Explaining Preference in Choice Modeling

Dissertation

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

By

Marc R. Dotson, B.S., M.Sc.

Graduate Program in Business Administration

The Ohio State University

2016

Dissertation Committee:

Greg M. Allenby, Advisor

Nino Hardt

Mingyu Joo

Abstract

Choice models and decision tools are broadly used in marketing to understand what consumers prefer and inform product development and product line optimization. However, explaining preference, understanding what drives consumer preferences, remains a challenge. Knowing why consumers prefer what they do helps inform targeting and promotion activities. In this dissertation, we address two related issues to improve our ability to explain preference in choice modeling. First, we incorporate variables describing the motivating state of the consumer and their beliefs about brands to identify which offerings are relevant to them and what effect this relevance has on choice. Second, we augment a choice model with a grade of membership model to account for high-level interactions among the drivers of preference to improve our ability to explain heterogeneity across consumer preferences. Together, these improvements help provide a way forward in terms of identifying the variables that describe the drivers of preference along with modeling how these variables interact.

To Rachel, Joseph, Thomas, and Sylvia

Acknowledgments

I have been the recipient of such grace and mercy over the last four years. Greg Allenby has demonstrated unfailing patience with and commitment to my development, professionally and personally. I will forever be grateful for an advisor that was willing and able to invest in the progress of his students, wherever they happened to be on the path. I may have entered the program with only one talent, but thanks to Greg I'm leaving with two.

Adam Smith has been a constant friend and confidant. Without his calm tutelage I wouldn't have learned half of what I have over these four years. Without his friendship I may not have survived them. Jeff Dotson has always been a trailblazer for me as his younger brother. His insight, instruction, and willingness to be a sounding board for the challenges of life as a husband, father, and doctoral student has saved me multiple times.

Roger Bailey, Dong Soo Kim, Nino Hardt, Mingyu Joo, and the other faculty have been gracious with their time and support. That is also true of the current and past students I have had the pleasure to associate with, including Tatiana Dyachenko, John Howell, Karthik Easwar, Chris Summers, Hyojin Lee, Amit Singh, Danny Zane, and Hyowon Kim. It has been an honor to develop relationships as colleagues and friends with this diverse cast of talented individuals.

My wife Rachel and our children – Joseph, Thomas, and Sylvia – have, as always, kept me grounded. Graduate school can at times be incredibly challenging and unkind, but they have made it all worth it. The encouragement of my parents, Paul and Wendy, and my siblings, Jeff, Jon, Sara, and Katie, has similarly kept me focused on what matters most.

Greg told me four years ago that a doctoral program is like a pressure cooker, that it requires a certain intensity applied consistently over time to produce something great. At the time I expected I would be the product – what I would learn, the skills I would acquire. However, I now realize that who I am is a small part of what this doctoral program has produced. The sacrifice and help of so many throughout these four years has instead made me part of a vast family of friends and colleagues. I am thankful to be part of this family and to have the privilege to prepare the way for others to join the fold.

Vita

November 26, 1983	Born – Phoenix, Arizona, USA
2008	B.S. *Summa Cum Laude* Political Science, Southern Utah University
2009	M.Sc. *with Merit* International Political Economy, The London School of Economics & Political Science, UK
2010-2012	Statistical Analyst and Research Associate, The Modellers, Salt Lake City, Utah, USA
2012-Present	Graduate Teaching and Research Associate, The Ohio State University

Fields of Study

Major Field: Business Administration

Table of Contents

List of Figures

List of Tables

Table **Page**

Chapter 1: Introduction

Choice models are used extensively in marketing to understand what consumers prefer. These insights have been leveraged in product development, allowing firms to create what consumers want. However, explaining preference, understanding why consumers prefer what they do and what drives their interest in particular products and product features, continues to be a challenge. Such insights are used for targeting and promotion, activities that are becoming increasingly important as firms seek to differentiate their products. In this dissertation, we look at two avenues to improve explaining preference in choice modeling. First, we incorporate variables describing product relevance. Second, we develop a method to improve our ability to explain preference heterogeneity.

In Essay 1 (Chapter 2), we account for the staged process of choice through an extended model of behavior and measure the effect of product relevance. A product is relevant if a respondent perceives it as being able to address his or her needs, where needs describe the motivating state of the individual. We investigate the effect of product relevance on choice through two mechanisms, the first in relation to preference coefficients directly and the second with respect to preference certainty. We find evidence that the effect of product relevance is manifest through preference certainty rather than its level.

In Essay 2 (Chapter 3), we address the issue of finding covariates for a model of preference heterogeneity that explain cross-sectional variation in the part-worths. We augment the choice model with a grade of membership model, part of the mixed membership class of models, to account for interactions among discrete variables that describe drivers of preference. We find improvement in model fit and inference using the covariates generated with the proposed model over competing models using standard discrete covariates.

Together, these essays improve our ability to explain preference using choice models and thus the ability of marketers to target and promote effectively. With increasing access to large-scale individual data, we need to both identify the variables that can improve our ability to explain preference as well as develop the methodology to handle the size and form of these variables. New data can help describe the current state of individuals and thus better inform what motivates them to engage in a market to begin with. The form of this new data, including text, is typically discrete, prime for the application of mixed membership models. In this regard, this dissertation provides a way forward in terms of both identifying the variables describing the drivers of preference and modeling the high-dimensional interactions among these variables.

Chapter 2: An Extended Model of Brand Choice: Incorporating Product Relevance and Perceived Efficacy

2.1 Introduction

Consumers proceed through a variety of stages prior to purchasing a product, including recognizing they have needs that may benefit from marketplace solutions, searching for appropriate offerings, and evaluating whether the offerings will help address the issues they face. Extended models of behavior that account for a staged decision process have been lauded theoretically, but have not been commonly executed. In this paper, we develop an extended model of behavior and investigate the role of relevance in utility formation and choice.

Extended models of behavior are difficult to implement because choice is a censored realization of the decision process, thus it is impossible to employ an extended model without additional data on the preceding stages. Extended models of behavior are often characterized by the principle of conditional independence, where two variables A and C are independent conditional on a third variable B. The simplest of these models employs conditional distributions to describe the joint distribution of the data. For example, $\pi(A, B, C) = \pi(A) \times \pi(B|A) \times \pi(C|B)$, which can be visualized as A $\rightarrow$ B $\rightarrow$ C where B acts as a mediating variable from A to C (Chandukala et al., 2011a). These types have been widely used in explaining the effects of advertising with variables as attention, interest, learning, and conviction as antecedents to purchase likelihood (Strong, 1925; Lavidge and Steiner, 1961; Vakratsas and Ambler, 1999). However, many models of choice involve relationships that are not easily factored into a series of conditional likelihoods, particularly when the variables in the extended model are related to other variables in a non-linear manner.

