Fall 2017

Models for bundle preference estimation using configuration data

I-Hsuan Shaine Chiu
University of Iowa

MODELS FOR BUNDLE PREFERENCE ESTIMATION
USING CONFIGURATION DATA

by

I-Hsuan (Shaine) Chiu

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration
in the Graduate College of
The University of Iowa

December 2017

Thesis Supervisors: Professor Gary J. Russell
Professor Thomas S. Gruca

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

I-Hsuan (Shaine) Chiu

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Business Administration at the December 2017 graduation.

Thesis committee: ______________________________

Gary J. Russell, Thesis Supervisor

____

Thomas S. Gruca, Thesis Supervisor

____

Jacob Oleson

____

Sanghak Lee

____

Ying Yang

ABSTRACT

Bundling is pervasive in the market; examples include desktop computer bundles, digital single-lens reflex camera kits and cookware sets, to name a few. The advancement in information technology allows more and more companies to provide customized bundles to customers. Wind and Mahajan (1997) recognize the importance of researching mass customization and suggest companies to use consumers' input "as a response (to a conjoint analysis-type task) that provides operational guidelines for the design of products to inventory for the segment that is not willing to pay the premium required for customized products".

In addition to conjoint analysis, researchers and practitioners are using a "build-your-own-bundle" or configuration approach. In a configuration study, participants are presented with a menu from which they can choose individual items to build up their desired product bundle. The process mimics the real decision process, is easy to implement, and is straight forward for participants to understand. However, as the size of the menu grows, the number of possible bundles grows geometrically. This results in computation difficulties.

This dissertation investigates the application of configuration approach, and examines if it extends and complements the choice-based conjoint (CBC) approach. We first develop an aggregate model for analyzing configuration data. We show analytically that the aggregate choice model consistent with configuration data has a closed form representation which takes the form of a Multivariate Logistic (MVL) model. We discuss the strengths

and weaknesses of the configuration approach.

Because configuration and conjoint data tasks have different strengths and weaknesses, taking advantages of these two choice tasks may improve the understanding of consumer preferences for bundles. A fundamental assumption in the data fusion literature is that the same decision making process is applied under different choice tasks. We examine whether consumer decision making process is the same under CBC and configuration studies by comparing the estimation results from CBC and conjoint studies. We show that these two procedures may not be fully comparable.

This dissertation makes two key contributions to the choice modeling literature. First, we show that the MVL model can be viewed as an aggregate -level bundle choice model. Second, we show that different data collection procedures (configuration and conjoint data tasks) can lead to different conclusions about the bundle preference distribution. Reconciling the differences is a topic for future research.

PUBLIC ABSTRACT

Bundling is pervasive in the market; examples include desktop computer bundles (i.e., computer, monitor, printer and/or scanner), season tickets for a performing arts center and prix fixe dinner menus. Companies that wish to provide bundles to the market need to understand consumers' preferences for bundles so that they can optimize their bundle design.

Researchers often use a configuration data approach ("build-your-own-bundle") to study bundle preferences. In a configuration study, participants are asked to create an ideal bundle from a given menu of various products. This procedure mimics the real-life decision process and is straightforward for participants. Researchers are interested in configuration studies because it is easy to design and implement.

In this research we propose an approach to analyze configuration data. Our proposed approach deals with the challenges of the configuration data, and can be analyzed using existing statistical software. It also provides managers with insight into consumer preference segmentation and bundling recommendation.

We also compare the configuration study with conjoint study, a widely used market research approach by practitioners. Because these two approaches have different strengths and weaknesses, researchers may improve their understanding of bundle preferences by using both configuration and conjoint approaches. We evaluate the comparability of these two data sources and provide suggestions with respect to research study design and data analysis.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

Bundling is the practice of selling two or more separate products in a single package for a special price (Basu and Vitharana, 2009; Ghosh and Balachander, 2007; Stremersch and Tellis, 2002). It is pervasive in the market place and receives a lot of attention from both economic and marketing researchers. The academic research on bundling by economists has predominantly been normative (Venkatesh and Mahajan, 2009). Several researchers examine the conditions under which a pure component strategy (consumers can buy products separately), a pure bundling strategy (consumers can only buy products as a bundle) or a mixed bundling strategy (consumers can choose to buy the products separately or as a bundle) would be optimal (Basu and Vitharana, 2009; Ghosh and Balachander, 2007; Venkatesh and Kamakura, 2003).

Researchers also investigate the rationale behind bundling from the perspectives of supply, demand and competition (see Venkatesh and Mahajan (2009) for a review). Supply-side rationale includes lower sorting and inventory holding costs, as well as increasing economies of scope (Eppen et al., 1991; Gilbert and Katz, 2001). Reasons on the demand side include variety seeking, savings from bundle purchase and reduced search cost (Harris and Blair, 2006; Yadav and Monroe, 1993). From the competition perspective, Eppen et al. (1991) and Kamakura et al. (2003) find that bundling increases customer switching costs and reduces customer turnover.

For marketing, there are two streams of bundling research. One directly extends the work done by economists. For example, researchers have shown that sellers benefit from bundling due to demand expansion (Eppen et al., 1991). Other studies show that bundle promotions can be an effective marketing tool for introducing new products or preventing price-sensitive customers from yielding to price promotions (Simonin and Ruth, 1995; Balachander et al., 2010). Marketing researchers complement this economic-oriented work by studying optimal bundle design and pricing. Some representative models include conjoint analysis (Goldberg et al., 1984) and balance modeling (Farquhar and Rao, 1976).

Understanding consumer preference for bundles is a key element of effective new product design. Conjoint analysis, while it is well established and widely used by practitioners, has a major drawback: it imposes significant cognitive strains on participants, resulting in decision fatigue that negatively impacts quality of the data. Thus, some practitioners have developed a "build-your-own-bundle" approach, also known as configuration analysis. In a configuration choice scenario, consumers are asked to create an ideal bundle from a given menu of various product options. This procedure mimics the real bundle decision making process, is straightforward and easy for participants to understand and does not cause participant fatigue.

The resulting choice data, however, poses two challenges for researchers. First, as the size of the menu grows, the number of possible bundles grows geometrically. Second, because each consumer only configures one ideal product bundle, researchers have only one choice observation per consumer. This limited amount of data from each respondent does not provide enough variation to obtain good estimates of model parameters.

The goal of this dissertation is twofold. We first develop a modeling approach to estimate bundle preferences using configuration data and evaluate the strengths and weaknesses of a configuration study. Next, we compare the bundle preferences collected via conjoint study with configuration study to access if consumer decision-making processes are compatible under these two different tasks. We believe that researchers can benefit from combining data from different measurement approaches if the decision-making processes are compatible.

The rest of the dissertation is organized as follows. Chapter 2 focuses on the preference analysis only based on configuration choice data. It reviews the literature on bundle models and develops a modeling framework that deals with the challenges of menu-based choice data, but which also provides clear insights into consumer preference segmentation. Chapter 3 analyzes the conjoint data and compares the result with the estimation obtained from configuration study to evaluate the decision-making process. It also proposes an approach to combine both conjoint and configuration data to further improve the bundle preference estimation. Chapter 4 concludes the findings from our study and discusses future research direction.

CHAPTER 2
ANALYSIS OF CONFIGURATION DATA

2.1 Introduction

Product bundling has been of interest to marketers and economists for decades (Stigler, 1963; Adams and Yellen, 1976; Yadav and Monroe, 1993). Researchers have created a typology of bundle types (Rao et al., 2017; Stremersch and Tellis, 2002), suggested various rationales for bundling and identified factors affecting optimal bundling strategies (Venkatesh and Mahajan, 2009; Venkatesh and Kamakura, 2003). In this research, we focus on bundle design and bundling strategy from a manufacturer's perspective. We define a bundle as a collection of individual products selected from one or more product categories that is sold for a discounted price. Classic examples include desktop computer bundles (i.e., computer, monitor, printer and/or scanner), season tickets for a performing arts center and prix fixe dinner menus.

2.1.1 Factors Affecting Bundling Strategy

Past analytical research has identified several factors that affect bundling strategy decisions. Since bundling is a form of price segmentation (i.e., price discrimination), the success of a bundling strategy depends on the reservation prices for the products in the bundle. Reservation prices for products may be positively-, negatively- or un-correlated. Stigler (1963) finds that pure bundling (firm only offers a bundle) is optimal for monopolists when reservation prices are perfectly negatively correlated. In his example, pure bundling reduces buyers heterogeneity in the reservation prices for the entire bundle (Venkatesh and Mahajan, 2009). Venkatesh and Kamakura (2003) find that different bundling strategies (mixed versus pure) are optimal depending on the levels of complementarity or substitutability among products. More recently, Armstrong and Vickers (2010) examine the impact of demand elasticity in a competitive market. They find that providing discounts on bundles actually reduces profits if demand is not sufficiently elastic (p. 45). In short, bundling strategies may depend on demand-side factors such as consumer heterogeneity, price elasticity, product interdependence (complementarity or substitutability). It may also be influenced by considerations such as firm cost structure and market competition (Venkatesh and Mahajan, 2009; Venkatesh and Kamakura, 2003).

2.1.2 Measuring Bundle Preference

For companies who want to implement a bundling strategy, it is critical to understand these demand-side factors. Conjoint analysis is a well-established and widely-used approach by practitioners to gather information on preference distributions, consumer heterogeneity and price elasticity. However, it has a major drawback. The repetitive and highly similar choice tasks can be a burden for participants resulting in decision fatigue and reduced data quality.

To address this concern, practitioners have developed the "build-your-own-bundle" (BYOB) or configuration approach to collect bundle preference data from consumers (Johnson et al., 2006; Rice and Bakken, 2006; Sambandam and Kumar, 2012). In a configuration study, participants are asked to create an ideal bundle from a given menu consisting of various product options.

The major benefit of configuration studies is the ease of implementation. This procedure is easy for researchers to design and implement in an online survey. For participants, the task is natural and very straightforward to understand. Often, the content of the menu does not change and participants only need to construct their ideal bundle once. Hence, they do not experience decision fatigue. However, collecting a single measurement may not provide sufficient information on key bundle attributes which may impact the estimation results.

2.1.3 Research Challenges

The resulting configuration data poses two analytical challenges for researchers. First, as the size of the menu (number of products that can be included in the bundle) grows, the number of possible bundles that a participant might construct grows geometrically. This implies a large number of alternatives in the choice set. Model estimation from a large choice set is computationally expensive and may even be impossible.

The easiest approach to analyze configuration data is counting analysis (Sambandam and Kumar, 2012; Johnson et al., 2006). However, counting analysis ignores the product interdependence in a bundle. Moreover, it cannot capture the preference heterogeneity across consumers. Furthermore, other proposed approaches such as Multinomial Logit (MNL) models (Johnson et al., 2006; Rice and Bakken, 2006) cannot capture product interdependencies. We assert that the interdependencies should not be ignored. Products in a bundle are usually complements or they may be partial substitutes (depending on the strategy of the firm). Therefore, counting analysis may result in a bundle that is impractical and/or unwanted.

The second analysis challenge arises form the small amount of data collected from each participant. In many instances, the participant configures only one ideal product bundle. So, researchers have only one choice observation per participant. This limited amount of data does not allow researchers to obtain estimates of model parameters at the individual level.

2.1.4 Overview of Research

In this chapter, we develop a modeling framework that deals with the challenges of configuration data. The goal is to provide marketers an approach to investigate consumer bundle preferences from a bundle configuration task. We develop a Mixed Logit model suitable for the analysis of configuration data. The preference heterogeneity distribution in our model takes into account the fact that participants for the bundle choice experiment are screened based on their interests in the product category under study. We show analytically that a Mixed Logit model consistent with the participant screening process takes the form of the multivariate logistic (MVL) choice model. The proposed Multivariate Logistic (MVL) model provides clear insights into product complementarity or substitutability as well as consumer preference heterogeneity.

The rest of this chapter is organized as follows. Following the literature review, we show analytically that the population choice model for bundle preferences takes the form of a MVL model with parameters reflecting consumer preference heterogeneity. We apply the methodology to a real-world configuration study of bundles of power tools. Finally, we conclude with a discussion of the methodology, the limitation of configuration studies and future directions.

2.2 Literature Review

Product bundles formed under a mixed bundling strategy typically have four characteristics. First, bundles consist of separate products. For example, a desktop computer bundle might include a computer, a monitor and a keyboard. Second, the probability that a bundle is chosen depends on the consumers' preferences for the individual products in a bundle. For instance, gamers are more likely to choose a desktop computer bundle with a high resolution monitor than regular users. Third, the interdependencies between products in a bundle influence the probability that a bundle is chosen. A desktop computer bundle including a computer and a monitor is more popular than a bundle consisting only of a monitor and a pair of computer speakers. Fourth, bundles are always sold for a special price. This special price may be a special "bundle" price that varies by items, a percentage-off discount or a quantity discount.

Bundles are conceptually similar to "shopping baskets" which are widely studied in the marketing literature. They share the first three characteristics. Therefore, shopping basket models can serve as a basis for modeling product bundle preferences. We next review modeling frameworks from both the bundling and shopping basket literatures.

2.2.1 Typology of Bundle Choice Models

Bundle choice models can be classified into two categories based on unit of analysis in the utility model. Both attribute-based and product-based (or component-based) approaches have appeared in the literature (Rao, 2004). The attribute-based approach models the utility in terms of the *attributes* all the products in the bundle. For example, Chung and Rao (2003) model the utility based on product attributes that are further classified into three comparability types: attributes that are common to all products, partially shared by products and unique to specific products. This approach captures the interactions among products by assessing the aggregation and dispersion of the attributes across products.

To estimate an attribute-based bundle choice model, a researcher needs to define the attributes of all products. In addition to bundle choice data, researchers have to collect the ratings of the product attributes from each consumer. If the products being considered for a bundle share very few attributes, the complexity of the data collection process and model estimation greatly increase.

In contrast, a product-based bundle choice model treats each *individual product* in a bundle as a unit of analysis. This approach does not require further utility decomposition of products into attributes. It is easier to implement when products in a bundle are highly heterogeneous. This type of model requires only bundle choice data. (The shopping basket literature (Manchanda et al., 1999; Russell and Petersen, 2000) naturally falls in this category as well.) Because our empirical study deals with bundles of highly heterogeneous products, we adopt a product-based approach, viewing products in a bundle as the base unit of bundle utility.

2.2.2 Product-Based Models

When the individual products in a bundle are viewed as the base unit for bundle utility analysis, several modeling approaches are possible. For the multinomial logit (MNL) approach (Johnson et al., 2006), the bundle-building process is viewed as a single choice task in which individuals are making decisions from 2^J bundles, where J is the number of products available to choose from. As J increases, the number of possible bundles can be so large that standard MNL software cannot calibrate the model.

To avoid this problem, practitioners assume independence among products. Thus, the simultaneous decisions for products are viewed as a set of separate choice tasks (Johnson et al., 2006; Rice and Bakken, 2006). A standard binary logit model is estimated for each product and the choice probability of a bundle is assumed to be the product of the choice probabilities of the products on the menu. Conceptually speaking, this approach is incorrect even though it is easy to implement. Assuming independence across products in a bundle is not consistent with the idea of a product bundle. In terms of methodology, Johnson et al. (2006) show that the results obtained from the binary logit approach are highly correlated with the estimates obtained from MNL model. Additionally the average part-worths are identical to the full MNL model to at least three decimal places. While binary logit approach is easy to handle and can be analyzed by a standard statistic program, the underlying assumption of independence of items is questionable. Moreover, we can show that MNL and binary logit models are nested in the multivariate logistic model (discussed in the subsequent paragraphs) with independence assumption. The derivation is provided in Appendix A.

2.2.3 Shopping Basket Models

In the shopping basket literature, there are two main approaches to deal with the interdependencies among products: the multivariate probit (MVP) model (Manchanda et al., 1999; Liechty et al., 2001) and the MVL model (Russell and Petersen, 2000; Kamakura and Kwak, 2014). In the MVP framework, an individual's latent utility for various products is assumed to have a joint normal distribution. The interdependencies of products are captured by the correlation structure in the multivariate normal distribution. This correlation structure implicitly assumes that the choice decision of putting one product in a bundle is influenced by all of the other products even if they are not in the bundle.

With roots in spatial statistics, the MVL approach develops marginal conditional distribution by taking influences from adjacent areas into consideration (Cressie, 1993). Thus, the MVL model explicitly incorporates parameters to capture the pairwise interdependencies among products. In a marketing context, the MVL model assumes that consumers choose a product to be in a bundle based on the presence (or absence) of other products in the bundle. Under appropriate assumptions, it has been shown that a joint distribution, based on the set of conditional distributions, can be derived. Furthermore, it has the form of the multivariate logistic distribution (Cox, 1972). Due to its explicit modeling of interdependencies, the MVL model has been applied to other areas, including studies of social influence on choice decisions (Moon and Russell, 2008; Yang et al., 2010).

2.2.4 MVL vs. MVP Models

In addition to their different approaches of modeling interdependencies among products, the major difference between MVL and MVP models is the ability to handle bundle-level attributes. A bundle-level attribute is related to the bundle as a whole, regardless of which products are included in a bundle. A quantity discount is an example of a bundle-level attribute. In practice, this usually means that the discount depends on the size of the bundle (e.g., buy two products get 20% off and buy three products get 25% off). In such as case, the discount does not depend on which products are included in the bundle. While both MVL and MVP models decompose the bundle utility into product utilities, the choice probability of a bundle in a MVP process assumes that the consumer makes simultaneous decisions for all products on the menu. However, in the actual configuration task, the respondent adds products to the bundle one by one. During the task, the consumer may not yet know what the final bundle will be if there is a quantity discount. Furthermore, due to the presence of bundle attributes, the choice probability for an individual product is actually undefined until the entire bundle is chosen. We provide more discussion in Appendix B.

In contrast, the MVL model has a specific expression for the utility of bundle attributes. For this reason, the MVL model is more suitable to analyze bundle choices when bundle-level attributes are present. Table 2.1 summarizes and compares the binary logit, MVP and MVL approaches.

Table 2.1: Summary of Modeling Approaches

Approach	Binary Logit	MVL	MVP
Analysis set up	View the configuration task as a process of making multiple independent choices with regard to products.	Individuals make multiple simultaneous choices of products.	Individuals make multiple simultaneous choices of products.
Interdependency	Do not allow interdependence among products.	Capture the interdependencies by explicitly incorporating interaction variables in the model. The utility of an product is influenced only by other products that are in the bundle.	Capture the interdependencies among products through a correlation matrix. The utility of an product is influenced by all other products even they are not in the bundle.
Utility formulation	An product will be will chosen if the utility for that product is above threshold.	An product will be included in the bundle if the utility for that product, after considering the impact from other products in the bundle, exceeds threshold.	An product will be included in the bundle if the utility for that product, after considering the impacts from all other products, exceeds threshold.
Consideration set assumption	All possible bundles are considered by subjects.	Choice can be restricted to a subset of bundles.	Choice can be restricted to a subset of bundles.
Model design	After constructing the chosen probability of each product given their characteristics, the probability of a bundle being chosen is the multiplication of the probability of each product.	Given characteristics of products, the model simultaneously predicts the chosen probability of a collection of products.	Given characteristics of products, the model simultaneously predicts the chosen probability of a collection of products.

2.3 Model Development

Our empirical study focuses on product bundles with bundle-level attributes. More specifically, we consider a situation where quantity discounts are offered on bundles. Following the discussion above, we start with the MVL choice model at the individual level. We then take into account the preference heterogeneity for products and derive the aggregate bundle choice model.

2.3.1 Individual Model

Suppose a menu consists of J products. Consumers can, therefore, assemble a total of 2^J different bundles, including an empty bundle consisting of no products. Let $b = 1, \ldots, B$ denote the bundle, where 1 represents the null bundle and $B = 2^J$ is the bundle consisting of all products. For individual i we define the direct utility function for the bundle category at time t as follows:

$$u(x) = \sum_b \psi_{bt} x_b$$

where x_b is the quantity for bundle b. Define P_j as the price for product j and $z_{jb} = 1$ if product j is in bundle b and 0 otherwise. The total price for bundle b, TP_b, is $TP_b = \sum_j P_j z_{jb}$.

Let d_b be the quantity discount for bundle b. Denote $BP_b = TP_b(1 - d_b)$ as the bundle price for bundle b after quantity discount. If E is the total expenditure on the bundles, the utility maximization decision is

$$\max\ u(x) \ \text{ subject to } \sum_b BP_b x_b \leq E$$

We define the marginal utility for bundle b as $\psi_{bt} = \psi_b e^{\varepsilon_{bt}}$. The term ψ_b is the deterministic component of the marginal utility. The term ε_{bt} is a random element which represents factors that influence the customer's choice but are unobserved to researchers. Applying the Kuhn-Tucker first-order conditions, the choice probability for bundle b is

$$\begin{aligned} Pr(b) &= Pr\left(\frac{\psi_{bt}}{BP_b} > \frac{\psi_{b't}}{BP_{b'}}\right) \qquad (2.1) \\ &= Pr\left(\ln\psi_b - \ln BP_b + \varepsilon_b > \ln\psi_{b'} - \ln BP_{b'} + \varepsilon_{b'} \text{ for any } b' \neq b\right) \end{aligned}$$

If we assume that ε_{bt} follows a Gumbel $(0,\, \delta)$ distribution and we integrate over ε_{bt}, the choice probability for bundle b is

$$Pr(b \mid \psi, \delta) = \frac{\exp\left\{\frac{\ln\psi_b - \ln BP_b}{\delta}\right)}{\sum_{b'}\exp\left\{\frac{\ln\psi_{b'} - \ln BP_{b'}}{\delta}\right\}} = \frac{\exp\left\{\frac{1}{\delta}\ln\psi_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\ln\psi_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \qquad (2.2)$$

The deterministic marginal utility has to be positive and show diminishing returns. Following Song and Chintagunta (2006), we assume that the deterministic marginal utility for bundle b is the exponential of the sum of the utilities of products in the bundle and their interactions. That is, $\ln\psi_b = \sum_j \alpha_j z_{jb} + \sum_j \sum_{j'} \omega_{jj'} z_{jb} z_{j'b} = \boldsymbol{\alpha}^T \mathbf{z}_b + \frac{1}{2}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b$, where α_j is individual i's preference for product j and $\omega_{jj'}$ captures the interdependency between products j and j'. Notice that the diagonals of the product interdependency matrix $\boldsymbol{\Omega}$ are zeros. Then Equation (2.2), the choice probability for bundle b, can be expressed as

$$\begin{aligned} Pr(b \mid \boldsymbol{\alpha}) &= \frac{\exp\left\{\frac{1}{\delta}\ln\psi_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\ln\psi_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \qquad (2.3) \\ &= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_b + \frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \end{aligned}$$

Without loss of generality, we set $\sum_b \boldsymbol{\alpha}^T \mathbf{z}_b = 0$ and mean center the log of bundle price $\ln BP_b^* = \ln BP_b - \frac{1}{B}\sum_{b'} \ln BP_{b'}$. Thus,

$$\sum_b \boldsymbol{\alpha}^T \mathbf{z}_b = \sum_b \sum_j \alpha_j z_{jb} = \sum_j \sum_b \alpha_j z_{jb} = \sum_j \alpha_j \sum_b z_{jb} = 0 \tag{2.4}$$

Because $\sum_b z_{jb} = 2^{J-1}$, it follows that $\sum_j \alpha_j = 0$.

2.3.2 Aggregate Model

Individuals have different preferences for the individual products and the resulting bundles (recall this latter assumption is the second defining characteristic of a bundle). We believe that bundles are formed for a specific application, and consumers have a consensus on the products needed for this application. Thus we assume that while participants have different preferences for products, they generally agree on the interactions between products. Therefore we only need to consider the distribution of $\boldsymbol{\alpha}$. The heterogeneity in preferences can be captured by assuming a multivariate probability distribution for $\boldsymbol{\alpha}$. Suppose the preference distribution is $g(\boldsymbol{\alpha})$. The aggregate choice probability of bundle b is a Mixed Logit model (Train, 2003):

$$Pr(b) = \int Pr(b \mid \boldsymbol{\alpha}) g(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \tag{2.5}$$

What should the preference distribution $g(\boldsymbol{\alpha})$ be? In other words, what is the preference distribution of those consumers who would buy bundles?

Let $f(\boldsymbol{\alpha})$ be the preference distribution for the population and assumed it follows a multivariate normal (MVN) distribution $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This assumption on population preference distribution implies that some consumers place high values on all of the products that might be bundled while some place extremely low values on some or all of these same products. However, we do not observe choices from consumers who are not interested in this product category. That is, consumers are screened for their interest in the product category. Hence $g(\boldsymbol{\alpha})$ should reflect the preference distribution for the screened sample (we call this group "heavy users").

2.3.3 Screening Process

To approximate the heavy user distribution, we define the relationship between the preference distributions for heavy users and the population as

$$\frac{g(\boldsymbol{\alpha})}{f(\boldsymbol{\alpha})} = w(\boldsymbol{\alpha}) \tag{2.6}$$

Here $w(\boldsymbol{\alpha})$ represents the screening process. The term $g(\boldsymbol{\alpha})$, the preference distribution for heavy users, may be written as $g(\boldsymbol{\alpha}) = w(\boldsymbol{\alpha})f(\boldsymbol{\alpha})$. We call this the screened multivariate normal (S-MVN) distribution. Furthermore, we define

$$w(\boldsymbol{\alpha}) = k \exp\{IV(\boldsymbol{\alpha})\} \tag{2.7}$$

where $k > 0$ is a scaling constant and $IV(\boldsymbol{\alpha})$ represents the inclusive value of the product category for individual i:

$$IV(\boldsymbol{\alpha}) = \log \sum_{b'} \exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}^*\right\} \tag{2.8}$$

The motivation for this specification of $w(\boldsymbol{\alpha})$ comes from the screening process. First, as explained earlier, heavy users are a subset of the population and this segment places higher value on the product category. We further assert that such consumers have more interest in buying a bundle. This suggests that the relationship between heavy user and population preference distributions is related to how attractive the product category is to the heavy users relative to the population. In the choice modeling literature, the attractiveness of a product category for individual i can be represented by an inclusive value function, the expected maximum utility of the alternatives in the set (Ben-Akiva and Lerman, 1985). Thus, we propose that $w(\boldsymbol{\alpha})$ is monotonically related to the inclusive value. Second, because both $g(\boldsymbol{\alpha})$ and $f(\boldsymbol{\alpha})$ are non-negative, $w(\boldsymbol{\alpha})$ has to be non-negative. Since inclusive value ranges from $-\infty$ to ∞, we define $w(\boldsymbol{\alpha})$ to be proportional to $\exp\{IV(\boldsymbol{\alpha})\}$. Lastly, the normalizing constant r ensures that the properties of the probability distribution $g(\boldsymbol{\alpha})$ are preserved under the specification.

2.3.4 S-MVN Distribution

From equations (2.6) and (2.7) we can obtain $g(\boldsymbol{\alpha})$:

$$g(\boldsymbol{\alpha}) = w(\boldsymbol{\alpha})f(\boldsymbol{\alpha}) = k\exp\{IV(\boldsymbol{\alpha})\}f(\boldsymbol{\alpha}) \tag{2.9}$$

Because $\int g(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \int k\exp\{IV(\boldsymbol{\alpha})\}f(\boldsymbol{\alpha})d\boldsymbol{\alpha} = 1$, the normalizing constant k for $w(\boldsymbol{\alpha})$ is

$$\begin{aligned}
k^{-1} &= \int \exp\{IV(\boldsymbol{\alpha})\}f(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&= \int \left(\sum_{b'} \exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}\right) f(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&= \sum_{b'} \int \exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} f(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&= \sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} \int \exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_{b'}\right\} f(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&= \sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}}\left(\frac{1}{\delta}\mathbf{z}_{b'}\right)
\end{aligned}$$

Thus we obtain the screening process $w(\boldsymbol{\alpha})$:

$$w(\boldsymbol{\alpha}) = k\exp\{IV(\boldsymbol{\alpha})\}$$

$$= \frac{\exp\{IV(\boldsymbol{\alpha})\}}{\sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}}\left(\frac{1}{\delta}\mathbf{z}_{b'}\right)}$$

where $M_{\boldsymbol{\alpha}}\left(\frac{1}{\delta}\mathbf{z}_{b'}\right) = \exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T \mathbf{z}_{b'} + \frac{1}{2\delta^2}\mathbf{z}_{b'}^T \boldsymbol{\Sigma} \mathbf{z}_{b'}\right\}$ is the moment generating function of $\boldsymbol{\alpha}$.

The S-MVN heavy user preference distribution is:

$$\begin{aligned} g(\boldsymbol{\alpha}) &= w(\boldsymbol{\alpha}) f(\boldsymbol{\alpha}) \\ &= \frac{\exp\{IV(\boldsymbol{\alpha})\} f(\boldsymbol{\alpha})}{\sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}}\left(\frac{1}{\delta}\mathbf{z}_{b'}\right)} \end{aligned} \tag{2.10}$$

Equation (2.10) shows that the heavy user preference distribution depends on the design of choice task. Moreover, when there is no preference heterogeneity, $\boldsymbol{\Sigma} = \mathbf{0}$, $\boldsymbol{\alpha} = \boldsymbol{\mu}$ and $w(\boldsymbol{\alpha}) = 1$. In such a circumstance the heavy user preference distribution and the population preference distribution are the same, $g(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha})$. If preference heterogeneity exists and $w(\boldsymbol{\alpha}) > 1$, individuals who have a higher expected maximum utility are more likely to be retained in the heavy user preference distribution. Similarly, $w(\boldsymbol{\alpha}) < 1$ means that individuals who have a lower expected maximum utility are more likely to be screened out.

2.3.5 MVL Choice Model

We can now model the aggregate choice probability as:

$$\begin{aligned} Pr(b) &= \int Pr(b \mid \boldsymbol{\alpha}) g(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \\ &= \int \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T \mathbf{z}_b + \frac{1}{2\delta}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b - \frac{1}{\delta}\ln BP_b\right\}}{\exp\{IV(\boldsymbol{\alpha})\}} \\ &\qquad \times \frac{\exp\{IV(\boldsymbol{\alpha})\}}{\sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}}(\frac{1}{\delta}\mathbf{z}_{b'})} f(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \\ &= \frac{\exp\left\{\frac{1}{2\delta}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}}(\frac{1}{\delta}\mathbf{z}_{b'})} \int \exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T \mathbf{z}_b\right\} f(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \\ &= \frac{\exp\left\{\frac{1}{2\delta}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b - \frac{1}{\delta}\ln BP_b\right\} M_{\boldsymbol{\alpha}}(\frac{1}{\delta}\mathbf{z}_b)}{\sum_{b'} \exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}}(\frac{1}{\delta}\mathbf{z}_b)} \end{aligned} \tag{2.11}$$

$$=\frac{\exp\left\{\frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\boldsymbol{\Sigma}\mathbf{z}_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_b\right\}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\boldsymbol{\Sigma}\mathbf{z}_{b'}\right\}}$$
$$=\frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\left(\delta\boldsymbol{\Omega}+\boldsymbol{\Sigma}\right)\mathbf{z}_b-\frac{1}{\delta}\ln BP_b^*\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\left(\delta\boldsymbol{\Omega}+\boldsymbol{\Sigma}\right)\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}^*\right\}}$$

Notice that equation (2.11) is a Mixed Logit model (Train 2003), which typically does not have a closed form expression. Remarkably, with the assumption of the S-MVN heavy user preference distribution, we are able to derive a closed form solution for the Mixed Logit model.

