolvedpairs.counts <- resolvedpairs.counts.update(cards.ranked.5)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

# TRANSITIVITY -------------------------------------------------------------

# To account for card 'wins' derived from observed 'wins' using the property of
#     transitivity, e.g., if 1>2 two times and 2>3 one time then 1>3 three times
resolvedpairs.counts <- transitivity(resolvedpairs.counts)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

# Count number of unresolved pairs
# which(): gives indices of matrix (by column) where == 0
unresolvedpairs_num = length(which(resolvedpairs==0))


# ----------------------------------------------------------------------------
# Choices set 6 : Compare "worst" choices from choice sets 1-4 ---------------
# ----------------------------------------------------------------------------

# Setup --------------------------------------------------------------------
```

```
# "Worst" choices from choice sets 1-4
cards.sets6.1 = bchoice.1_4.4

# Attributes of best cards
X.6.1 = data.matrix(subset(data.frame(attribdata[,3:12])
      , (Card == bchoice.1_4.4[1,1])|(Card == bchoice.1_4.4[1,2])
      |(Card == bchoice.1_4.4[1,3])|(Card == bchoice.1_4.4[1,4])
      ,-c(Card)),rownames.force = NA)

# Best Choices ------------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.6.1 = prob.best(X.6.1,beta,1,4)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.6.1 = draw.y(X.6.1,prob.6.1,1,4)

# Identify Best Choices
bchoice.6.1 = id.best(cards.sets6.1,Y.6.1,1,4)

# Remove best choices (Y.b_vec==0 for other than best choices)
# Convert obs matrix to a vector by row (use t())
X.6.2 = remove.choicesX(Y.6.1,X.6.1)
cards.sets6.2 = remove.choicesS(Y.6.1,cards.sets6.1)

# 2nd Best Choices --------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.6.2 = prob.best(X.6.2,beta,1,3)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.6.2 = draw.y(X.6.2,prob.6.2,1,3)

# Identify Best Choices
bchoice.6.2 = id.best(cards.sets6.2,Y.6.2,1,3)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.6.3 = remove.choicesX(Y.6.2,X.6.2)
cards.sets6.3 = remove.choicesS(Y.6.2,cards.sets6.2)

# 3rd Best Choices --------------------------------------------------------
```

```
# Construct Linear Predictor and Probabilities
prob.6.3 = prob.best(X.6.3,beta,1,2)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.6.3 = draw.y(X.6.3,prob.6.3,1,2)

# Identify Best Choices
bchoice.6.3 = id.best(cards.sets6.3,Y.6.3,1,2)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.6.4 = remove.choicesX(Y.6.3,X.6.3)
cards.sets6.4 = remove.choicesS(Y.6.3,cards.sets6.3)

# Worst Choices ------------------------------------------------------------
bchoice.6.4=t(cards.sets6.4)


# RESULTS ------------------------------------------------------------------

# Display cards in order presented by choice set ---------------------------
cards.sets6.1

# Identify cards in order presented by choice set --------------------------
Y.bw.6 = id.bestworst(Y.6.1,cards.sets6.1,bchoice.6.4)

# Display cards in ranked order by choice set ------------------------------
cards.ranked.6 = t(rbind(bchoice.6.1,bchoice.6.2,bchoice.6.3,bchoice.6.4))


# ACCOUNTING: Resolved Pairs -----------------------------------------------

# Update resolved pair counts
resolvedpairs.counts <- resolvedpairs.counts.update(cards.ranked.6)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)


# TRANSITIVITY -------------------------------------------------------------

# To account for card 'wins' derived from observed 'wins' using the property of
#     transitivity, e.g., if 1>2 two times and 2>3 one time then 1>3 three times
resolvedpairs.counts <- transitivity(resolvedpairs.counts)
```

```
# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

# Count number of unresolved pairs
# which(): gives indices of matrix (by column) where == 0
unresolvedpairs_num = length(which(resolvedpairs==0))


# ------------------------------------------------------------------------
# Choices set 7 : Compare "middle" choices from choice sets 1 and 2 ----------
# ------------------------------------------------------------------------

# Setup ------------------------------------------------------------------

# "Middle" choices from choice sets 1 and 2
cards.sets7.1 = matrix(cards.ranked.1_4[1:2,2:3], nrow=1, ncol=4)

# Attributes of best cards
X.7.1 = data.matrix(subset(data.frame(attribdata[,3:12])
      , (Card == cards.sets7.1[1,1])|(Card == cards.sets7.1[1,2])
      |(Card == cards.sets7.1[1,3])|(Card == cards.sets7.1[1,4])
      ,-c(Card)),rownames.force = NA)

# Best Choices -----------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.7.1 = prob.best(X.7.1,beta,1,4)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.7.1 = draw.y(X.7.1,prob.7.1,1,4)

# Identify Best Choices
bchoice.7.1 = id.best(cards.sets7.1,Y.7.1,1,4)

# Remove best choices (Y.b_vec==0 for other than best choices)
# Convert obs matrix to a vector by row (use t())
X.7.2 = remove.choicesX(Y.7.1,X.7.1)
cards.sets7.2 = remove.choicesS(Y.7.1,cards.sets7.1)

# 2nd Best Choices -------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.7.2 = prob.best(X.7.2,beta,1,3)
```

```
# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.7.2 = draw.y(X.7.2,prob.7.2,1,3)

# Identify Best Choices
bchoice.7.2 = id.best(cards.sets7.2,Y.7.2,1,3)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.7.3 = remove.choicesX(Y.7.2,X.7.2)
cards.sets7.3 = remove.choicesS(Y.7.2,cards.sets7.2)

# 3rd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.7.3 = prob.best(X.7.3,beta,1,2)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.7.3 = draw.y(X.7.3,prob.7.3,1,2)

# Identify Best Choices
bchoice.7.3 = id.best(cards.sets7.3,Y.7.3,1,2)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.7.4 = remove.choicesX(Y.7.3,X.7.3)
cards.sets7.4 = remove.choicesS(Y.7.3,cards.sets7.3)

# Worst Choices ---------------------------------------------------------
bchoice.7.4=t(cards.sets7.4)


# RESULTS ---------------------------------------------------------------

# Display cards in order presented by choice set ----------------------------
cards.sets7.1

# Identify cards in order presented by choice set ---------------------------
Y.bw.7 = id.bestworst(Y.7.1,cards.sets7.1,bchoice.7.4)

# Display cards in ranked order by choice set -------------------------------
cards.ranked.7 = t(rbind(bchoice.7.1,bchoice.7.2,bchoice.7.3,bchoice.7.4))
```

```
# ACCOUNTING: Resolved Pairs ----------------------------------------------

# Update resolved pair counts
resolvedpairs.counts <- resolvedpairs.counts.update(cards.ranked.7)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)


# TRANSITIVITY ------------------------------------------------------------

# To account for card 'wins' derived from observed 'wins' using the property of
#     transitivity, e.g., if 1>2 two times and 2>3 one time then 1>3 three times
resolvedpairs.counts <- transitivity(resolvedpairs.counts)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

# Count number of unresolved pairs
# which(): gives indices of matrix (by column) where == 0
unresolvedpairs_num = length(which(resolvedpairs==0))


# ---------------------------------------------------------------------------
# Choices set 8 : Compare "middle" choices from choice sets 3 and 4 ----------
# ---------------------------------------------------------------------------

# Setup -------------------------------------------------------------------

# "Middle" choices from choice sets 1 and 2
cards.sets8.1 = matrix(cards.ranked.1_4[3:4,2:3], nrow=1, ncol=4)

# Attributes of best cards
X.8.1 = data.matrix(subset(data.frame(attribdata[,3:12])
       , (Card == cards.sets8.1[1,1])|(Card == cards.sets8.1[1,2])
       |(Card == cards.sets8.1[1,3])|(Card == cards.sets8.1[1,4])
       ,-c(Card)),rownames.force = NA)

# Best Choices ------------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.8.1 = prob.best(X.8.1,beta,1,4)

# Draw Y as a Multinomial Response
```

```
# Apply random multinomial function to each row of prob matrix
Y.8.1 = draw.y(X.8.1,prob.8.1,1,4)