We develop an extended model of choice using data from two conjoint experiments in the pre-packaged dinner category. The first conjoint experiment has alternatives composed of benefit bundles and the second conjoint experiment is a brand-price tradeoff. Our goal is to investigate the role of product relevance on preferences and brand choice, where relevance is measured in terms of the benefits sought by an individual and their belief that the brand is capable of providing the benefit. The brand belief data is used with the benefit importances to form the brand intercept in the second conjoint experiment. Thus, our extended model of behavior employs brand belief data to both model brand choice and to identify relevant brands.

We examine two mechanisms of how product relevance might effect choice. The first mechanism specifies the effect of product relevance on preference. Depending on whether the product is relevant or not, respondents are expected to have high or low preference, respectively. The second mechanism specifies the effect of product relevance on preference certainty by modifying the scale of the random utility model's error term. This mechanism allows for an individual to be more or less certain about the preferences they form for a given product based on whether or not the product is relevant. We find evidence for the second mechanism. In other words, asking respondents about their preferences for things they don't need leads to noisier responses, not responses that indicate consistently low preference.

The intended contribution of our paper is fourfold. First, we develop an extended model of behavior and demonstrate its application. Second, we add to the literature on the importance of motivating conditions, or needs, in models of choice using stated preference data. Third, we add another behavioral explanation to the literature on heteroscedastic errors. Fourth, our analysis indicates that traditional screening criteria, where candidate respondents are identified based on category participation, may not be sufficient for product research.

The remainder of the paper will be organized as follows. We develop our extended model in Section 2.2. Section 2.3 provides details on our empirical application. In Section 2.4, we compare results from our proposed extended model and alternative extended models. Concluding remarks and research extensions are offered in Section 2.5.

2.2 Model Development

Consumers value products for the benefits they provide, and seek out products with benefits they believe will address their needs – i.e., products that are relevant (Gutman, 1982; Griffin and Hauser, 1993; Kim et al., 2009). The benefits each product provides depends on consumer beliefs. We assume that a product is relevant to an individual if it is perceived as being able to address their needs. We measure needs as the concerns and interests of a respondent engaged in behaviors related to a product category, and use brand belief data to determine which brands are viewed as efficacious, or relevant.

Before we can more fully describe our model, we first need to consider two mechanisms through which product relevance might effect choice in terms of the random utility model. We then develop our extended model for brand choice, and conclude the section with a validation of the proposed model using a simulation experiment.

2.2.1 Product Relevance and Choice

Marketing is concerned with understanding the needs of consumers and developing products that people will want to buy. Failing to understand the needs of individuals as they enter a market provides an incomplete picture regarding how products, offerings, and promotions should be designed. We examine the role of needs, or benefits sought by consumers, using the framework described in Fennell and Allenby (2014) and used by Chandukala et al. (2011b) to examine unmet demand, and Yang et al. (2002) to study the role of environmental context on choice. As illustrated below, we measure needs in specific, concrete terms that correspond to specific beliefs about the brands (cf. Bagozzi and Dholakia, 1999; van Osselaer and Janiszewski, 2012).

We investigate two mechanisms by which product relevance might impact brand choice. Consider the random utility model:

$$U_{jh} = V_{jh} + \epsilon_{jh}. \tag{2.1}$$

The utility respondent h has for alternative j consists of two components. The deterministic component is often modeled as $V_{jh} = \sum_{m=1}^{M} \beta_{mh} x_{mj}$, a function of the M attributes x_j and the associated M coefficients β_h. The random component ϵ_{jh} is thought to arise from unobservables. It is typically assumed that ϵ_{jh} is independent and identically distributed $EV(0, \sigma)$. This scale parameter for the random component represents overall preference certainty. Note that U_{jh} is a function of both the preferences β_h and the scale of the error term σ. This σ is a model parameter, though it is typical to set $\sigma = 1$ (Swait and Louviere, 1993).

The first mechanism specifies the effect of product relevance on preference as measured by the part-worths β_h. This is accomplished within the deterministic component of the random utility model by allowing the part-worths to be cross-sectionally related to a set of covariates through a random-effects model:

$$\beta_h = \Delta' z_h + \xi_h. \tag{2.2}$$

If these z_h covariates consist of a binary vector indicating active needs for respondent h, it is expected that some elements of the Δ coefficient matrix would be positive and would predict large positive coefficients in β_h when the associated needs are addressed. The matrix of coefficients Δ in Equation (2.2) maps cross-sectional variation in the variable z_h to variation in the coefficients β_h, and provides a flexible model describing correlates of preference.

The inclusion of covariates in the random-effects model after this fashion is commonplace. Allenby and Ginter (1995b) specify z_h as the demographic variables of age, income, and gender and examine their relationship to part-worths in a conjoint analysis. Rossi et al. (1996) examine the information contents of demographics in general. Lenk et al. (1996) examine the role of expertise and other variables on personal computer purchases. Chandukala et al. (2011b) examine the role of needs in explaining variation in β_h. In practice, the influence of z_h on explaining heterogeneity in β_h has not met with much success. Rossi et al. (1996) show that the inclusion of demographic variables as covariates only explains between 7 and 33% of the variability in β_h. Horsky et al. (2006) only see a 5% improvement in the log-marginal density when moving from an intercept model to a model that includes covariates, as in Equation (2.2). Similarly, Chandukala et al. (2011b) only see a 1% improvement of the log-marginal density when moving from an intercept model to a model that includes covariates. Heterogeneity in model coefficients has largely been explained by unobservable factors (i.e., ξ_h) rather than observable factors (i.e., $\Delta' z_h$).