Let $\boldsymbol{\sigma}=\left[\sigma_1^2,\ldots,\sigma_J^2\right]^T$ and $\boldsymbol{\Sigma}_D=diag(\boldsymbol{\sigma})$. Since $z_{jb}=0$ or 1, $z_{jb}^2=z_{jb}$, $\mathbf{z}_b^T\boldsymbol{\Sigma}_D\mathbf{z}_b=\boldsymbol{\sigma}^T\mathbf{z}_b$. We can rearrange the terms in equation (2.11):

$$Pr(b)=\frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\left(\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_D+\boldsymbol{\Sigma}_D\right)\mathbf{z}_b-\frac{1}{\delta}\ln BP_{b'}\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_{b'}+\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\left(\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_D+\boldsymbol{\Sigma}_D\right)\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}} \tag{2.12}$$
$$=\frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b\frac{1}{2\delta^2}\mathbf{z}_b^T\boldsymbol{\Sigma}^\dagger\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\boldsymbol{\Sigma}_D\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_{b'}+\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\boldsymbol{\Sigma}^\dagger\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\boldsymbol{\Sigma}_D\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}}$$
$$=\frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T(\boldsymbol{\Omega}+\boldsymbol{\Sigma}^\dagger)\mathbf{z}_b+\frac{1}{2\delta^2}\boldsymbol{\sigma}^T\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T(\boldsymbol{\Omega}+\boldsymbol{\Sigma}^\dagger)\mathbf{z}_{b'}+\frac{1}{2\delta^2}\boldsymbol{\sigma}^T\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}}$$
$$=\frac{\exp\left\{\left(\frac{1}{\delta}\boldsymbol{\mu}+\frac{1}{2\delta^2}\boldsymbol{\sigma}\right)^T\mathbf{z}_b+\frac{1}{2}\mathbf{z}_b^T(\boldsymbol{\Omega}+\boldsymbol{\Sigma}^\dagger)\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\left(\frac{1}{\delta}\boldsymbol{\mu}+\frac{1}{2\delta^2}\boldsymbol{\sigma}\right)^T\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T(\boldsymbol{\Omega}+\boldsymbol{\Sigma}^\dagger)\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}}$$
$$=\frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{\dagger T}\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\left(\delta\boldsymbol{\Omega}+\boldsymbol{\Sigma}^\dagger\right)\mathbf{z}_b-\frac{1}{\delta}\ln BP_b^*\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{\dagger T}\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\left(\delta\boldsymbol{\Omega}+\boldsymbol{\Sigma}^\dagger\right)\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}^*\right\}} \tag{2.13}$$

where $\boldsymbol{\mu}^\dagger=\boldsymbol{\mu}+\frac{1}{2\delta^2}\boldsymbol{\sigma}$ and $\boldsymbol{\Sigma}^\dagger=\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_D$. The resulting aggregate choice probability takes the form of a MVL model (Russell and Petersen 2000; Kwak, Duvvuri, and Russell 2015).

Kwak, Duvvuri, and Russell (2015) discuss the properties of the MVL model from the perspective of marketing science literature. In our application, equation (2.12) specifies four major aspects that affect the aggregate choice probability for a given bundle b: 1) the overall preferences for products $\boldsymbol{\mu}$, 2) the product interdependencies $\boldsymbol{\Omega}$, 3) the preference correlations between products $\mathbf{\Sigma}$ and 4) the discounted bundle price.

We make several remarks regarding this result. First, when consumer preferences are homogeneous, $\mathbf{\Sigma} = \mathbf{0}$ and $\boldsymbol{\alpha} = \boldsymbol{\mu} = \boldsymbol{\mu}^{\dagger}$. In this case equation (2.12) simplifies to equation (2.3). Second, if all products are independent of each other, $\mathbf{\Omega} = \mathbf{0}$. Thus the bundle choice probability is influenced only by preference heterogeneity across consumers, in addition to bundle price. Third, if $\delta \to 0$, decision makers are becoming more deterministic in the decision making process. Consequently the preference heterogeneity has a greater influence on the aggregate choice probability. Lastly, the model specification allows us to recover population preference parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. This allows the researchers to infer the properties of the population level MVN distribution, if desired.

2.3.6 Lack of Identification for $\mathbf{\Omega}$

We emphasize that while both $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are interaction matrices, they represent different sources of interactions. The $\mathbf{\Omega}$ matrix represents the interdependencies between products, i.e., substitutability or complementarity. Hence $\mathbf{\Omega}$ is a *within individual* measure. The $\mathbf{\Sigma}$ matrix, on the other hand, is a *between individual* measure because it captures the product preference heterogeneity and preference correlation across consumers. In our empirical configuration study, we do not have repeated measures for each individual. Thus we are not able to identify $\mathbf{\Omega}$. For this study, without loss of generality we set $\mathbf{\Omega} = \mathbf{0}$. Note, however, that the aggregate choice model in equation (2.12) is a MVL model, regardless of the value of $\mathbf{\Omega}$.

Table 2.2: Summary Statistics of Simulated Preference Distributions

	Distribution	Min.	Median	Mean	Max.
$IV(\boldsymbol{\alpha})$	$f(\boldsymbol{\alpha})$	2.08	5.20	5.36	13.70
	$g(\boldsymbol{\alpha})$	2.18	6.66	6.70	13.70
α_1	$f(\alpha_1)$	-4.27	2.93	2.95	9.28
	$g(\alpha_1)$	-4.27	4.19	4.19	9.28
α_2	$f(\alpha_2)$	-0.86	1.97	1.98	5.40
	$g(\alpha_2)$	-0.60	2.33	2.32	5.40
α_3	$f(\alpha_3)$	-13.68	-4.95	-4.93	3.96
	$g(\alpha_3)$	-13.68	-6.49	-6.50	-3.96

$f(\boldsymbol{\alpha})$ = MVN distribution. $g(\boldsymbol{\alpha})$ = S-MVN distribution.

2.3.7 Heavy Users vs. General Population

To illustrate the effects of respondent screening on the parameters of the MVL model, we simulated the preference distribution for a population that included an identified "heavy user" segment. For this purpose, we focus on the product preference parameters $\boldsymbol{\alpha}$ using $J = 3$ products as our example. To simplify the presentation of results, we ignore product prices and other bundle attributes (e.g. quantity discount). Consistent with the restriction $\sum_j \alpha_j = 0$, the true parameters in the simulation dataset also sum to zero. Lastly, we assume $\delta = 1$.

The true parameters are

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \sim MVN\left(\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 4.0 & 0.3 & -4.3 \\ 0.3 & 1.0 & -1.3 \\ -4.3 & -1.3 & 5.6 \end{bmatrix}\right)$$

The summary statistics of the heavy user and population preference distributions are shown in Table 2.2.

As stated earlier, the screening process retains individuals who place a higher value on the products in the category. From the simulation study, we can see that the proposed function $w(\boldsymbol{\alpha})$ indeed reflects the relationship between the product category attractiveness and the product preferences of heavy users as well as the population. Figure 2.1 shows the inclusive value distributions for both heavy users and the population. The preference distributions are shown in Figures 2.2. Those who do not think the product category is attractive are more likely to be screened out. Table 2.2 shows that the means of $g(\alpha_1)$ and $g(\alpha_2)$ are higher than $f(\alpha_1)$ and $f(\alpha_2)$, respectively. The mean of $g(\alpha_3)$ is also more extreme than $f(\alpha_3)$ due to the zero sum constraint on preference. In short, the heavy user preference distributions are more skewed than the population preference distributions. To conclude, the simulation results presented here suggest that our proposed approximation to the screening process is reasonable.

2.4 Model Calibration

The aggregate model as shown in equation (2.11) has some restrictions and challenges. The restriction $\sum_j \alpha_j = 0$ imposes restrictions on the parameters. Moreover, it is clear to see from equation (2.12) that the variances are confounded with the means. We also need to address the major challenge that our configuration data presents - one observation per participant and the "curse of dimensionality." The dimensionality problem is twofold here: the number of parameters to be estimated and the size of the choice set in the denominator. We discuss each of those issues subsequently.

2.4.1 Parameter Restrictions

Because $\sum_j \alpha_j = 0$, without loss of generality we set $\alpha_J = -\sum_{j=1}^{J-1} \alpha_j = -\mathbf{1}_{J-1}^T \boldsymbol{\alpha}^*$ where $\boldsymbol{\alpha}^* = \left[\alpha_1, \ldots, \alpha_{J-1}\right]^T$ and $\mathbf{1}_l$ is a vector of length l with 1 as every element. Define $\mathbf{T} = \left[\mathbf{I}_{J-1\times J-1} \quad -\mathbf{1}_{J-1}\right]^T$. Thus we have $\boldsymbol{\alpha} = \left[\boldsymbol{\alpha}^* \quad \alpha_J\right]^T = \mathbf{T}\boldsymbol{\alpha}^*$. Define $\boldsymbol{\mu}^* =$

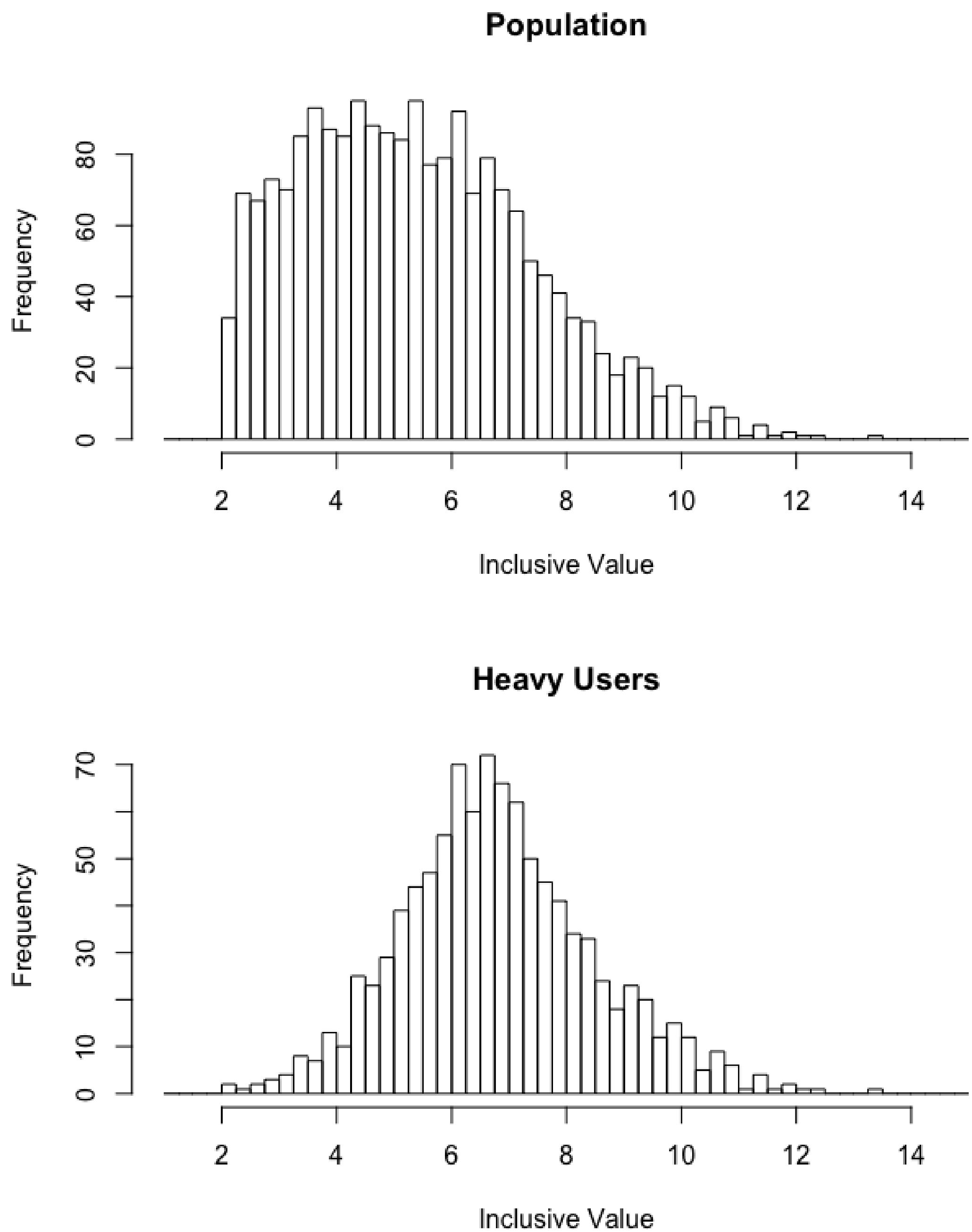


Figure 2.1: Weighted Sample and Population Inclusive Value Distributions

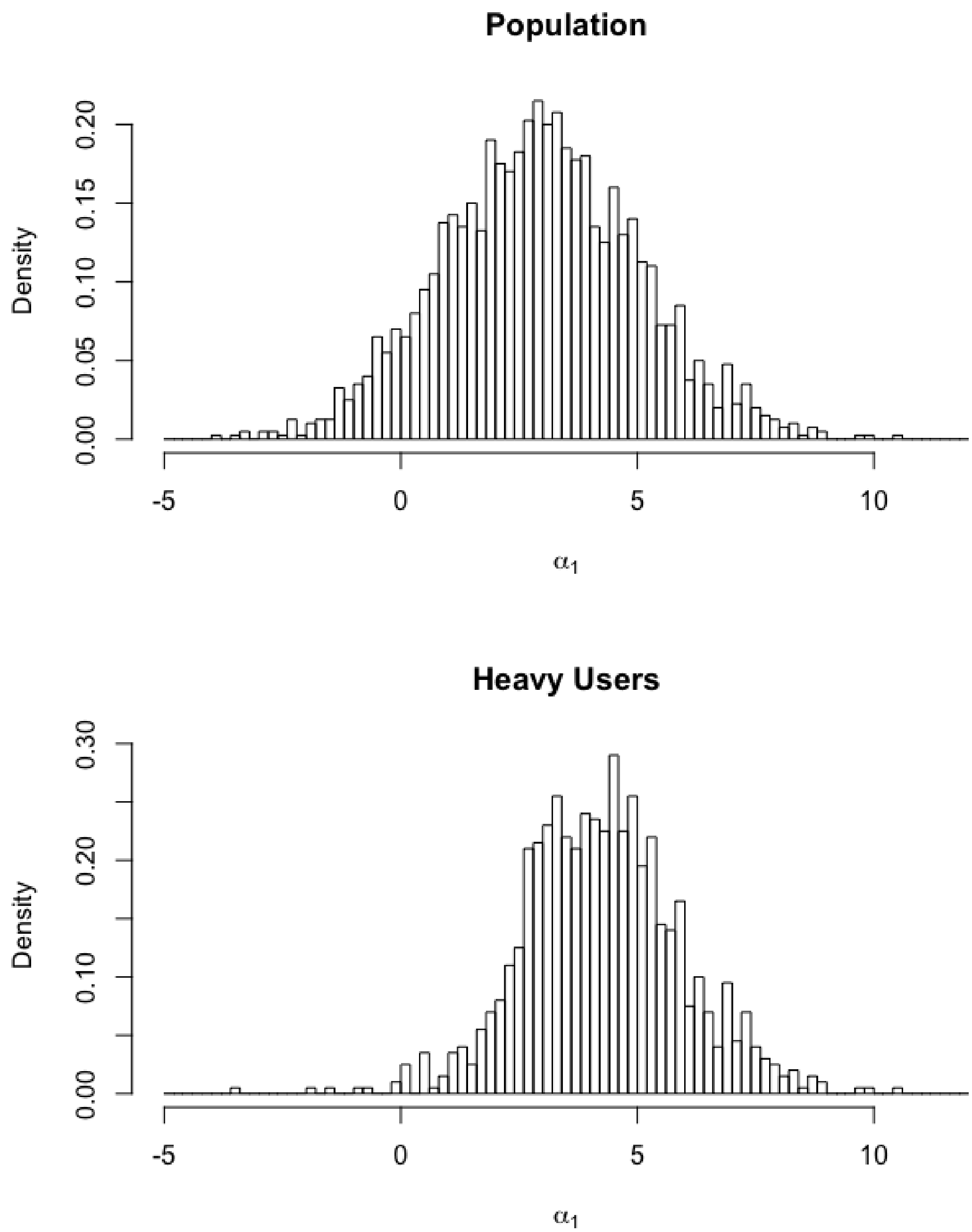
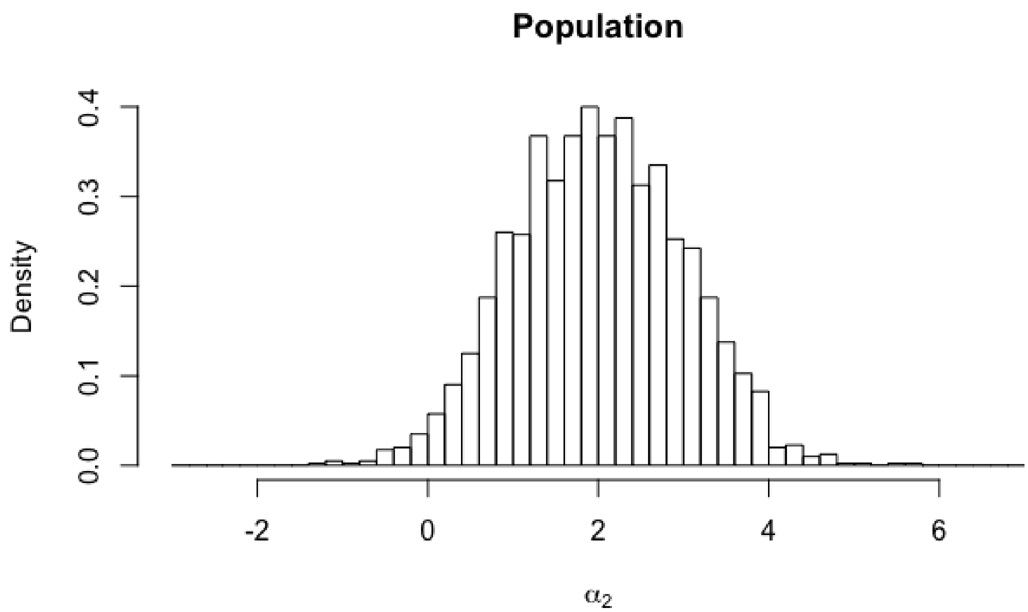


Figure 2.2: Weighted Sample and Population Preference Distributions - α_1



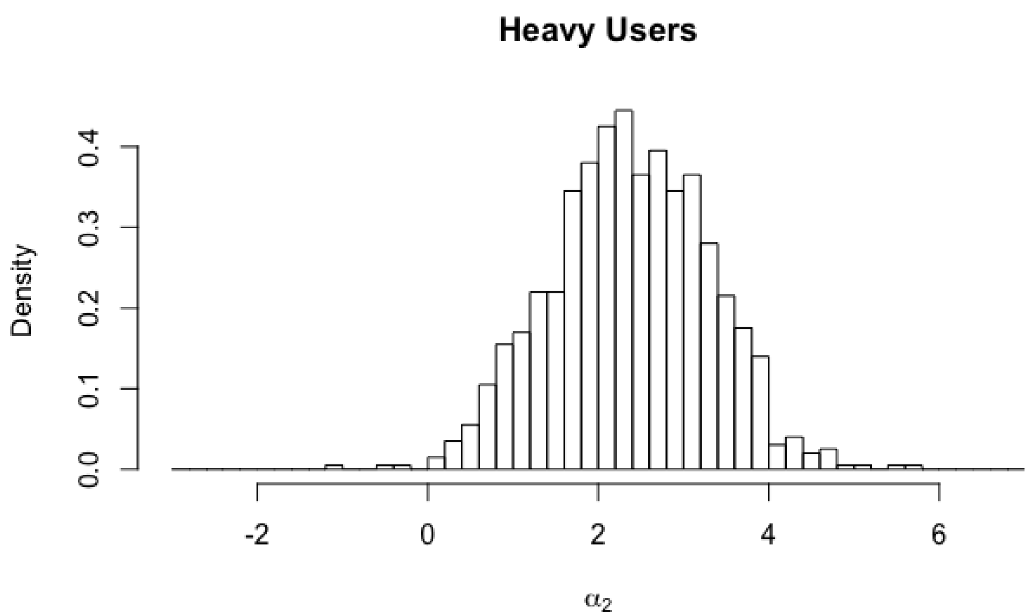


Figure 2.3: Weighted Sample and Population Preference Distributions - α_2

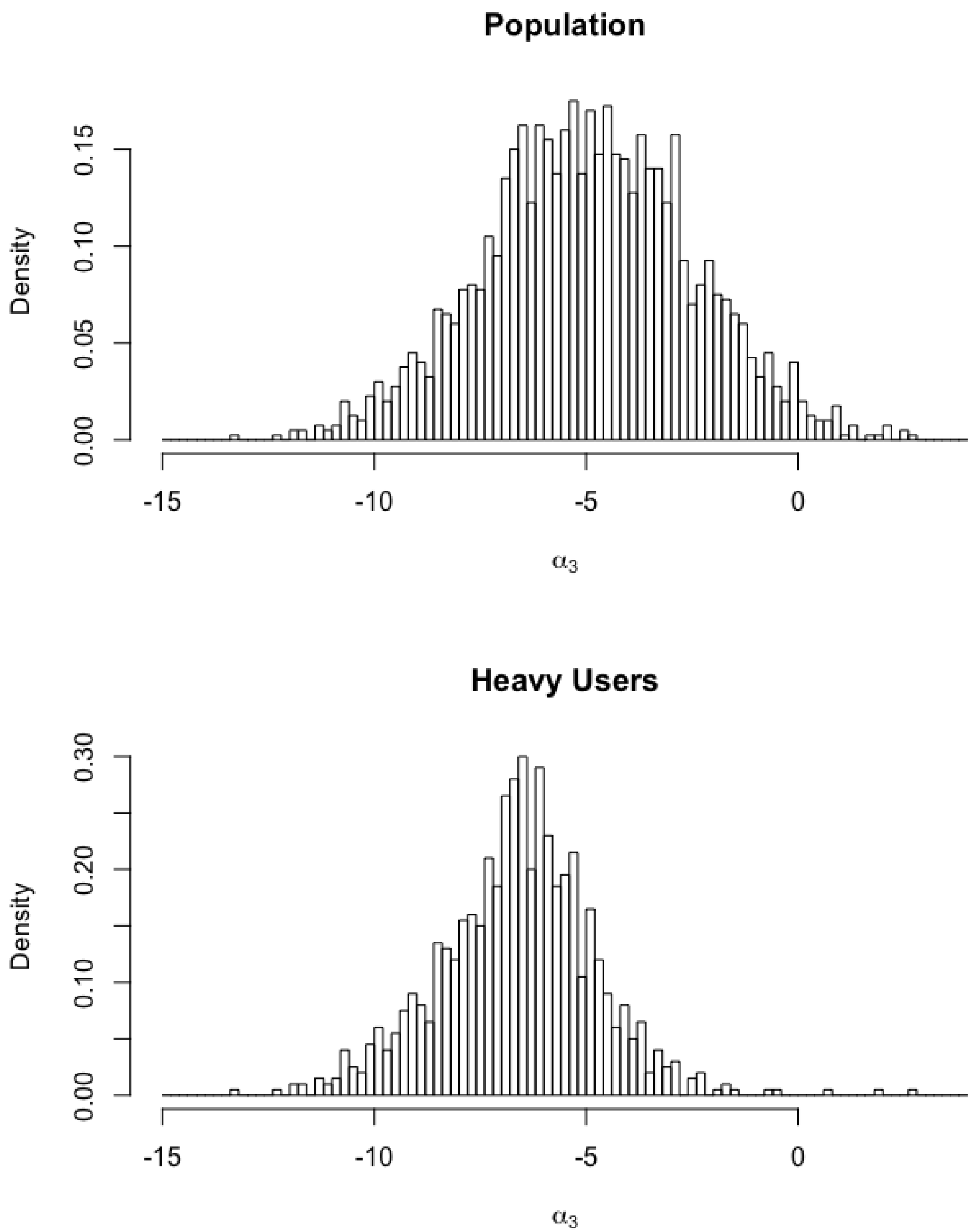


Figure 2.4: Weighted Sample and Population Preference Distributions - α_3

$\begin{bmatrix}\mu_1 & \dots & \mu_{J-1}\end{bmatrix}^T$ and $\mathbf{\Sigma}^*$ as the first $J-1$ rows and columns of $\mathbf{\Sigma}$. The distribution for $\boldsymbol{\alpha}$ can be written as

$$\boldsymbol{\alpha} = \mathbf{T}\boldsymbol{\alpha}^* \sim MVN\left(\boldsymbol{\mu} = \mathbf{T}\boldsymbol{\mu}^* = \begin{bmatrix}\boldsymbol{\mu}^* \\ \mu_J\end{bmatrix}, \mathbf{\Sigma} = \mathbf{T}\mathbf{\Sigma}^*\mathbf{T}^T = \begin{bmatrix}\mathbf{\Sigma}^* & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}^T & \sigma_J^2\end{bmatrix}\right) \tag{2.14}$$

where $\mu_J = -\mathbf{1}_{J-1}^T\boldsymbol{\mu}^*$, $\boldsymbol{\sigma} = \begin{bmatrix}\sigma_{1J}, \dots, \sigma_{J-1,J}\end{bmatrix}^T = -\mathbf{\Sigma}^*\mathbf{1}_{J-1}$ and $\sigma_J^2 = \mathbf{1}_{J-1}^T\mathbf{\Sigma}^*\mathbf{1}_{J-1}$. Thus equation (2.11) can be rewritten as

$$\begin{aligned} Pr(b) &= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\mathbf{T}^T\mathbf{z}_b + \frac{1}{2\delta^2}\mathbf{z}_b^T\mathbf{T}\mathbf{\Sigma}^*\mathbf{T}^T\mathbf{z}_b - \frac{1}{\delta}\ln BP_{b'}\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\mathbf{T}^T\mathbf{z}_{b'} + \frac{1}{2\delta^2}\mathbf{z}_{b'}^T\mathbf{T}\mathbf{\Sigma}^*\mathbf{T}^T\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \\ &= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_b^* + \frac{1}{2\delta^2}\tilde{\mathbf{z}}_b^{*T}\mathbf{\Sigma}^*\tilde{\mathbf{z}}_b^* - \frac{1}{\delta}\ln BP_{b'}^*\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{1}{2\delta^2}\tilde{\mathbf{z}}_{b'}^{*T}\mathbf{\Sigma}^*\tilde{\mathbf{z}}_{b'}^* - \frac{1}{\delta}\ln BP_{b'}^*\right\}} \end{aligned} \tag{2.15}$$

where $\tilde{\mathbf{z}}_b^* = \mathbf{T}^T\mathbf{z}_b = \begin{bmatrix}\tilde{z}_{1b} & \dots & \tilde{z}_{J-1,b}\end{bmatrix}^T$ and $\tilde{z}_{jb} = z_{jb} - z_{Jb}$.

2.4.2 Assumptions for Scaling Parameters

There are two scaling parameters in Equation (2.15): the scaling parameter δ from the individual logit model and the variance matrix $\mathbf{\Sigma}^*$ from the heterogeneity distribution. As shown in equation (2.12), the variances are confounded with means; we are unable to distinguish the means and the variances. For the equation to be identifiable, we assume the variances for all products are constant: $\sigma_1^2 = \cdots = \sigma_{J-1}^2 = \sigma^2$ and rewrite equation (2.15) as follows:

$$\begin{aligned} Pr(b) &= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_b^* + \frac{1}{2\delta^2}\tilde{\mathbf{z}}_b^{*T}\mathbf{\Sigma}^*\tilde{\mathbf{z}}_b^* - \frac{1}{\delta}\ln BP_{b'}\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{1}{2\delta^2}\tilde{\mathbf{z}}_{b'}^{*T}\mathbf{\Sigma}^*\tilde{\mathbf{z}}_{b'}^* - \frac{1}{\delta}\ln BP_{b'}\right\}} \\ &= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_b^* + \frac{\sigma^2}{2\delta^2}\tilde{\mathbf{z}}_b^{*T}\mathbf{\Theta}^*\tilde{\mathbf{z}}_b^* - \frac{1}{\delta}\ln BP_{b'}\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{\sigma^2}{2\delta^2}\tilde{\mathbf{z}}_{b'}^{*T}\mathbf{\Theta}^*\tilde{\mathbf{z}}_{b'}^* - \frac{1}{\delta}\ln BP_{b'}\right\}} \\ &= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_b^* + \frac{\sigma^2}{2\delta^2}\tilde{\mathbf{z}}_b^{*T}\mathbf{\Theta}^*\tilde{\mathbf{z}}_b^* - \beta\ln BP_{b'}^*\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{\sigma^2}{2\delta^2}\tilde{\mathbf{z}}_{b'}^{*T}\mathbf{\Theta}^*\tilde{\mathbf{z}}_{b'}^* - \beta\ln BP_{b'}^*\right\}} \end{aligned} \tag{2.16}$$

where $\mathbf{\Theta}^*$ is a $J-1 \times J-1$ correlation matrix, and $\beta = \frac{1}{\delta} > 0$.

2.4.3 Number of Parameters to be Estimated

Applying equation (2.16) to a dataset with J products we need to estimate $\frac{J(J+1)}{2}+1$ parameters. This clearly imposes some estimation difficulty. To reduce the number of parameters, we project the correlation matrix $\mathbf{\Theta}^*$ onto a two-dimensional space. Let

$$\mathbf{V}^* = \begin{bmatrix} \vec{v}_1 \\ \vdots \\ \vec{v}_{J-1} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ \vdots & \vdots \\ v_{J-1,1} & v_{J-1,2} \end{bmatrix}$$

Define $\mathbf{\Theta}^* = \mathbf{V}^*\mathbf{V}^{*T}$, where $\vec{v}_j\vec{v}_j^T = 1$ and $|\vec{v}_j\vec{v}_k^T| \leq 1$ for $j \neq k$. We define $v_{j1} = \cos(\theta_j)$, $v_{j2} = \sin(\theta_j)$ because trigonometric functions have properties that are consistent with the restrictions on $\vec{v}_j$. However, there are infinitely many solutions for θ_j that satisfy the restrictions. To make sure that we have only one unique solution, we further define $\theta_j = \frac{2\pi}{e^{a_j}+1}$, where a_j is the parameter to estimate on the real line. Thus θ_j is always between 0 and 2π. Next, to fix the scale and the direction of the map, we set $v_{11} = 1$ and $v_{12} = 0$. Lastly, to prevent the axes to be flipped, we set $v_{21} = \cos(\frac{\pi}{e^{a_2}+1})$.