# Identify Best Choices
bchoice.8.1 = id.best(cards.sets8.1,Y.8.1,1,4)

# Remove best choices (Y.b_vec==0 for other than best choices)
# Convert obs matrix to a vector by row (use t())
X.8.2 = remove.choicesX(Y.8.1,X.8.1)
cards.sets8.2 = remove.choicesS(Y.8.1,cards.sets8.1)

# 2nd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.8.2 = prob.best(X.8.2,beta,1,3)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.8.2 = draw.y(X.8.2,prob.8.2,1,3)

# Identify Best Choices
bchoice.8.2 = id.best(cards.sets8.2,Y.8.2,1,3)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.8.3 = remove.choicesX(Y.8.2,X.8.2)
cards.sets8.3 = remove.choicesS(Y.8.2,cards.sets8.2)

# 3rd Best Choices ------------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.8.3 = prob.best(X.8.3,beta,1,2)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.8.3 = draw.y(X.8.3,prob.8.3,1,2)

# Identify Best Choices
bchoice.8.3 = id.best(cards.sets8.3,Y.8.3,1,2)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.8.4 = remove.choicesX(Y.8.3,X.8.3)
cards.sets8.4 = remove.choicesS(Y.8.3,cards.sets8.3)
```

```
# Worst Choices ------------------------------------------------------------
bchoice.8.4=t(cards.sets8.4)


# RESULTS ------------------------------------------------------------------

# Display cards in order presented by choice set ---------------------------
cards.sets8.1

# Identify cards in order presented by choice set --------------------------
Y.bw.8 = id.bestworst(Y.8.1,cards.sets8.1,bchoice.8.4)

# Display cards in ranked order by choice set ------------------------------
cards.ranked.8 = t(rbind(bchoice.8.1,bchoice.8.2,bchoice.8.3,bchoice.8.4))


# ACCOUNTING: Resolved Pairs -----------------------------------------------

# Update resolved pair counts
resolvedpairs.counts <- resolvedpairs.counts.update(cards.ranked.8)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)


# TRANSITIVITY -------------------------------------------------------------

# To account for card 'wins' derived from observed 'wins' using the property of
#     transitivity, e.g., if 1>2 two times and 2>3 one time then 1>3 three times
resolvedpairs.counts <- transitivity(resolvedpairs.counts)

# Update indicator matrix of resolved pairs
resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

# Count number of unresolved pairs
# which(): gives indices of matrix (by column) where == 0
unresolvedpairs_num = length(which(resolvedpairs==0))


# ---------------------------------------------------------------------------
# Choices sets 9+ : Resolve remaining pairs  -------------------------------
# ---------------------------------------------------------------------------

cards.sets.all = rbind(cards.sets1_4.1,cards.sets5.1,cards.sets6.1
          ,cards.sets7.1,cards.sets8.1)
```

```
Y.bw.all = rbind(Y.bw.1_4,Y.bw.5,Y.bw.6,Y.bw.7,Y.bw.8)
cards.ranked.all = rbind(cards.ranked.1_4,cards.ranked.5,cards.ranked.6
          ,cards.ranked.7,cards.ranked.8)
X.all = rbind(X.1_4.1,X.5.1,X.6.1,X.7.1,X.8.1)

# CHECK IF ADDITIONAL CHOICE SETS NEEDED ------------------------------------
# List unresolved pairs -----------------------------------------------------
checklist = listunresolvedpairs(resolvedpairs)

# If list of unresolved pairs is not empty then attempt to create more choice sets
if (length(checklist) != 0) {

  # CREATE ADDITIONAL CHOICE SETS -------------------------------------------

  repeat{

    # Create a list of unresolved card pairs ----------------------------------
    pairlist = listunresolvedpairs(resolvedpairs)
    if (length(pairlist) == 0) {break}

    # Input: list of pairs
    # Output: new choice set
    ncs.1=matrix(c(newchoiceset(pairlist)), ncol = 4)

    # If new choice set contains a set of unique items
    if (sum(duplicated(c(ncs.1))) == 0){

      # Attributes of cards in new choice set
      X.1 = data.matrix(subset(data.frame(attribdata[,3:12])
            , (Card == ncs.1[1,1])|(Card == ncs.1[1,2])
            |(Card == ncs.1[1,3])|(Card == ncs.1[1,4])
            ,-c(Card)),rownames.force = NA); X.1

      # Best Choices ----------------------------------------------------------

      # Construct Linear Predictor and Probabilities
      prob.1 = prob.best(X.1,beta,1,4)
      rowSums(prob.1)

      # Draw Y as a Multinomial Response
      # Apply random multinomial function to each row of prob matrix
      Y.1 = draw.y(X.1,prob.1,1,4)

      # Identify Best Choices
      bchoice.1 = id.best(ncs.1,Y.1,1,4)
```

```
# Remove best choices (Y.b_vec==0 for other than best choices)
# Convert obs matrix to a vector by row (use t())
ncs.2 = remove.choicesS(Y.1,ncs.1)
X.2 = remove.choicesX(Y.1,X.1)

# 2nd Best Choices ----------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.2 = prob.best(X.2,beta,1,3)
rowSums(prob.2)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.2 = draw.y(X.2,prob.2,1,3)

# Identify Best Choices
bchoice.2 = id.best(ncs.2,Y.2,1,3)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
ncs.3 = remove.choicesS(Y.2,ncs.2)
X.3 = remove.choicesX(Y.2,X.2)

# 3rd Best Choices ----------------------------------------------------

# Construct Linear Predictor and Probabilities
prob.3 = prob.best(X.3,beta,1,2)
rowSums(prob.3)

# Draw Y as a Multinomial Response
# Apply random multinomial function to each row of prob matrix
Y.3 = draw.y(X.3,prob.3,1,2)

# Identify Best Choices
bchoice.3 = id.best(ncs.3,Y.3,1,2)

# Remove best choices
# Convert obs matrix to a vector by row (use t())
X.4 = remove.choicesX(Y.3,X.3)
ncs.4 = remove.choicesS(Y.3,ncs.3)

# Worst Choices -------------------------------------------------------
bchoice.4=t(ncs.4)
```

```
    # RESULTS ------------------------------------------------------------

    # Display cards in order presented by choice set ---------------------
    ncs.1

    # Identify cards in order presented by choice set --------------------
    Y.bw = id.bestworst(Y.1,ncs.1,bchoice.4)

    # Display cards in ranked order by choice set ------------------------
    cards.ranked = t(rbind(bchoice.1,bchoice.2,bchoice.3,bchoice.4))

    # Update data of choice sets presented, best and worst choices
    #      and ranked choices

    cards.ranked.all = rbind(cards.ranked.all,cards.ranked)
    cards.sets.all = rbind(cards.sets.all,ncs.1)
    Y.bw.all = rbind(Y.bw.all,Y.bw)
    X.all = rbind(X.all,X.1)

    # ACCOUNTING: Resolved Pairs -----------------------------------------

    # Update resolved pair counts
    resolvedpairs.counts <- resolvedpairs.counts.update(cards.ranked)

    # Update indicator matrix of resolved pairs
    resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

    # TRANSITIVITY -------------------------------------------------------

    # To account for card 'wins' derived from observed 'wins' using the
    #     property of transitivity, e.g., if 1>2 two times and 2>3 one time
    #     then 1>3 three times
    resolvedpairs.counts <- transitivity(resolvedpairs.counts)

    # Update indicator matrix of resolved pairs
    resolvedpairs <- resolvedpairs.update(resolvedpairs.counts,resolvedpairs)

    # Count number of unresolved pairs
    # which(): gives indices of matrix (by column) where == 0
    unresolvedpairs_num = length(which(resolvedpairs==0))

  } else if (sum(duplicated(c(ncs.1))) > 0) {break}

}
```

```
}

# ------------------------------------------------------------------------
# Score the winner of each combination of cards ----------------------------
# ------------------------------------------------------------------------

# Remove "NAs" from count matrix diagonal for arithmetic
for (j in 1:16){
  resolvedpairs.counts[j,j] = 0
}

# Setup an empty scoring matrix
scores = matrix(0, nrow = 16, ncol = 16)

# Score the combination of cards {i,j}
# To handle conflicting wins:
# Element scores[i,j] indicates whether or not (1/0) card i
# won or lost to card j by majority rule
# Card pairs that are unresolved get a score of 0.5

for (i in 1:15){
  for (j in (i+1):16) {
    denom = resolvedpairs.counts[i,j] + resolvedpairs.counts[j,i]
    win_percent = resolvedpairs.counts[i,j]/denom

    if (win_percent>0.5 & denom !=0){
      scores[i,j] = 1
    } else if (win_percent==0.5 & denom !=0){
      scores[i,j] = 0.5
    } else {
      scores[i,j] = 0
    }
    scores[j,i]=1-scores[i,j]
  }
}