The second mechanism specifies the effect of product relevance on preference certainty by modifying the scale term σ. This is accomplished within the random component of the random utility model by specifying the scale parameter for alternative j and respondent h as a function of i) the presence of the respondent's needs and ii) the belief that the brand for alternative j is able to address those needs:

$$\sigma_{jh} = \exp\left[\gamma \cdot I\left(\sum_{m=1}^{M} z_{mh} \geq 1\right) \cdot I\left(\sum_{m=1}^{M} b_{mjh} z_{mh} = \sum_{m=1}^{M} z_{mh}\right)\right]. \tag{2.3}$$

The covariates z_h are again a binary vector indicating active needs. B_h is a binary matrix of respondent h's brand beliefs regarding benefits, where b_{jh} is the vector from that matrix for beliefs about brand j. We assume an *a priori* one-to-one mapping between the M needs and M benefits, thus for every need included in the analysis there is a corresponding benefit that satisfies it. The indicator functions specify that when respondent h has active needs and that the product in question has a brand that the respondent believes is able to address all of their needs, γ will measure the effect of relevance on preference certainty. We exponentiate the expression in Equation (2.3) to ensure that the scale term is positive. If either of the indicator functions don't hold then $\sigma_{jh} = 1$ as is typically assumed. A negative value of γ would produce a small

σ_{jh} and thus indicate that relevance leads to more preference certainty. We expect choices among relevant options to be associated with greater preference certainty, therefore we expect the estimate of γ to be negative.

With the scale term specific to respondents and alternatives, the random component in the random utility model is no longer identically distributed (i.e., it is heteroscedastic). Extant research employs heteroscedastic errors to account for unspecified factors influencing choice probabilities (Allenby and Ginter, 1995a), to alleviate the IIA property (Bhat, 1995), to facilitate the combination of multiple data sets (Swait and Louviere, 1993; Louviere et al., 2002), and to account for other sources of heterogeneity (Fiebig et al., 2010). Modeling the scale of the error in the random utility model isn't as commonplace. The work that models the scale parameter as a function of observables focuses largely on heteroscedasticity as a behavioral phenomenon in terms of how choice experiments are conducted (Louviere et al., 2002; Salisbury and Feinberg, 2010; Dellaert et al., 2012). Our model provides another behavioral interpretation by specifying the unobserved variability of the scale parameter in terms of the effect of product relevance on preference consistency within the framework of an extended model of behavior.

2.2.2 An Extended Model of Brand Choice

Our extended model has to account for both mechanisms through which product relevance might effect choice. Without assuming identically distributed error terms, the standard multinomial logit model used for discrete choice no longer has a closed form expression. Additionally, we expect that in the conjoint survey the relevant choices will be the ones that are picked first. To ensure that we have enough data to identify γ, we anticipate using ranked data rather than first-choice only. To make use of ranked data, we employ the exploded multinomial logit model (Chapman and Staelin, 1982). The exploded multinomial logit decomposes each choice task with K alternatives into $K-1$ independent choice tasks, each with successively fewer alternatives. Ranking the Kth alternative is deterministic given the previous $K-1$.

With ranked data, we are now interested in the probability that the first *ranked* alternative, denoted by $U_{(1)}$, has a utility expression that is greater than or equal to the second *ranked* alternative, $U_{(2)}$, and so on. Thus the random utility components for the ith ranked alternative are denoted $V_{(i)h}$ and $\epsilon_{(i)h}$. The exploded multinomial logit assumes that individuals rank their most preferred alternative first, their second preferred alternative second, and so on, and that the choice probabilities for each ranking are independent. The probability of a single observed sequence of choices from K alternatives is:

$$
\begin{aligned}
&Pr(U_{(1)} > U_{(2)} > \cdots > U_{(K)})_h \\
&\qquad = \prod_{i=1}^{K-1} Pr\left(V_{(i)h} + \epsilon_{(i)h} > V_{(k)h} + \epsilon_{(k)h} \text{ for } k = i+1, \ldots, K\right) \\
&\qquad = \prod_{i=1}^{K-1} \int_{-\infty}^{\infty} \left[\prod_{k=i}^{K} F([V_{(i)h} - V_{(k)h} + \epsilon_{(i)h}]/\sigma)\right] f(\epsilon_{(i)h}/\sigma) d\epsilon_{(i)h} \\
&\qquad = \prod_{i=1}^{K-1} \frac{\exp[V_{(i)h}/\sigma]}{\sum_{k=i}^{K} \exp[V_{(k)h}/\sigma]}.
\end{aligned}
\tag{2.4}
$$

Relaxing the assumption of homoscedastic errors leads to the heteroscedastic exploded multinomial logit model. The probability of a single observed sequence of choices from K alternatives no longer has a closed form:

$$
\begin{aligned}
&Pr(U_{(1)} > U_{(2)} > \cdots > U_{(K)})_h \\
&\qquad = \prod_{i=1}^{K-1} Pr\left(V_{(i)h} + \epsilon_{(i)h} > V_{(k)h} + \epsilon_{(k)h} \text{ for } k = i+1, \ldots, K\right) \\
&\qquad = \prod_{i=1}^{K-1} \int_{-\infty}^{\infty} \left[\prod_{k=i}^{K} F([V_{(i)h} - V_{(k)h} + \epsilon_{(i)h}]/\sigma_{(k)h})\right] f(\epsilon_{(i)h}/\sigma_{(i)h}) d\epsilon_{(i)h}
\end{aligned} \tag{2.5}
$$

where each $\sigma_{(k)h}$ follows the notation for ranked alternatives:

$$
\sigma_{(k)h} = \exp\left[\gamma \cdot I\left(\sum_{m=1}^{M} z_{mh} \geq 1\right) \cdot I\left(\sum_{m=1}^{M} b_{m(k)h} z_{mh} = \sum_{m=1}^{M} z_{mh}\right)\right]. \tag{2.6}
$$

We now have the components needed to specify our extended model of brand choice. Both needs z_h and brand beliefs B_h are included as additional data in our model. The benefit evaluation (i.e., the first likelihood or L_1) utilizes the exploded multinomial logit as detailed in Equation (2.4). For S observed choice sequences from J alternatives, the likelihood is:

$$
L_1(\beta_{zh} | X^{L_1}, z_h) = \prod_{s=1}^{S} \prod_{i=1}^{J-1} \frac{\exp[\sum_{m=1}^{M} \beta_{mzh} x_{m(i)s}^{L_1}]}{\sum_{j=i}^{J} \exp[\sum_{m=1}^{M} \beta_{mzh} x_{m(j)s}^{L_1}]} \tag{2.7}
$$

where the resulting part-worths β_{zh} are benefit-specific such that $dim(\beta_{zh}) = dim(z_h)$ $= M$. The brand-price evaluation (i.e., the second likelihood or L_2) utilizes the heteroscedastic exploded multinomial logit as detailed in Equation (2.5) with the structured scale term as detailed in Equation (2.6). To create an extended model, we use brand beliefs to provide a bridge from the benefit evaluation L_1 to the brand-price tradeoff L_2 so that:

$$
L_2(\beta_{0h}, \beta_{ph}, \gamma | X^{L_2}, \beta_{zh}, z_h, B_h) \text{ where } \beta_{0(j)h} = \sum_{m=1}^{M} \beta_{mzh} b_{m(j)h}. \tag{2.8}
$$