Replacing $\mathbf{\Theta}^*$ with $\mathbf{V}^*\mathbf{V}^{*T}$, equation (2.16) becomes

$$\begin{aligned} & Pr(b) \\ =& \frac{\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_b^* + \frac{k^2}{2}\tilde{\mathbf{z}}_b^{*T}\mathbf{\Theta}^*\tilde{\mathbf{z}}_b^* - \beta\ln BP_{b'}\right\}}{\sum_{b'}\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{k^2}{2}\tilde{\mathbf{z}}_{b'}^{*T}\mathbf{\Theta}^*\tilde{\mathbf{z}}_{b'}^* - \beta\ln BP_{b'}\right\}} \\ =& \frac{\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_b^* + \frac{k^2}{2}\tilde{\mathbf{z}}_b^{*T}\mathbf{V}^*\mathbf{V}^{*T}\tilde{\mathbf{z}}_b^* - \beta\ln BP_{b'}\right\}}{\sum_{b'}\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{k^2}{2}\tilde{\mathbf{z}}_{b'}^{*T}\mathbf{V}^*\mathbf{V}^{*T}\tilde{\mathbf{z}}_{b'}^* - \beta\ln BP_{b'}\right\}} \\ =& \frac{\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_b^* + \frac{k^2}{2}\left(\sum_j^{J-1}\tilde{z}_{jb}^2\vec{v}_j\vec{v}_j^T + 2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb}\tilde{z}_{kb}\vec{v}_j\vec{v}_k^T\right) - \beta\ln BP_b\right\}}{\sum_{b'}\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{k^2}{2}\left(\sum_j^{J-1}\tilde{z}_{jb'}^2\vec{v}_j\vec{v}_j^T + 2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb'}\tilde{z}_{kb'}\vec{v}_j\vec{v}_k^T\right) - \beta\ln BP_{b'}\right\}} \\ =& \frac{\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_b^* + \frac{k^2}{2}\sum_j^{J-1}\tilde{z}_{jb}^2 + k^2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb}\tilde{z}_{kb}\vec{v}_j\vec{v}_k^T - \beta\ln BP_b^*\right\}}{\sum_{b'}\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{k^2}{2}\sum_j^{J-1}\tilde{z}_{jb'}^2 + k^2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb'}\tilde{z}_{kb'}\vec{v}_j\vec{v}_k^T - \beta\ln BP_{b'}^*\right\}} \end{aligned} \tag{2.17}$$

2.4.4 Size of Choice Set

In the configuration task, the number of the possible bundles is 2^J. As the size of the menu grows, the number of possible bundles grows geometrically. To address this problem, researchers in the transportation science literature have developed a solution known as "sampling of alternatives." This approach, developed for MNL model, replaces the full choice set with a subset of all alternatives in the denominator (McFadden, 1978). Doing so still provides consistent estimates, except for the intercept. Furthermore, the intercept bias can be corrected by adding an additional term to the model. Since the MVL model can be characterized as a simple logit model defined in terms of bundles and has the independence of irrelevant alternatives (IIA) property (Kwak et al., 2015), sampling of alternatives is an appropriate option for our formulation. We explain the sampling of the alternatives approach in the following.

Let υ_b be the deterministic utility for alternative b, and $\mathcal{B}$ is the set containing all the alternatives. Based on random utility theory, the probability of choosing alternative b is defined as

$$Pr(Y_b = 1) = \frac{\exp(\upsilon_b)}{\sum_{b' \in \mathcal{B}} \exp(\upsilon_{b'})} \tag{2.18}$$

A procedure for sampling of alternatives assigns to an individual i a subset of the alternatives, denoted by $\mathcal{B}'$ which includes the chosen alternative b. Let $\pi_i(\mathcal{B}'|b)$ be the conditional probability of constructing the set $\mathcal{B}'$ for subject i, given the chosen alternative b. For ease of readability, we suppress the individual subscript i in the following discussion. The joint probability of a chosen alternative and a subset of alternatives, $\mathcal{B}'$, is

$$\pi(b, \mathcal{B}') = \pi(\mathcal{B}'|b) Pr(Y_b = 1)$$

By Bayes theorem, the conditional probability of alternative b being chosen given a subset of alternatives, $\mathcal{B}'$, is

$$\pi(b|\mathcal{B}') = \frac{\pi(\mathcal{B}'|b)Pr(Y_b = 1))}{\sum_{b' \in \mathcal{B}'} \pi(\mathcal{B}'|b')Pr(Y_{b'} = 1)} \tag{2.19}$$

Substituting the choice probabilities $Pr(Y_b = 1)$ in equation (2.19) with equation (2.18), we have

$$\pi(b|\mathcal{B}') = \frac{\exp\{\upsilon_b + \ln \pi(\mathcal{B}'|b)\}}{\sum_{b' \in \mathcal{B}'} \exp\{\upsilon_{b'} + \ln \pi(\mathcal{B}'|b')\}} \tag{2.20}$$

McFadden (1978) proves the maximization of the conditional log likelihood function of equation (2.20) yields consistent estimates of the unknown parameters under normal regularity conditions. Notice that equation (2.20) includes an additive alternative-specific correction for the bias introduced by the sampling of alternatives.

Two different sampling strategies have been proposed for alternative sampling: simple random sampling and importance sampling. We apply importance sampling due to the consideration of efficiency. The sampling with replacement procedure suggested by Ben-Akiva and Lerman (1985) is as follows: Draw a sample of size k from the set $\mathcal{B}$ with probability w_b for each bundle b at each draw. Delete the duplicate alternatives and add the chosen alternative if it was not sampled. Then the probability to obtain the set $\mathcal{B}'$ is

$$\pi(\mathcal{B}'|b) = \prod_{b \in \mathcal{B}' \& b' \neq b} q_{b'} (\sum_{b' \in \mathcal{B}'} q_{b'})^{k+1-k'} = \frac{1}{q_b} \prod_{b' \in \mathcal{B}'} q_{b'} (\sum_{b' \in \mathcal{B}'} q_{b'})^{k+1-k'} = \frac{1}{q_k} Q(\mathcal{B}') \tag{2.21}$$

where $Q(\mathcal{B}') = \prod_{b' \in \mathcal{B}'} q_{b'} (\sum_{b' \in \mathcal{B}'} q_{b'})^{k+1-k'}$. Combining equation (2.21) and equation (2.20), we have

$$\pi(b|\mathcal{B}') = \frac{\exp(u_b + \log \frac{1}{q_b} Q(B'))}{\sum_{b' \in \mathcal{B}'} \exp(u_{b'} + \log \frac{1}{q_{b'}} Q(B'))} = \frac{\exp(u_b - \ln q_b)}{\sum_{b' \in \mathcal{B}'} \exp(u_{b'} - \ln q_{b'})}$$

Based on the sampling of alternatives approach described above, equation (2.17) becomes

$$Pr_{\mathcal{B}_i}(b) \tag{2.22}$$
$$= \frac{\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_b^* + \frac{k^2}{2}\sum_j^{J-1}\tilde{z}_{jb}^2 + k^2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb}\tilde{z}_{kb}\vec{v}_j\vec{v}_k^T - \beta\ln BP_b - \ln q_b\right\}}{\sum_{b'\in\mathcal{B}_i}\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{k^2}{2}\sum_j^{J-1}\tilde{z}_{jb'}^2 + k^2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb'}\tilde{z}_{kb'}\vec{v}_j\vec{v}_k^T - \beta\ln BP_{b'} - \ln q_{b'}\right\}}$$

where q_b is the weight of bundle b and $\mathcal{B}_i$ is the sample choice set for consumer i.

We implement sampling of alternative procedure for model estimation. The procedure is as follows:

1. Exclude the baskets with size 0 and size 1 from the total set of the bundles. Denote the new set that contains 502 bundles as $\mathcal{B}$.
2. Count the frequency of each bundle being chosen by the participants.
3. Adjust the frequency of each bundle being chosen as follows:

$$freq_b = freq_b^* + \frac{1}{2}$$

   where $freq_b^*$ represents number of times the bundle being chosen by the participants in the dataset.
4. Calculate the probability of each bundle being chosen based on the data:

$$Prob_b = \frac{freq_b}{\sum_{b'\in\mathcal{B}} freq_{b'}}$$

5. For each participant i, we apply sampling of alternatives as follows:
   (a) Using sampling with replacement, we get 90 bundles from the set $\mathcal{B}$ based on the probability $Prob_b$.

(b) Eliminate the duplicate bundles. Denote the set of alternatives as $\mathcal{B}_i$.

(c) Check if the chosen bundle of subject i is in $\mathcal{B}_i$. If not, add the chosen bundle to the set $\mathcal{B}_i$.

2.4.5 Model Estimation

Because of the closed-formed expression of the aggregate choice model, we are able to use maximum likelihood estimation procedure for model estimation. The log likelihood is:

$$
\begin{aligned}
LL &= \log\left(\prod_i^n Pr_{\mathcal{B}_i}(b)\right) = \sum_i^n \log Pr_{\mathcal{B}_i}(b) \qquad (2.23)\\
&= \sum_i^n \left\{ \tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_b^* + \frac{k^2}{2}\sum_j^{J-1}\tilde{z}_{jb}^2 + k^2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb}\tilde{z}_{kb}\vec{v}_j\vec{v}_k^T - \beta\ln BP_b^* - \ln q_b \right\}\\
&\quad - \sum_i^n \log\left\{\sum_{b'\in\mathcal{B}_i}\exp\left\{\tilde{\boldsymbol{\mu}}^{*T}\tilde{\mathbf{z}}_{b'}^* + \frac{k^2}{2}\sum_j^{J-1}\tilde{z}_{jb'}^2 + k^2\sum_j^{J-2}\sum_{k>j}^{J-1}\tilde{z}_{jb'}\tilde{z}_{kb'}\vec{v}_j\vec{v}_k^T - \beta\ln BP_{b'}^* - \ln q_{b'}\right\}\right\}
\end{aligned}
$$

2.5 Empirical Application

A real-life marketing problem motivated our modeling efforts using configuration data. An unnamed power tool manufacturer wanted to know what would be the best bundle to introduce into a non-US distribution channel. They wanted a way to determine the size of the bundle as well as the exact set of tools that would maximize purchases of the bundle.

Qualitative research suggested that a bundle of power tools would appeal to two segments of customers. The first is made up of professionals who use power tools every day as part of their job and are involved in the purchase decisions for their personal tools. The second segment consists of serious do-it-yourself (DIY) consumers who are more involved in home repairs, woodworking, etc. than the average person. These segments are designated Professionals and Hobbyists respectively. The online survey screened out any potential respondents that did not fall into one of these two segments. Therefore, the data reported here comes from heavy users in the category as defined in equation (2.10).

After answering several other survey questions (including purchase intentions for individual tools), the participants entered the configuration task. Participants were told to create their ideal bundle from a menu consisting of nine tools, subject to the restriction that the bundle had to contain at least two tools.

All the tools are generic and unbranded. The individual price of each tool was different. Participants in the same group faced the same pricing structure. To enhance the incentive of constructing a bundle, each group was offered three different bundle discount levels – at bundle of size two, three, and four and above. The discount went deeper when the bundle size increases. After configuring a tool bundle, the discounted total price was shown. Participants were given the opportunity to revise their bundle. This process continued until the participant was satisfied with the bundle and the total price.

2.5.1 Data Description

A total of 301 participants, of which 150 are Professionals and 151 are Hobbyists, passed the screening questions and successfully completed the configuration task. Summary statistics are presented in Table 2.3. For confidentiality, the tools are presented in disguised form. We classified the tools into three categories based on their purpose. The first three (D1, D2 and D3) are used for drilling or driving. Four tools (C1, C2, C3 and

Table 2.3: Data Summary - Bundles

(a) Professionals

Bundle Size Class	Number of Possible Bundles	Number of Participants in Bundle Size Class	Number of Distinct Bundles Chosen by Participants
2	36	7	6
3	84	16	14
4	126	43	32
5	126	22	19
6	84	9	8
7	36	4	3
8	9	0	0
9	1	49	1
Total	502	150	83

(b) Hobbyists

Bundle Size Class	Number of Possible Bundles	Number of Participants in Bundle Size Class	Number of Distinct Bundles Chosen by Participants
2	36	5	5
3	84	34	24
4	126	50	33
5	126	21	21
6	84	9	9
7	36	3	3
8	9	0	0
9	1	29	1
Total	502	151	96

C4) are used for cutting. The remaining two tools (A1, A2) serve different purposes. Both can be viewed as accessories. Note that tools in the same category are not perfect substitutes. While the main purpose may seem very similar, there are situations that require, for example, a very specific drilling or cutting tool.

From the nine tools, respondents could have constructed one of 512 different bundles. In our data, a total of 152 bundles were assembled at least once by the participants - 83 for Professionals and 96 for Hobbyists.

Nearly one third of the Professionals and one fifth of the Hobbyists included all

Table 2.4: ANOVA Analysis

	Status			Size			Status × Size		
	Sum of Squares	F-value	p-value	Sum of Squares	F-value	p-value	Sum of Squares	F-value	p-value
D1	5.0	3.25	0.07	2.4	1.52	0.22	0.3	0.21	0.65
D2	7.0	4.70	0.03*	3.0	2.02	0.16	1.2	0.80	0.37
D3	6.9	5.43	0.02*	23.6	18.72	<0.001*	0.4	0.29	0.59
C1	5.7	4.29	0.04*	14.8	11.08	<0.001*	0.3	0.24	0.62
C2	7.5	5.97	0.02*	12.1	9.59	0.002*	0.1	0.09	0.77
C3	8.2	6.46	0.01*	11.2	8.80	0.003*	3.4	2.68	0.10
C4	0.0	0.01	0.91	10.8	7.89	0.005*	3.2	2.253	0.13
A1	3.3	2.03	0.16	13.1	7.94	0.005*	0.1	0.04	0.84
A2	5.7	3.85	0.05*	18.2	12.18	0.001*	0.5	0.336	0.56

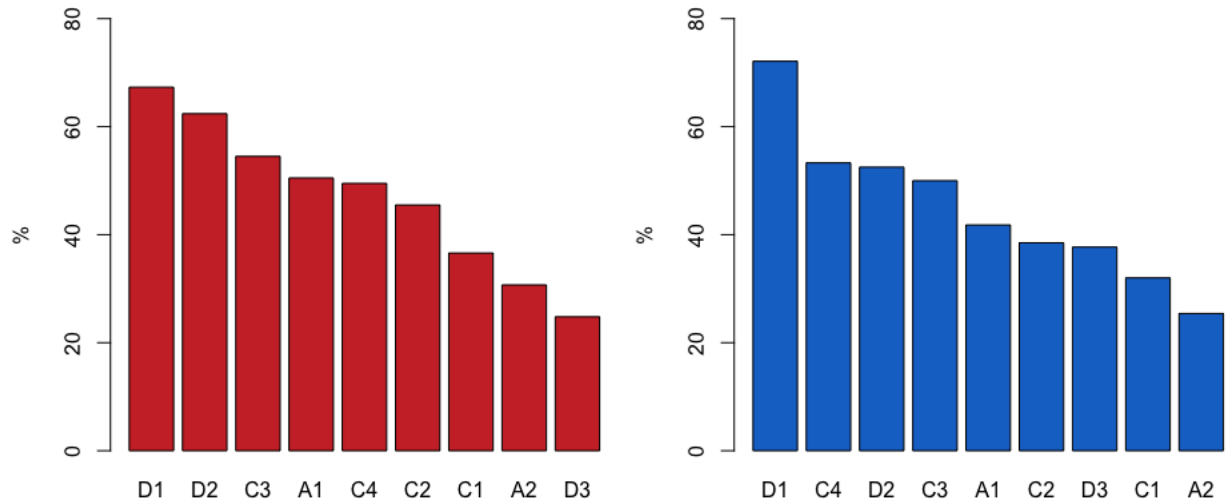
Status = Professionals/Hobbyists

Size = Bundle Size = 9/Bundle Size < 9

∗ significant at $p = 0.05$

nine tools in their ideal bundle. This is a relatively high proportion of all bundles in the sample. We suspected that participants who included the tools in their ideal bundle may have had a different mindset from the rest of the sample. To test this conjecture, we divided the participants into two groups – those who included all of the tools in their ideal bundle and those whose ideal bundle fell between two and eight tools. We performed a two-way ANOVA on the two groups' answers to the survey purchase intention question – "How likely are you to purchase" each tool at a given retail (non-discounted) price? These questions were answered on a 7-point Likert scale. Results are presented in Table 2.4.

None of the interactions is significant. The results suggest that participants who included all of the tools in their ideal bundle were more significantly likely to buy the individual products at the given retail price. This finding seems to suggest that this group has higher reservation prices for all of the products and hence are less price sensitive. Since the primary purpose of bundling is to segment consumers on the basis of price sensitivity, we excluded these participants from our data. The frequencies of the tools being chosen by



(a) Professionals (b) Hobbyists

Figure 2.5: Frequencies of Tools Being Chosen

remaining participants are shown in Figure 2.5. The distributions differ slightly between Professionals and Hobbyists. The top three tools are D1, D2 and C3 for Professionals. For Hobbyists, the top 3 tools include D1, C4 and D2.

2.5.2 Model Fit

The fit statistics of equation (2.22), including the log likelihood of the models as well as the Pearson's χ^2 test and Kolmogorov–Smirnov test results on the expected count of bundle sizes vs. the true distribution of bundle sizes are shown in Table 2.5. For both Professionals and Hobbyists we can see some lack of fit for bundles of size four, but overall the expected counts of different bundle sizes have similar patterns as the data. Moreover,

Table 2.5: Model Fit

		Professionals			Hobbyists		
		Model (LL=-490.958)		Data	Model (LL=-587.947)		Data
		Probability	Exp. Ct.	Count	Probability	Exp. Ct.	Count
Bundle Size	2	.085	9	7	.110	13	5
	3	.196	20	16	.232	28	34
	4	.292	29	43	.304	37	50
	5	.254	26	22	.226	28	21
	6	.129	13	9	.099	12	9
	7	.038	4	4	.025	3	3
	8	.006	0	0	.004	1	0
	Avg. Size		4.28	4.22		4.06	4.03
Pearson's χ^2 Test	χ^2	9.85			13.16		
	df	5			5		
	p-value	0.08			0.02		
K-S Test	D	0.08			0.09		
	p-value	0.55			0.27		

LL = log likelihood. Exp. Ct. = expected count. K-S test = KolmogorovSmirnov test.

the p-values from Kolmogorov–Smirnov test are .55 and .27 for Professionals and Hobbyists, respectively, suggesting that the model fit the data well.

2.5.3 Estimates for Preference, Price and Scaling Parameters

The parameter estimates for equation 2.22 are shown in Table 2.6. The signs of β and k are positive, as we expect. Because the inverse of β is the scale parameter of the individual utility process, we are able to recover the preference heterogeneity distribution parameters by setting $\mu_j = \tilde{\mu}_j \times \delta = \frac{\tilde{\mu}_j}{\beta}$ for $j = 1, \ldots, J-1$ and $\sigma = k \times \delta = \frac{k}{\beta}$. By construction, μ_{A2} – the preference parameter for A2 – is not estimated. It is inferred based on equation (2.14). The standard error for μ_{A2} is obtained from using the delta method. The population parameters are shown in Table 2.7.

We will first discuss the preference parameters as well as price and scaling parameters, followed by the discussion of the parameters related to the correlation matrix in the next section.

Table 2.6: Estimation Result

Parameters	Professionals Estimate	Professionals Std. Err.	Hobbyists Estimate	Hobbyists Std. Err.
$\tilde{\mu}_{D1}$	0.823 *	0.218	1.187 *	0.198
$\tilde{\mu}_{D2}$	0.633 *	0.220	0.595 *	0.199
$\tilde{\mu}_{D3}$	-0.810 *	0.225	-0.099	0.196
$\tilde{\mu}_{C1}$	-0.376	0.233	-0.447 *	0.199
$\tilde{\mu}_{C2}$	-0.313	0.213	-0.314	0.202
$\tilde{\mu}_{C3}$	0.203	0.205	0.142	0.190
$\tilde{\mu}_{C4}$	0.319	0.218	0.358	0.210
$\tilde{\mu}_{A1}$	0.204	0.224	-0.357	0.224
a_{D2}	-0.058	0.561	-1.142 *	0.515
a_{D3}	-0.722 *	0.306	0.173	0.191
a_{C1}	0.300	0.301	-0.625 *	0.238
a_{C2}	1.481 *	0.516	1.151 *	0.296
a_{C3}	2.436 *	1.185	1.392 *	0.302
a_{C4}	0.130	0.280	-2.079 *	0.808
a_{A1}	-1.341 *	0.436	-1.284 *	0.354
β	0.686 *	0.300	1.199 *	0.301
k	0.695 *	0.070	0.757 *	0.061

* Significant at $p = 0.05$

Table 2.7: MVL Model Parameter Estimates

Parameters		Professionals Estimate	Professionals Std. Err.	Hobbyists Estimate	Hobbyists Std. Err.
Overall Preference	μ_{D1}	1.200†	.619	.990*	.304
	μ_{D2}	.923†	.487	.496*	.194
	μ_{D3}	-1.181†	.627	-.083	.167
	μ_{C1}	-.548	.432	-.373†	.195
	μ_{C2}	-.456	.392	-.262	.189
	μ_{C3}	.296	.311	.465*	.160
	μ_{C4}	.465	.353	.299	.182
	μ_{A1}	.297	.382	-.298	.183
	μ_{A2}	-.996*	.452	-.888*	.248
Two Dimensional Space Parameters	a_{D2}	-.058	.561	-1.142*	.515
	a_{D3}	-.722*	.306	.173	.191
	a_{C1}	.300	.301	-.625*	.238
	a_{C2}	1.481*	.516	1.151*	.296
	a_{C3}	2.436*	1.185	1.392*	.302
	a_{C4}	.130	.280	-2.079*	.808
	a_{A1}	-1.341*	.436	-1.28*	.354
Price Coefficient	β	.686*	.300	1.199*	.301
Gumbel Distribution Parameter	δ	1.458*	.638	.834*	.209
Preference Heterogeneity	σ	1.013*	.451	.631*	.163

* Significant at $p = 0.05$

† Significant at $p = 0.10$

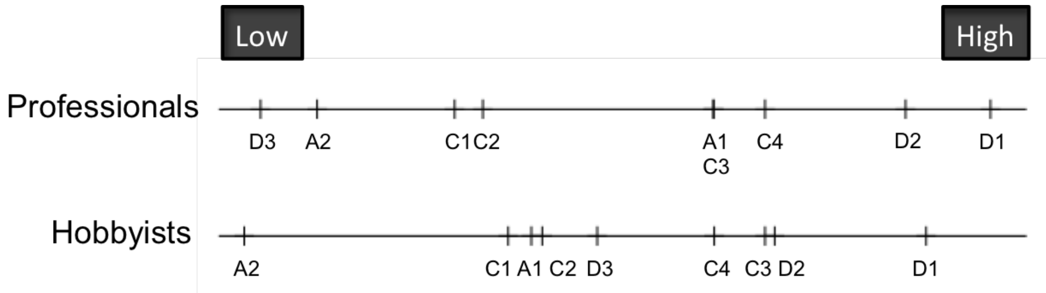


Figure 2.6: Preference Scale for Tools

We plot the preference estimation results for each tool on a one dimensional preference scale, see Figure 2.6. We can see that for both Professionals and Hobbyists, D1 and D2 are the most preferred tools, while A2 is close to the bottom of the list. Professionals and Hobbyists show different preferences for the rest of the tools.

If we were to design a bundle based on these preferences for tools, we would select the set of most preferred tools. This approach, however, ignores the interdependencies among tools and may result in an unreasonable recommendation (as we discuss in our subsequent analysis).

From Table 2.7, the price coefficient β for Professionals is smaller than Hobbyists, suggesting that Professionals are less price sensitive than are Hobbyists. On the other hand, the preference variance σ^2 for Professionals are larger than for Hobbyists, indicating that preference heterogeneity is greater for Professionals than for Hobbyists. We believe that this is a reasonable result because professionals come from different industries and these differences drive their greater preference heterogeneity.

The modeling framework developed here takes into account two sources of variation: the preference differences between choice decisions for an individual (within-individual variation), and the preference differences between individuals at each choice decision (between-individual variation). Within-individual variation can be represented by the variance of the individual's utility distribution: $Var_W = \frac{\pi^2}{6}\delta^2$. Between-individual variation is captured by the variance of the preference heterogeneity distribution: $Var_B = \sigma^2$. The ratio of the two variations is:

$$R_v = \frac{Var_B}{Var_W} = \frac{\sigma^2}{\frac{\pi^2}{6}\delta^2} = \frac{\sigma^2}{\delta^2}\frac{6}{\pi^2} = k^2\frac{6}{\pi^2} \approx 0.61k^2$$

Thus the relative magnitude of the two variations is proportional to k^2. For Professionals and Hobbyists $R_v = 0.29$ and 0.35, respectively. That is, the between-individual variation is smaller than the within-individual variation for both Professionals and Hobbyists. In addition, from Table 2.7 we can see that the preference heterogeneity (σ) is greater for Professionals than for Hobbyists.

2.5.4 Correlations in Tool Preferences

The preference correlation matrix for all tools is obtained from equations (2.4.3) and (2.14). The resulting correlation matrices, however, are not positive definite. To obtain positive definite correlation matrices, we make the following adjustment: we add 0.002 and 0.004 to the diagonal of the correlation matrices for Professionals and Hobbyists, respectively (Schott, 1997). Then, we normalize each matrix to the constraints of a correlation matrix. The consequence of this adjustment is the slight decrease in the absolute value of the correlation coefficients. The adjusted correlation matrices are shown in Tables 2.8 and 2.9. The adjusted correlation coefficients are the same as the unadjusted correlation (to the first decimal place) for Professionals. The same is true for 31 of 36 ad-

Table 2.8: Preference Correlation Estimates – Professionals

	D1	D2	D3	C1	C2	C3	C4	A1	A2
D1	1.00	-0.05	-0.46	-0.89*	0.40	0.87	-0.98*	0.26	-0.13
D2		1.00	-0.86*	0.49*	0.90*	0.44	0.25	-0.97*	-0.98*
D3			1.00	0.02	-1.00*	-0.83*	0.28	0.73*	0.93*
C1				1.00	0.06	-0.56	0.96*	-0.67*	-0.33
C2					1.00	0.79*	-0.20	-0.78*	-0.96*
C3						1.00	-0.76	-0.24	-0.59
C4							1.00	-0.45	-0.08
A1								1.00	0.91*
A2									1.00

* Significant at $p = 0.05$

Table 2.9: Preference Correlation Estimates – Hobbyists

	D1	D2	D3	C1	C2	C3	C4	A1	A2
D1	1.00	-0.72*	-0.96*	-0.58*	0.06	0.31	0.76	0.21	-0.16
D2		1.00	0.88*	-0.14	0.64*	0.42	-0.99*	-0.82*	-0.53
D3			1.00	0.34	0.21	-0.05	-0.91*	-0.46	-0.11
C1				1.00	-0.84*	-0.95*	0.08	0.67	0.83
C2					1.00	0.96*	-0.59*	-0.96*	-0.94*
C3						1.00	-0.37	-0.86*	-0.93*
C4							1.00	0.78*	0.46
A1								1.00	0.86*
A2									1.00

* Significant at $p = 0.05$

justed correlation coefficients in the Hobbyist results.

We plot the two dimensional space obtained from equations (2.4.3) and (2.14). Figure 2.7 summarizes the examples of correlation patterns. On this mapping, a vector represents a tool and the angle between two vectors represents the correlation between tools. If two tools are independent, the two vectors representing the tools will be perpendicular to each other. If the tools are positively correlated, the angle will be smaller than 90°. If the tools are negatively correlated, the angel will be greater than 90°. The plots for Professionals and Hobbyists are shown in Figure 2.8. Professionals and Hobbyists display some similar bundling behavior. For example, preferences for accessories (A1 and A2) are pos-

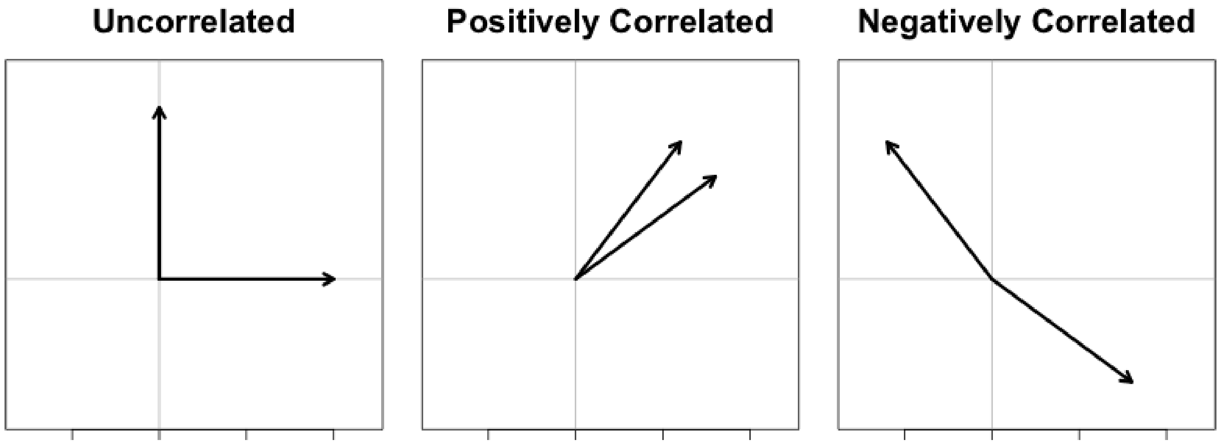
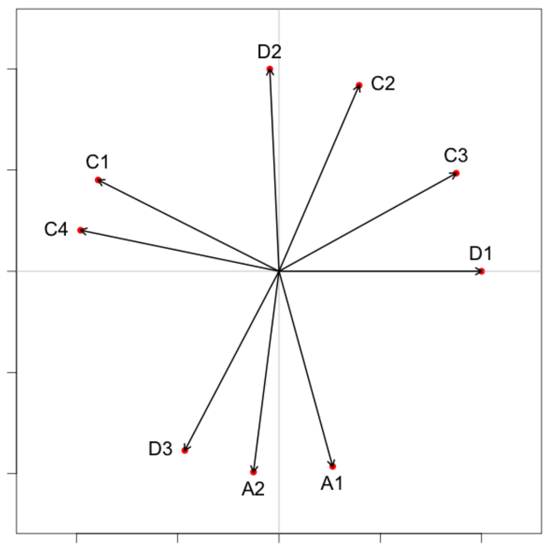


Figure 2.7: Examples of Preference Correlation Patterns

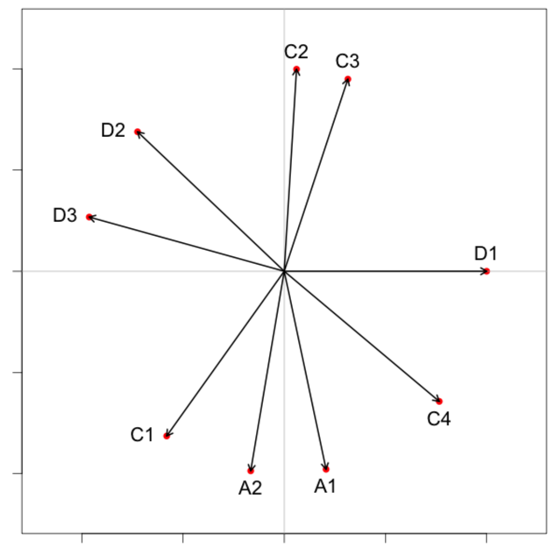
itively correlated. Within the set of cutting tools, C2 and C3 have positively correlated preferences. Similarly, respondents who prefer C1 usually prefer C4 as well. The major difference between Professionals and Hobbyists lies in the drilling/driving tools. For Professionals D2 and D3 have negatively correlated preferences. In addition, the choices of D1 and D2 are nearly independent. For Hobbyists preference for D1 is negatively correlated with preferences for D2 as well as D3. For many tasks, D1 can fulfill the functions of D2 and D3. Thus, we believe the resulting map has some face validity. In terms of cutting tools, for example, C1 and C2 as well as C1 and C3 are close to being uncorrelated for Professionals. For Hobbyists, C1 is negatively correlated with C2 and with C3.