# Check unsolved pairs and score as 0.5
for (i in 1:15){
  for (j in (i+1):16) {
    if (resolvedpairs[i,j]==0){ # card pair unresolved
      scores[i,j] = 0.5
    }
    scores[j,i]=1-scores[i,j]
  }
}
```

```
  # Count number of times a card beat other cards
  Ycounts<-rowSums(scores, na.rm=TRUE)

  # ------------------------------------------------------------------------
  # Return DCE data ---------------------------------------------------------
  # ------------------------------------------------------------------------

  # Display unresolved pairs
  unresolvedpairs = listunresolvedpairs(resolvedpairs)
  # Convert X.all to a matrix
  X.all = matrix(X.all, ncol=9)

  output<-list(Y=Y.bw.all, cardsets = cards.sets.all, X=X.all, beta=beta
       , ranked = cards.ranked.all, Yscores=Ycounts)
  return(output)

}
```

B.1.2 (002) Simulate Data - Random Effects.R

```
# Author: Anna Liza Antonio
# Developed using R version 3.3.1
# Contingencies: None

# Program Description ---------------------------------------------------
# The purpose of this code is to:
# Code to generate random effects to be used in script, "(003) Simulate Data.R"

# Load the Required Libraries -------------------------------------------
library(MASS)

# Specifications of the Data  -------------------------------------------

nsubj = 100            # number of subjects
nattrib = 9            # number of attributes which describe each alternative
ndatasets = 1000       # number of datasets to generate

# Define means for multivariate normal distribution of heterogeneity --------

## Population Parameters
popbeta = c(1,.3,1,1.5,.75,1.25,.5,.02,.8)   # example
popbetamat = matrix(rep(popbeta,nsubj), nrow = nsubj, ncol = nattrib, byrow=TRUE)

# Random Effects for Each Subject ---------------------------------------
# We generate nsubj*ndatasets rows of random effects and use every set of 100 rows
# in the matrix (beta) to generate each of the 1000 datasets

# Variance-covariance matrix for normal distribution of heterogeneity
a = 1  # example
resigma = matrix(c(rep(c(a, rep(0, nattrib)), nattrib-1), a), ncol = nattrib)
remean = c(0,0,0,0,0,0,0,0,0)
set.seed(1234)
rebeta = mvrnorm(n=nsubj, remean, resigma)
beta = popbetamat+rebeta     # use random effects betas