The resulting parameters include the brand intercepts β_{0h}, the price coefficient β_{ph}, and the measure of the effect of product relevance on preference certainty γ. The brand intercepts are a sum of the benefit part-worths β_{zh} from the benefits respondent h believes each brand provides as indicated by their brand beliefs B_h. The complete likelihood expression is:

$$
\begin{aligned}
L(\beta_{0h}, \beta_{zh}, \beta_{ph}, \gamma | X^{L_1}, X^{L_2}, z_h, B_h) = {} & L_1(\beta_{zh} | X^{L_1}, z_h) \\
& \times \; L_2(\beta_{0h}, \beta_{ph}, \gamma | X^{L_2}, \beta_{zh}, z_h, B_h).
\end{aligned}
\tag{2.9}
$$

To be clear, we are making the following assumptions. We have a one-to-one mapping between needs and benefits. The brand intercepts or overall liking of each brand for each respondent is determined by the benefits each brand is perceived to provide. For an alternative to be relevant, the respondent must have one or more needs and the associated brand must be perceived as being able to address those needs. Relevance only applies when making a choice in the brand-price tradeoff data (L_2), not when evaluating benefits (L_1). Finally, this model is identified because respondents have different beliefs about each brand, making $\sum_{m=1}^{M} \beta_{mzh} b_{m(j)h}$ different for each ranked alternative.

We use a diffuse normal prior for γ and diffuse standard conjugate priors for a multivariate normal distribution of heterogeneity on $\beta_h' = [\beta_{0h}, \beta_{zh}, \beta_{ph}]$. The extended model can be expressed as a sequence of conditional distributions:

$$
\begin{aligned}
y_h^{L_1} | X^{L_1}, \beta_{zh}, z_h &\sim \text{Exploded MNL} \\
y_h^{L_2} | X^{L_2}, \beta_{0h}, \beta_{zh}, \beta_{ph}, \gamma, z_h, B_h &\sim \text{Heteroscedastic Exploded MNL} \\
\beta_h | W, \Delta, V_\beta &\sim \text{MVN}(\Delta W, V_\beta) \\
\text{vec}(\Delta | V_\beta, \overline{\Delta}, A_\Delta) &\sim \text{Normal}(\text{vec}(\overline{\Delta}), V_\beta \otimes A_\Delta^{-1}) \\
V_\beta | \nu, V &\sim \text{IW}(\nu, V) \\
\gamma | \overline{\gamma}, \sigma_\gamma^2 &\sim \text{Normal}(\overline{\gamma}, \sigma_\gamma^2)
\end{aligned}
\tag{2.10}
$$

The variables used in the extended model are summarized in Table 2.1.

Table 2.1: Extended Model Variable Descriptions

Variables	Description
M	number of needs and benefits.
N	number of brands.
B_h	$N \times M$ matrix of indicating respondent h's brand beliefs.
z_h	M-dim vector of binary variables indicating respondent h's needs.
β_h	vector of part-worths for respondent h where $\beta_h' = [\beta_{0h}, \beta_{zh}, \beta_{ph}]$.
β_{0h}	N-dim vector of brand intercepts as defined in Equation (2.8).
β_{zh}	M-dim vector of benefit part-worths.
β_{ph}	1-dim price part-worth.
γ	1-dim coefficient measuring the effect of product relevance.
W	covariates for the random effects distribution.
Δ	$W \times M + 1$ mean matrix of the random effects distribution.
V_β	$M + 1 \times M + 1$ covariance matrix of the random effects distribution.
$\overline{\Delta}$	mean for normal prior on Δ.
A_Δ	precision matrix for normal prior on Δ.
ν	degrees of freedom for IW prior on V_β.
V	scale matrix for IW prior on V_β.
$\overline{\gamma}$	mean for normal prior on γ.
σ_γ^2	variance for normal prior on γ.

2.2.3 Simulation Experiment

We validate our extended model of behavior by generating data according to the model and recovering parameters using a random-walk Metropolis-Hastings estimation algorithm. We employ Simpson's rule to numerically integrate for the heteroscedastic exploded multinomial logit. The simulation experiment matches the dimensions used in our empirical application: a single γ and 31 β_h's for each of 567 respondents. Details on generating data and the estimation procedure are provided in Appendices A and B.

We can see in Figure 2.1 that after 60,000 iterations the Markov chain converges to the true stationary (i.e., posterior) distribution. In Table 2.2 we demonstrate that we have recovered the true parameter values for γ and the mean of the model of heterogeneity over β_h, where each parameter estimate is within or near the bounds of a 95% credible interval.

2.3 Empirical Application

We employ data from a national survey of preferences for pre-packaged dinners conducted by a major packaged goods manufacturer. Because of the proprietary nature of the data, we are restricted from revealing information about the specific brands studied in the survey. A total of 567 respondents provided information on needs, benefits sought, brand beliefs, and preferences expressed in two conjoint experiments. One of the authors was involved with the sponsoring company in designing the survey and conjoint studies to be able to explore the kind of issues we address in this paper. In particular, the exploratory work that was employed to generate needs