2.5.5 Maximizing Preference Share

After the data were collected, the company decided to limit the size of the retail bundle to a set of three tools. Based on the MVL model, we simulated the choice probability for all bundles. Then, we normalized the choice probability of all bundles of size three.



(a) Professionals



(b) Hobbyists

Figure 2.8: Tool Preference Correlation Maps

Table 2.10: Top Three Bundle Recommendations

Professionals		Hobbyists	
Bundles	Probability	Bundles	Probability
(D1, D2, C3)	0.077	(D1, C4, A1)	0.094
(D1, C3, A1)	0.053	(D1, C3, C4)	0.066
(D1, D2, C2)	0.044	(D1, D2, C3)	0.055

The top three bundles and their choice probabilities are shown in Table 2.10. Interestingly, the most preferred bundle for Hobbyists contains tools from each of the categories. Because Hobbyists use limited tools to accomplish most of their tasks, a diverse tool bundle is more desirable.

Using the single dimensional preference scale from Figure 2.6 above, we show the most preferred bundles from the Professional and Hobbyist segments in Figure 2.9. It is clear from Figure 2.9 that we will provide different bundle recommendation if we do not

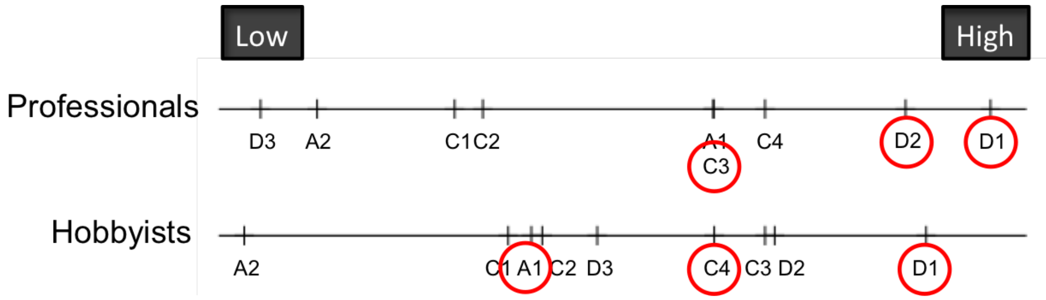


Figure 2.9: Most Preferred Bundles

take into account preference correlation. Ignoring preference correlations, we would recommend (D1, D2, C4) and (D1, D2, C3) for Professionals and Hobbyists, respectively. It is important to note that the approach is comparable to the counting method often used in practice. Notice that these particular bundles contain tools with negatively correlated preferences. For example, D1 and C4 are negatively correlated for Professionals while D1 and D2 are negatively correlated for Hobbyists. Therefore, designing a most preferred bundle must take into account the preference correlations for individual products.

2.6 Maximizing Bundle Revenue

A key objective of bundling strategy is to maximize revenue. We present our scenario and simulation method, followed by the simulation results and discussions. We conclude this section with a brief summary.

2.6.1 Scenario and Method

We consider a situation in which only a specific bundle of size three is provided to the market. Any combinations of tools in which the specific bundle is a subset will enjoy some savings due to the discount from the bundle. Any other combinations of tools that do not include this bundle will cost the consumers the sum of the tool prices. For example, suppose (D1, D2 ,C3) is offered as a bundle at discount. The total price for (D1, D2, C3, A1) will be the discounted bundle price for (D1, D2, C3) plus the price of A1. The total price for (D1, C3, C4) will be the sum of the prices from each tool without any discount. Suppose b^* is the promoted bundle, that is, the only bundle with discounted price. Under this scenario, the total revenue can be expressed as follows:

$$R = N \times \left[\sum_{b \in \mathcal{C}_1} TP_b \times Pr(b) + \sum_{b \in \mathcal{C}_2} BP_b \times Pr(b) \right] \tag{2.24}$$

where $\mathcal{C}_1$ is the set of bundles that do not include b^* as a subset, $\mathcal{C}_2$ is the set of bundles that have b^* as a subset. N is the market size, TP_b is the total bundle price, BP_b is the discounted bundle price and $Pr(b)$ is the bundle choice probability.

We assume that a bundle of size three is offered for Professionals and Hobbyists each. There are 84 bundles of size three in total, resulting in 84 different revenue simulation scenarios. Those bundles are offered with the same pricing and discount strategy as in the empirical study. For each scenario, We calculate the bundle choice probability $Pr(b)$ from equation (2.12) given the prices of products and the special price for the bundle offered. Lastly, we assume the market size $N = 10000$. The simulated revenues from the most popular bundles are presented in Table 2.11. Table 2.11 also includes the simulated revenues when no discount is offered and the top five highest estimated revenues and their corresponding bundles if discounts are applied.

Table 2.11: Revenue Simulation

	Professionals		Hobbyists	
	Bundle	Revenue	Bundle	Revenue
No bundle	–	\$4,374,326	–	\$2,023,699
Top five bundles	(D3, C2, A2)	\$4,361,047	(C1, C2, A2)	\$2,020,384
	(C1, C2, A2)	\$4,358,454	(C1, A1, A2)	\$2,020,124
	(D3, C1, A1)	\$4,356,908	(C1, C3, A2)	\$2,019,136
	(C1, A1, A2)	\$4,355,543	(D3, A1, A2)	\$2,018,976
	(D3, C1, C2)	\$4,355,185	(D3, C2, A2)	\$2,018,501
Most preferred bundle	(D1, D2, C3)	\$4,196,693	(D1, C4, A1)	\$1,975,211

2.6.2 Revenue Analysis

Our simulation results show that those bundles with highest choice probability do not generate the highest revenue. The top five bundles generating the highest revenue include tools of which preferences are negatively correlated. More importantly, our simulation results suggest that providing bundles to the market does not increase the revenue for the company.

Recently Armstrong and Vickers (2010) discussed demand elasticity in a competitive environment. They show that in a competitive environment with unit demand, providing discounts to bundles decreases profit if demand is not sufficiently elastic (see Armstrong and Vickers (2010), Proposition 4). We now discuss the impacts of price elasticities on the revenue.

2.6.3 Analysis of Bundle Price Elasticity

To understand the price elasticity, we derive the own price elasticity of bundle b^* based on equation (2.11):

$$\frac{\partial \log Pr(b^*)}{\partial \log BP_{b^*}} = \beta \times BP_{b^*} \times \left(\sum_{c \in \mathcal{C}_2} \frac{Pr(c)}{BP_c} - \frac{1}{BP_b^*} \right) = \beta \times \left(\sum_{b \in \mathcal{C}_2} \frac{BP_{b^*}}{BP_c} \times Pr(b) - 1 \right) \quad (2.25)$$

Table 2.12: Own Price Elasticity of the Most Preferred Bundle

	Price Elasticity		
	Lower Bound	Exact Value	Upper Bound
Professionals	-0.686	-0.567	-0.487
Hobbyists	-1.199	-1.103	-1.028

Equation (2.25) suggests that the price elasticity is determined by three components: the price coefficient β, the choice probability of those bundles in $\mathcal{C}_2$, and the bundle prices. Moreover, because $Pr(b) \geq 0$ and $\frac{BP_{b^*}}{BP_b} \leq 1$, we can derive the following relationship:

$$-\beta \leq \frac{\partial \log Pr(b^*)}{\partial \log BP_{b^*}} \leq \beta \times \left(\sum_{b \in \mathcal{C}_2} Pr(b) - 1 \right) \tag{2.26}$$

From equation (2.26) we can see that for the demand for b^* to be elastic, β has to be larger than 1.

Table 2.12 reports the exact value, the upper and lower bounds of price elasticity based on equation (2.26) of the most preferred bundles for Professionals and Hobbyists. Professionals have an inelastic demand for the bundles. This makes sense because they rely on those tools at work. Hobbyists show a slightly elastic demand, compared to Professionals. We plot the choice probability at different discount level for the most preferred bundles, as shown in Figure 2.10. While Hobbyists show a more elastic demand than Professionals, the demand for the most preferred bundles for Professionals and Hobbyists are not very elastic.

In short, the demand for bundles is not sufficiently elastic and it is not profitable to provide bundles to the market.

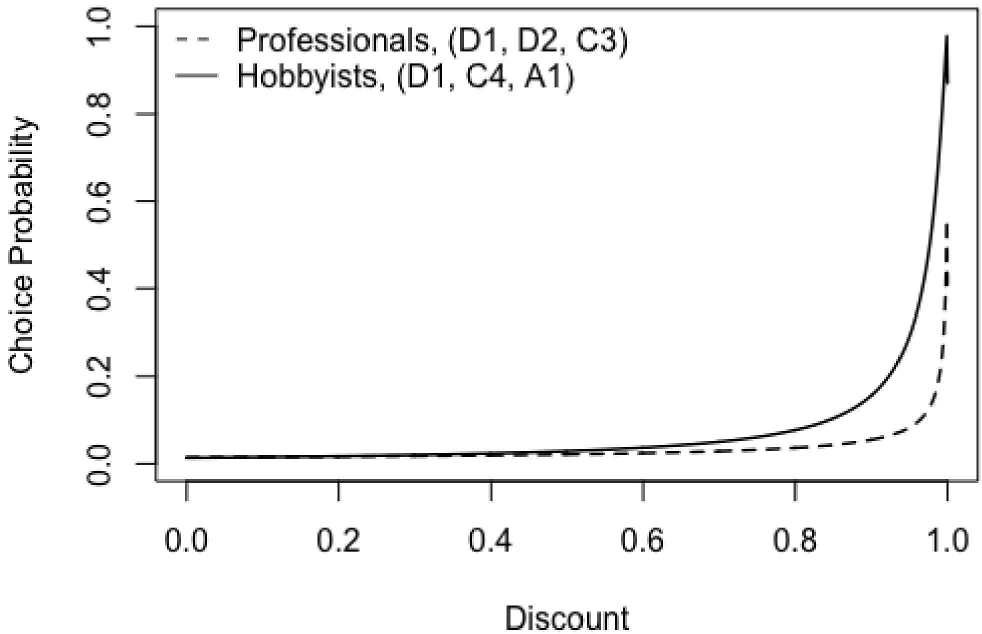


Figure 2.10: Demand Curve for the Most Preferred Bundles

2.6.4 Summary

To develop optimal bundling strategy, we consider two different marketing objectives: maximizing preference share and maximizing revenue. We show that different managerial objectives lead to different bundle recommendation. Moreover, we demonstrate that bundling strategy depends on factors such as demand elasticity, and bundling is not always optimal.

2.7 Conclusions

In this chapter, we develop a new approach for analyzing bundle preference data collected via a configuration task. This type of study spurred us to develop a generalized model of multiple product purchasing that can be used for configuration data and other applications including shopping basket analysis. We discuss the methodological contributions of our study next.

2.7.1 Methodological Contribution

In practice, researchers limit their study samples to include only respondents who may be interested in the focal product/service of the survey. In the case of product bundles, participants are likely to be restricted to those more likely to buy bundles in the market place, i.e., the heavy users discussed above. Consequently, the resulting preference distribution based on the bundle choices of heavy users is a screened distribution. We propose that the screened distribution be modeled as a function of the population preference distribution and a screening process driven by the inclusive value of the product category. These combine in our proposed screened multivariate normal (S-MVN) distribution for heavy users.

Incorporating the S-MVN into our aggregate Mixed Logit choice model for bundles, we show there is a closed-form solution, which is generally not possible with Mixed Logit models. Furthermore, the resulting formulation is a type of MVL model which is well-known in the spatial and choice modeling literatures (Russell and Petersen, 2000; Song and Chintagunta, 2006; Kwak et al., 2015). The model is relatively straight-forward to estimate and has other desirable properties. In addition to providing insights into preference heterogeneity and product interdependencies from configuration data, the model developed for this paper can be readily adapted to data from shopping basket studies and other cross-category purchasing situations.

2.7.2 Strengths and Weaknesses of Configuration Data

As discussed in the introduction, a typical single task configuration study is very easy to implement for marketing researchers. For participants, the task is natural and straightforward to understand. However, if the menu of options does not change across participants, researchers are left with limited information upon which to base their analysis.

In our empirical example, all participants faced the same product prices. The only variation was the bundle discount which was affected by the size of the bundle. Despite different levels of bundle discounts, the variation may not be sufficient and thus a single task configuration study may underestimate consumer price sensitivity.

Unfortunately, the recommended remedy is to expand the number of BYOB tasks for each individual respondent. This requires a well-constructed experimental design as well as survey participants willing to complete multiple, highly similar tasks with equal attention. The former requirement reduces the attractiveness of this approach for some practitioners. The latter issue of respondent attention and its counter-part of decision fatigue are problems common to many research methods. However, this exact problem was supposed to be avoided by using a configuration task for collecting bundle preference data. Moreover, it must be noted that the reliability (test-retest, parallel forms, etc.) of repeated configuration tasks, even for identical menus of options, has yet to be reported in the literature.

As noted above, the empirical study we analyzed had fixed individual product prices but varying levels of discounts depending on the bundle size. This limited our ability to measure the price elasticity of individual products. To obtain such data, some researchers design the bundle menu so that some or all of the prices vary across participants or across repeated menus given to the same participant (Orme, 2013). These menu-based choice studies seem to provide sufficient information to estimate the impact of price changes. However, it must be noted that the conjoint literature has identified changing a single attribute (such as price) across tasks to be problematic. The range of a product attribute (how much it varies across tasks) as well as the number of attribute levels increases the *measured* importance of that attribute (Wittink et al., 1990). Consequently, changing the prices across menus may elicit unwanted attention to prices and thus *overestimate* price sensitivity.

Another limitation of a configuration study is that we are not able to analyze the interependencies among products at the individual level due to a lack of repeated observations from participants. Distinguishing the product interdependencies at the individual level from the preference heterogeneity could be useful in developing a multi-bundle strategy. Understanding the interdependencies among products and the preference heterogeneity allows companies to devise different bundles for different segments, and maximize customer value and profits.

2.7.3 Future Direction

2.7.3.1 Incorporating individual covariates

Some individual measures were collected in this survey. The proposed model in this chapter does not consider those individual demographic factors. The next step is to develop a modeling approach that delivers individual preference estimates. Marketers are interested in understanding individual consumer preferences so that they can generate most profit by tailoring their products and services to target consumers. Preference heterogeneity gives rise to market segmentation and product differentiation. It also serves as the basis for targeted communication and promotion strategy. Hence in many marketing contexts researchers wishes to estimate individual level parameters. The individual variables are observed heterogeneity that contribute to the preference variance across respondents. Thus doing so would allow researchers to understand the market better and improve model prediction ability. Appendix C illustrates the fundamental setup of the model that incorporates individual covariates. Further model identification derivation is needed for the model to be able to apply to configuration dataset.

2.7.3.2 Choice-based conjoint versus configuration data

For measuring bundle preferences, choice-based conjoint (CBC) analysis is an alternative approach because it provides repeated observations. However, CBC analysis is limited in the situation where bundles can be constructed from a large assortment of products. As the number of products increases, the number of choice tasks required increases and participants are more likely to experience information overload. This will adversely affect the quality of data.

Researchers often use fractional factorial or incomplete block designs to reduce the number of choice tasks. However, this results in the possibility that participants are not presented with ideal (or close to ideal) products throughout the choice tasks. This creates a high proportion of none choices resulting in reduced validity of the parameter estimates.

Both the configuration task and the CBC approach use hypothetical bundles. For the BYOB task, the respondent determines which products go into their bundle. In the CBC, the respondent chooses one (or none) of the bundles from a pre-specified set. These tasks differ in the demands put on the respondent and the type of decision making they employ. The configuration task seems to rely more on sequential processing of product choices. On the other hand, the CBC tasks require holistic evaluations of entire bundles and a choice from these few presented options. An interesting area of future research would be to explore how the results of these methods differ and whether they could be used in combination to improve results for managers.

CHAPTER 3
COMPARISON OF CONFIGURATION AND CONJOINT DATA

3.1 Introduction

Building on the previous chapter, this chapter examines the consumer decision-making process under choice-based conjoint (CBC) and configuration studies.

We learned in Chapter 2 that researchers have limited information in a configuration study, and may underestimate price sensitivity. Choice Based Conjoint (CBC) studies, on the other hand, are designed to provide sufficient variations in key attributes throughout different choice tasks. The design of a CBC study for measuring bundle preferences can be very complicated as the number of products available increases, and the data quality may be negatively affected if the choice tasks required in a CBC study is large. Due to the different strengths and weaknesses of the configuration and CBC studies, researchers may improve the understanding of bundle preference by combining these two different sources. In marketing, a rich literature has been dedicated to combining different data sources. In this stream of research, an utility maximization, compensatory decision process is assumed to be the underlying mechanism for all choices (Louviere et al., 1999). If consumers adopt utility maximization and compensatory decision rules for both CBC and configuration studies, researchers can benefit from combining CBC and conjoint data and apply the traditional tool directly.

It is possible, however, that consumers use different decision-making processes under configuration or CBC study. Consumer behavior researchers have studied extensively different decision making processes under different conditions (Gaeth et al., 1991; Levin et al., 2002; Park et al., 2000). If this is the case we need to develop a model that reflects the different psychological process while taking advantage of these two data sources. Because the model for combining the two data sources depends on the underlying decision-making process, it is important for us to evaluate the underlying assumptions before we proceed to data fusion.

To evaluate and compare the underlying decision making process of CBC and configuration studies, we start by assuming that consumers adopt utility maximization compensatory decision process for a CBC study. This assumption directly follows from Chapter 2, and allows us to make a reasonable comparison between the estimation results from CBC and configuration study. Moreover, analyzing CBC studies using utility maximization models is widely used by marketing practitioners. This assumption also allows us to provide an evaluation of current approaches.

This chapter is organized as follows. We provide the hierarchical Bayesian Multinomial Logit model for CBC data analysis following the review of the literature on data fusion and the psychological aspects of bundling. After describing the CBC data collected from the power tool study, we present the estimation results and compare it with the result from Chapter 2. We conclude this chapter with conclusions and future direction.

3.2 Literature Review

3.2.1 Literature on Data Fusion

Researchers have been interested in combining different data sources since the 1990s. The underlying belief is that one data source may only reflect certain aspects of decision-making process and is limited in its ability to understand the choice and judgment process. Thus, integrating different sources of data each of which possesses different strengths and weaknesses would significantly benefit researchers (Louviere et al., 1999; Agarwal et al., 2015).

One major research area is combining choice experiment and market data (Ben-Akiva et al., 1994; Hensher and Bradley, 1993; Louviere et al., 1993; Swait and Andrews, 2003; Hensher, 2008; Louviere et al., 2008; Horsky et al., 2006; Feit et al., 2010). Because market data are real choices made by consumers who spent their resources to make the decisions, it has high reliability and face validity. Unfortunately, in the real world it is quite common that market data often do not have enough information, such as insufficient variation in key attributes of the products offered in the market for parameter estimation or collinearity between attributes due to product design constraint (Brownstone et al., 2000). Choice models will not be able to detect the effect of retail prices and other marketing activities and cannot help managers to make decisions in response to competitors actions. On the other hand, choice experiments have proven to be useful in a variety of context (Green et al., 2001). In addition, experimental data are rich in trade-off information, thus are more robust to analyze structural change. The drawback is that experiment data are hypothetical and may not take into account all possible market constraints. Studies have also found that at times the preferences estimated from choice experiments diverge from those inferred market data (Brownstone et al., 2000). Due to their complementary characteristics, researchers are interested in exploiting the strengths and ameliorating the weaknesses by combining the two data sources.

Studies in this area assume that the underlying factors governing the decision processes in different contexts are the same. A key insight is that pooling the two data sources requires a scaling parameter to accommodate differences in the unobserved factors between the two sources (Morikawa, 1989; Ben-Akiva and Morikawa, 1990; Swait and Louviere, 1993; Louviere et al., 2000). The concept of this data enrichment approach is summarized as follows.

Denote the market data as RP (revealed preference) and choice experiment data as SP (stated preference). Let $\mathcal{C}^{RP}$ and $\mathcal{C}^{SP}$ be the choice set observed in market and choice experiment data, respectively. Suppose alternative b can be described by variables X_b^{RP} and X_b^{SP}, the common attributes observed in $\mathcal{C}^{RP}$ and $\mathcal{C}^{SP}$, respectively and Z_b and W_b, the unique attributes observed for each data source. The utility for alternative b in both data sources are:

$$u_b^{RP} = \alpha_b^{RP} + \beta^{RP} X_b^{RP} + \gamma Z_b + \varepsilon_b^{RP} \quad \forall\, b \in \mathcal{C}^{RP} \tag{3.1}$$

$$u_b^{SP} = \alpha_b^{SP} + \beta^{SP} X_b^{SP} + \delta W_b + \varepsilon_b^{SP} \quad \forall\, b \in \mathcal{C}^{SP} \tag{3.2}$$

If we assume the error terms in equations (3.1) and (3.2) are independently and identically distributed as a Gumbel distribution with scaling parameters λ^{RP} and λ^{SP}, the corresponding choice probability can be written as:

$$Pr(b)^{RP} = \frac{\exp\left\{\lambda^{RP}\left(\alpha_b^{RP} + \beta^{RP} X_b^{RP} + \gamma Z_b\right)\right\}}{\sum_{b' \in \mathcal{C}^{RP}} \exp\left\{\lambda^{RP}\left(\alpha_{b'}^{RP} + \beta^{RP} X_{b'}^{RP} + \gamma Z_{b'}\right)\right\}} \quad \forall\, b \in \mathcal{C}^{RP} \tag{3.3}$$

$$Pr(b)^{SP} = \frac{\exp\left\{\lambda^{SP}\left(\alpha_b^{SP} + \beta^{SP} X_b^{SP} + \delta W_b\right)\right\}}{\sum_{b' \in \mathcal{C}^{SP}} \exp\left\{\lambda^{SP}\left(\alpha_{b'}^{SP} + \beta^{SP} X_{b'}^{SP} + \delta W_{b'}\right)\right\}} \quad \forall\, b \in \mathcal{C}^{SP} \tag{3.4}$$

Because the underlying process governing the decision processes in two data sources is the same, the parameters for the common attributes are assumed to be equal, i.e., $\beta^{RP} = \beta^{SP} = \beta$. As shown in equations (3.3) and (3.4) the scaling parameter λ is confounded with β. Thus, to impose the equality condition on β we need to control for the differences in scale between data sources. It is conventional to normalize the scaling parameter for the experiment data (set $\lambda^{RP} = 1$), and estimate λ^{SP} in a relative scale with respect to λ^{RP}. The log likelihood for the pool data is:

$$\mathcal{L}(\lambda^{SP}, \alpha^{RP}, \alpha^{SP}, \beta, \gamma, \delta) = \sum_{i \in R} y_{ib} \ln Pr_i(b)^{RP} + \sum_{i \in S} y_{ib} \ln Pr_i(b)^{SP}$$

where $y_{ib} = 1$ if participant i chooses alternative b and 0 otherwise.

In short, if the decision making processes are consistent under CBC and configuration studies, the only difference will be the scaling parameter. Thus, when we compare the estimates of the scaling parameters from the two datsets, we expect to see that the estimates are proportional to each other.

3.2.2 Literature on Psychology of Bundling

Consumer behavior researchers have studied extensively with respect to the psychological process of bundle preference formation and choices. Several researchers study the preference formation process from the aspect of anchoring and adjusting (Levin and

Gaeth, 1988; Park et al., 2000). Levin et al. (2002) examine the bundle formation processes in which consumers were asked to construct their ideal bundle using building up (adding products) or scaling down (removing products) approaches, and found that consumers formed a bigger bundle at a higher cost when adopting the scaling down approach. In line with this notion, Yadav (1994) finds that consumers anchored their evaluation on the item perceived as most important and then made adjustments based on the evaluations of the remaining items in the bundle. Gaeth et al. (1991) also show similar results.

Literature has different assumptions about how consumers perceive the value of the bundle and make the comparison among bundles. In addition to deriving the value of the bundle based on the products in the bundle, researchers assume that consumers can integrate the values of product attributes to form attribute inventories, which are used to form the bundle value and are served as the bundle comparison basis. Rao et al. (2017) summarize four different strategies to compare bundles based on different aforementioned assumptions.

The above decision strategies may be adopted under different bundle choice conditions – if the bundles are presented to consumers (static bundles) or if the bundles are formed by consumers (sequential bundles) – and hence result in different preference formation with respect to the same bundle and the bundle choices. Kim and Rao (2016) examine the choices made under sequential and static bundle conditions. Their result suggests that consumers behave differently for these two types of bundle choices.

In the configuration study, participants are involved in a dynamic bundle formation process whereas in the conjoint study consumers compare different static bundles. It is possible that consumers may use different decision rules and utility maximization assumption does not hold for both datasets. Thus in this chapter we evaluate the comparability between two studies under utility maximization assumption which serves as the foundation for developing a data fusion model for CBC and configuration studies.

3.3 Model Development

We choose to use a Hierarchical Bayeisan Multinomial Logit (HB MNL) model to analyze the CBC data for two reasons. First, MNL model holds the same utility maximization assumption as the MVL model in Chapter 2. Second, conjoint studies are a familiar tool for marketing practitioners, and the standard approach to analyze the conjoint data in practice is HB MNL approach. Comparing the result from HB MNL model with the MVL model allows us to provide a commentary on these two approaches.

CBC analysis provides repeated measurement on consumer preferences, and thus allows us to distinguish the product interdependencies and preference correlation matrices. We decide to simplify the comparison by setting product interdependency matrix to zero, following the assumption in Chapter 2 and focus on the aggregate parameter analysis.

Let $j = 1, \ldots, J$ represent tools, and $b = 1, \ldots, B$ denote the bundle, where 1 represents the null bundle and $B = 2^J$ is the bundle consisting of all products. For individual i, the choice probability for bundle b at time $t$is

$$Pr(b \mid \boldsymbol{\alpha}) = \frac{\exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_b - \beta_i \ln BP_b\right\}}{\sum_{b' \in \mathcal{C}_{it}} \exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \ln BP_{b'}\right\}} \tag{3.5}$$

where $\mathcal{C}_{it}$ is the choice set the individual i faces at time time t.

Following equation (2.4), we set $\sum_j^J \alpha_{ij} = 0$ without loss of generality. The preference heterogeneity distribution for the $J - 1$ α's is assumed to follow normal distribution

$$\psi_i = (\boldsymbol{\alpha}_i, \beta_i) \sim f(\psi) = MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.6}$$

The prior distribution for $\boldsymbol{\mu}$ is $MVN(\mathbf{0}, 100\mathbf{I})$, where I is the identity matrix. The prior distribution for $\boldsymbol{\Sigma}$ follows Inverse-Wishart distribution $IW(\nu, \nu I)$, where ν is set to be the number of the parameters plus 3.

One major difference of the HB MNL model and the MVL model in Chapter 2 is the preference distribution assumption. The MVL model takes into account the screening process and assumes the preference distribution for heavy users follows S-MVN distribution. The HB MNL model assumes the normal distribution instead. Because of the difference, we expect the estimation results from these two models should reflect the difference.

We use Markov Chain Monte Carlo (MCMC) procedure for parameter estimation. More specifically, we conduct the model estimation process in `R` (R Core Team, 2015) using the `rhierMnlRwMixture` function in `bayesm` package (Rossi, 2015). We apply the HB MNL to the CBC power tool data. In the following section, we summarize the CBC data and present the estimation results.

3.4 Empirical Application

3.4.1 Summary of the CBC Study on Power Tools

In the current power tool study, the participants complete a choice-based conjoint (CBC) study before the configuration task. The CBC study consists of six choice tasks for each participant. The bundles in the CBC study always have a base unit, labeled as A3 in the subsequent analysis. Each choice task includes four options – three of which are bundles of sizes two, three or four in addition to A3. The fourth option is the "no-choice" option (null bundle). All the drilling and cutting tools are present in the CBC study. Tool A1 never appears in a bundle of size two, and tool A2 is excluded from the CBC study. The price for each tool is the same as in the configuration study. The total bundle price is the sum of all the tool prices after applying the quantity discount. The discount rule applied to the configuration study also applies to the CBC study. Because tool prices remains the same across choice tasks, bundle prices do not vary across different choice tasks and participants.

Table 3.1 summarizes the number of no-choice options in the CBC study for Professionals and Hobbyists. The column "Partial" refers to the participants who do not construct a full size bundle in the configuration study, and "Complete" refers to the group that selects all the tools. Overall 40.7% of Professionals and 52.3% of Hobbyists choose the no-choice option at least once. 10.7% of Professionals and 18.5% of Hobbyists choose the no-choice option across all six choice tasks.

We did a chi-squared test to see if the choices of the complete group is different from the partial group. For Professionals and Hobbyists, the test results are $X^2 = 22.72$, $p = 0.001$ and $X^2 = 4.16$, $p = 0.65$, respectively. This suggests that Professionals in these two groups behaved differently, while Hobbyists did not differ from each other. To be consistent with Chapter 2, the complete group is removed from the dataset used in the analysis.