# Save the data ---------------------------------------------------------------
write.csv(beta,file="truebetas_n100000.csv")
```

Appendix C

Proofs

This appendix presents work completed prior to the development of the Bayesian hierarchical model for discrete choice data of Chapter 6.

Statement 1. ***If*** X ***is distributed according to a standard Minimum Extreme Value, Type I Distribution(***$\mu = 0$ ***and*** $\beta = 1$***), then*** $-X$ ***is distributed according to a Maximum Extreme Value, Type I Distribution (***$\mu = 0$ ***and*** $\beta = 1$***).***

Proof. The Gumbel (Extreme Value, Type I) distribution has two forms: the minimum extreme value distribution and the maximum extreme value distribution. If X is a random variable having the standard minimum extreme value distribution, then X has probability density function

$$f(x) = \exp(x)\exp(-\exp(x)) \tag{C.1}$$

and the cumulative density function

$$F(X) = 1 - \exp(-\exp(X)). \tag{C.2}$$

If X is a random variable having the standard maximum extreme value distribution, then X has probability density function

$$f(x) = \exp(-x)\exp(-\exp(-x)) \tag{C.3}$$

and the cumulative density function

$$F(X) = \exp(-\exp(-X)). \tag{C.4}$$

We now show that if X is a random variable having the standard minimum extreme value distribution, then -X is a random variable having the standard maximum extreme value distribution.

Let Y = -X. Then

$$\begin{aligned} F_Y(y) &= F(Y \leq y) = F(-X \leq y) = F(X \geq -y) && \text{(C.5)} \\ &= 1 - F(X \leq -y) && \text{(C.6)} \\ &= 1 - (1 - \exp(-\exp(-y))) && \text{(C.7)} \\ &= \exp(-\exp(-y)) && \text{(C.8)} \end{aligned}$$

and

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_y(y) && \text{(C.9)} \\ &= \frac{d}{dy}[1 - (1 - \exp(-\exp(-y)))] && \text{(C.10)} \\ &= \frac{d}{dy}(\exp(-\exp(-y))) && \text{(C.11)} \\ &= \exp(-y)\exp(-\exp(-y)). && \text{(C.12)} \end{aligned}$$

Hence, we see that Y = -X is a random variable having the standard maximum extreme value distribution.

□

Statement 2. ***The probability that respondent*** i ***chooses alternative*** j ***as worst alternative from a choice set containing*** n ***alternatives is***

$$P(Y_i = j) = \frac{\exp(-V_{ij})}{\sum_{k=1}^{n} \exp(-V_{ik})}. \tag{C.13}$$

Consider the random utility model. If we assume that the random components of utility, the ϵ_{ik}'s, are iid with a standard Gumbel distribution (minimum) then the probability of individual i choosing alternative j as worst or least preferred is

$$\begin{aligned} P(Y_i = j) &= P(U_{ij} \leqslant U_{ik}, \text{ for all } k \neq j) && \text{(C.14)} \\ &= P(V_{ij} + \epsilon_{ij} \leqslant V_{ik} + \epsilon_{ik}, \text{ for all } k \neq j) && \text{(C.15)} \\ &= P(-\epsilon_{ik} \leqslant V_{ik} - V_{ij} - \epsilon_{ij}, \text{ for all } k \neq j). && \text{(C.16)} \end{aligned}$$

Then conditional on ϵ_{ij}

$$P(Y_i = j|\epsilon_{ij}) = P(-\epsilon_{ik} \leqslant V_{ik} - V_{ij} + \epsilon_{ij}, \text{ for all } k \neq j) \tag{C.17}$$

is the cumulative distribution for each ϵ_{ik} evaluated at $V_{ij} - V_{ik} + \epsilon_{ij}$. Because the ϵ_{ik}'s are assumed to be independent, the cumulative distribution over all $k \neq j$ is the product of the individual cumulative distributions,

$$P(Y_i = j|\epsilon_{ij}) = \prod_{k=1, k\neq j}^{n} P(-\epsilon_{ik} \leqslant V_{ik} - V_{ij} + \epsilon_{ij}). \tag{C.18}$$

Then by Statement 1,

$$P(Y_i = j|\epsilon_{ij}) = \prod_{k=1, k\neq j}^{n} \exp\{-\exp[-(V_{ik} - V_{ij} - \epsilon_{ij})]\}. \tag{C.19}$$

Because the ϵ_{ij}'s are not given, the choice probability is the integral of $P(Y_i = j|\epsilon_{ij})$ over

all values of ϵ_{ij} weighted by the density, $f(\epsilon_{ij})$. Then

$$\begin{aligned} P(Y_i = j) &= \int_{-\infty}^{\infty} P(Y_i = j|\epsilon_{ij}) f(\epsilon_{ij}) d\epsilon_{ij} \\ &= \int_{-\infty}^{\infty} \prod_{k=1,k\neq j}^{n} \exp\{-\exp[-(V_{ik} - V_{ij} - \epsilon_{ij})]\} \\ &\quad \cdot \exp[-(-\epsilon_{ij})] \exp\{-\exp[-(-\epsilon_{ij})]\} d\epsilon_{ij} \\ &= \int_{-\infty}^{\infty} \prod_{k=1,k\neq j}^{n} \exp[-\exp(V_{ij} - V_{ik}) \exp(\epsilon_{ij})] \\ &\quad \cdot \exp(\epsilon_{ij}) \exp[-\exp(\epsilon_{ij})] d\epsilon_{ij}. \end{aligned} \tag{C.20}$$

Let $u = -\exp(\epsilon_{ij})$. Then $du = -\exp(\epsilon_{ij}) d\epsilon_{ij}$ and the limits change from $\epsilon_{ij} = -\infty$ to $u = 0$ and $\epsilon_{ij} = \infty$ to $u = -\infty$. Now we have that