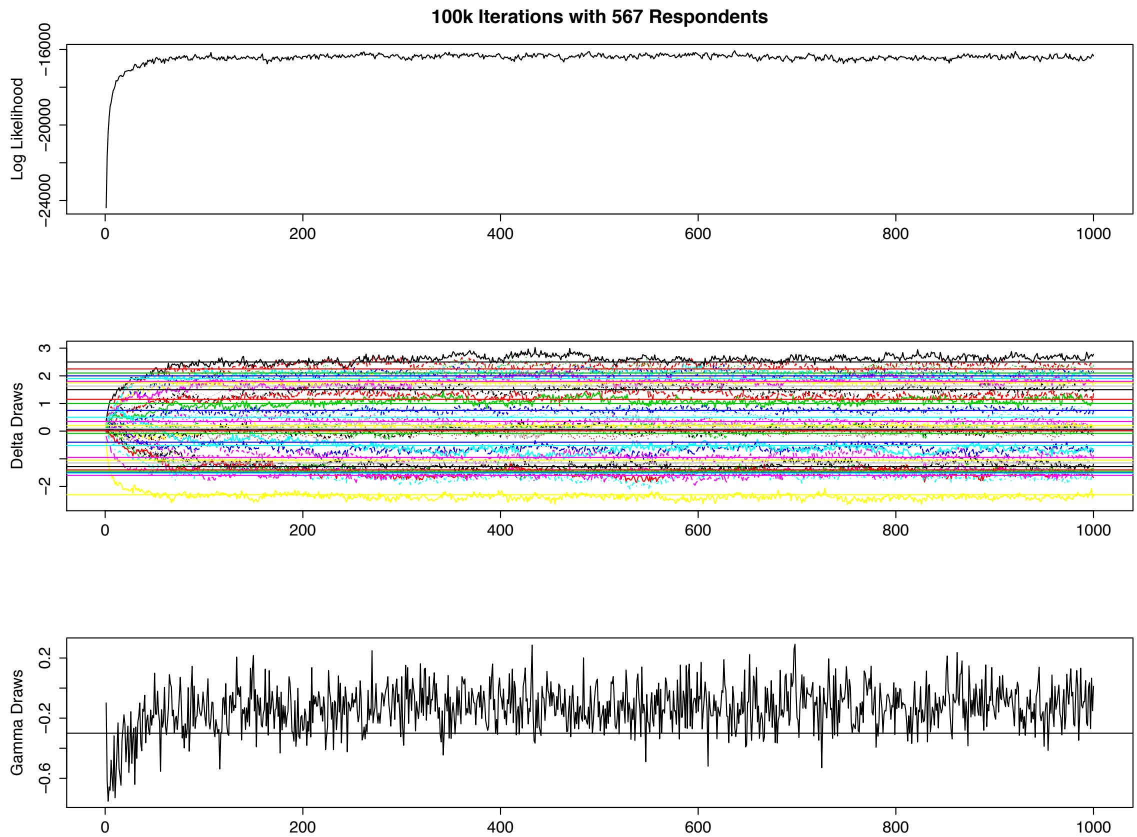

Figure 2.1: Simulation Experiment Traceplots

Table 2.2: Simulation Results and 95% Credible Intervals

Parameter	True Value	Posterior Mean	Lower	Upper
γ	-0.30	-0.11	-0.36	0.16
δ_1	2.50	2.64	2.46	2.84
δ_2	2.25	2.36	2.11	2.56
δ_3	2.10	2.24	1.99	2.45
δ_4	2.00	2.08	1.85	2.30
δ_5	1.90	2.05	1.88	2.23
δ_6	1.80	1.84	1.63	2.07
δ_7	1.75	1.79	1.61	1.97
δ_8	1.63	1.57	1.40	1.75
δ_9	1.50	1.47	1.23	1.69
δ_{10}	1.15	1.26	1.08	1.43
δ_{11}	1.00	1.10	0.91	1.28
δ_{12}	0.75	0.77	0.61	0.91
δ_{13}	0.50	0.71	0.52	0.89
δ_{14}	0.35	0.34	0.15	0.52
δ_{15}	0.20	0.16	-0.04	0.40
δ_{16}	0.10	0.12	-0.16	0.34
δ_{17}	0.05	0.05	-0.16	0.26
δ_{18}	0.01	-0.07	-0.23	0.09
δ_{19}	-0.08	0.05	-0.14	0.27
δ_{20}	-0.40	-0.63	-0.86	-0.41
δ_{21}	-0.52	-0.66	-0.88	-0.46
δ_{22}	-0.95	-0.91	-1.11	-0.68
δ_{23}	-1.05	-1.27	-1.45	-1.08
δ_{24}	-1.15	-1.08	-1.28	-0.90
δ_{25}	-1.28	-1.20	-1.42	-0.97
δ_{26}	-1.40	-1.48	-1.70	-1.25
δ_{27}	-1.45	-1.42	-1.58	-1.22
δ_{28}	-1.50	-1.46	-1.67	-1.24
δ_{29}	-1.52	-1.66	-1.83	-1.47
δ_{30}	-1.60	-1.51	-1.72	-1.28
δ_{31}	-2.30	-2.39	-2.57	-2.22

and map them to benefits makes it an ideal setting to study the effect of product relevance on choice within an extended model framework.

Prior to the conjoint experiments, respondents rated 30 potential motivating conditions or needs associated with pre-packaged dinners on a 5-point rating scale, from Not at All (1) to Completely (5) describing the respondent. We operationalize active needs for a respondent by their providing a top-box indication of the need (e.g., a "5" on a 5-point scale). We conducted a sensitivity analysis and found no difference between a top box and a top-two box indicator.

Table 2.3 lists each of the 30 needs and the 30 corresponding benefits. The needs are concrete and specific to the given purchase context without being category or even brand-specific. The needs are generated within the motivational classification framework discussed in Fennell and Allenby (2014). The class structure helps to identify qualitatively distinct types of motivating conditions within the given context. There are 7 different classes within the framework, with overarching groups of classes representing moving away from an undesirable state (classes 1 through 3), moving toward the source of motivation (classes 4 and 5), and avoiding expected excessive cost or harm (classes 6 and 7). The framework is used only to generate candidate items for inclusion in the survey. Once the needs data are collected, the general framework is not used and analysis proceeds with the responses alone.