Figure 3.1 summarizes the frequencies of tools included in the chosen bundles (not including the null bundle). Because A3 is always included in the bundle, participants in

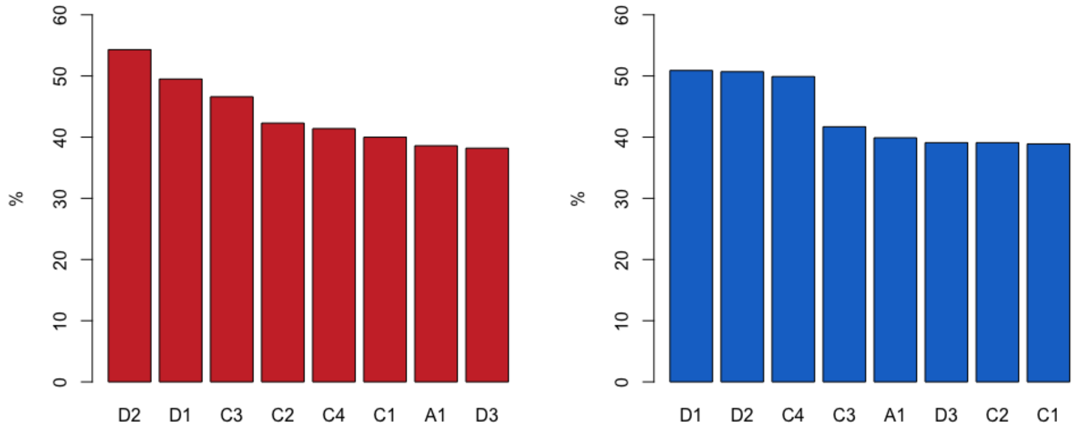
Table 3.1: Summary of Null Bundle in CBC Study

(a) Professionals

Number of Null Bundle Selected	Professionals					
	Partial		Complete		Total	
	Count	%	Count	%	Count	%
0	50	49.5	39	79.7	89	59.3
1	9	8.9	1	2.0	10	6.7
2	11	10.9	0	0.0	11	7.3
3	12	11.9	2	4.1	14	9.3
4	6	5.9	0	0.0	6	4.0
5	3	3.0	1	2.0	4	2.7
6	10	9.9	6	12.2	16	10.7
Total	101	100.0	49	100.0	150	100.0

(b) Hobbyists

Number of Null Bundle Selected	Hobbyists					
	Partial		Complete		Total	
	Count	%	Count	%	Count	%
0	56	45.9	16	55.2	72	47.7
1	14	11.5	3	10.4	17	11.3
2	11	9.0	1	3.4	12	7.9
3	7	5.7	3	10.4	10	6.6
4	8	6.6	1	3.4	9	6.0
5	2	1.6	1	3.4	3	2.0
6	24	19.7	4	13.8	28	18.5
Total	122	100.0	29	100.0	151	100.0

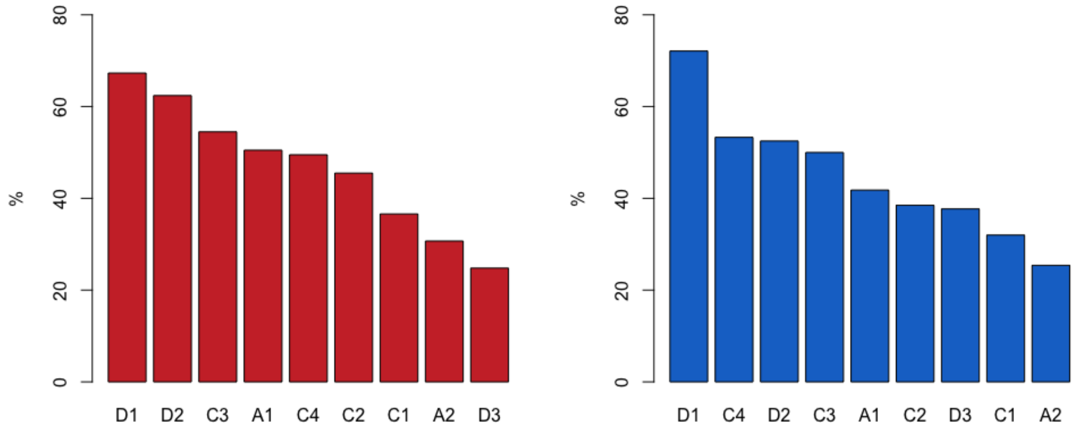


(a) Professionals

(b) Hobbyists

Figure 3.1: Frequencies of Tools Being Chosen - Conjoint Data

the conjoint choice are restricted to choose A3 regardless the bundle, except for the null bundle. Thus we do not show the chosen frequency of A3 here. Comparing to the frequencies in the configuration data, as shown in Figure 3.2, the results share some similarities. Not considering A2 in the configuration data, the top three most frequently chosen tools for Professionals and Hobbyists are the same – (D1, D2, C3) for Professionals and (D1, D2, C4) for Hobbyists – with slightly different rank order. The least frequently chosen tool are the same across two datasets – D3 for Professionals and C1 for Hobbyists. The rank order for the rest of the tools are slightly different. The differences in the preferences for tools are stronger in the configuration study than in the conjoint study. Overall the result suggests a certain degree of preference consistency across the two different preference measurement.



(a) Professionals (b) Hobbyists

Figure 3.2: Frequencies of Tools Being Chosen - Configuration Data

Because participants are limited in choosing A3, we decided to exclude A3 from the dataset to reflect true preferences. In order to achieve this, we need to drop null bundle from the dataset as well. This results in 91 Professionals with 404 observations and 98 Hobbyists with 489 observations. We then apply the HB MNL model to these observations.

3.4.2 Estimation Results

We adopt a thinning procedure and obtain 10000 samples by keeping every 5 samples out of 50000 samples. The trace plots are shown in Appendix D, showing that the MCMC algorithm had converged. We use the first 2000 samples as burn-in and use the remaining of the 8000 samples for estimation. Table 3.2 provides the estimation result from the HB MNL model.

Table 3.2: Conjoint Analysis - HB MNL Model

Parameters		Professionals			Hobbyists		
		Mean	Std. Err.	95% HDI*	Mean	Std. Err.	95% HDI†
Overall preference	μ_{D1}	0.697	0.964	(-1.194, 2.613)	0.970	1.327	(-1.618, 3.639)
	μ_{D2}	0.787	1.229	(-1.635, 3.247)	0.363	1.589	(-2.804, 3.513)
	μ_{D3}	-0.763	1.699	(-4.147, 2.628)	-0.981	1.140	(-3.248, 1.261)
	μ_{C1}	-0.459	1.073	(-2.569, 1.680)	-0.285	1.228	(-2.717, 2.163)
	μ_{C2}	-0.391	1.103	(-2.567, 1.796)	-1.319	1.350	(-3.985, 1.343)
	μ_{C3}	0.184	1.073	(-1.946, 2.300)	0.021	1.151	(-2.223, 2.336)
	μ_{C4}	-0.317	1.245	(-2.800, 2.149)	0.502	1.497	(-2.471, 3.466)
Price Coefficient	β	4.111	3.923	(-3.873, 12.908)	7.480	2.043	(3.333, 11.441)

† HDI = Highest Density Interval

Because $\sum_j \alpha_{ij} = 0$, we can infer $\mu_{A1} = 0.262$ for Professionals and $\mu_{A1} = 0.729$ for Hobbyists. The estimation results suggest that D1 and D2 are the top two most frequently chosen tools for Professionals, which is consistent with Figure 3.1. For Hobbyists, the top two tools are D1 and A1. The result is quite different for Hobbyists since A1 ranks the fifth most frequently chosen tools among chosen bundles in the counting analysis, as shown in Figure 3.1.

Because of the scaling parameter is confounded with the parameters in the MNL model, we cannot directly compare parameters between Professionals and Hobbyists. We leave the discussion about elasticity to Section 3.5 when we compare estimations with the MVL results.

Based on the 8000 samples, we calculate the preference correlations for the tools. We first calculate the correlation among tools for each sample. We then calculate the means and standard errors for the correlation estimates based. The correlation estimates as well

Table 3.3: Preference Correlation Estimates

Correlation	Professionals			Hobbyists		
	Estimate	Std. Err.	95% HDII†	Estimate	Std. Err.	95% HDI†
(D1, D2)	-0.146	0.234	(-0.595, 0.303)	-0.270	0.219	(-0.673, 0.160)
(D1, D3)	0.052	0.260	(-0.445, 0.528)	-0.115	0.231	(-0.553, 0.329)
(D1, C1)	-0.234	0.208	(-0.625, 0.185)	0.118	0.235	(-0.346, 0.553)
(D1, C2)	-0.093	0.228	(-0.505, 0.358)	-0.406	0.187	(-0.745, -0.039)
(D1, C3)	0.028	0.233	(-0.411, 0.482)	-0.079	0.231	(-0.517, 0.367)
(D1, C4)	-0.164	0.238	(-0.607, 0.310)	0.017	0.228	(-0.418, 0.474)
(D2, D3)	-0.372	0.216	(-0.771, 0.041)	0.159	0.222	(-0.284, 0.561)
(D2, C1)	-0.108	0.227	(-0.522, 0.345)	-0.266	0.216	(-0.642, 0.194)
(D2, C2)	0.166	0.229	(0.278, 0.585)	0.093	0.205	(-0.310, 0.476)
(D2, C3)	-0.154	0.221	(-0.564, 0.283)	-0.061	0.222	(-0.489, 0.380)
(D2, C4)	0.144	0.237	(-0.297, 0.602)	-0.483	0.178	(-0.788, -0.123)
(D3, C1)	-0.060	0.252	(-0.534, 0.426)	-0.060	0.225	(-0.500, 0.358)
(D3, C2)	-0.369	0.210	(-0.731, 0.053)	-0.036	0.236	(-0.491, 0.419)
(D3, C3)	-0.304	0.221	(-0.706, 0.133)	-0.289	0.208	(-0.687, 0.104)
(D3, C4)	-0.485	0.189	(-0.816, -0.110)	-0.172	0.230	(-0.600, 0.260)
(C1, C2)	0.072	0.228	(-0.360, 0.509)	-0.061	0.228	(-0.509, 0.374)
(C1, C3)	-0.037	0.224	(-0.465, 0.390)	0.023	0.232	(-0.426, 0.467)
(C1, C4)	-0.074	0.231	(-0.519, 0.379)	-0.010	0.218	(-0.436, 0.404)
(C2, C3)	-0.048	0.234	(-0.495, 0.386)	0.061	0.234	(-0.402, 0.499)
(C2, C4)	0.188	0.221	(-0.253, 0.584)	-0.017	0.233	(-0.432, 0.450)
(C3, C4)	0.169	0.238	(-0.297, 0.597)	-0.050	0.226	(-0.466, 0.396)

† HDI = Highest Density Interval

as standard errors and highest density interval are shown in Table 3.3. Using the constraint $\sum_j \alpha_{ij} = 0$ we can infer the correlation between A1 and the rest of the tools. The preference correlation matrices for Professionals and Hobbyists are shown in Tables 3.4 and 3.5.

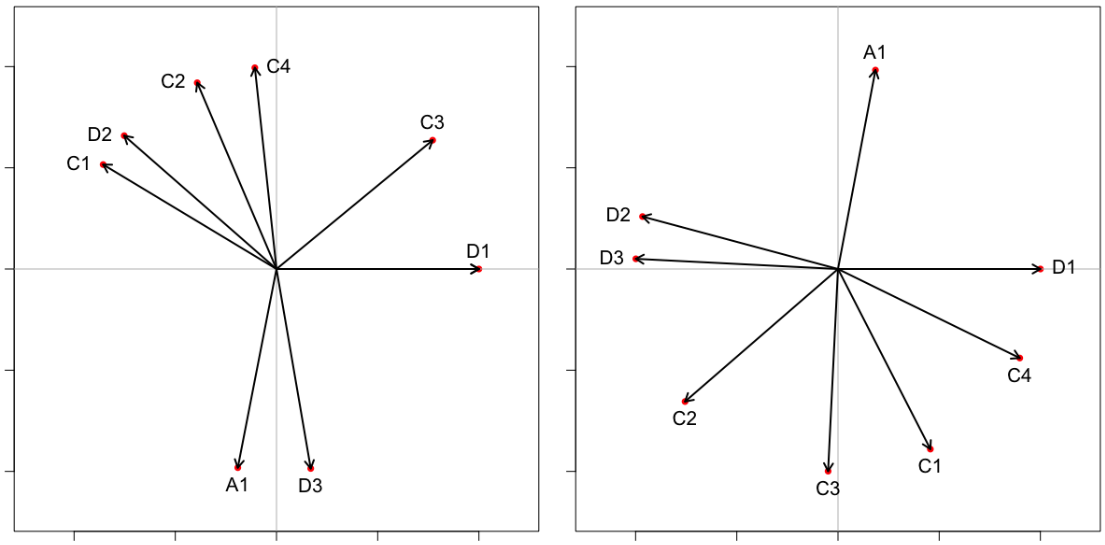
We project the preference correlation matrices onto a two dimensional map, as shown in Figure 3.3. Professionals and Hobbyists show different patterns. For example, preferences for D1 and A1 are positively correlated for Professionals but negatively corre-

Table 3.4: Preference Correlation Matrix – Professionals

	D1	D2	D3	C1	C2	C3	C4	A1
D1	1.000	-0.146	0.052	-0.234	-0.093	0.028	-0.164	-0.242
D2		1.000	-0.372	-0.108	0.166	-0.154	0.144	-0.290
D3			1.000	-0.060	-0.369	-0.304	-0.485	0.295
C1				1.000	0.072	-0.037	-0.074	-0.306
C2					1.000	-0.048	0.188	-0.501
C3						1.000	0.169	-0.357
C4							1.000	-0.426
A1								1.000

Table 3.5: Preference Correlation Matrix – Hobbyists

	D1	D2	D3	C1	C2	C3	C4	A1
D1	1.000	-0.270	-0.115	0.118	-0.406	-0.079	0.017	-0.148
D2		1.000	0.159	-0.266	0.093	-0.061	-0.483	-0.096
D3			1.000	-0.060	-0.036	-0.289	-0.172	-0.273
C1				1.000	-0.061	0.023	-0.010	-0.416
C2					1.000	0.061	-0.017	-0.355
C3						1.000	-0.050	-0.339
C4							1.000	-0.160
A1								1.000



(a) Professionals

(b) Hobbyists

Figure 3.3: Preference Correlation Maps

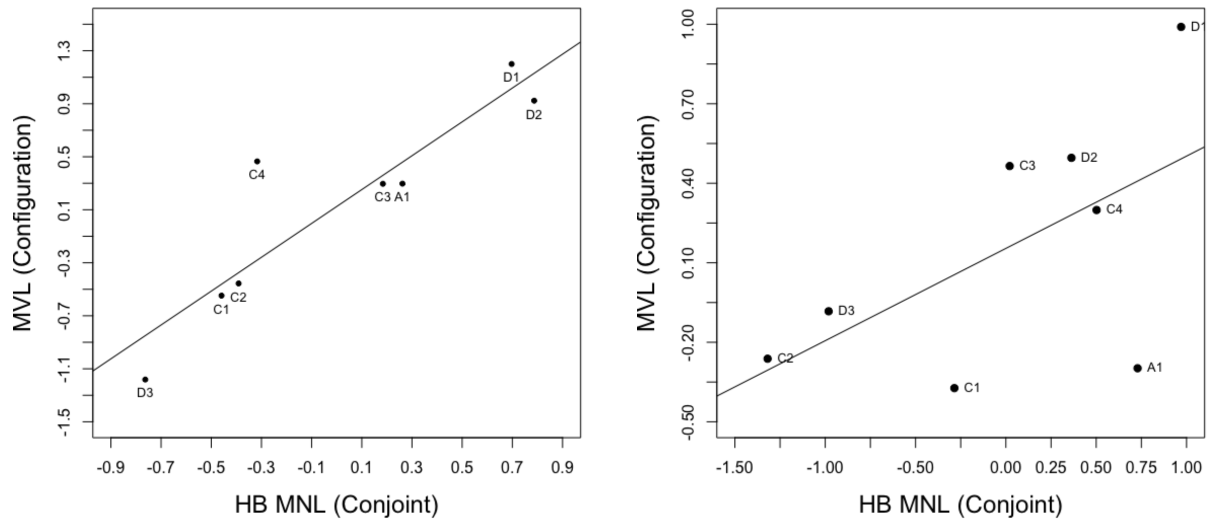
lated for Hobbyists. Similarly, D1 and C3 are positively correlated for Professionals but negatively correlated for Hobbyists.

3.5 Comparison of CBC and Configuration Studies

We now compare the estimation results between MVL and HB MNL models. Because A2 does not appear in the conjoint study, we only focus on parameters for D1 to A1.

3.5.1 Comparison of Overall Preferences

To evaluate the comparability between preference estimates, we plot the results from MVL and HB MNL models, as shown in Figure 3.4. If the parameters are comparable, we expect them to be aligned on a straight line. If the estimates are exactly the same, then we expect to see a 45° line.



(a) Professionals

(b) Hobbyists

Figure 3.4: Overall Preference Comparison

As we can see from Figure 3.4, the estimates are more comparable for Professionals than for Hobbyists. The correlations are 0.91 and 0.56 for Professionals and Hobbyists, respectively. For Professionals, the top two and bottom three tools are consistent across different approaches. The rank order for the rest of the tools are differently. For Hobbyists, D1 is the most frequently chosen tool across two choice tasks. A1 is least preferred in the configuration dataset while it is the second most preferred tool based on the conjoint analysis result. In short, Professionals show a more consistent preference among the two choice tasks while Hobbyists show greater variation.

As mentioned before, these two models have different preference distribution assumptions – S-MVN distribution for heavy users (normal distribution for the population) in the MVL model and normal distribution in the HB MNL model. The normal distribution parameters recovered in the MVL model represent the overall preferences and preference correlations for the population, not the heavy users. Thus the difference in Figure 3.4 may be due to the different segments.

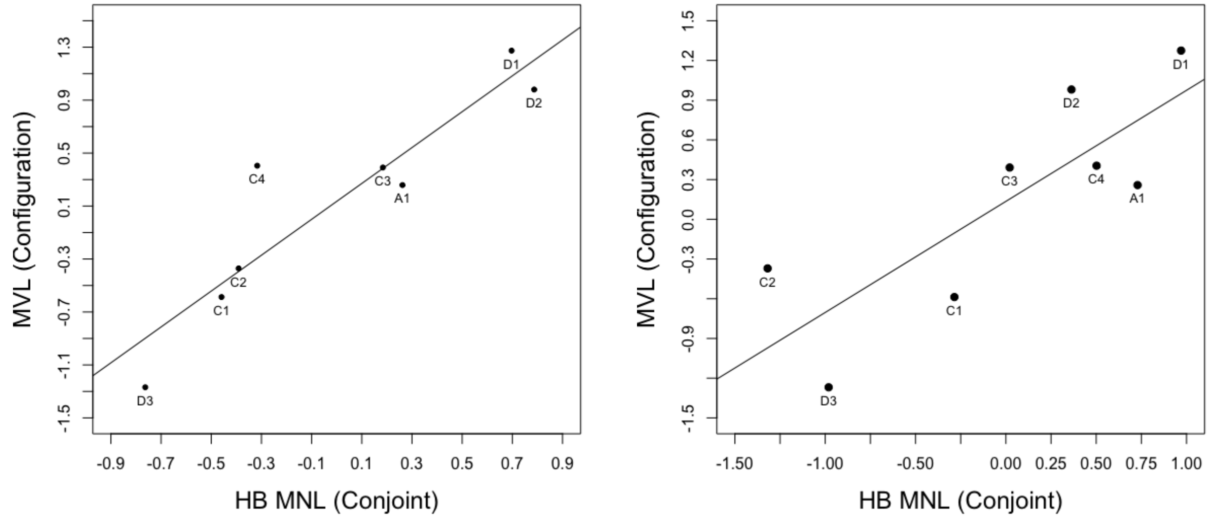
To compare the overall preferences for heavy users, we calculate the overall preference for heavy users in the MVL model. The procedure is as follows. We first draw 15000 samples from the MVN distribution in Chapter 2. We then calculate $w(\boldsymbol{\alpha})$ for those samples. We draw 10000 out of 15000 samples based on $w(\boldsymbol{\alpha})$, and obtain the average of the overall preferences from the 10000 samples.

Figure 3.5 shows the comparison between the overall preferences among heavy users across the two approaches. The correlations for Professionals and Hobbyists are 0.92 and 0.82, respectively. The difference between Figures 3.4 and 3.5 suggests that heavy users show different preference correlations from the population in in the Hobbyists segment.

Overall Professionals show a more consistent preference than Hobbyists. Notice that almost 20% of Hobbyists in the conjoint study choose the null bundle across all six choice tasks (see Table 3.1). It is possible that they do not find the bundles in the conjoint study appealing, or they do not want A3 in their ideal bundle. Regardless the reasons, the high proportion of choosing null bundle in the conjoint study may contribute to the differences in overall preferences across different approaches.

3.5.2 Comparison of Preference Correlations

Figure 3.6 plots the preference correlations from MVL and HB MNL model. The correlations are .475 and .337 for Professionals and Hobbyists, respectively. It suggests



(a) Professionals

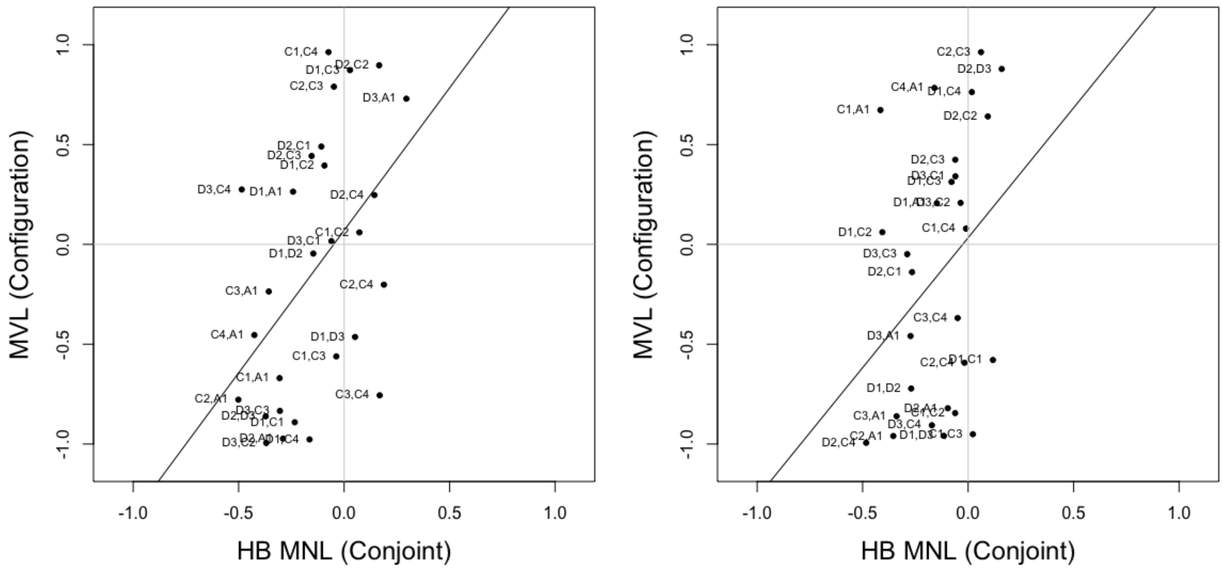
(b) Hobbyists

Figure 3.5: Overall Preference Comparison – S-MVN vs MVN

that the preference correlation patterns are different between configuration and conjoint approaches.

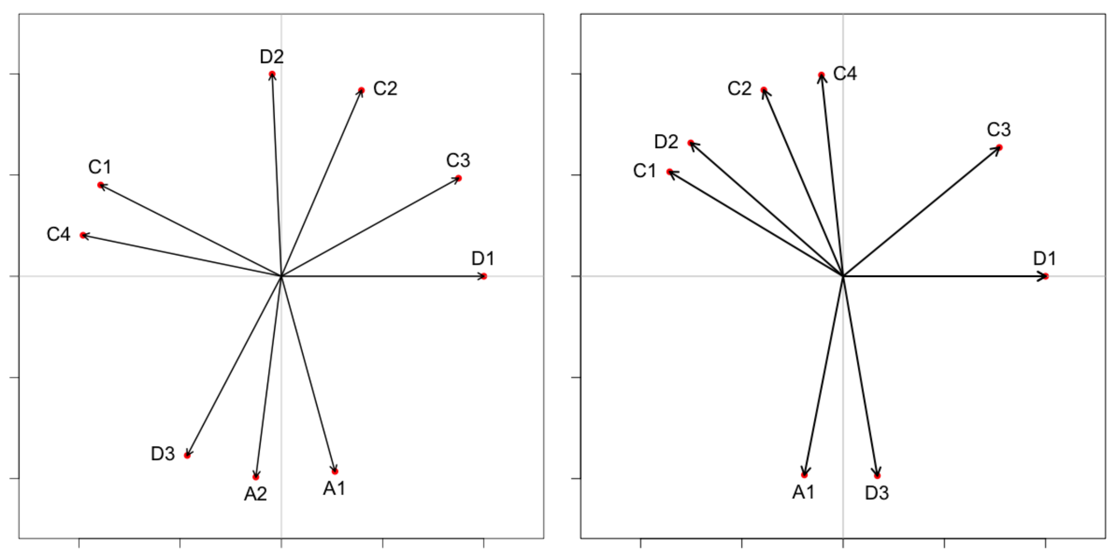
Figures 3.7 and 3.8 shows the two dimensional maps of preference correlations from configuration and conjoint studies. For Professionals, for example, preferences for C3 and C4 are negatively correlated in the configuration study, but positively correlated in the conjoint study. On the other hand, D1 and C2 are positively correlated in the configuration study but negatively correlated in the conjoint study. Similar results can be found in Figure 3.8. D1 and C3 are positively correlated in the configuration study but negatively correlated in the conjoint study, and the opposite for D1 and C1.

Following the previous section, we also use the 10000 draws from S-MVN distribution to calculate the preference correlations, and plot them against the conjoint anal-



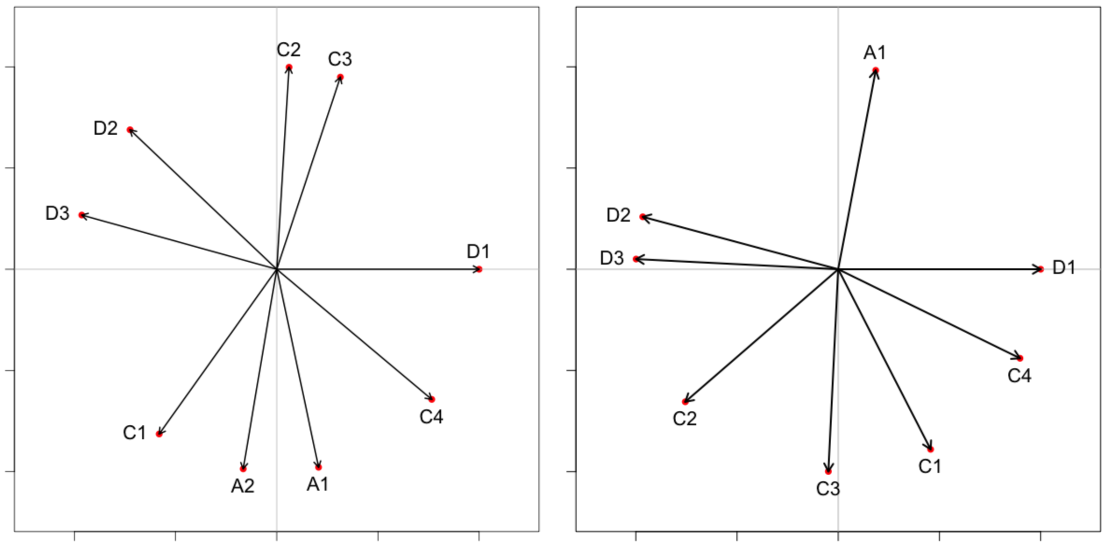
(a) Professionals (b) Hobbyists

Figure 3.6: Comparison of Preference Correlation Estimates



(a) MVL Model (Configuration) (b) HB MNL Model (Conjoint)

Figure 3.7: Preference Correlation Comparison – Professionals



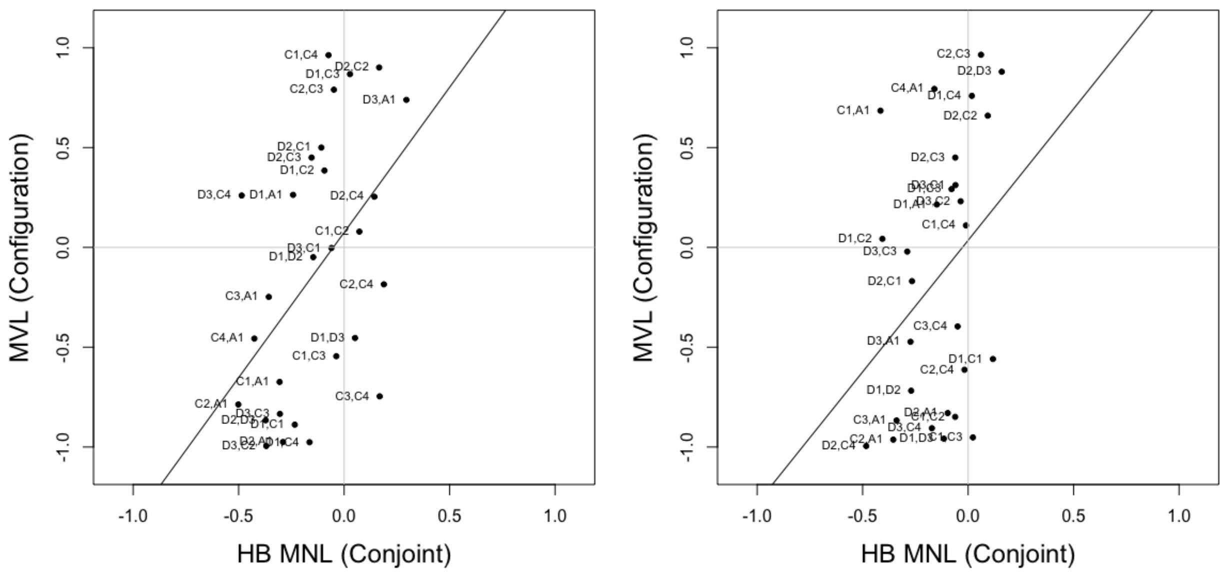
(a) MVL Model (Configuration) (b) HB MNL Model (Conjoint)

Figure 3.8: Preference Correlation Comparison – Hobbyists

ysis result in Figure 3.9. The correlations are .483 and .339 for Professionals and Hobbyists, respectively. We observe some improvements, but correlation coefficients are still low. From the two-dimensional maps (Figures 3.10 and 3.11), we can see that the preference estimates from these two approaches are very different.