$$P(Y_i = j) = -\int_{0}^{-\infty} \prod_{k=1,k\neq j}^{n} \Big\{\exp\Big[u \exp(V_{ij} - V_{ik})\Big]\Big\} \exp(u) du \tag{C.21}$$

$$= \int_{-\infty}^{0} \exp\Big\{u\Big[1 + \sum_{k=1,k\neq j}^{n} \exp(V_{ij} - V_{ik})\Big]\Big\} du. \tag{C.22}$$

Finally, if we let $w = u\Big[1 + \sum_{k=1,k\neq j}^{n} \exp(V_{ij} - V_{ik})\Big]$, then
$dw = \Big[1 + \sum_{k=1,k\neq j}^{n} \exp(V_{ij} - V_{ik})\Big] du$ and the limits of integration change from $u = 0$ to $w = 0$ and $u = -\infty$ to $w = -\infty$. We now have that

$$P(Y_i = j) = \frac{1}{1 + \sum_{k=1,k\neq j}^{n} \exp(V_{ij} - V_{ik})} \int_{-\infty}^{0} \exp(w) dw \tag{C.23}$$

$$= \frac{1}{1 + \sum_{k=1,k\neq j}^{n} \exp(V_{ij} - V_{ik})} \tag{C.24}$$

$$= \frac{1}{1 + \exp(V_{ij}) \sum_{k=1,k\neq j}^{n} \exp(-V_{ik})} \cdot \frac{\exp(-V_{ij})}{\exp(-V_{ij})} \tag{C.25}$$

$$= \frac{\exp(-V_{ij})}{\exp(-V_{ij}) + \sum_{k=1,k\neq j}^{n} \exp(-V_{ik})} \tag{C.26}$$

$$= \frac{\exp(-V_{ij})}{\sum_{k=1}^{n} \exp(-V_{ik})}. \tag{C.27}$$

Thus, the probability respondent i chooses alternative j as worst alternative from a choice set containing n alternatives is given by the closed form expression,

$$P(Y_i = j) = \frac{\exp(-V_{ij})}{\sum_{k=1}^{n} \exp(-V_{ik})}. \tag{C.28}$$

Statement 3. ***The probability of a preference order or full ranking for the rank ordered logit model and the sequential best-worst logit is not the same for all values of*** $V_j = X_j\beta$***.***

Proof. Consider the alternatives A,B,C and D and suppose that $A > B > C > D$ is the observed preference order. Under the rank ordered logit model, the probability of observing this preference order for a single respondent is

$$P_{ROL} = \frac{\exp(V_A)}{\sum_{j\in\{A,B,C,D\}}\exp(V_j)}\frac{\exp(V_B)}{\sum_{j\in\{B,C,D\}}\exp(V_j)}\frac{\exp(V_C)}{\sum_{j\in\{C,D\}}\exp(V_j)}\frac{\exp(V_D)}{\exp(V_D)}. \tag{C.29}$$

Under the sequential best-worst logit model, the probability of observing the same preference order is

$$P_{SBW} = \frac{\exp(V_A)}{\sum_{j\in\{A,B,C,D\}}\exp(V_j)}\frac{\exp(-V_D)}{\sum_{j\in\{B,C,D\}}\exp(-V_j)}\frac{\exp(V_B)}{\sum_{j\in\{B,C\}}\exp(V_j)}\frac{\exp(-V_C)}{\exp(-V_C)}. \tag{C.30}$$

If we assume that $P_{ROL} = P_{SBW}$, then

$$\frac{\exp(V_A)}{\sum_{j\in\{A,B,C,D\}}\exp(V_j)}\frac{\exp(V_B)}{\sum_{j\in\{B,C,D\}}\exp(V_j)}\frac{\exp(V_C)}{\sum_{j\in\{C,D\}}\exp(V_j)} = \tag{C.31}$$

$$\frac{\exp(V_A)}{\sum_{j\in\{A,B,C,D\}}\exp(V_j)}\frac{\exp(-V_D)}{\sum_{j\in\{B,C,D\}}\exp(-V_j)}\frac{\exp(V_B)}{\sum_{j\in\{B,C\}}\exp(V_j)}. \tag{C.32}$$

After expanding the sums in the denominators and rearranging the terms we have that

$$\exp(V_C)(\exp(-V_B)+\exp(-V_C)+\exp(-V_D))(\exp(V_B)+\exp(V_C)) = \tag{C.33}$$

$$\exp(-V_D)(\exp(V_B)+\exp(V_C)+\exp(V_D))(\exp(V_C)+\exp(V_D)). \tag{C.34}$$

Using the distributive property and properties of exponents yields

$$\exp(2V_C - V_B) = \exp(V_D). \tag{C.35}$$

If we equate exponents, then

$$2V_C - V_B = V_D \tag{C.36}$$

which implies that

$$V_C = \frac{V_B + V_D}{2}. \tag{C.37}$$

Thus, the probabilities of observing the observed preference order ($A > B > C > D$) under the ROL and the SBW logit models are equivalent only in the case when V_C is the arithmetic average of V_B and V_D. □

Example. Let $V_A = 4$,$V_B = 3$,$V_C = 2$,$V_D = 1$. So, $V_C = \frac{V_B+V_D}{2} = 2$. Then

$$P_{ROL} = 0.313 = P_{SBW}. \tag{C.38}$$

Now let $V_A = 4$,$V_B = 3$,$V_C = 2$,$V_D = 1.5$. So, $V_C = 2$ is not equal to $\frac{V_B+V_D}{2} = 2.5$. Then

$$P_{ROL} = 0.247 \neq 0.252 = P_{SBW}. \tag{C.39}$$

Appendix D

Additional Tables and Figures

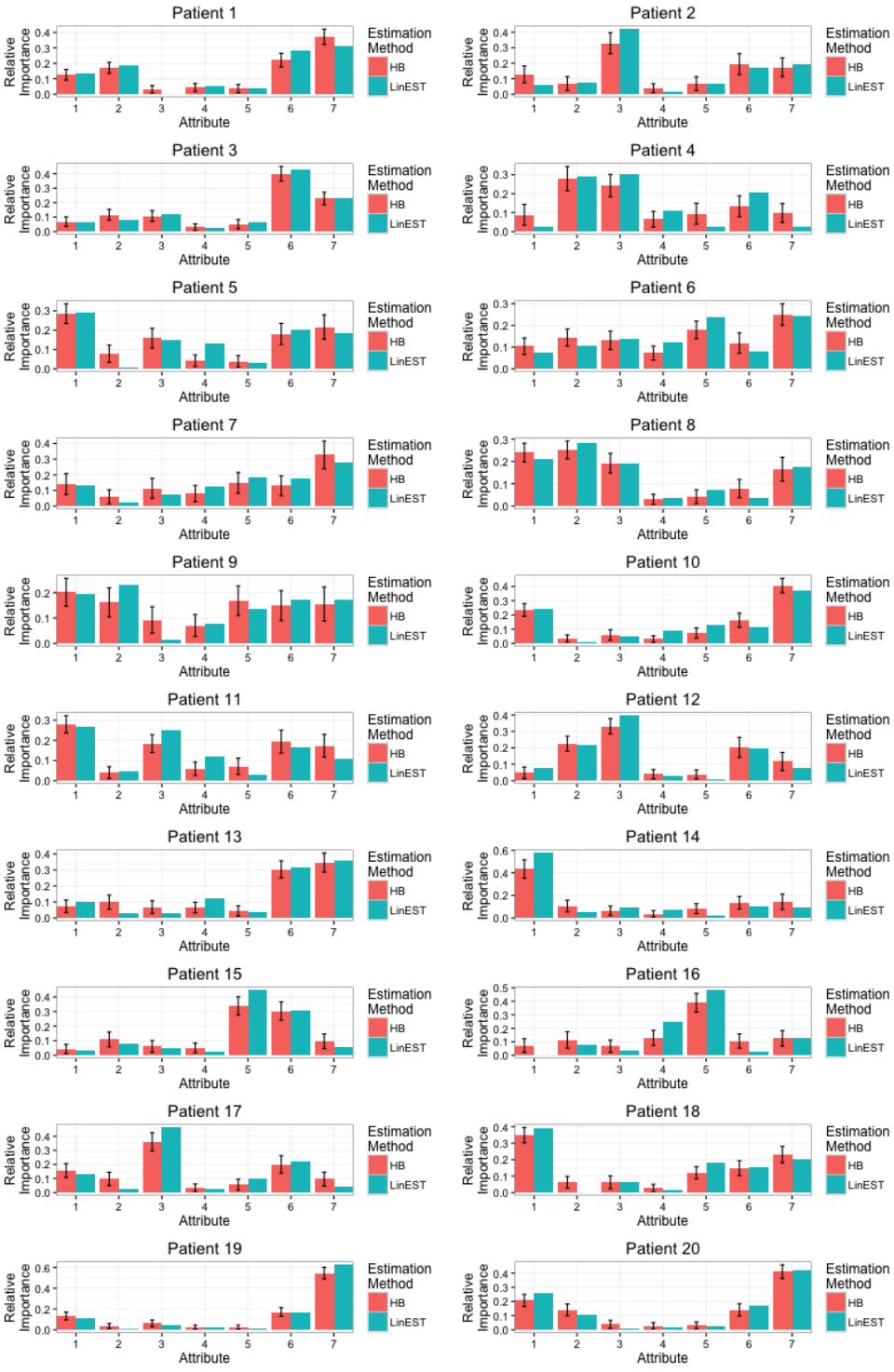

Figure D.1: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

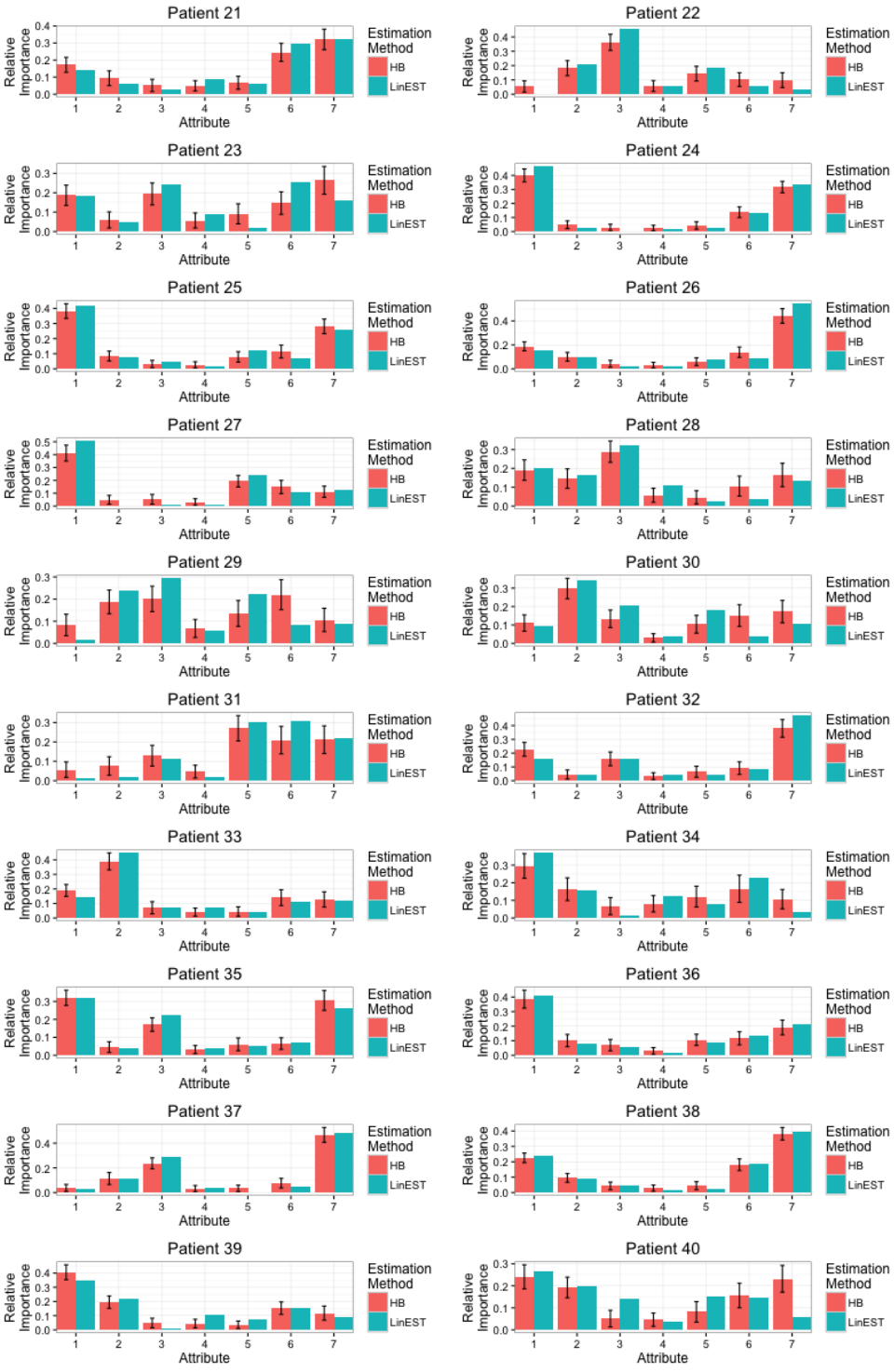

Figure D.2: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

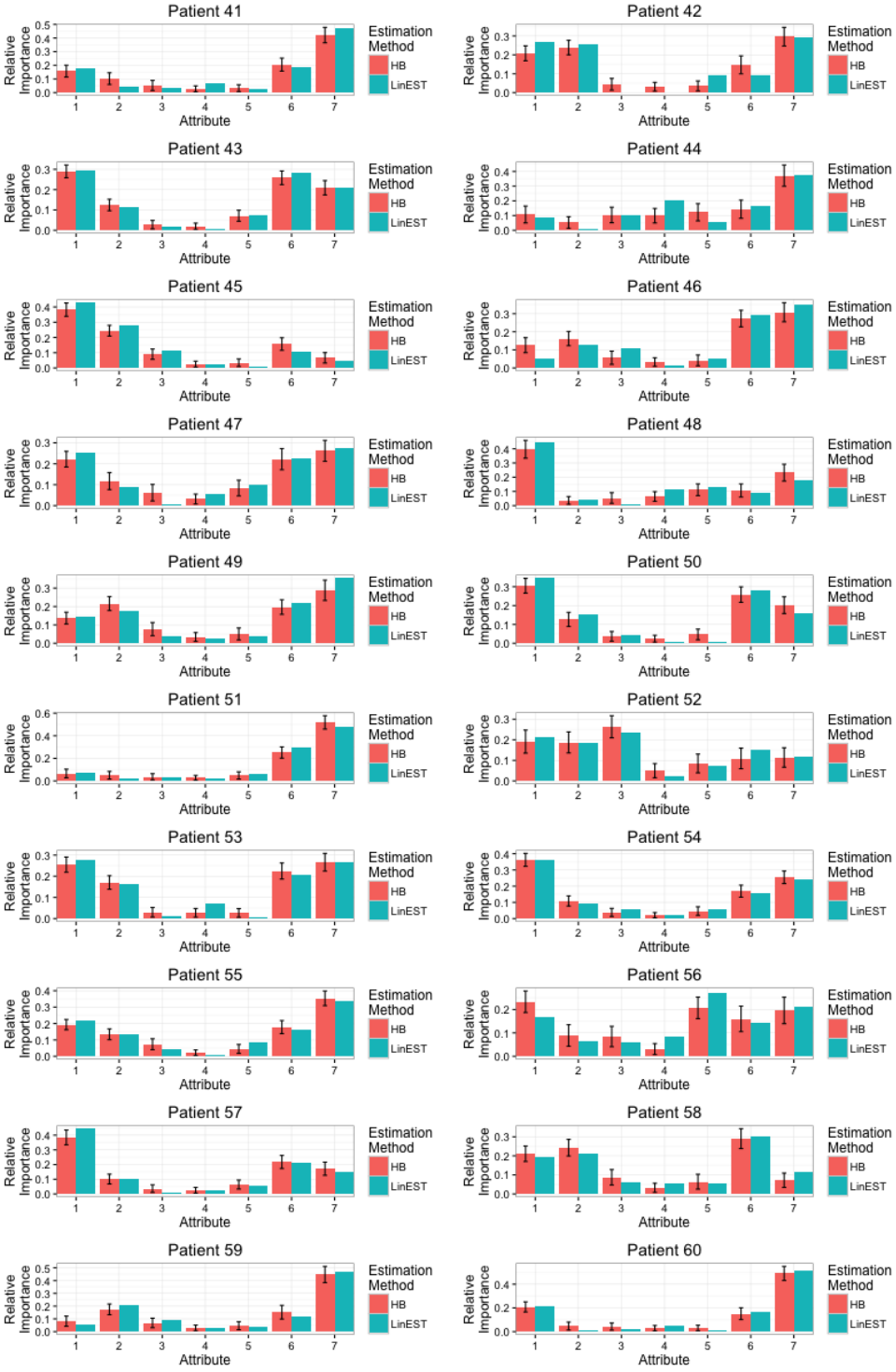

Figure D.3: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

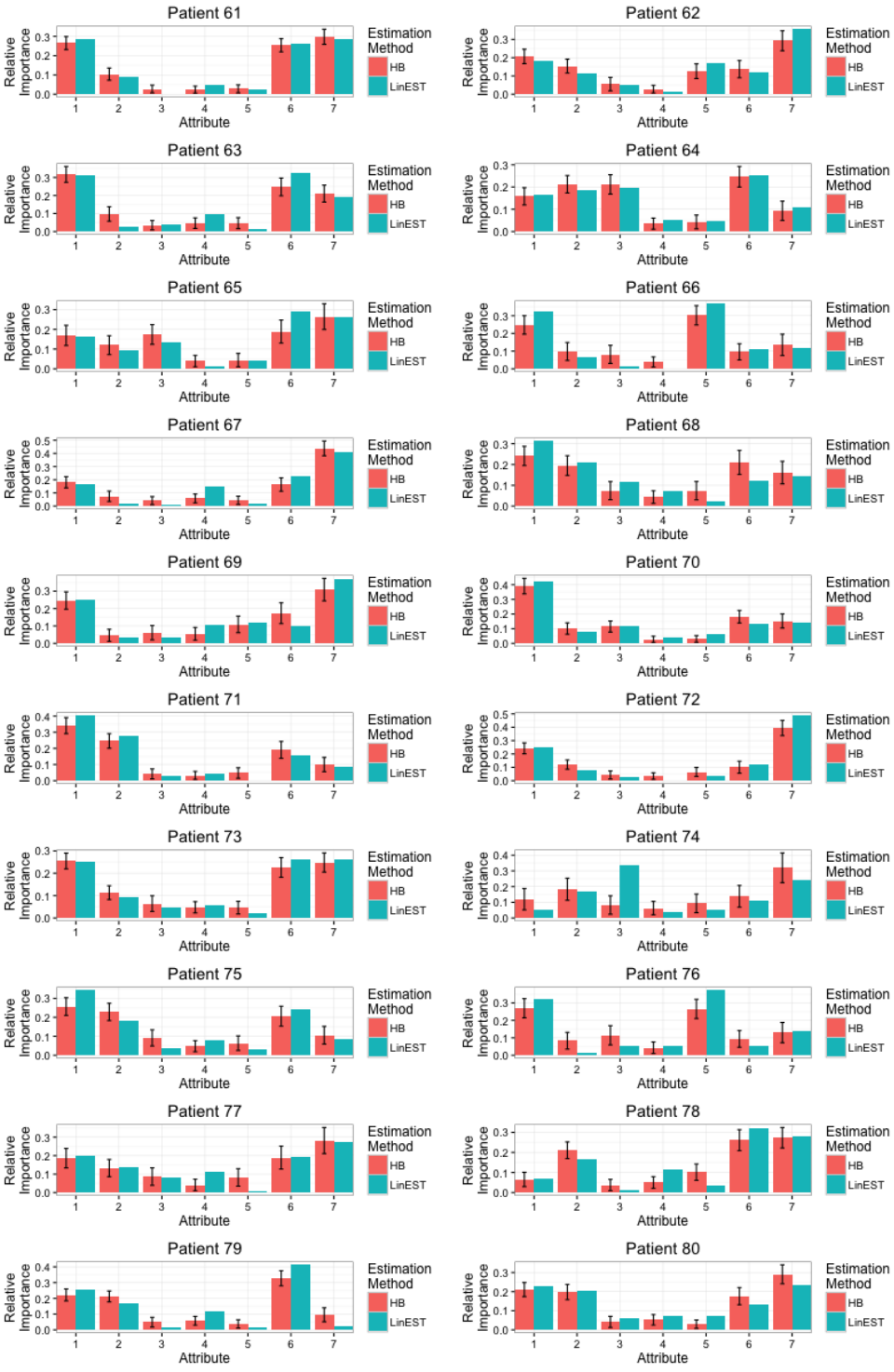

Figure D.4: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

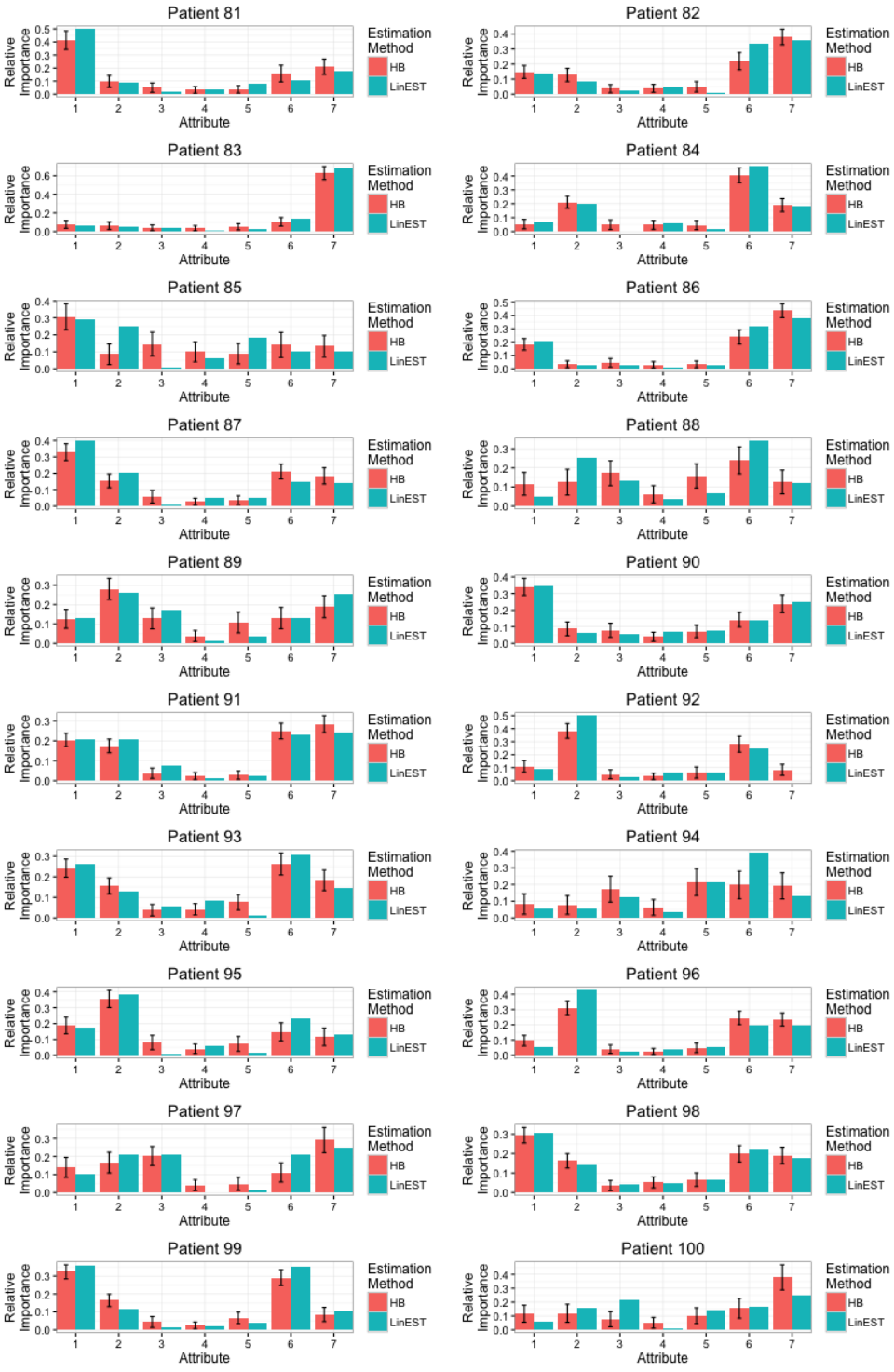

Figure D.5: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

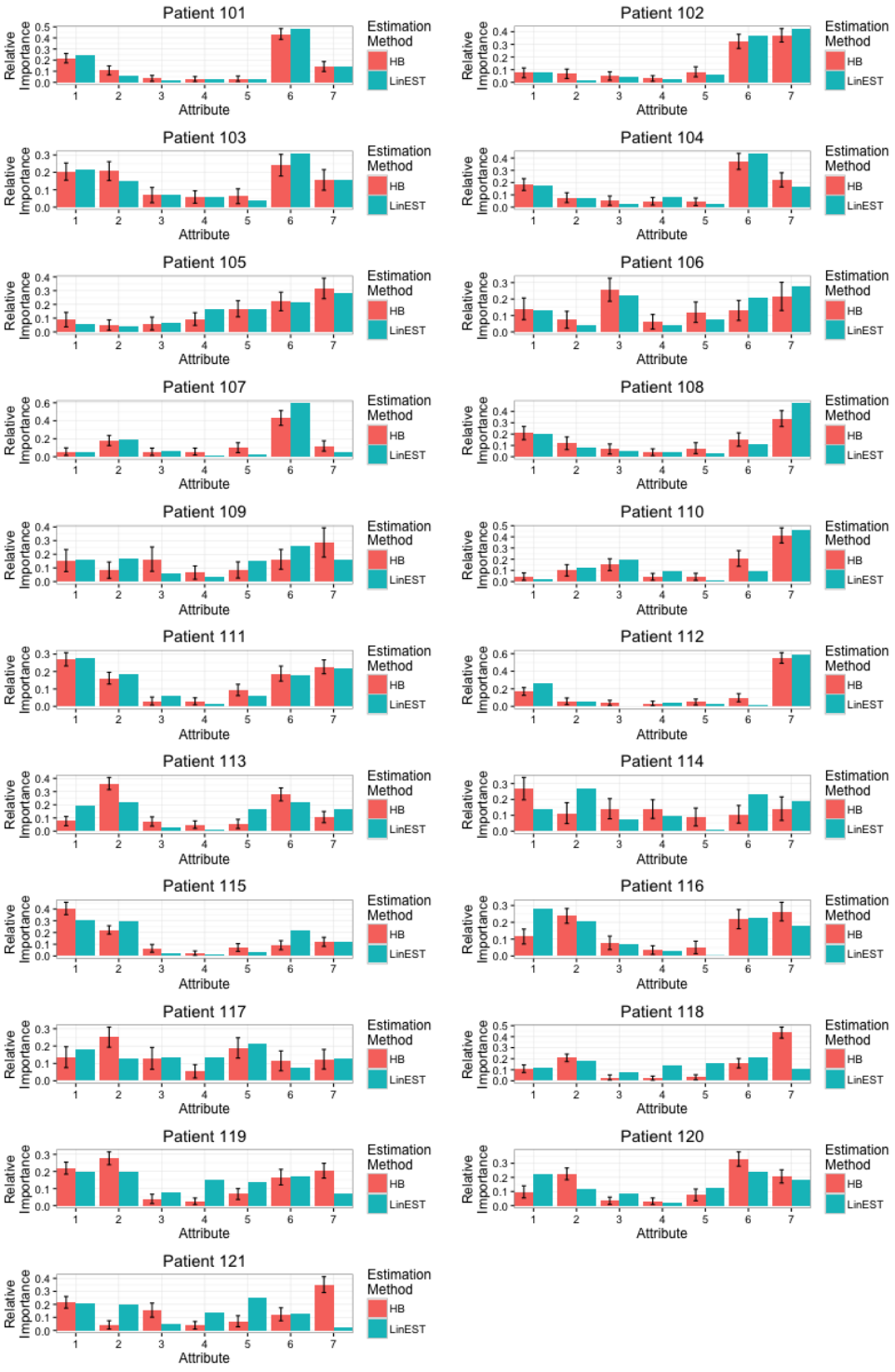

Figure D.6: Posterior mean estimates of relative importance ± 1 SD and the LinEst estimates of relative importance for ten randomly selected men. Attributes IDs correspond to attribute names as follows: 1 = Lifespan, 2 = Bowel Issues, 3 = Cutting, 4 = Taking Action, 5 = Others' Support, 6 = Urinary Functioning, 7 = Sexual Functioning.