Table 2.3: Needs and Corresponding Benefits

No.	Needs	Benefits
1	I was worried that I hadn't anything available to make a dinner.	On your shelf, always available to make a dinner.
2	I was pressed for time to make dinner.	Helps make dinner when you're pressed for time.
3	It was a day when I just didn't feel like making dinner.	Makes dinner on days when you don't feel like making dinner.
4	I was worried that I was running out of menu ideas.	Ready to hand, when you've run out of menu ideas.
5	I was too rushed/pressured preparing dinner to enjoy eating it.	Makes a dinner you can enjoy, even when you're too rushed to hope to enjoy eating it.
6	I felt it a strain to have no relief from being the person to plan/cook dinner.	Shares the burden of being the one person responsible to plan/cook dinner.
7	I felt I'd be letting myself/my family down if I didn't provide a nutritious dinner.	Reassures me I'm providing nutritious dinners.
8	I felt I'd be letting myself/my family down if I didn't provide a tasty dinner.	Reassures me I'm providing tasty dinners.
9	I felt that preparing dinner is one way I show I'm a good family person.	Reassures me I'm a good family person by preparing family dinner.
10	I felt I'd be letting myself/my family down if I didn't give each family member their choice of what to eat for dinner.	Reassures me I'm giving each family member their choice of what to eat for dinner.
11	I felt I'd be letting myself/my family down if I didn't provide a substantial dinner.	Reassures me I'm providing substantial dinners.
12	I felt I'd be letting myself/my family down if I didn't provide a dinner that includes salad/veggies.	Reassures me I'm providing dinners that include veggies/salads.
13	I felt I'd be letting myself/my family down if I didn't provide a home cooked dinner.	Reassures me I'm providing home cooked dinners.
14	I felt I'd be letting myself/my family down if I didn't provide a dinner that includes meat.	Reassures me I'm providing dinners that include meat.
15	I felt I'd be letting myself/my family down unless everyone including my kids and spouse loved what I'd make.	Reassures me I'm providing dinners that everyone–kids and spouse–love.
16	I felt that preparing weekday dinner is just a matter of routine.	Suits my view that preparing weekday dinner is just a routine matter.
17	I felt that the conversation around the table at dinner would interest me.	Allows me appreciate dinner table conversation that interests me.
18	It interested me to make a dinner from different kinds of food, day to day.	Supports my interest in making many different kinds of food for dinner.
19	It interested me to tweak favorite family dinner recipes.	Supports my interest in tweaking favorite family dinner recipes.
20	I was enjoying making dinner with foods of different textures.	Allows me provide different textures of food for dinner.
21	I was relishing the added enjoyment of appetizing smells from a home prepared dinner.	Allows appetizing smells add to the enjoyment of dinner prepared at home.
22	I was concerned that my family would leave the dinner uneaten.	Helps ensure my family won't leave the dinner uneaten.
23	I was concerned that the kids would complain and refuse to eat dinner.	Helps ensure that kids don't complain and refuse to eat dinner.
24	I was concerned about burdensome clean-up afterwards.	Ensures there's no burdensome clean-up for me.
25	I was concerned about the burden of complicated or lengthy preparation.	Assures me I'm not burdened by complicated or lengthy preparation.
26	High cost kept me from serving a better dinner.	Allows me serve better dinners without high cost.
27	I was concerned about the problem too much salt in food would cause me/my family.	Guards against my family getting too much salt .
28	I was concerned not to prepare a depressing same old dinner.	Fights depressing same old thing dinner every time.
29	I was upset to think of having dinner food left over.	Reassures me there won't be food left over to upset me.
30	I was upset to think there wouldn't be enough food for dinner.	Reassures me there will be enough food.

Classes 1 through 3 represent moving away from an undesirable state currently being experienced (e.g., need 5, “I was too rushed/pressured preparing dinner to enjoy eating it”), an undesirable state in the future (e.g., need 11, “I felt I’d be letting myself/my family down if I didn’t provide a substantial dinner”), and a “default” undesirable state (e.g., need 16, “I felt that preparing weekday dinner is just a matter of routine”). Classes 4 and 5 represent an interest in mental exploration (e.g., need 19, “It interested me to tweak favorite family dinner recipes”) and sensory enjoyment (e.g., need 20, “I was enjoying making dinner with foods of different textures”). Classes 6 and 7 represent avoiding expected excessive cost (e.g., need 26, “High cost kept me from serving a better dinner”) and expected dissatisfaction (e.g., need 30, “I was upset to think there wouldn’t be enough food for dinner”). The items included as needs in the study were generated from focus groups and packaging claims in the pre-packaged dinner category, using the above classification system as a guide. It is important to note, as show in Table 2.3, that the needs relate to the person while the benefits relate to the product. Once the items are generated, the structure used to guide their elicitation is ignored. Details of the motivational classes and their elicitation are provided in Fennell and Allenby (2014).

After rating the 30 needs, each respondent completed 10 choice tasks each with 4 alternatives, where alternatives were benefit bundles. The alternatives were ranked, with 1 being the most preferred alternative. We thus explode the rank ordering to a depth of 3, which is at the recommended limit in Chapman and Staelin (1982). The “brand” of the pre-packaged dinners was fixed across choice tasks, so only the benefits changed. Only 3 of the 30 attributes were active for each of the 4 alternatives in each choice task. See Figure 2.2 for an example of a single benefit-bundle choice task.

Pre-Packaged Dinner 53	Pre-Packaged Dinner 28	Pre-Packaged Dinner 31	Pre-Packaged Dinner 75
• On my shelf, always available to make a dinner.	• Helps make dinner when I'm pressed for time.	• Makes dinner on days when I don't feel like making dinner.	• Ready to hand, when you've run out of menu ideas.
• Makes a dinner I can enjoy, even when I'm too rushed to hope to enjoy eating it.	• Shares the burden of being the one person responsible for planning/cooking dinner.	• Reassures me I'm providing nutritious dinners.	• Reassures me I'm providing tasty dinners.
• Reassures me I'm a good family person by preparing family dinner.	• Reassures me I'm giving each family member their choice of what to eat for dinner.	• Reassures me I'm providing substantial dinners.	• Reassures me I'm providing dinners that include veggies/salads.

Figure 2.2: Example Benefit-Bundle Choice Task

Respondents then indicated their beliefs regarding 6 brands in the pre-packaged dinner category. They indicated in a pick any/J format whether each brand provided each of the same 30 benefits used in the benefit-bundle conjoint, which also map one-to-one to the needs as shown in Table 2.3. After indicating their brand beliefs, each respondent completed 8 choice tasks each with 3 to 6 alternatives, depending on which brands they had either purchased or indicated were part of their consideration set. Each alternative in this second conjoint was a pair of one of the 6 brands and price. The alternatives were again ranked, with 1 being the most preferred alternative. We again explode the rank ordering to a depth of 3, the recommended limit in Chapman and Staelin (1982).