We run a regression analysis of the preference correlation coefficients from the configuration study on the correlation coefficients from the conjoint study. Table 3.6 summarizes the regression analysis results. The regression lines are shown in Figures 3.6 and 3.9. The slopes of those regression lines deviate from a 45° line. This implies the scaling parameters are different across configuration and conjoint studies. Because the slopes are greater than one, it suggests that the scaling parameter is larger in Conjoint study.

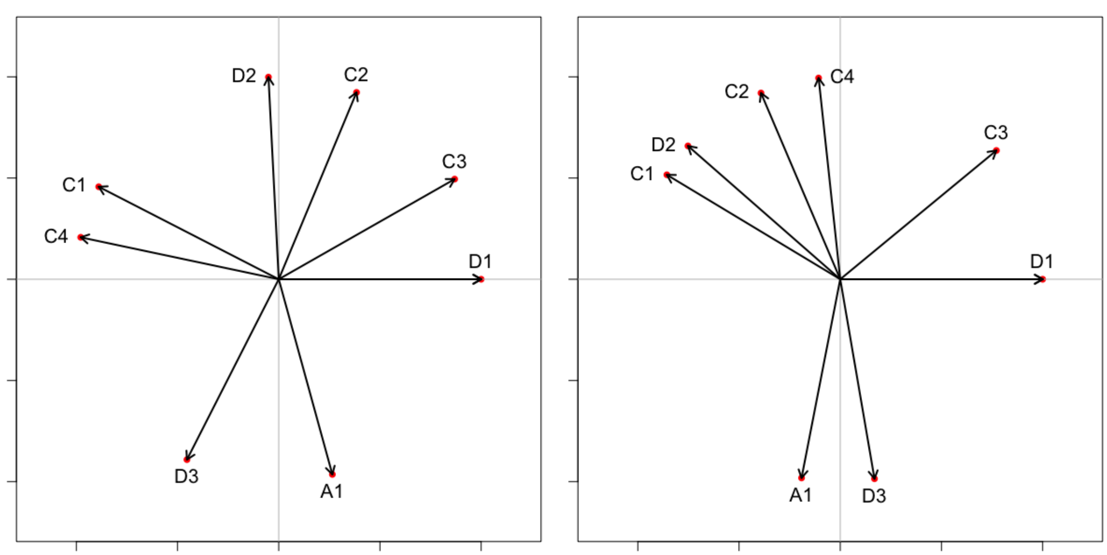
In summary, these two approaches provide different results in terms of preference



(a) Professionals

(b) Hobbyists

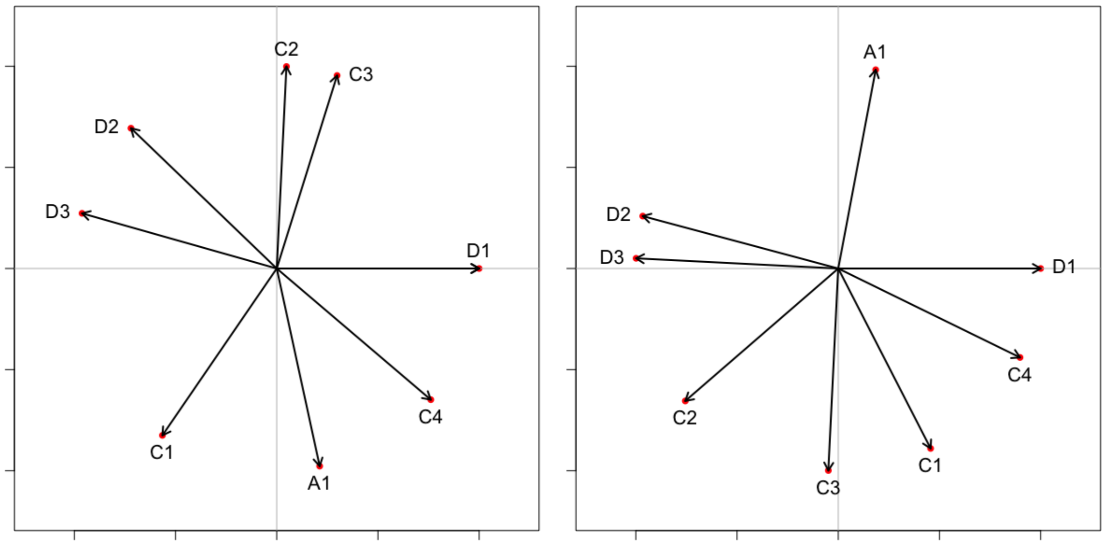
Figure 3.9: Comparison of Preference Correlation Estimates – Heavy Users



(a) MVL Model (Configuration)

(b) HB MNL Model (Conjoint)

Figure 3.10: Preference Correlation Comparison for Heavy Users – Professionals



(a) MVL Model (Configuration)   (b) HB MNL Model (Conjoint)

Figure 3.11: Preference Correlation Comparison for Heavy Users – Hobbyists

Table 3.6: Regression Analysis of Correlation Coefficients

	Professionals				Hobbyists			
	Population		Heavy Users		Population		Heavy Users	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Intercept	0.07	0.13	0.07	0.13	0.03	0.15	0.04	0.15
Conjoint Correlations	1.43	0.52†	1.45	0.52 ‡	1.30	0.71	1.32	0.72

Est. = Estimate; Std. Err. = Standard Error

† Significant at $p = 0.01$

‡ Significant at $p = 0.001$

correlations, suggesting these two approaches are not comparable.

3.5.3 Comparison of Price Elasticity

In Chapter 2 we learn that configuration study may underestimate price elasticity due to limited information. We expect that the price elasticity from conjoint study will be greater than the configuration study.

Because in the HB MNL model the coefficients are confounded with the scaling parameter and cannot be compare directly, we compare the price elasticity across different approaches. The derivation of the price elasticity is in Appendix E.

To compare the price elasticity, we calculate the price elasticity for the most preferred bundle offered in Chapter 2. More specifically, we consider a situation in which all possible combinations of all tools except A2 and A3 are available. This results in 255 possible bundles excluding the null bundle. We then calculate the price elasticity for conjoint study using equation (E.4). The result is summarized in Table 3.7.

We can see that Professionals are less price elastic than Hobbyists across different approaches. In the configuration study, we concluded that heavy users are not price elastic. Conjoint study, however, shows that they are price elastic and thus providing bundles may be profitable. In short, the price elasticity from conjoint study is greater than the configuration study for both Professionals and Hobbyists, which is consistent with what we expect.

Table 3.7: Price Elasticity Comparison

Segment	Bundle	Configuration	Conjoint
Professionals	(D1, D2, C3)	-0.567	-1.986
Hobbyists	(D1, C4, A1)	-1.103	-7.267

3.6 Conclusions

Our goal in this chapter is to evaluate if participants use the same decision-making process under CBC and configuration studies. We start by assuming that participants use utility maximization to make choices under both CBC and configuration studies. If the assumption holds, the parameter estimates from both datasets should be comparable and proportional to each other. If the parameters are not comparable, the conventional utility maximization model may not be appropriate for combining two datasets.

To achieve the goal, we apply a HB MNL model to analyze bundle preferences using choice-based conjoint data. This modeling approach also assumes utility maximization decision process, which is consistent with the model in Chapter 2. In addition, it is the standard approach used by practitioners. Comparing these two approaches allows us to provide a commentary on the current standard method and the MVL model.

We compare the conjoint estimation results with the results obtained from Chapter 2. The overall preference estimates for Professionals show a higher consistency than Hobbyists. We suspect that this is because hobbyists are less satisfied with the options in the conjoint study and hence more likely to select null choice option. In terms of preference correlations, configuration and conjoint studies give very different results, suggesting that these two approaches may not be comparable. The price elasticity from conjoint study is consistent with the configuration study. Moreover, it is greater than the configuration study, as we expect.

From a managerial perspective, the differences in these two approaches will lead to different managerial recommendations. We have shown in Chapter 2 that bundle recommendation should take into account preference heterogeneity as well as product interdependency. While configuration and conjoint studies provide consistent product preferences, the preference heterogeneity from these two approaches is very different. This results in different consumer segmentation and marketing strategies. From the perspective of marketing research practice, further investigation of these two approaches is needed so that marketing researchers can adopt the most appropriate approach that captures the actual choices.

In summary, at this moment it may not be appropriate to combine the two datasets without further understanding of the preference correlations. More research will be needed to understand the difference in terms of preference correlations across configuration and conjoint study, such as the influence of the design of the choice tasks and the psychological process on preference correlations.

However, if these two datasets are comparable under utility maximization assumption, Appendix F outline a data fusion model for bundle preference estimation to take advantages of configuration and conjoint analyses.

CHAPTER 4
CONCLUSION AND FUTURE WORK

This dissertation analyzes and evaluates the approach to measure bundle preferences via configuration study. In Chapter 2, we develop an aggregate MVL model to analyze configuration data. In Chapter 3 we examine the underlying consumer decision-making process of configuration and CBC studies. This provides a foundation for model development that takes advantages of these two different measure approaches. We use a power tool dataset to investigate the above issues.

4.1 Summary of the Dissertation

In Chapter 2, we develop an aggregate MVL model that takes into account preference heterogeneity as well as product interdependencies. We show that this proposed MVL model is the closed-form expression of a Mixed Logit model when the preference distribution follows S-MVN distribution. Our estimation result suggests that configuration data may underestimates key attributes due to limited information. Lastly, we demonstrate that optimal bundling strategy depends on the managerial goals as well as other factors, such as price elasticity.

In Chapter 3, we investigate the comparability of different bundle preference measurement approaches. That is, if consumers adopt same decision-making process under these two choice studies. We assume that in the conjoint study participants also adopt utility maximization strategy, same assumption as in Chapter 3. We apply a HB MNL model to the conjoint data, and compare the results with the parameter estimates from

Chapter 2. In terms of overall preferences for products, we find that preferences for most and least frequently chosen products are consistent across two approaches, and different rank order for products in the middle. This suggests a certain degree of consistency. The price elasticity from the two approaches is consistent and as we expect, the conjoint study show a greater elastic city than the configuration study. However, these two approaches do not provide consistent result in terms of preference correlations. Thus we do not recommend to combine these two datasets for preference estimation without further investigation of the preference correlations. We propose a data fusion model for both conjoint and configuration studies should these two datasets prove to be comparable under utility maximization assumption.

4.2 Contribution of the Dissertation

This dissertation contributes to the literature in two ways. In terms of methodology, we establish the relation between Mixed Logit and MVL models. Traditionally Mixed Logit is known for no closed-form expression. We are able to show that when the mixing distribution is S-MVN distribution, Mixed Logit has a closed-form expression which takes the form of MVL model. This closed-form expression allows researchers to estimate a Mixed Logit model using MLE procedure, which can be estimated using existing statistical software. Researchers may find this approach attractive and suitable for a variety of research areas.

This dissertation also enriches the configuration data analysis literature. Due to the information limitation, currently counting analysis is the mostly commonly used approach. This approach ignores product interdpendencies and cannot take into account preference heterogeneity. Our proposed MVL model is able to provide rich insights into bundle preferences. We also demonstrate that the limited information results in underestimation of key variables. We investigate the possibility to take advantages of configuration and CBC data by examining the fundamental assumption: if participants use the same decision-making process for these two studies. More specifically, following the assumption in Chapter 3 we examine if in a CBC study participants also adopt utility maximization. This investigation provide a venue for researchers to consider enriching configuration data with CBC data.

4.3 Future Research Directions

Configuration data approach is easy to implement for researchers and straightforward for participants when measuring bundle preferences. However, as shown in Chapter 2 configuration data analysis may underestimate key attributes. While conjoint data may complement configuration data, these two datasets may not be fully comparable. Below we discuss some future research direstions.

1. Design of configuration data task

   To extend the configuration study, several approaches are possible. One direct extension is repeated measurement – participants repeat the configuration tasks several times with menus in which key attributes such as price or products vary. However, in the conjoint community it is known that participants are more sensitive to attributes that have more levels (Wittink et al., 1997). Similarly, participants may be alerted to the changing prices or products on a menu and hence become more sensitive to those changes. Evaluating the effect of repeated measures of configuration studies on the alternation of consumer preferences will contribute to the application of configuration data task.

Another approach is to present different menus with different products and prices to participants. This approach may improve the aggregate parameter estimates and reduce the collinearity, but still is limited to aggregate data analysis. An investigation to the improvement of preference measurement will greatly contribute to the practice of configuration data analysis.

2. External validity of configuration data study

   Because of the limited information in the configuration data study, we do not reserve holdout samples or taking other approaches to examine the external validity of the configuration data analysis in Chapter 2. We learn from Chapter 3 that configuration and conjoint data analyses provide very different preference correlation patterns. Because different strengths and weaknesses of configuration and conjoint studies, it is not clear which approach better captures the true choices. Subsequently it is difficult to conclude if conjoint data task serves to validate the configuration data task, or vice versa. More research about the decision making process under these two choice tasks and comparing the results from the two studies with consumers' true choices will help us gain a better understanding.

3. Combine configuration and conjoint data

   As mentioned before, configuration data task is easy to design and implement for researchers and is easy to participate for consumers. The major weakness of the configuration data task is underestimation of key attributes due to lack of repeated measurement. Also, the configuration data task limits researchers to aggregate data analysis. Conjoint study, on the other hand, provides sufficient variations in key attributes. The choice tasks, however, are not easy to design. As seen in our conjoint data, participants may find the bundles not appealing resulting in high proportion of no choice options. This may negatively affect the data quality. Thus taking advantages of these two data source – configuration tasks supplemented by some conjoint tasks – can be an effective approach to measure bundle preferences. Moreover, conjoint studies provide repeated measures, which allow us to distinguish product interdependency matrix from preference correlation matrix and provide better understanding of consumers. Understanding the different sources of product interactions will help managers make better decisions.

4. Data fusion when data sources are not fully comparable

   In our research, we raise the question about data fusion when data sources may not be fully comparable. This does not mean that we should not combine different data sources if they are not comparable. Rather, it calls for a more appropriate data fusion approach. Future research on combing data sources when they are not fully comparable can be valuable to managers.

APPENDIX A
BINARY LOGIT, MNL AND MVL MODELS UNDER INDEPENDENCE ASSUMPTION

The goal of this appendix is to show the relation between MVL, MNL and binary logit - under the assumption that all the items are independent of each other, MVL is equivalent to MNL and binary logit.

Let j denote an item, and b_k denote a bundle. Without loss of generality, consider a situation where a bundle can be constructed from two items, i.e., $j = 2$, which result in four possible bundles: both items are chosen ($b_1 = \{1, 1\}$), either one of them is chosen ($b_2 = \{1, 0\}$ or $b_3 = \{0, 1\}$), or none of them are chosen ($b_4 = \{0, 0\}$). Let v_j be the utility for item j and u_{b_k} be the utility for bundle b_k. We further assume that the utility for a bundle entirely depends on the items that are included in the bundle:

$$u_{b_k} = \sum_{j=1}^{2} z_{jb_k} v_j \tag{A.1}$$

where $z_{jb_k} = 1$ if item j is in bundle b_k.

This is a legitimate assumption since the bundle is constructed entirely based on the items. Moreover, we can always set the utility of no choice option as the baseline utility and set it to zero, i.e., $u_{b_4} = 0$ without loss of generality. The MNL model then can be expressed as

$$Pr(b_k) = \frac{\exp(u_{b_k})}{\sum_{k'=1}^{4} \exp(u_{b_{k'}})} = \frac{\exp(u_{b_k})}{\exp(u_{b_1}) + \exp(u_{b_2}) + \exp(u_{b_3}) + \exp(u_{b_4})} \tag{A.2}$$

and the probability of choosing a bundle modeled by the binary logit model is

$$Pr(b_k) = \prod_{j=1}^{2} P(z_{jb_k} = 1)^{z_{jb_k}} (1 - P(z_{jb_k} = 1))^{1 - z_{jb_k}} = \prod_{j=1}^{2} \frac{\exp(z_{jb_k} v_j)}{1 + \exp(v_j)} \tag{A.3}$$

To show that MNL and MVL are equivalent under the assumption of independence between items, we start from writing out the MVL model:

$$
\begin{aligned}
Pr(b_k) &= \frac{\exp\left(\sum_{j=1}^{2} z_{jb_k}(v_j + \sum_{j'>j} z_{j'b_k}\theta_{jj'})\right)}{\sum_{k'=1}^{4}\exp\left(\sum_{j=1}^{2} z_{jb_{k'}}(v_j + \sum_{j'>j} z_{j'b_{k'}}\theta_{jj'})\right)} \qquad \text{(A.4)}\\
&= \frac{\exp\left(\sum_{j=1}^{2} z_{jb_k}v_j\right)}{\sum_{k'=1}^{4}\exp\left(\sum_{j=1}^{2} z_{jb_{k'}}v_j\right)} \ (\theta_{jj'} = 0 \text{ due to the independence assumption})\\
&= \frac{\exp\left(\sum_{j=1}^{2} z_{jb_k}v_j\right)}{\sum_{i=1}^{4}\exp\left(z_{1b_{k'}}v_1 + z_{2b_{k'}}v_2\right)}\\
&= \frac{\exp\left(\sum_{j=1}^{2} z_{jb_k}v_j\right)}{\begin{aligned}&\exp\left(z_{1b_1}v_1 + z_{2b_1}v_2\right) + \exp\left(z_{1b_2}v_1 + z_{2b_2}v_2\right)\\ &\quad + \exp\left(z_{1b_3}v_1 + z_{2b_3}v_2\right) + \exp\left(z_{1b_4}v_1 + z_{2b_4}v_2\right)\end{aligned}}
\end{aligned}
$$

Based on equation (A.1),

$$
\begin{aligned}
\text{Equation } (A.4) &= \frac{\exp\left(u_{b_k}\right)}{\exp\left(u_{b_1}\right) + \exp\left(u_{b_2}\right) + \exp\left(u_{b_3}\right) + \exp\left(u_{b_4}\right)}\\
&= \text{equation } (A.2),\ \text{the MNL model}
\end{aligned}
$$

We can also rewrite equation (A.4) in another way:

$$
\begin{aligned}
\text{Equation } (A.4) &= \frac{\exp\left(z_{1b_k}v_1 + z_{2b_k}v_2\right)}{\exp\left(v_1 + v_2\right) + \exp\left(v_1\right) + \exp\left(v_2\right) + 1}\\
&= \frac{\exp\left(z_{1b_k}v_1\right)\exp\left(z_{2b_k}v_2\right)}{\exp\left(v_1\right)\exp\left(v_2\right) + \exp\left(v_1\right) + \exp\left(v_2\right) + 1}\\
&= \frac{\exp\left(z_{1b_k}v_1\right)\exp\left(z_{2b_k}v_2\right)}{(1 + \exp\left(v_1\right))(1 + \exp\left(v_2\right))}\\
&= \frac{\exp\left(z_{1b_k}v_1\right)}{1 + \exp\left(v_1\right)} \times \frac{\exp\left(z_{2b_k}v_2\right)}{1 + \exp\left(v_2\right)}\\
&= \text{equation } (A.3),\ \text{the product of a set of binary logit}
\end{aligned}
$$

APPENDIX B
BUNDLE LEVEL ATTRIBUTES AND MVP DECISION PROCESS

A bundle level attribute is defined as an attribute that is related to the bundle as a whole and is not specific to any products. Examples include buy two get one free, spend \$40 or more and get 15% off, or 20% off when purchase three or more products. Thus the bundle level attribute applies to all products in the bundle. In a standard MVP model, a bundle choice probability is defined in terms of the choice probability of products. The underlying utility for product j, $j = 1, \cdots, J$, is written as:

$$u_j = \alpha_j - \beta_j Price_j + \epsilon_j \tag{B.1}$$

The link between the observed bundle choice and the latent utility for product j is represented as:

$$y_{jb} = \begin{cases} 1, & \text{if } u_j > 0, \\ 0, & \text{if } u_j \leq 0. \end{cases} \tag{B.2}$$

Assuming ϵ's follows multivariate normal distribution $f(\boldsymbol{\epsilon}) = MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ results in a MVP model with the probability of observing a bundle $\mathbf{y}_b = \{y_{1b}, \cdots, y_{Jb}\}$ as

$$Pr(\mathbf{y}_b) = \int_{s_1} \cdots \int_{s_J} f(\epsilon_1, \cdots, \epsilon_J) d\epsilon_1 \cdots d\epsilon_J \tag{B.3}$$

where $\epsilon_j = u_j - \alpha_j + \beta_j Price_j$ and

$$s_j = \begin{cases} (-\infty, 0), & \text{if } y_{jb} = 0, \\ (0, \infty), & \text{if } y_{jb} = 1. \end{cases} \tag{B.4}$$

The formulation above describes a decision process in which an individual makes simultaneous, interdependent choice of many products. To incorporate bundle level attributes into the decision making process, one possible approach is to define the product level utility as

$$u_j = \alpha_j - \beta_j Price_j + \gamma_{jb} B_{jb} + \epsilon_j \tag{B.5}$$

where B_{jb} represents the bundle level attribute. We argue that this formulation, however, is not consistent with the decision making process. The value for a bundle level attribute can only be known *after* the bundle is formed. Thus conceptually this individual cannot derive the utilities for products based on known value of the bundle level attribute.

One might argue that a researcher can create an as-if MVP model by replacing variables, such as price, with the actual price levels found in the selected bundle. This approach is invalid for two reasons. First, the modeling approach is equivalent to making price an endogenous variable. In effect, the model structure would assume that the consumer first creates his or her own set of prices and then makes a decision. At a minimum, the researcher would need to use some type of endogeneity correction while estimating the MVP model. Second, it is not at all clear that such a model would have any managerial relevance. It does not make sense to create an implausible consumer model and then attempt to make optimal bundle strategy recommendations.

Finally, the bulk of the bundling literature assumes that consumer forms global attributes from the characteristics of products in the bundle (Rao et al. 2017). These global attributes play a major role in the choice process, particularly when the products in the bundle are jointly used to serve some consumer goal. The MVP model implicitly makes two key assumptions. First, global attributes are not important in the choice process. Second, the attributes of products not in the bundle alter the probability that a certain bundle is selected. Although a researcher might argue that these assumptions are justified in certain applications (such as market basket analysis), these assumptions are not reasonable for the application in our research.

In summary, the MVL choice model and the MVP choice model are not interchangeable. The MVL is not a restricted form of the MVP model. Each model has its particular strengths and weaknesses. Researchers need to take care that choice model assumptions match the intended application.

APPENDIX C
INCORPORATING INDIVIDUAL COVARIATES

Following Allenby and Ginter (1995), we assume the preferences for tools are as follows:

$$\boldsymbol{\alpha} = \boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x} + \boldsymbol{\eta} \tag{C.1}$$

where

$\boldsymbol{\mu}$ = the overall preference for tools that are constant across participants

$\mathbf{x}$ = vector of individual covariates that account for observed heterogeneity

$\boldsymbol{\gamma}$ = vector of coefficients for the observed heterogeneity

$\boldsymbol{\eta}$ = unobserved population preference heterogeneity and is assumed to be multivariate normal, $f(\eta) \sim (0, \boldsymbol{\Sigma})$

The marginal utility can be rewritten as

$$\psi_b = \exp\left\{\boldsymbol{\alpha}^T \mathbf{z}_b + \frac{1}{2}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b\right\} = \exp\left\{(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_b + \boldsymbol{\eta}^T \mathbf{z}_b + \frac{1}{2}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b\right\}$$

and the choice probability for bundle b of participant i is

$$Pr(b \mid \boldsymbol{\eta}) = \frac{\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_b + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_b + \frac{1}{2\delta}\mathbf{z}_b^T \boldsymbol{\Omega} \mathbf{z}_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'} \exp\left\{\frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}}$$

Next, we consider the screening process implemented in the study. As in Chapter 2,

$$w(\boldsymbol{\eta}) = k \exp\{IV(\boldsymbol{\eta})\}$$

and

$$IV(\boldsymbol{\eta}) = \log \sum_{b'} \exp\left\{\frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}$$

. The normalizing constant k can be derived as follows:

$$
\begin{aligned}
k^{-1} &= \int \exp\left\{IV(\boldsymbol{\eta})\right\} f(\boldsymbol{\eta}) d\boldsymbol{\eta} \\
&= \int \left(\sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} \right) f(\boldsymbol{\eta}) d\boldsymbol{\eta} \\
&= \sum_{b'} \int \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} f(\boldsymbol{\eta}) d\boldsymbol{\eta} \\
&= \sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} \int \exp\left\{ \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b'} \right\} f(\boldsymbol{\eta}) d\boldsymbol{\eta} \\
&= \sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'} \right)
\end{aligned}
$$

Thus the screening process $w(\boldsymbol{\alpha})$ is:

$$
\begin{aligned}
w(\boldsymbol{\eta}) &= k \exp\left\{IV(\boldsymbol{\eta})\right\} \\
&= \frac{\exp\left\{IV(\boldsymbol{\eta})\right\}}{\sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'} \right)}
\end{aligned}
$$

The sample preference distribution $g(\boldsymbol{\eta})$ is

$$
\begin{aligned}
g(\boldsymbol{\eta}) &= w(\boldsymbol{\eta}) f(\boldsymbol{\eta}) \\
&= \frac{\exp\left\{IV(\boldsymbol{\alpha})\right\} f(\boldsymbol{\eta})}{\sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'} \right)}
\end{aligned}
$$

The unconditional choice probability $Pr(b)$ for participant i is

$$
\begin{aligned}
Pr(b) &= \int Pr(b \mid \boldsymbol{\eta}) g(\boldsymbol{\eta}) d\boldsymbol{\eta} \\
&= \int \frac{\exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b} + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b} + \frac{1}{2\delta}\mathbf{z}_{b}^T \boldsymbol{\Omega} \mathbf{z}_{b} - \frac{1}{\delta}\ln BP_{b} \right\}}{\exp\left\{IV(\boldsymbol{\eta})\right\})} \\
&\quad \times \frac{\exp\left\{IV(\boldsymbol{\alpha})\right\} f(\boldsymbol{\eta})}{\sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'} \right)} d\boldsymbol{\eta} \\
&= \int \frac{\exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b} + \frac{1}{\delta}\boldsymbol{\eta}^T \mathbf{z}_{b} + \frac{1}{2\delta}\mathbf{z}_{b}^T \boldsymbol{\Omega} \mathbf{z}_{b} - \frac{1}{\delta}\ln BP_{b} \right\}}{\sum_{b'} \exp\left\{ \frac{1}{\delta}(\boldsymbol{\mu} + \boldsymbol{\gamma}^T \mathbf{x})^T \mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T \boldsymbol{\Omega} \mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'} \right\} M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'} \right)} f(\boldsymbol{\eta}) d\boldsymbol{\eta}
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'}\right)}\int\exp\left\{\frac{1}{\delta}\boldsymbol{\eta}^T\mathbf{z}_b\right\}f(\boldsymbol{\eta})d\boldsymbol{\eta}\\
=&\frac{\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_b\right)}{\sum_{b'}\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}M_{\boldsymbol{\eta}}\left(\frac{1}{\delta}\mathbf{z}_{b'}\right)}\\
=&\frac{\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Omega}\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\boldsymbol{\Sigma}\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Omega}\mathbf{z}_{b'}+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\boldsymbol{\Sigma}\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}}\\
=&\frac{\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_b^T\big(\delta\boldsymbol{\Omega}+\boldsymbol{\Sigma}\big)\mathbf{z}_b-\frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}(\boldsymbol{\mu}+\boldsymbol{\gamma}^T\mathbf{x})^T\mathbf{z}_b+\frac{1}{2\delta^2}\mathbf{z}_{b'}^T\big(\delta\boldsymbol{\Omega}+\boldsymbol{\Sigma}\big)\mathbf{z}_{b'}-\frac{1}{\delta}\ln BP_{b'}\right\}}
\end{aligned}
$$

Because of configuration data has only one observation per participant, we set $\boldsymbol{\Omega}=\mathbf{0}$. Following Chapter 2, we will have to calibrate the above model so that it is identifiable.