Bibliography

Afifi, A., May, S., and Clark, V. (2003). *Computer-Aided Multivariate Analysis, Fourth Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Allenby, G. (1994). Introduction to hierarchical Bayes modeling. In *Advanced Research Techniques Forum*, number D-05 in Tutorial Notes. American Marketing Association.

Allenby, G. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research* 32, 152–162.

Allenby, G. and Lenk, P. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association* 89, 1218–1231.

Allenby, G. and Rossi, P. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics* 89, 57–78.

Allenby, G., Rossi, P., and McCulloch, R. (2005). Hierarchical Bayes models: a practitioners guide. SSRN Working Paper.

Allison, P. and Christakis, N. (1994). Logit models for sets of ranked items. *Sociological Methodology* 24, 199–228.

Ben-Akiva, M. and Bolduc, D. (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper.

Bergland, O. (1994). Estimation of stated preferences from incomplete ranking. In *Discussion Papers*, number D-05 in Discussion Paper Series. Department of Economics and Social Sciences, Agricultural University of Norway.

Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies* 27, 87–101.

Boyd, J. and Mellmand, J. (1980). The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis. *Transportation Research* 14A, 367–378.

Bradlow, E. T., Weiss, R. E., and Cho, M. (1998). Bayesian identification of outliers in computerized adaptive tests. *Journal of the American Statistical Association* 93, 910–919.

Brownstone, D. and Train, K. (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89, 109–129.

Campbell, D. and Hess, S. (2010). Outlying sensitivities in discrete choice data: Consequences and remedies, working paper.

Cardell, N. and Dunbar, F. (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research* 14A, 423–434.

Carson, R. and Louviere, J. (2011). A common nomenclature for stated preference elicitation approaches. *Enviromental and Resource Economics* 49, 539–559.

Chaloner, K. (1991). Bayesian residual analysis in the presence of censoring. *Biometrika* 78, 637–644.

Chaloner, K. (1994). Residual analysis and outliers in Bayesian hierarchical models. In Smith, A. and Freeman, P., editors, *Aspects of Uncertainty*, pages 149–157. Chichester Wiley.

Chaloner, K. and Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika* 75, 651–659.

Chapman, R. and Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research* 19, 288–301.

Cheng, S. and Long, J. S. (2007). Testing for iia in the multinomial logit model. *Sociological Methods & Research* 35, 583–600.

Crabbe, M. and Vandebroek, M. (2011). Improving the efficiency of individualized designs for the mixed logit choice model by including covariates. *Computational Statistics and Data Analysis* 56, 2059–2072.

Dahan, E. and Saigal, C. (2012). The Voice of the Patient. In *Proceedings of the Sawtooth Software Conference*, Orlando, Florida.

DeBekker-Grob, E. W., Ryan, M., and Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21, 145–172.

Dey, D. K., Chen, M.-H., and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* 53, 1239–1252.

Dowsey, M. M., Scott, A., Nelson, E. A., Li, J., Sundararajan, V., Nikpour, M., and Choong, P. F. M. (2016). Using discrete choice experiments as a decision aid in total knee arthroplasty: study protocol for a randomised controlled trial. *Trials* 17, 1–10.

Farrel, P., Groshen, S., MacGibbon, B., and Tomberlin, T. (2012). Outlier detection for a hierarchical Bayes model in a study of hospital variation in surgical procedures. *Statistical Methods in Medical Research* 19, 601–619.

Finn, A. and Louviere, J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy and Marketing* 11, 12–25.

Geisser, S. (1980). Discussion of sampling and Bayes inference in scientific modelling and robustness(box, 1980). *Journal of the Royal Statistical Society: Series A* 143, 416417.

Geisser, S. (1987). Influential observations, diagnostics and discovery tests. *Journal of Applied Statistics* 14, 133–142.

Geisser, S. (1989). Predictive discordancy tests for exponential observations. *Canadian Journal of Statistics* 17, 19–26.