Following the structure specified in Equation (2.6), $\sigma_{(j)h} = \exp(\gamma)$ if the respondent perceives the brand for the jth ranked alternative as being able to provide the benefits that address their needs, with $\sigma_{(j)h} = 1$ otherwise. For example, and using Table 2.3 as reference, if a respondent only stated that the third need was applicable

Table 2.4: Model-Free Evidence

Rank	Proportion of Relevant Choices
1	0.089
2	0.066
3	0.058
4	0.027

to him the last time he used a pre-packaged meal (i.e., he ranked "It was a day when I just didn't feel like making dinner" as the only need Completely (5) describing him) and he believes that the alternative's brand is able to provide the corresponding benefit – "Makes dinner on days when you don't feel like making dinner" – than that alternative would be relevant.

We offer the following model-free evidence for using our model in the context of this empirical application. If needs matter for preference certainty, as described above, one would expect the proportion of relevant choices for respondents with one or more needs should be largest for those items ranked first. We include the proportion of relevant choices for each rank in Table 2.4. Not only is the proportion of relevant choices highest for the first rank, but the next-largest proportions match for each subsequent rank as well. These proportions are small because of our strict definition of relevance, where a product is relevant only when it is perceived to address all of a respondent's active needs. Less strict definitions of relevance (e.g., a product is relevant when it is perceived to address one or more of a respondent's active needs) lead to larger proportions. However, such definitions risk polluting the interpretation of the estimated effect with concerns about which needs are being addressed, with some needs potentially more important than others. We use the strict definition of relevance and cover all active needs to ensure the meaning of the estimated effect is clear.

2.4 Results

Our results indicate differences in preference certainty within the extended model framework for responses where needs are addressed versus responses where needs are not addressed. We compare three models. First, we estimate a standard exploded multinomial logit model, without including consumer needs, to serve as a baseline comparison. Second, we estimate an exploded multinomial logit model with the needs included as covariates in the random-effects specification of heterogeneity, as in Equation (2.2). This upper-level model tests the first mechanism for the effect of product relevance on choice. We mean-center the covariates in this model so that the intercept of Δ can be interpreted as the part-worths for the average respondent, making it comparable to the posterior means from the proposed model. Third, we estimate the proposed heteroscedastic exploded multinomial logit model with a fixed-effect γ to test the second mechanism for the effect of product relevance on choice. For each model, we reserved the last choice task for out-of-sample fit. We ran each model for 80,000 iterations, using the final 4,000 iterations for inference.

Model fit is detailed in Table 2.5. Log-marginal density is a Bayesian measure of in-sample model fit. Hit probability is the posterior mean of the predicted probability for the observed ranking in the brand-price tradeoff hold-out data (see Appendix C). The proposed heteroscedastic exploded multinomial logit model outperforms the baseline and upper-level models in that it has the smallest absolute value log-marginal

Table 2.5: Model Fit

Model	γ	In-Sample LMD	Hit Probabilities L_2
Baseline ($\sigma = 1$)	-	-16,107.39	0.431
Upper-Level ($\sigma = 1$, $\beta_h = \Delta' z_h + \xi_h$)	-	-16,144.37	0.429
Heteroscedastic ($\sigma_{(j)h} = exp(\gamma)$ or 1)	-0.378 (0.157)	-15,895.56	0.435

density and the largest hit probability. The baseline model also performs better than the upper-level model. Both serve as evidence that the mechanism through which product relevance impacts choice is indeed with respect to preference certainty rather than part-worth heterogeneity. Additionally, Table 2.5 includes the estimate of γ for the heteroscedastic model. The effect is negative, as expected. Thus product relevance leads to greater preference certainty.

To illustrate how product relevance leads to greater preference certainty, note that we are comparing two groups of responses. We have responses where the part-worths are scaled by $\sigma_{(j)h} = 1$ and responses where the part-worths are scaled by $\sigma_{(j)h} = \exp(-0.378) = 0.685$. Responses with $\sigma_{(j)h} = 0.685$ exhibit greater preference certainty since a smaller scale term gives more weight to the deterministic component of random utility. Responses where $\sigma_{(j)h} = 1$ exhibit less preference certainty since a larger scale term gives more weight to the random component of random utility.

We plot the posterior means of the random-effects distributions in pairwise comparisons of the proposed heteroscedastic model with the baseline and upper-level models. We can see in Figure 2.3 that in both pairwise comparisons the estimates are highly correlated. This is expected. Highly correlated estimates between the

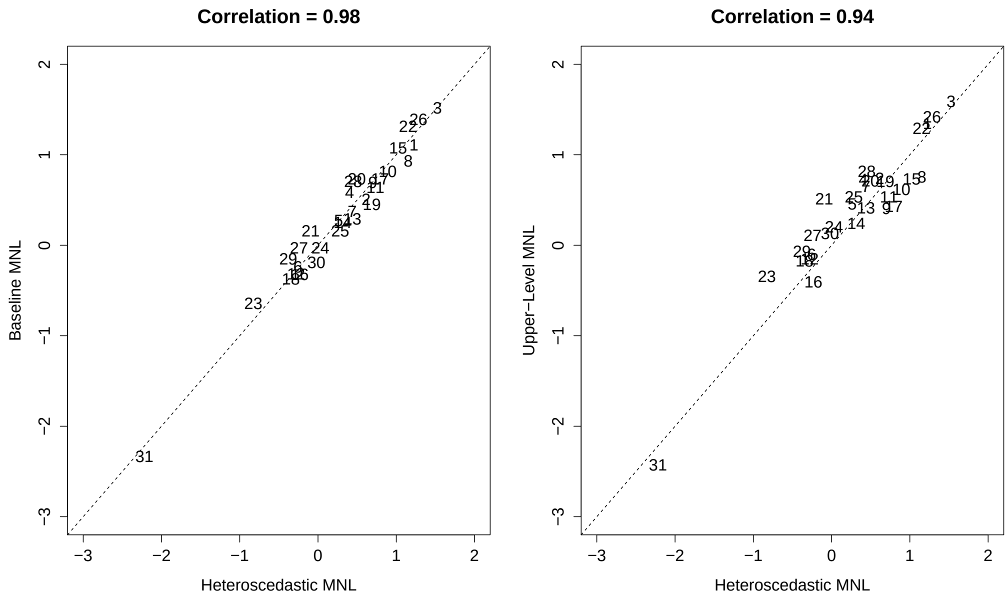

Figure 2.3: Comparison of Posterior Means

proposed and alternative models help confirm that the effect of product relevance is in terms of preference certainty and not the part-worths themselves. The high correlation between the baseline and heteroscedastic models provides evidence that the effect of product relevance is truly kept to the scale of random utility. Moreover, the benefit conjoint data (L_1) provides the majority of information for the determination of part-worth estimates plotted in Figure 2.3. The difference is model fit reported in Table 2.5 is primarily due to the brand-price tradeoff data (L_2). This illustrates the importance of using an extended model of behavior rather than simply assuming that the estimated part-worths can accurately predict demand for marketplace offerings. Table 2.6 provides additional comparison of model estimates.