APPENDIX D
TRACE PLOTS FOR HB MNL MODEL

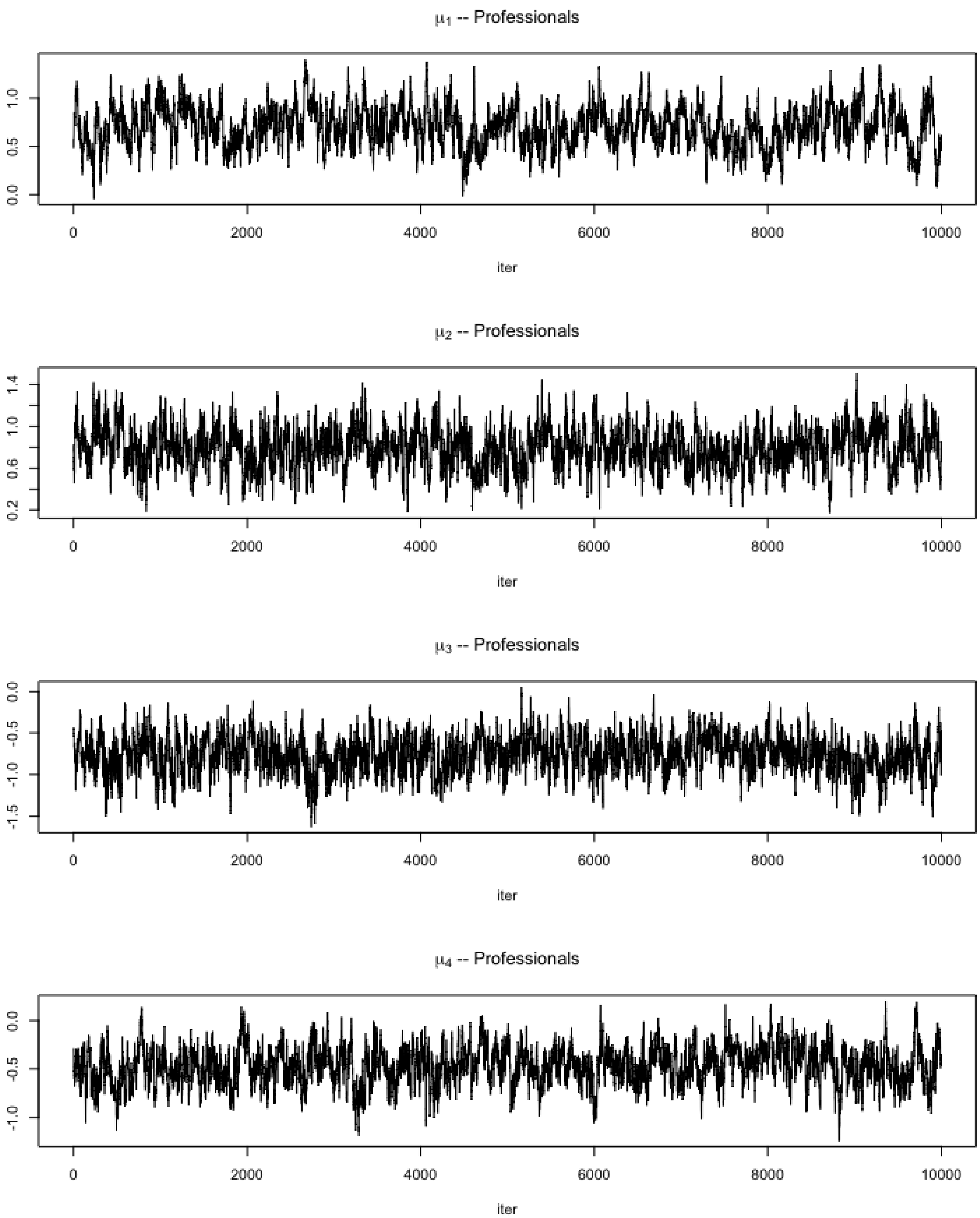


Figure D.1: Trace Plots for $\boldsymbol{\mu}$ – Professionals

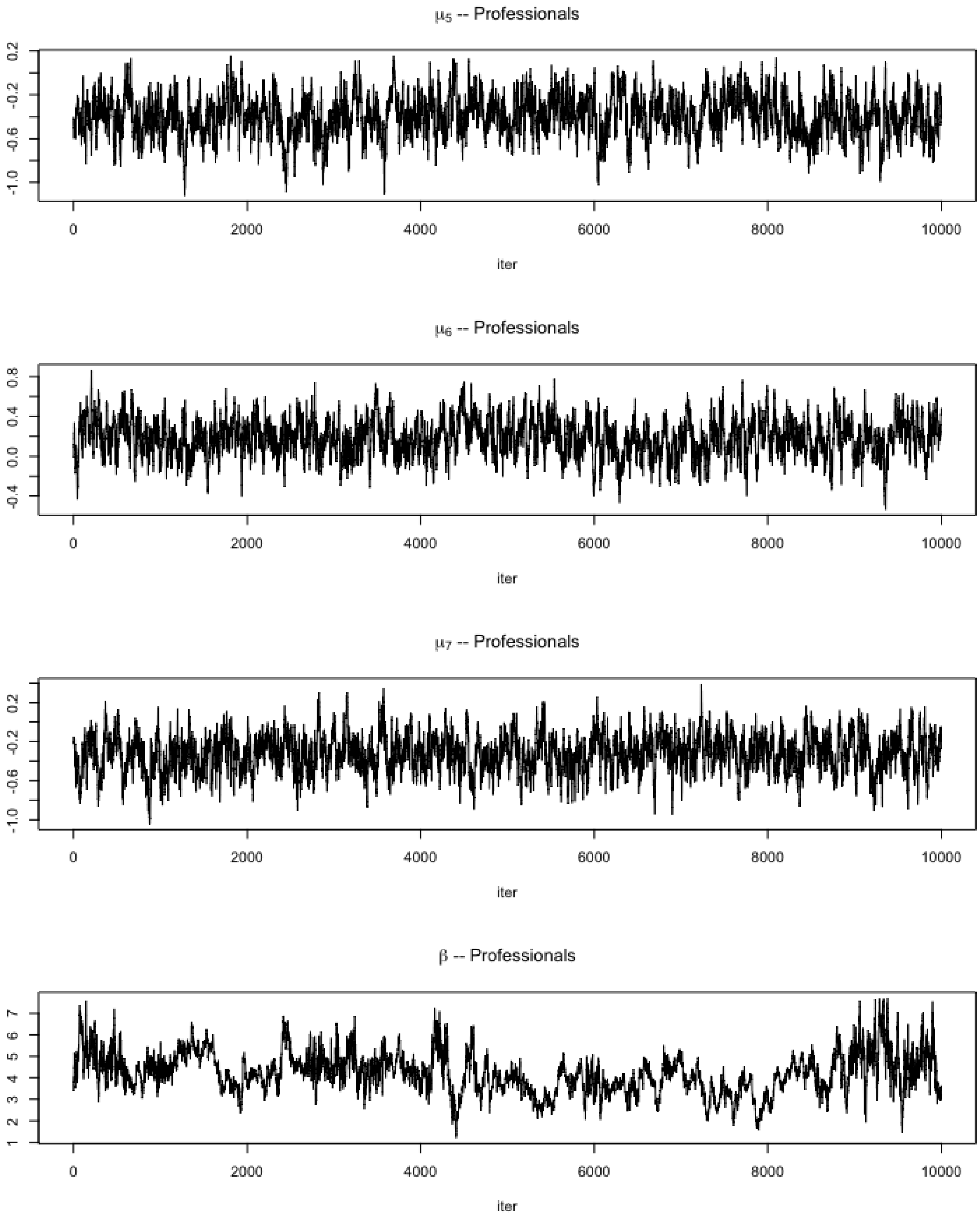


Figure D.2: Trace Plots for $\boldsymbol{\mu}$ and β – Professionals

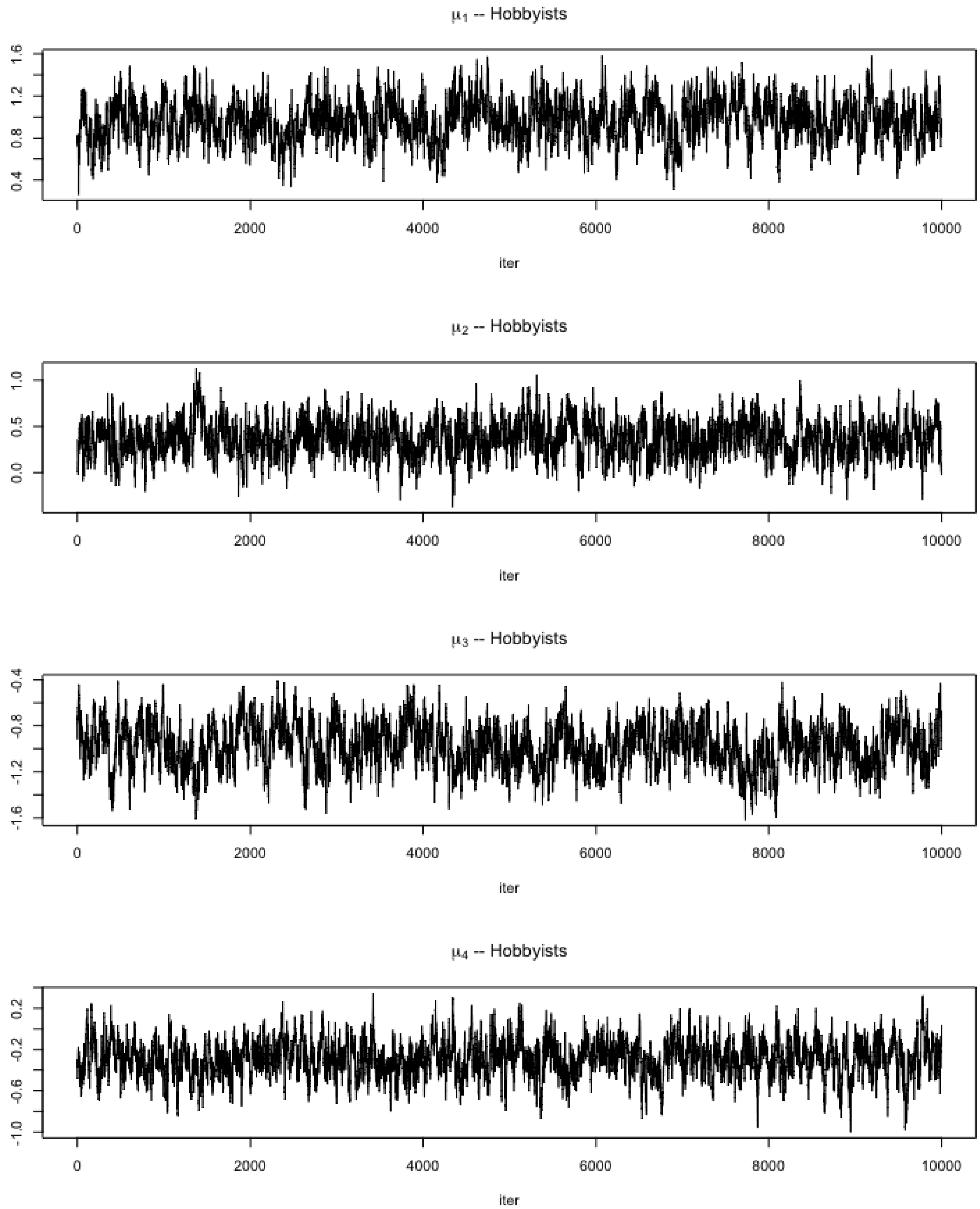


Figure D.3: Trace Plots for $\boldsymbol{\mu}$ – Hobbyists

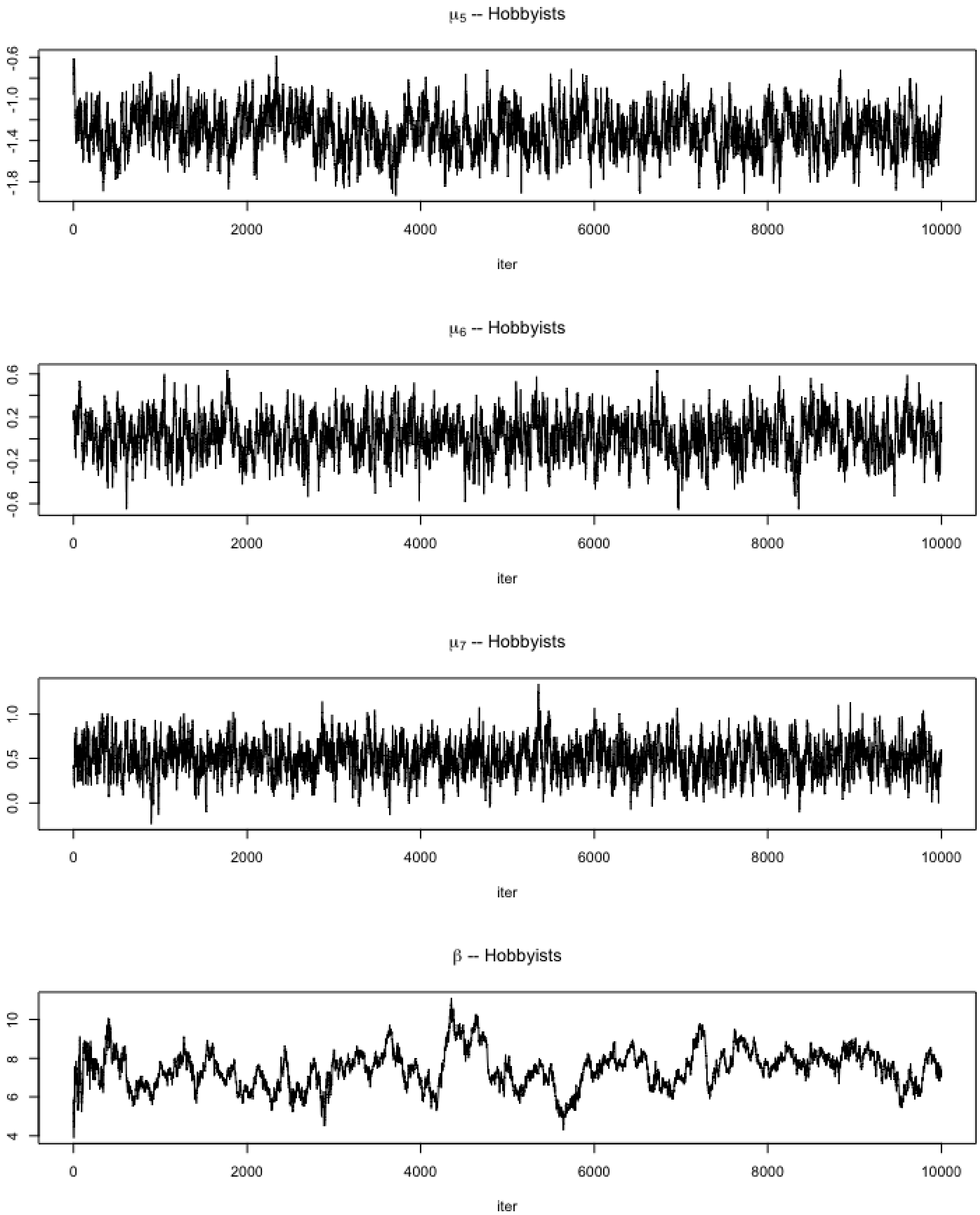


Figure D.4: Trace Plots for $\boldsymbol{\mu}$ and β – Hobbyists

APPENDIX E
DERIVATION OF BUNDLE PRICE ELASTICITY IN MNL MODEL

In the HB MNL model, the choice probability of bundle b for individual i at time t is

$$\begin{aligned} Pr_{it}(b) &= \frac{\exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_b - \beta_i \log BP_b\right\}}{\sum_{b' \in \mathcal{C}_{it}} \exp\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\}} \\ &= \frac{\exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_b - \beta_i \log BP_b\right\}}{\sum_{b' \in \mathcal{C}_{it}^1} \exp\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\} + \sum_{b' \in \mathcal{C}_{it}^0} \exp\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\}} \\ &= \frac{\exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_b - \beta_i \log BP_b\right\}}{\sum_{b' \in \mathcal{C}_{it}^1} \exp\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\} + C} \end{aligned} \tag{E.1}$$

where $\boldsymbol{\alpha}_i$ is the preferences for products in the bundle, $\mathbf{z}_b$ is a vector of binary variables representing the bundle content, and $\log BP_b$ is the log of bundle price for bundle b. $\mathcal{C}_{it}$ is the choice set individual i faces at time t. Denote $\mathcal{C}_{it}^1$ as the set of bundles that contain bundle b as as subset, and $\mathcal{C}_{it}^0$ as the set of bundles that do not include bundle b as a subset. Because the utility for bundles in $\mathcal{C}_{it}^0$ is not affected by bundle b's price change, it is viewed as constant to $\log BP_b$ and is denoted as C.

The own elasticity of bundle b is

$$\begin{aligned} e_{it}(b) &= \frac{\partial \log Pr_{it}(b)}{\partial \log BP_b} \\ &= \frac{\partial}{\partial \log BP_b} \left(\boldsymbol{\alpha}_i^T \mathbf{z}_b - \beta_i \log BP_b - \log \left(\sum_{b' \in \mathcal{C}_{it}^1} \exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\right\} + C \right) \right) \\ &= -\beta_i + \beta_i \frac{\sum_{b' \in \mathcal{C}_{it}^1} \exp\left\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\right\}}{\sum_{b' \in \mathcal{C}_{it}^1} \exp\{\boldsymbol{\alpha}_i^T \mathbf{z}_{b'} - \beta_i \log BP_{b'}\} + C} \\ &= -\beta_i \left(1 - Pr_{it}(b)\right) \end{aligned} \tag{E.2}$$

Define the aggregate choice probability for bundle b as

$$Pr_t(b) = \frac{\sum_i Pr_{it}(b)}{N} \tag{E.3}$$

where N is the number of individuals. The elasticity with respect to the aggregate choice share at time t is

$$\begin{aligned} e_t(b) = \frac{\partial \log Pr_t(b)}{\partial \log BP_b} &= \frac{1}{Pr_t(b)} \frac{\partial Pr_t(b)}{\partial \log BP_b} \\ &= \frac{1}{Pr_t(b)} \frac{\partial \left(\sum_i Pr_{it}(b)\right)}{N \partial \log BP_b} \\ &= \frac{1}{\sum_i Pr_{it}(b)} \sum_i \frac{\partial Pr_{it}(b)}{\partial \log BP_b} \\ &= \frac{1}{\sum_i Pr_{it}(b)} \sum_i Pr_{it}(b) e_{it}(b) \\ &= -\frac{\sum_i \beta_i Pr_{it}(b) \left(1 - Pr_{it}(b)\right)}{\sum_i Pr_{it}(b)} \end{aligned} \tag{E.4}$$

APPENDIX F
MODEL FOR COMBINING CBC AND CONFIGURATION DATA

We consider a consumer's decision marking process shown in Figure F.1. This procedure assumes that the bundle choices under both configuration and conjoint choice tasks are influenced by the same factors - preferences for products, product interdependencies and prices, as well as the bundle discount.

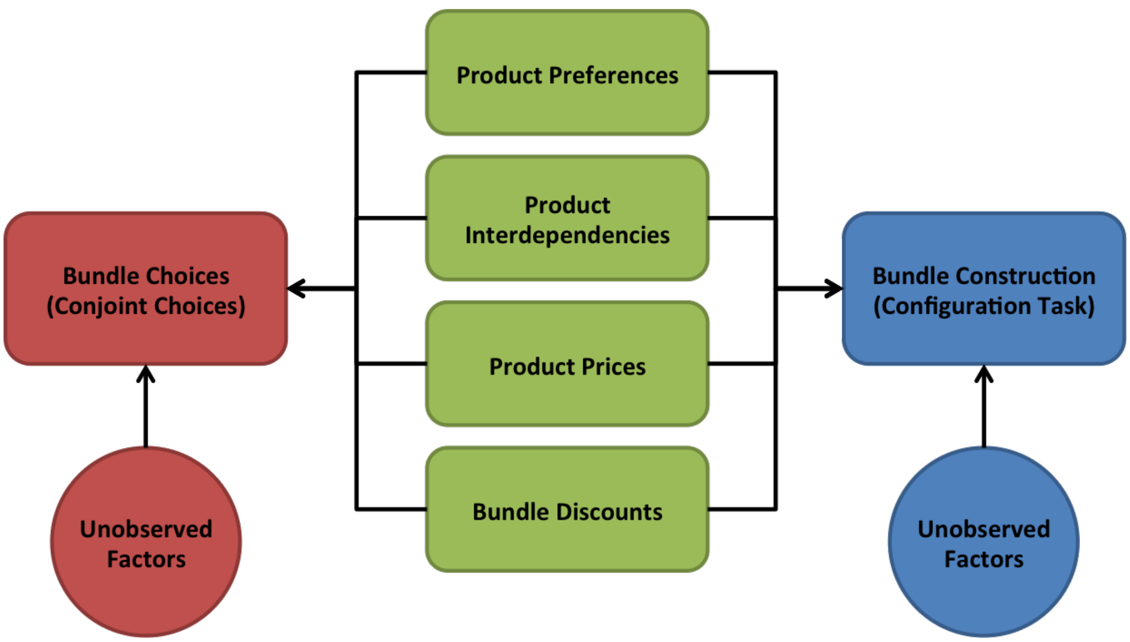


Figure F.1: Decision Making Process

This process assumes that consumers evaluate a bundle based on his preferences for the products in the bundle, the interactions among those products, and the price. He then chooses the bundle that gives him the maximized utility. The underlying mechanism governing the decision process is the same across all choice tasks. The interactions among products are the same across consumers. That is, consumers share a consensus about how the products interact with each other. The random errors for each decision context are different and are independent to each other. We now first derive the model for the general decision making process, then we calibrate the model under different choice tasks.

Suppose there are J products to choose from, resulting in a total of 2^J bundles, including the bundle consists of nothing. Let $b = 1, \ldots, B$ denote the bundle, where $b = 1$ represents the null bundle and $B = 2^J$ is the total number of bundles. For consumer i we define the direct utility function for the bundle category at time t as follows:

$$u(x) = \sum_b \psi_{bt} x_b$$

where x_b is the quantity for bundle b. Define P_j as the price for tool j and $z_{jb} = 1$ if tool j is in bundle b and 0 otherwise. The total price for bundle b, TP_b, is $TP_b = \sum_j P_j z_{jb}$. Let d_b be the quantity discount for bundle b. Denote $BP_b = TP_b(1-d_b) = (\sum_j P_j z_{jb})(1-d_b)$ as the bundle price for bundle b after quantity discount. If E is the total expenditure on the bundles, the utility maximization decision is:

$$\max\ u(x) \ \text{ subject to } \sum_b BP_b x_b \leq E$$

Define the marginal utility for bundle b as $\psi_{bt} = \psi_b e^{\varepsilon_{bt}}$. ψ_b is the deterministic component of the marginal utility, and ε_{bt} is a random element which represents factors that influence the consumer's choice but are unobserved to researchers. Applying the Kuhn-Tucker first-order condition, the choice probability for bundle b is

$$Pr(b) = Pr\left(\frac{\psi_{bt}}{BP_b} > \frac{\psi_{b't}}{BP_{b'}}\right)$$

$$= Pr\left(\ln\psi_b - \ln BP_b + \varepsilon_b > \ln\psi_{b'} - \ln BP_{b'} + \varepsilon_{b'} \text{ for any } b' \neq b\right)$$

We assume that ε_{bt} follows a Gumbel (0, δ) distribution, and integrate over ε_{bt}, the choice probability for bundle b thus is

$$Pr(b \mid \psi, \delta) = \frac{\exp\left\{\frac{\ln\psi_b - \ln BP_b}{\delta}\right)}{\sum_{b'}\exp\left\{\frac{\ln\psi_{b'} - \ln BP_{b'}}{\delta}\right\}} = \frac{\exp\left\{\frac{1}{\delta}\ln\psi_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\ln\psi_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \tag{F.1}$$

Suppose consumers acknowledge the interactions among tools. Following Song and Chintagunta (2006), the deterministic marginal utility for bundle b is defined as

$$\psi_b = \exp\left\{\sum_j \alpha_j z_{jb} + \sum_{j'>j}\sum \theta_{jj'} z_{jb} z_{j'b}\right\}$$

where $\theta_{jj'}$ is the interaction between tools j and j' with the properties $\theta_{jj} = 0$ and $-\infty < \theta_{jj'} = \theta_{j'j} < \infty$. With the above expression we rewrite equation (F.1) as

$$\begin{aligned}
Pr(b \mid \boldsymbol{\alpha}) &= \frac{\exp\left\{\frac{1}{\delta}\ln\psi_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\ln\psi_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \\
&= \frac{\exp\left\{\frac{1}{\delta}\sum_j \alpha_j z_{jb} + \frac{1}{\delta}\sum_{j'>j}\sum \theta_{jj'} z_{jb} z_{j'b} - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\sum_j \alpha_j z_{jb'} + \frac{1}{\delta}\sum_{j'>j}\sum^J \theta_{jj'} z_{jb'} z_{j'b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \\
&= \frac{\exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_b + \frac{1}{2\delta}\mathbf{z}_b^T\boldsymbol{\Theta}\mathbf{z}_b - \frac{1}{\delta}\ln BP_b\right\}}{\sum_{b'}\exp\left\{\frac{1}{\delta}\boldsymbol{\alpha}^T\mathbf{z}_{b'} + \frac{1}{2\delta}\mathbf{z}_{b'}^T\boldsymbol{\Theta}\mathbf{z}_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}} \\
&= \frac{\exp\{A_b\}}{\sum_{b'}\exp\{A_{b'}\}}
\end{aligned}$$

The above equation is the basic individual choice model for bundles. Because not all products are presented in both conjoint and configuration studies. Let $\boldsymbol{\tau}$ represent the preferences for products that are common across studies. Also, let $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ be the preferences for products that are unique to configuration and conjoint studies, respectively. Thus $\boldsymbol{\alpha} =$

$\begin{bmatrix}\boldsymbol{\tau}^T & \boldsymbol{\lambda}^T & \boldsymbol{\eta}^T\end{bmatrix}^T$. Consequently, the interaction matrix $\boldsymbol{\Theta}$ can be decomposed as

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_{\boldsymbol{\tau}} & \boldsymbol{\Theta}_{\boldsymbol{\tau\lambda}} & \boldsymbol{\Theta}_{\boldsymbol{\tau\eta}} \\ & \boldsymbol{\Theta}_{\boldsymbol{\lambda}} & \boldsymbol{\Theta}_{\boldsymbol{\lambda\eta}} \\ & & \boldsymbol{\Theta}_{\boldsymbol{\eta}} \end{bmatrix}$$

$\boldsymbol{\Theta}_{\boldsymbol{\tau}}$, $\boldsymbol{\Theta}_{\boldsymbol{\lambda}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\eta}}$ are limited to be symmetric with 0 as the diagonal elements. $\boldsymbol{\Theta}_{\boldsymbol{\tau\lambda}}$, $\boldsymbol{\Theta}_{\boldsymbol{\tau\eta}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\lambda\eta}}$ are matrices without any restrictions.

With the above setup, A_b can be expressed as

$$\begin{aligned} & A_b \\ =& \frac{1}{\delta}\begin{bmatrix}\boldsymbol{\tau}^T & \boldsymbol{\lambda}^T & \boldsymbol{\eta}^T\end{bmatrix}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b} \\ \mathbf{z}_{\boldsymbol{\lambda}b} \\ \mathbf{z}_{\boldsymbol{\eta}b}\end{bmatrix} + \frac{1}{2\delta}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b}^T & \mathbf{z}_{\boldsymbol{\lambda}b}^T & \mathbf{z}_{\boldsymbol{\eta}b}^T\end{bmatrix}\begin{bmatrix} \boldsymbol{\Theta}_{\boldsymbol{\tau}} & \boldsymbol{\Theta}_{\boldsymbol{\tau\lambda}} & \boldsymbol{\Theta}_{\boldsymbol{\tau\eta}} \\ & \boldsymbol{\Theta}_{\boldsymbol{\lambda}} & \boldsymbol{\Theta}_{\boldsymbol{\lambda\eta}} \\ & & \boldsymbol{\Theta}_{\boldsymbol{\eta}} \end{bmatrix}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b} \\ \mathbf{z}_{\boldsymbol{\lambda}b} \\ \mathbf{z}_{\boldsymbol{\eta}b}\end{bmatrix} - \frac{1}{\delta}\ln BP_b \end{aligned}$$

In the configuration study, we can only observe $\boldsymbol{\tau}$, $\boldsymbol{\lambda}$, $\boldsymbol{\Theta}_{\boldsymbol{\tau}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\tau\lambda}}$, thus the individual choice probability for the configuration study is

$$Pr(Y_b^M = 1) = \frac{\exp\left\{A_b^M\right\}}{\sum_{b' \in \mathcal{B}^M} \exp\left\{A_{b'}^M\right\}} \tag{F.2}$$

where

$$\begin{aligned} A_b^M &= \frac{1}{\delta}\begin{bmatrix}\boldsymbol{\tau}^T & \boldsymbol{\lambda}^T\end{bmatrix}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b} \\ \mathbf{z}_{\boldsymbol{\lambda}b}\end{bmatrix} + \frac{1}{2\delta}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b}^T & \mathbf{z}_{\boldsymbol{\lambda}b}^T\end{bmatrix}\begin{bmatrix} \boldsymbol{\Theta}_{\boldsymbol{\tau}} & \boldsymbol{\Theta}_{\boldsymbol{\tau\lambda}} \\ & \boldsymbol{\Theta}_{\boldsymbol{\lambda}} \end{bmatrix}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b} \\ \mathbf{z}_{\boldsymbol{\lambda}b}\end{bmatrix} - \frac{1}{\delta}\ln BP_b \\ &= \frac{1}{\delta}\left(\boldsymbol{\alpha}^M\right)^T \mathbf{z}_b^M + \frac{1}{2\delta}\left(\mathbf{z}_b^M\right)^T \boldsymbol{\Theta}^M \mathbf{z}_b^M - \frac{1}{\delta}\ln BP_b \end{aligned} \tag{F.3}$$

Similarly for the conjoint study,

$$Pr(Y_b^C = 1) = \frac{\exp\left\{A_b^C\right\}}{\sum_{b' \in \mathcal{B}^C} \exp\left\{A_{b'}^C\right\}} \tag{F.4}$$

where

$$A_b^C = \frac{1}{\tilde{\delta}}\begin{bmatrix}\boldsymbol{\tau}^T & \boldsymbol{\eta}^T\end{bmatrix}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b} \\ \mathbf{z}_{\boldsymbol{\eta}b}\end{bmatrix} + \frac{1}{2\tilde{\delta}}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b}^T & \mathbf{z}_{\boldsymbol{\eta}b}^T\end{bmatrix}\begin{bmatrix} \boldsymbol{\Theta}_{\boldsymbol{\tau}} & \boldsymbol{\Theta}_{\boldsymbol{\tau\eta}} \\ & \boldsymbol{\Theta}_{\boldsymbol{\eta}} \end{bmatrix}\begin{bmatrix}\mathbf{z}_{\boldsymbol{\tau}b} \\ \mathbf{z}_{\boldsymbol{\eta}b}\end{bmatrix} - \frac{1}{\tilde{\delta}}\ln BP_b \tag{F.5}$$

$$= \frac{1}{\tilde{\delta}} \left(\boldsymbol{\alpha}^C\right)^T \mathbf{z}_b^C + \frac{1}{2\tilde{\delta}} \left(\mathbf{z}_b^C\right)^T \boldsymbol{\Theta}^C \mathbf{z}_b^C - \frac{1}{\tilde{\delta}} \ln BP_b$$

Consumers have different preferences for products. However, we assume that consumers hold the same notion with respect to the product interdependencies. That is, the view on what products work best together toward a specific goal is agreed by all consumers. Therefore the heterogeneity is reflected on the preference parameters $\boldsymbol{\alpha}$. Following the previous definition of the preference distribution for the sample which includes only bundle fans, we assume that the preference distribution of bundle fans $g(\boldsymbol{\alpha})$ is weighted by $w(\boldsymbol{\alpha})$, which is proportional to consumers' inclusive value.

One issue to consider here is that the weighting function $w(\boldsymbol{\alpha})$ depends on the inclusive value, which depends on the definition of the choice sets. Here the choice sets in configuration and conjoint datasets are different, due to some products uniquely appear in one of the dataset. This give rise to three different definitions of the inclusive value: 1). The inclusive value defines in terms of all possible bundles in both configuration and conjoint choice datasets $IV(\boldsymbol{\alpha})$; 2). The inclusive value defines in terms of the choice sets in configuration study $IV(\boldsymbol{\alpha}^M)$; and 3). The inclusive value defines in terms of the choice sets in conjoint study $IV(\boldsymbol{\alpha}^C)$. Previously in the configuration data analysis we define inclusive value using the second definition $IV(\boldsymbol{\alpha}^M)$. We will use the same definition here in order to be consistent with previous analysis. We can conduct a sensitive analysis to see if different definitions of inclusive values give rise to different estimation results.