Geisser, S. (1993). *Predictive Inference*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Gelfand, A. (1996). Model determination using sampling-based methods. In Gilks, W., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice, Chapter 9*, pages 145–161. Chapman & Hall, Boca Raton, FL.

Gelfand, A., Smith, A., and Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs samplings. *Journal of the American Statistical Association* 87, 523–532.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1, 515–533.

Giergiczny, M., Valasiuk, S., Czajkowski, M., De Salvo, M., and Signorello, G. (2012). Including cost income ratio into utility function as a way of dealing with "exploding" implicit prices in mixed logit models. *Journal of Forest Economics* 18, 370–380.

Green, P. and Rao, V. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research* 8, 355–363.

Greene, W. H., Hensher, D. A., and Rose, J. (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B: Methodological* 40, 75–92.

Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics* 7, 137–152.

Halbrendt, C., Wang, Q., Fraiz, C., and O'Dierno, L. (1995). Marketing problems and opportunities in Mid-Atlantic seafood retailing. *American Journal of Agricultural Economics* 77, 1313–1318.

Harris, I. R. and Burch, B. D. (2005). Measuring relative importance of sources of variation without using variance. *The American Statistician* 59, 217–222.

Hauber, A. B., Gonzalez, J. M., Groothuis-Oudshoorn, C. G., Prior, T., Marshall, D. A., Cunningham, C., IJzerman, M. J., and Bridges, J. F. (2016). Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. *Value in Health* 19, 300–315.

Hauser, J. and Rao, V. (2002). Conjoint analysis, related modeling and applications. *Marketing Research and Modeling: Progress and Prospects International Series in Quantitative Marketing* 14, 141–168.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica* 52, 1219–1240.

Hensher, D. and Greene, W. (2003). The mixed logit model: The state of practice. *Transportation* 30, 133–176.

Hernandez-Alava, M., Brazier, J., Rowen, D., and Tsuchiya, A. (2013). Common scale valuations across difference preference-based measures: Estimation using rank data. *Medical Decision Making* 6, 839–852.

Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge University Press, Cambridge, United Kingdom.

Jaynes, J., Wong, W.-K., and Xu, H. (2016). Using blocked fractional factorial designs to construct discrete choice experiments for healthcare studies. *Statistics in medicine* .

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., and Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health* 16, 3 –13.

Johnson, J. W. and Lebreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods* 7, 238–257.

Johnson, R. (2000). Understanding HB: An intuitive approach. *Sawtooth Software Research Paper Series* .

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific Publishing Company, 1st edition.

Kruk, M. E., Riley, P. L., Palma, A. M., Adhikari, S., Ahoua, L., Arnaldo, C., Belo, D. F., Brusamento, S., Cumba, L. I. G., Dziuban, E. J., El-Sadr, W. M., Gutema, Y., Habtamu, Z., Heller, T., Kidanu, A., Langa, J., Mahagaja, E., McCarthy, C. F., Melaku, Z., Shodell, D., Tsiouris, F., Young, P. R., and Rabkin, M. (2016). How can the health system retain women in HIV treatment for a lifetime? A discrete choice experiment in Ethiopia and Mozambique. *PLoS ONE* 11, 1–14.

Kruskal, W. and Majors, R. (1989). Concepts of relative importance in recent scientific literature. *The American Statistician* 43, 2–6.

Lancsar, E. and Louviere, J. (2008). Estimating individual level discrete choice models and welfare measures using best worst choice experiments and sequential best worst MNL. In *CenSoC Working Paper Series*, number 08-003.

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

Louviere, J. (1998). Conjoint analysis modeling of stated preferences: A review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy* 31, 375–142.

Louviere, J., Street, D., Burgess, L., Wasi, N., Islam, T., and Marley, A. (2008). Modelling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1, 128–163.

Louviere, J. J., Pihlens, D., and Carson, R. (2011). Design of discrete choice experiments: A discussion of issues that matter in future applied research. *Journal of Choice Modelling* 4, 1 – 8.

Luce, R. and Tukey, J. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology* 1, 1–27.