Suppose the market preference distribution for products $f(\boldsymbol{\alpha})$ follows multivariate normal distribution. The derivation of $w(\boldsymbol{\alpha}^M)$ is shown in Appendix A. We can obtain the bundle fan preference distribution as follows:

$$\boldsymbol{\alpha} \sim g(\boldsymbol{\alpha}) = w(\boldsymbol{\alpha}^M) f(\boldsymbol{\alpha}) \tag{F.6}$$

$$f(\boldsymbol{\alpha}) = MVN\left(\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\tau}} \\ \boldsymbol{\mu}_{\boldsymbol{\lambda}} \\ \boldsymbol{\mu}_{\boldsymbol{\eta}} \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\tau} & \boldsymbol{\Sigma}_{\tau\boldsymbol{\lambda}} & \boldsymbol{\Sigma}_{\tau\boldsymbol{\eta}} \\ & \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} & \boldsymbol{\Sigma}_{\boldsymbol{\lambda}\boldsymbol{\eta}} \\ & & \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \end{bmatrix}\right) \tag{F.7}$$

$$w(\boldsymbol{\alpha}^M) = \frac{\sum_{b' \in \mathcal{B}^M} \exp\left\{\frac{1}{\delta}\left(\boldsymbol{\alpha}^M\right)^T \mathbf{z}_{b'}^M + \frac{1}{2\delta}\left(\mathbf{z}_{b'}^M\right)^T \boldsymbol{\Theta}^M \mathbf{z}_{b'}^M - \frac{1}{\delta}\ln BP_{b'}\right\}}{\sum_{b' \in \mathcal{B}^M} \exp\left\{\frac{1}{2\delta}\left(\mathbf{z}_{b'}^M\right)^T \boldsymbol{\Theta}^M \mathbf{z}_{b'}^M - \frac{1}{\delta}\ln BP_{b'}\right\} M_{\boldsymbol{\alpha}^M}\left(\frac{1}{\delta}\mathbf{z}_{b'}^M\right)} \tag{F.8}$$

where $\mathcal{B}^M$ represents the set the contains all possible bundles in a configuration study.

Because the decision making processes are assumed to be the same for configuration and conjoint studies, parameters $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ in equations (F.3) and (F.5) are also the same. However, the estimation cannot tell apart those parameters with scaling parameters; they are confounded with δ and $\tilde{\delta}$. To impose the equality condition on those parameters we need to control for the differences in the scaling parameters between data sources so that the model is estimable. Without loss of generality, we normalize $\tilde{\delta}$ to be 1.

We also need to calibrate the preference correlation matrix $\boldsymbol{\Sigma}$ and the product interdependence matrix $\boldsymbol{\Theta}$. Because $\boldsymbol{\Sigma}$ can take on infinitely many values resulting the same preference distribution, we normalized it to be a correlation matrix to make the model identifiable. Thus the diagonal entries of $\boldsymbol{\Sigma}$ are set to 1 and the rest of the elements are limited to between -1 and 1. We rewrite $\boldsymbol{\Theta} = \rho\boldsymbol{\Theta}^*$ where $\rho > 0$ and $\boldsymbol{\Theta}^*$ is a symmetric matrix with 0s on the diagonal and $-1 \leq \theta^*_{jj'} = \theta^*_{j'j} \leq 1$.

At this point, the number of parameters for estimation is large. Let $n_{\boldsymbol{\tau}}$, $n_{\boldsymbol{\lambda}}$ and $n_{\boldsymbol{\eta}}$ represent the number of parameters that are common and unique to configuration and conjoint datasets, respectively. There will be $n_{\boldsymbol{\tau}} + n_{\boldsymbol{\lambda}} + n_{\boldsymbol{\eta}}$ mean preference parameters, $\frac{(n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}})(n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}}-1)}{2}$ preference correlation parameters, $\frac{(n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}})(n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}}-1)}{2} + 1$ product interdependence parameters, and one scaling parameters for estimation. If $n_{\boldsymbol{\tau}} = n_{\boldsymbol{\lambda}} = n_{\boldsymbol{\eta}} = 1$, this model will result in ten parameters for estimation. To further reduce the number of parameters, we project the preference correlation matrix onto a two dimensional map – the same approach used in configuration data analysis.

Let

$$\mathbf{V} = \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vec{v}_J \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ \vdots & \vdots \\ v_{J1} & v_{J2} \end{bmatrix}$$

Define $\mathbf{\Sigma} = \mathbf{V}\mathbf{V}^T$, where $\vec{v}_j\vec{v}_j^T = 1$ and $|\vec{v}_j\vec{v}_k^T| \leq 1$ for $j \neq k$. We define $v_{j1} = \cos(\phi_j^V)$, $v_{j2} = \sin(\phi_j^V)$ because trigonometric functions have properties that are consistent with the restrictions on $\vec{v}_j$. However, there are infinitely many solutions for ϕ_j^V that satisfy the restrictions. To make sure that we have only one unique solution, we further define $\phi_j^V = \frac{2\pi}{\exp(a_j^V)+1}$, where a_j^V is the parameter to estimate on the real line. Thus ϕ_j^V is always between 0 and 2π. Next, to fix the scale and the direction of the map, we set $v_{11} = 1$ and $v_{12} = 0$. Lastly, to prevent the axes to be flipped, we set $\phi_2^V = \frac{\pi}{\exp(a_2^V)+1}$. Replacing $\mathbf{\Sigma}$ with $\mathbf{V}\mathbf{V}^T$, we reduce the number of parameters from $\frac{(n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}})(n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}}-1)}{2}$ to $n_{\boldsymbol{\tau}}+n_{\boldsymbol{\lambda}}+n_{\boldsymbol{\eta}}-1$.

Similarly for $\mathbf{\Theta}^*$, we define $\mathbf{\Theta}^* = \mathbf{W}\mathbf{W}^T$, where

$$\mathbf{W} = \begin{bmatrix} \vec{w}_1 \\ \vec{w}_2 \\ \vdots \\ \vec{w}_J \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ \vdots & \vdots \\ w_{J1} & w_{J2} \end{bmatrix}$$

We let $w_{11} = 1$, $w_{12} = 0$, $w_{j1} = \cos\left(\phi_j^W\right)$ and $w_{j2} = \sin(\phi_j^W)$. $\phi_2^W = \frac{\pi}{\exp(a_2^W)+1}$ and $\phi_j^W = \frac{2\pi}{\exp(a_j^W)+1}$ for $j > 2$. Lastly, we force $\vec{w}_j\vec{w}_j^T = 0$ for all j.

In the configuration task, the number of the possible bundles is 2^J. That is, as the size of the menu grows, the number of possible bundles grows geometrically. We adopt the sampling of alternatives approach proposed by McFadden (1978). Let $\tilde{\mathcal{B}}^M$ be the subset from the full choice set. Thus equation (F.2) becomes

$$Pr(Y^M_{b|\tilde{\mathcal{B}}^M} = 1) = \frac{\exp\left\{A^M_b - \ln\gamma_b\right\}}{\sum_{b'\in\tilde{\mathcal{B}}^M}\exp\left\{A^M_{b'} - \ln\gamma_{b'}\right\}} \tag{F.9}$$

The full model after model calibration is:

$$\begin{aligned}
Pr(Y^M_{b|\tilde{\mathcal{B}}^M} = 1) &= \frac{\exp\left\{A^M_b - \ln\gamma_b\right\}}{\sum_{b'\in\tilde{\mathcal{B}}^M}\exp\left\{A^M_{b'} - \ln\gamma_{b'}\right\}} \\
A^M_b &= \frac{1}{\delta}\left(\boldsymbol{\alpha}^M\right)^T\mathbf{z}^M_b + \frac{\rho}{2\delta}\left(\mathbf{z}^M_b\right)^T\mathbf{W}^M\left(\mathbf{W}^M\right)^T\mathbf{z}^M_b - \frac{1}{\delta}\ln BP_b \\
Pr(Y^C_b = 1) &= \frac{\exp\left\{A^C_b\right\}}{\sum_{b'\in\mathcal{B}^C}\exp\left\{A^C_{b'}\right\}} \\
A^C_b &= \left(\boldsymbol{\alpha}^C\right)^T\mathbf{z}^C_b + \frac{\rho}{2}\left(\mathbf{z}^C_b\right)^T\mathbf{W}^C\left(\mathbf{W}^C\right)^T\mathbf{z}^C_b - \ln BP_b \\
g(\boldsymbol{\alpha}) &= w(\boldsymbol{\alpha}^M)f(\boldsymbol{\alpha}) \\
w(\boldsymbol{\alpha}^M) &= \frac{\sum_{b'\in\mathcal{B}^M}\exp\left\{\frac{1}{\delta}\left(\boldsymbol{\alpha}^M\right)^T\mathbf{z}^M_{b'} + \frac{\rho}{2\delta}\left(\mathbf{z}^M_{b'}\right)^T\mathbf{W}^M\left(\mathbf{W}^M\right)^T\mathbf{z}^M_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}}{\sum_{b'\in\mathcal{B}^M}\exp\left\{\frac{\rho}{2\delta}\left(\mathbf{z}^M_{b'}\right)^T\mathbf{W}^M\left(\mathbf{W}^M\right)^T\mathbf{z}^M_{b'} - \frac{1}{\delta}\ln BP_{b'}\right\}M_{\boldsymbol{\alpha}^M}\left(\frac{1}{\delta}\mathbf{z}^M_{b'}\right)} \\
f(\boldsymbol{\alpha}) &= MVN\left(\boldsymbol{\mu}, \mathbf{V}\mathbf{V}^T\right) \\
M_{\boldsymbol{\alpha}^M}\left(\frac{1}{\delta}\mathbf{z}^M_{b'}\right) &= \exp\left\{\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{z}_{b'} + \frac{1}{2\delta^2}\mathbf{z}^T\mathbf{V}\mathbf{V}^T\mathbf{z}\right\}
\end{aligned}$$

The likelihood for consumer i is given by

$$L(Y^C, Y^M_{\tilde{\mathcal{B}}^M}) = Pr(Y^M_{b|\tilde{\mathcal{B}}^M} = 1)\left[\prod_t Pr_t(Y^C_b = 1)\right] \tag{F.10}$$

We use a hierarchical Bayesian approach for model estimation

To estimate the model, we need to specify the prior distributions for $\boldsymbol{\mu}$, $\vec{a}^V$, $\vec{a}^W$, ρ and δ. We assume those parameters are independent to each other and use diffuse priors. We assume that $\boldsymbol{\mu} \sim f(\boldsymbol{\mu}) = MVN(0, 1000I)$, where I is an identity matrix with appropriate dimension. a^V_j and a^W_j are also assumed to follow normal distribution, $f(\vec{a}^V) =$

$f(\vec{a}^W) = MVN(0, 1000I)$ Lastly, the scaling parameters ρ and δ have to be greater than 0, and thus are assumed to follow Gamma distribution, $h(\rho) = h(\delta) = Gamma(1/1000, 1000)$.

The full posterior distribution is

$$
\begin{aligned}
&\pi(\boldsymbol{\alpha}, \boldsymbol{\mu}, \vec{a}^V, \vec{a}^W, \rho, \delta \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}) \\
&\quad \propto L(Y^C, Y^M_{\tilde{\mathcal{B}}^M} \mid \boldsymbol{\alpha}, \vec{a}^W, \rho, \delta) g(\boldsymbol{\alpha} \mid \boldsymbol{\mu}, \vec{a}^V, \vec{a}^W, \rho, \delta) f(\boldsymbol{\mu}) f(\vec{a}^V) f(\vec{a}^W) h(\rho) h(\delta)
\end{aligned}
\tag{F.11}
$$

We use Metropolis-Hastings algorithm for parameter estimation because all the posterior distributions are not standard distribution. Moreover, we use the normal random walk proposal to generate draws.

Step 0. Initialize values for $\boldsymbol{\mu}, \vec{a}^V, \vec{a}^W, \rho$ and δ

We set the initial values $\boldsymbol{\mu}^0, \vec{a}^{V,0}, \vec{a}^{W,0}, \rho^0$ and δ^0. Then we generate $\boldsymbol{\alpha}^0$ for each individual i according to the model $f(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$.

For each iteration $r = 1, \ldots, R$, we do the following:

Step 1. Draw $\boldsymbol{\alpha}^r$

(a) For each individual i calculate $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}^{r-1} + \mathbf{e}_{\boldsymbol{\alpha}}$, where $\mathbf{e}_{\boldsymbol{\alpha}}$ is the random dram from $MVN\left(\mathbf{0}, \sigma^2_{\boldsymbol{\alpha}} \mathbf{I}\right)$

(b) Calculate $\pi\left(\boldsymbol{\alpha}^*, \mid Y^C, \boldsymbol{\mu}^{r-1}, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)$

$$
\begin{aligned}
&\pi\left(\boldsymbol{\alpha}^* \mid Y^C, \boldsymbol{\mu}^{r-1}, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right) \propto \\
&\qquad L\left(Y^C \mid \boldsymbol{\alpha}^*, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right) g\left(\boldsymbol{\alpha}^* \mid \boldsymbol{\mu}^{r-1}, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)
\end{aligned}
$$

(c) Calculate $\kappa_{\boldsymbol{\alpha}}$

$$
\kappa_{\boldsymbol{\alpha}} = \min\left\{1, \frac{\pi\left(\boldsymbol{\alpha}^*, \mid Y^C, \boldsymbol{\mu}^{r-1}, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)}{\pi\left(\boldsymbol{\alpha}^{r-1}, \mid Y^C, \boldsymbol{\mu}^{r-1}, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)}\right\}
$$

(d) Draw a random variable u from $U(0,1)$

(e) Set $\boldsymbol{\alpha}^r = \boldsymbol{\alpha}^*$ if $\kappa_{\boldsymbol{\alpha}} \geq u$, otherwise $\boldsymbol{\alpha}^r = \boldsymbol{\alpha}^{r-1}$

Step 2. Draw $\boldsymbol{\mu}^r$

(a) Calculate $\boldsymbol{\mu}^* = \boldsymbol{\mu}^{r-1} + \mathbf{e}_{\boldsymbol{\mu}}$, where $\mathbf{e}_{\boldsymbol{\mu}}$ is the random dram from $MVN\left(\mathbf{0}, \sigma^2_{\boldsymbol{\mu}}\mathbf{I}\right)$

(b) Calculate $\pi\left(\boldsymbol{\mu}^* \mid Y^C, \boldsymbol{\alpha}^r, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)$

$$\pi\big(\boldsymbol{\mu}^* \mid Y^C, \boldsymbol{\alpha}^r, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\big) \propto$$

$$L\left(Y^C \mid \boldsymbol{\alpha}^r, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right) g\left(\boldsymbol{\alpha}^{r-1} \mid \boldsymbol{\mu}^*, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right) f\left(\boldsymbol{\mu}^*\right)$$

(c) Calculate $\kappa_{\boldsymbol{\mu}}$

$$\kappa_{\boldsymbol{\mu}} = \min\left\{1, \frac{\pi\left(\boldsymbol{\mu}^* \mid Y^C, \boldsymbol{\alpha}^r, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)}{\pi\left(\boldsymbol{\mu}^{r-1} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \vec{a}^{V,r-1}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right)}\right\}$$

(d) Draw a random variable u from $U(0,1)$

(e) Set $\boldsymbol{\mu}^r = \boldsymbol{\mu}^*$ if $\kappa_{\boldsymbol{\mu}} \geq u$, otherwise $\boldsymbol{\mu}^r = \boldsymbol{\mu}^{r-1}$

Step 3. Draw $\vec{a}^{V,r}$

(a) Calculate $\vec{a}^{V,*} = \vec{a}^{V,r-1} + \mathbf{e}_{\vec{a}^V}$, where $\mathbf{e}_{\vec{a}^V}$ is the random dram from $MVN\left(\mathbf{0}, \sigma^2_{\vec{a}^V}\mathbf{I}\right)$

(b) Calculate $\pi(\vec{a}^{V,*} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1})$

$$\pi(\vec{a}^{V,*} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}) \propto g\left(\boldsymbol{\alpha}^r \mid \boldsymbol{\mu}^r, \vec{a}^{V,*}, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1}\right) f\left(\vec{a}^{V,*}\right)$$

(c) Calculate $\kappa_{\vec{a}^V}$

$$\kappa_{\vec{a}^V} = \min\left\{1, \frac{\pi(\vec{a}^{V,*} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1})}{\pi(\vec{a}^{V,r-1} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{W,r-1}, \rho^{r-1}, \delta^{r-1})}\right\}$$

(d) Draw a random variable u from $U(0,1)$

(e) Set $\vec{a}^{V,r} = \vec{a}^{V,*}$ if $\kappa_{\vec{a}^V} \geq u$, otherwise $\vec{a}^{V,r} = \vec{a}^{V,r-1}$

Step 4. Draw $\vec{a}^{W,r}$

(a) Calculate $\vec{a}^{W,*} = \vec{a}^{W,r-1} + \mathbf{e}_{\vec{a}^W}$, where $\mathbf{e}_{\vec{a}^W}$ is the random dram from $MVN\left(\mathbf{0}, \sigma^2_{\vec{a}^W}\mathbf{I}\right)$

(b) Calculate $\pi(\vec{a}^{W,*} \mid Y^C, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \rho^{r-1}, \delta^{r-1})$

$$\pi(\vec{a}^{W,*} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \rho^{r-1}, \delta^{r-1}) \propto$$
$$L\left(Y^C, Y^M_{\tilde{\mathcal{B}}^M} \mid \boldsymbol{\alpha}^r, \vec{a}^W, \rho^{r-1}, \delta^{r-1}\right) g\left(\boldsymbol{\alpha}^r \mid \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^W, \rho^{r-1}, \delta^{r-1}\right) f\left(\vec{a}^W\right)$$

(c) Calculate $\kappa_{\vec{a}^W}$

$$\kappa_{\vec{a}^W} = \min\left\{1, \frac{\pi(\vec{a}^{W,*} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \rho^{r-1}, \delta^{r-1})}{\pi(\vec{a}^{W,r-1} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \rho^{r-1}, \delta^{r-1})}\right\}$$

(d) Draw a random variable u from $U(0,1)$

(e) Set $\vec{a}^{W,r} = \vec{a}^{W,*}$ if $\kappa_{\vec{a}^W} \geq u$, otherwise $\vec{a}^{W,r} = \vec{a}^{W,r-1}$

Step 5. Draw ρ^r

(a) Calculate $\rho^* = \rho^{r-1} + \mathbf{e}_\rho$, where $\mathbf{e}_\rho$ is the random dram from $MVN\left(\mathbf{0}, \sigma^2_\rho\mathbf{I}\right)$

(b) Calculate $\pi(\rho^* \mid Y^C, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \delta^{r-1})$

$$\pi(\rho^* \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \delta^{r-1}) \propto$$
$$L\left(Y^C, Y^M_{\tilde{\mathcal{B}}^M} \mid \boldsymbol{\alpha}^r, \vec{a}^{W,r}, \rho^*, \delta^{r-1}\right) g\left(\boldsymbol{\alpha}^r \mid \boldsymbol{\mu}^r, \vec{a}^V, \vec{a}^W, \rho^*, \delta^{r-1}\right) h\left(\rho^*\right)$$

(c) Calculate κ_ρ

$$\kappa_\rho = \min\left\{1, \frac{\pi(\rho^* \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \delta^{r-1})}{\pi(\rho^{r-1} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \delta^{r-1})}\right\}$$

(d) Draw a random variable u from $U(0,1)$

(e) Set $\rho^r = \rho^*$ if $\kappa_\rho \geq u$, otherwise $\rho^r = \rho^{r-1}$

Step 6. Draw δ^r

(a) Calculate $\delta^* = \delta^{r-1} + \mathbf{e}_\delta$, where $\mathbf{e}_\delta$ is the random dram from $MVN\left(\mathbf{0}, \sigma_\delta^2 \mathbf{I}\right)$

(b) Calculate $\pi(\delta^* \mid Y^C, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \rho^r)$

$$\pi(\delta^* \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \rho^r) \propto$$
$$L\left(Y^C, Y^M_{\tilde{\mathcal{B}}^M} \mid \boldsymbol{\alpha}^r, \vec{a}^{W,r}, \rho^r, \delta^*\right) g\left(\boldsymbol{\alpha}^r \mid \boldsymbol{\mu}^r, \vec{a}^V, \vec{a}^W, \rho^r, \delta^*\right) h\left(\delta^*\right)$$

(c) Calculate κ_δ

$$\kappa_\delta = \min\left\{1, \frac{\pi(\delta^* \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \rho^r)}{\pi(\delta^{r-1} \mid Y^C, Y^M_{\tilde{\mathcal{B}}^M}, \boldsymbol{\alpha}^r, \boldsymbol{\mu}^r, \vec{a}^{V,r}, \vec{a}^{W,r}, \rho^r)}\right\}$$

(d) Draw a random variable u from $U(0,1)$

(e) Set $\delta^r = \delta^*$ if $\kappa_\delta \geq u$, otherwise $\delta^r = \delta^{r-1}$

We repeat steps 1 to 6 until the estimation converges. The choice of the spread, or scale, of the candidate-generating density has implications for the efficiency of the algorithm (Chib and Greenberg, 1995). Considering the dimensionality in our application, We choose $\mathbf{e}_{\boldsymbol{\alpha}}$, $\mathbf{e}_{\boldsymbol{\mu}}$, $\mathbf{e}_{\vec{a}^V}$, $\mathbf{e}_{\vec{a}^W}$, $\mathbf{e}_\rho$ and $\mathbf{e}_\delta$ so that the acceptance rate is around .30.

REFERENCES

Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *The Quarterly Journal of Economics*, 475:475–498.

Agarwal, J., Desarbo, W. S., Malhotra, N. K., and Rao, V. R. (2015). An interdisciplinary review of research in conjoint analysis: Recent developments and directions for future research. *Customer Needs and Solutions*, 2(1):19–40.

Allenby, G. M. and Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research*, 32(4):392 – 403.

Armstrong, M. and Vickers, J. (2010). Competitive non-linear pricing and bundling. *Review of Economic Studies*, 77(1):30 – 60.

Balachander, S., Ghosh, B., and Stock, A. (2010). Why bundle discounts can be a profitable alternative to competing on price promotions. *Marketing Science*, 29(4):624 – 638.

Basu, A. and Vitharana, P. (2009). Impact of customer knowledge heterogeneity on bundling strategy. *Marketing Science*, 28(4):792 – 801.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., and Rao, V. (1994). Combining revealed and stated preferences data. *Marketing Letters*, 5(4):335 –349.

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, volume 19. MIT Press.

Ben-Akiva, M. and Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, 24(6):485 – 495.

Brownstone, D., Bunch, D. S., and Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5):315 – 338.

Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327 – 335.

Chung, J. and Rao, V. R. (2003). A general choice model for bundles with multiple-category products: Application to market segmentation and optimal pricing for bundles. *Journal of Marketing Research*, 40(2):115 – 130.

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied Statistics (Journal of the Royal Statistical Society, Series C)*, 21(2):113 – 120.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Joen Wiley and Sons, New York.

Eppen, G. D., Hanson, W. A., and Martin, R. K. (1991). Bundling – new products, new markets, low risk. *MIT Sloan Management Review*, 32(4):7.

Farquhar, P. H. and Rao, V. R. (1976). A balance model for evaluating subsets of multiattributed items. *Management Science*, 2(5):528 – 539.

Feit, E. M., Beltramo, M. A., and Feinberg, F. M. (2010). Reality check: Combining choice experiments with market data to estimate the importance of product attributes. *Management Science*, 56(5):785 – 800.

Gaeth, G. J., Levin, I. P., Chakraborty, G., and Levin, A. M. (1991). Consumer evaluation of multi-product bundles: An information integration analysis. *Marketing letters*, 2(1):47–57.

Ghosh, B. and Balachander, S. (2007). Competitive bundling and counterbundling with generalist and specialist firms. *Management Science*, 55(1):159 – 168.

Gilbert, R. J. and Katz, M. L. (2001). An economist's guide to us v. microsoft. *The Journal of Economic Perspectives*, 15(2):25 – 44.

Goldberg, S. M., Green, P. E., and Wind, Y. (1984). Conjoint analysis of price premiums for hotel amenities. *Journal Of Business*, 57(1):111 – 132.

Green, P. E., Krieger, A. M., and Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3_supplement):S56–S73.

Harris, J. and Blair, E. A. (2006). Consumer preference for product bundles: The role of reduced search costs. *Journal of the Academy of Marketing Science*, 34(4):506 – 513.

Hensher, D. A. (2008). Empirical approaches to combining revealed and stated preference data: Some recent developments with reference to urban mode choice. *Research in Transportation Economics, Transit Economics*, 23(1):23 – 29.

Hensher, D. A. and Bradley, M. (1993). Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters*, 4(2):139 – 151.

Horsky, D., Misra, S., and Nelson, P. (2006). Observed and unobserved preference heterogeneity in brand-choice models. *Marketing Science*, 25(4):322–335.

Johnson, R., Orme, B., and Pinnell, J. (2006). Simulating market preference with "Build Your Own" data. In *Proceedings of the Sawtooth Software Conference*, pages 239 – 253, Sequim, WA. Sawtooth Software, Inc.

Kamakura, W. A. and Kwak, K. (2014). Menu-choice modeling. Working paper.

Kamakura, W. A. M. W., de Rosa, F., and Mazzon, J. A. (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, 20(1):45 – 65.

Kim, H. and Rao, V. R. (2016). A conjoint study of sequential vs. simultaneous bundle choices. Working paper.

Kwak, K., Duvvuri, S. D., and Russell, G. J. (2015). An analysis of assortment choice in grocery retailing. *Journal of Retailing*, 91(1):19 – 33.

Levin, I. P. and Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of consumer research*, 15(3):374–378.

Levin, I. P., Schreiber, J., Lauriola, M., and Gaeth, G. J. (2002). A tale of two pizzas: Building up from a basic product versus scaling down from a fully-loaded product. *Marketing Letters*, 13(4):335–344.

Liechty, J., Ramaswamy, V., and Cohen, S. (2001). Choice menus for mass customization: An experimental approach for analyzing customer demand with an application to a web-based information service. *Journal of Marketing Research*, 38(2):183 – 196.

Louviere, J. J., Fox, M. F., and Moore, W. L. (1993). Cross-task validity comparisons of stated preference choice models. *Marketing Letters*, 4(3):205–213.

Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge Univerisyt Press, United Kingdom.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D., and Irwin, J. (1999). Combining sources of preference data for modeling complex decision processes. *Marketing Letters*, 10(3):205 – 217.

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

Manchanda, P., Ansari, A., and Gupta, S. (1999). The 'shopping basket': A model for multicategory purchase incidence decision. *Marketing Science*, 18(2):95 – 114.

McFadden, D. (1978). Modeling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J. W., editors, *Spatial Interaction Theory and Planning Models*, pages 75 – 96. North-Holland Publishing Company, Amsterdam.

Moon, S. and Russell, G. J. (2008). Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science*, 54(1):71 – 82.

Morikawa, T. (1989). *Incorporating stated preference data in travel demand analysis*. PhD thesis, Massachusetts Institute of Technology.

Orme, B. K. (2013). Menu-based choice (mbc) for multi-check choice experiments. Retrieved March 28, 2013, from Sawtooth Software Inc: `https://www.sawtoothsoftware.com/download/mbcbooklet.pdf`.

Park, C. W., Jun, S. Y., and MacInnis, D. J. (2000). Choosing what i want versus rejecting what i do not want: An application of decision framing to product option choice decisions. *Journal of Marketing Research*, 37(2):187–202.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, V. R. (2004). Bundles of multi-attributed items: Modeling perceptions, preferences, and choice. In Albers, S., editor, *Cross-Functional Innovation Management: Perspectives from Different Disciplines*, pages 177 – 190. Gabler Verlag, New York.

Rao, V. R., Russell, G. J., Bhargava, H., Cooke, A., Derdenger, T., Kim, H., Kumar, N., Levin, I., Yu Ma, N. M., Pracejus, J., and Vehkatesh, R. (2017). Emerging trends in product bundling: Investigating consumer choice and firm behavior. Working paper.

Rice, J. and Bakken, D. G. (2006). Estimating attribute level utilities from “Design Your Own Product” data – chapter 3. In *Proceedings of the Sawtooth Software Conference*, pages 229 – 238, Sequim, WA. Sawtooth Software, Inc.

Rossi, P. (2015). *bayesm: Bayesian Inference for Marketing/Micro-Econometrics*. R package version 3.0-2.

Russell, G. J. and Petersen, A. (2000). Analysis of cross category dependence in market basket selection. *Journal of Retailing*, 76(4):467 – 392.

Sambandam, R. and Kumar, P. (2012). Product configuration: A research approach for the times. Retrieved April 28, 2017, from `http://www.trchome.com/research-knowledge/white-paper-library/450-product-configuration-a-research-approach-for-the-times-complete`.

Schott, J. R. (1997). *Matrix Analysis for Statistics*. Wiley and Sons, New YorK.

Simonin, B. L. and Ruth, J. A. (1995). Bundling as a strategy for new product introduction: Effects on customers’ reservation prices for the bundle, the new product, and its tie-in. *Journal of Business Research*, 33(3):219 – 230.

Song, I. and Chintagunta, P. K. (2006). Measuring cross-category price effects with aggregate store data. *Management Science*, 52(10):1594–1609.

Stigler, G. (1963). United states vs. loews' inc.: A note on block-booking. *Supreme Court Review*, pages 152–157.

Stremersch, S. and Tellis, G. J. (2002). Strategic bundling of products and prices: A new synthesis for marketing. *Journal of Marketing*, 66(1):55 – 72.

Swait, J. and Andrews, R. L. (2003). Enriching scanner panel models with choice experiments. *Marketing Science*, 22(4):442 – 460.

Swait, J. and Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30(3):305–314.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Venkatesh, R. and Kamakura, W. (2003). Optimal bundling and pricing under a monopoly: Contrasting complements and substitutes from independently valued products. *Journal of Business*, 76(2):211 – 231.

Venkatesh, R. and Mahajan, V. (2009). Design and pricing of product bundles: A review of normative guidelines and practical approaches. In Rao, V. R., editor, *Handbook of Pricing Research in Marketing*, pages 232 – 257. Edward Elgar Publishing Company, Northampton, MA.

Wind, J. and Mahajan, V. (1997). Editorial: issues and opportunities in new product development: an introduction to the special issue. *Journal of Marketing Research*, 34(1):1–12.

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(2):113–123.

Wittink, D. R., McLauchlan, W. G., and Seetharaman, P. (1997). Solving the number-of-attribute-levels problem in conjoint analysis. In *Proceedings of the Sawtooth Software Conference*, pages 227 – 240, Sequim, WA. Sawtooth Software, Inc.

Yadav, M. S. (1994). How buyers evaluate product bundles: A model of anchoring and adjustment. *Journal of Consumer Research*, 21(2):342–353.

Yadav, M. S. and Monroe, K. B. (1993). How buyers perceive savings in a bundle price: An examination of a bundle's transaction value. *Journal of Marketing Research*, 30(3):350 – 358.

Yang, S., Zhao, Y., Erdem, T., and Zhao, Y. (2010). Modeling the intrahousehold behavioral interaction. *Journal of Marketing Research*, 47(3):470 – 484.