Lusk, J. L. and Norwood, F. B. (2005). Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics* 87, 771–785.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Economics*, pages 105–142. Wiley, New York.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Economics* 5, 447–470.

Mehndiratta, S. (1996). *Time-of-day Effects in Inter-city Business Travel*. PhD thesis, Berkley, California.

Orme, B. (2010). *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Research Publishers LLC, Madison, WS, 2nd edition. 29-37.

Orme, B. and Howell, J. (2009). Application of covariates within Sawtooth Software's CBC/HB program: theory and practical example. In *Sawtooth Software Conference Papers (2009)*, Sequoia, WA. Sawtooth Software.

Paul E. Green, V. S. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research* 5, 103–123.

Pettit, L. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society, Series B* 52, 175–184.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Rao, V. R. (2008). Developments in conjoint analysis. In Wierenga, B., editor, *Handbook of Marketing Decision Models*, pages 23–53. Springer US, Boston, MA.

Retzer, J., Soofi, E., and Soyer, R. (2009). Information importance of predictors: Concept, measures, Bayesian inference, and applications. *Computational Statistics & Data Analysis* 53, 2363–2377.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics* 53, 647–657.

Revelt, D. and Train., K. (2000). Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. Economics Working Papers E00-274, University of California at Berkeley.

Rigby, D. and Burton, M. (2006). Modeling disinterest and dislike: A bounded Bayesian mixed logit model of the UK market for GM food. *Environmental Resource Economics* 33, 485–509.

Rossi, P. and Allenby, G. (1993). A Bayesian approach to estimating household parameters. *Journal of Marketing Research* 30, 171–182.

Ryan, M., Gerard, K., and Amaya-Amaya, M. (2008). *Using Discrete Choice Experiments to Value Health and Healthcare*. Springer, Dordrecht, The Netherlands, 1st edition.

Saigal, C. and Dahan, E. (2012). Voice of the Patient. In *Proceedings of the Sawtooth Software Conference 2012*, pages 153–164. Sawtooth Software, Inc.

Soofi, E. S., Retzer, J. J., and Yasai-Ardekani, M. (2000). A framework for measuring the importance of variables with applications to management research and decision models. *Decision Sciences* 31, 595–625.

Sun, M., Lebanon, G., and Kidwell, P. (2012). Estimating probabilities in recommendation systems. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61, 471–492.

Taneva, B., Giesen, J., Zolliker, P., and Mueller, K. (2008). Choice based conjoint analysis: Discrete choice models vs direct regression. In *Proceedings of the ECML PKDD Workshop on Preference Learning*.

Train, K. (2001). A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. *Working Paper No. E00-278, University of California, Berkeley* .

Train, K. (2009). *Discrete Choice Methods With Simulation*. Cambridge University Press, New York, NY, 2nd edition.

van Dijk, J. D., Groothuis-Oudshoorn, C. G. M., Marshall, D. A., and IJzerman, M. J. (2016). An empirical comparison of discrete choice experiment and best-worst scaling to estimate stakeholders: risk tolerance for hip replacement surgery. *Value in Health* 19, 316–322.

Velandia, M., Lambert, D., Mendieta, M., Roberts, R., Larson, J., English, B., Rejesus, R., and Mishra, A. (2011). Factors influencing cotton farmers' perceptions about the importance of information sources in precision farming decisions. In *Proceedings of the AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July, 2011*, Selected Paper Series. Agricultural & Applied Economics Association.

Viney, R., Lancasa, E., and Louviere, J. (2002). Discrete choice experiments to measure consumer preferences for health and healthcare. *Pharmacoeconomics Outcomes Research* 4, 89–96.

Weiss, R. (1994). Pediatric pain, predictive inference, and sensitivity analysis. *Evaluation Review* 18, 651–677.

Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 739–750.

Wittink, D. R., Krishnamurthi, L., and Nutter, J. B. (1982). Comparing derived importance weights across attributes. *Journal of Consumer Research* 8, 471–474.

Wittink, D. R., Krishnamurthi, L., and Reibstein, D. J. (1990). The effect of differences in the number of attribute levels on conjoint results. *Marketing Letters* 1, 113–123.