Table 2.6: Comparison of Model Estimates

	Baseline		Upper-Level		Heteroscedastic	
Parameter	Mean	SD	Mean	SD	Mean	SD
δ_1	1.11	0.15	1.35	0.15	1.23	0.19
δ_2	0.51	0.11	0.74	0.13	0.62	0.19
δ_3	1.52	0.14	1.59	0.17	1.53	0.12
δ_4	0.59	0.08	0.73	0.09	0.40	0.12
δ_5	0.27	0.09	0.46	0.08	0.26	0.10
δ_6	-0.24	0.18	-0.10	0.12	-0.26	0.16
δ_7	0.37	0.07	0.65	0.13	0.44	0.11
δ_8	0.93	0.12	0.75	0.10	1.15	0.13
δ_9	0.69	0.15	0.41	0.10	0.70	0.16
δ_{10}	0.81	0.12	0.62	0.06	0.89	0.11
δ_{11}	0.64	0.10	0.54	0.10	0.73	0.14
δ_{12}	-0.32	0.09	-0.15	0.08	-0.28	0.12
δ_{13}	0.29	0.11	0.41	0.12	0.44	0.15
δ_{14}	0.25	0.18	0.24	0.14	0.32	0.15
δ_{15}	1.08	0.10	0.73	0.10	1.02	0.13
δ_{16}	-0.32	0.27	-0.41	0.12	-0.23	0.19
δ_{17}	0.74	0.09	0.43	0.09	0.79	0.12
δ_{18}	-0.37	0.14	-0.17	0.13	-0.35	0.14
δ_{19}	0.45	0.15	0.70	0.08	0.69	0.14
δ_{20}	0.73	0.11	0.71	0.07	0.50	0.10
δ_{21}	0.16	0.17	0.52	0.11	-0.09	0.23
δ_{22}	1.32	0.15	1.29	0.10	1.15	0.14
δ_{23}	-0.64	0.13	-0.34	0.09	-0.83	0.13
δ_{24}	-0.03	0.12	0.21	0.10	0.03	0.11
δ_{25}	0.16	0.10	0.53	0.09	0.28	0.09
δ_{26}	1.39	0.13	1.42	0.09	1.28	0.14
δ_{27}	-0.03	0.12	0.11	0.08	-0.24	0.13
δ_{28}	0.70	0.14	0.82	0.10	0.45	0.20
δ_{29}	-0.15	0.12	-0.07	0.07	-0.38	0.09
δ_{30}	-0.19	0.13	0.13	0.08	-0.02	0.10
δ_{31}	-2.33	0.21	-2.43	0.17	-2.22	0.18

2.5 Discussion

In this paper we demonstrate within an extended model framework that respondents are less certain stating what they want when considering things they don't need. We accomplish this by developing an extended model of brand choice that allows us to estimate the effect of product relevance, when a product is perceived as being able to address a respondent's needs, on choice. Our results indicate that the effect of product relevance is manifest through the scale of random utility and not its location. The effect of product relevance on preference certainty reaffirms both the construction of extended models and the inclusion of needs, which describe where consumers are coming from, in models of choice. The proposed model also adds another behavioral explanation to the literature on heteroscedastic errors in choice models.

Our results also suggest that traditional screening criteria may be insufficient when studying aspects of an offering that might not be relevant to all respondents engaged in activities related to a product category. A study of outboard marine engines, for example, might focus on people owning boats for pleasure and recreation, and would naturally include features such as horsepower, acceleration, and fuel efficiency. Concerns about durability, however, might be more prevalent among people engaged in fishing where running over submerged logs is more likely. Obtaining accurate measures of preference for durability requires respondents for whom the issue of durability is relevant in their pursuits. Furthermore, traditional screening based on category participation may simply exclude consumers that may need what the category provides but are simply not currently engaged with the category. Researchers should consider identifying the needs of particular category prospects and using those needs to help screen qualified respondents. We estimated the proposed model again after screening respondents who didn't have any of the 30 needs and saw a marginal improvement in predictive fit.

Research extensions include determining whether it is possible to fix or condition upon needs within a conjoint study like we currently fix purchase occasion, studying the relationship between the current state of consumers (i.e., needs) and their imagined desired state (i.e., goals) within the context of choice, and determining how consumer needs are mapped to perceived benefits. This last item relates to the literature regarding the construction of "meta-attributes" – bundles of physical attributes that together define subjective characteristics or perceived benefits (Luo et al., 2008; Netzer et al., 2008; Kim et al., 2014). Our model assumes that needs are matched to perceived efficacy in terms of brand beliefs. However, if attributes other than brand are included, we need to be clear how they effect perceptions of efficacy. Furthermore, there is no reason to expect that the mapping between needs and benefits must always be one-to-one, as we have assumed, or that needs are consistent across purchase occasions. Future research can help resolve these issues by both identifying perceived benefits and mapping needs to these benefits.

We find that an extended model of behavior is needed to translate part-worth estimates from a benefit-based conjoint study to predict demand for branded products. Product relevance and perceived brand efficacy are shown to play important roles in the translation. As research progresses in the area of extended choice models, similar constructs requiring additional data will need to be incorporated into models of choice.

Chapter 3: Explaining Preference Heterogeneity with Mixed Membership Modeling

3.1 Introduction

The fact that consumers are heterogeneous in their preferences gives rise to marketing as a discipline and an industry. Choice models and associated decision tools that account for this heterogeneity allow firms to better understand what consumers prefer and have become a standard for product development and product line optimization. However, explaining preference heterogeneity remains an elusive problem. In this paper we develop an expanded choice model that improves our ability to explain preference heterogeneity by employing a novel approach to model discrete data, including binary and ratings survey data, that describe the drivers of consumer preference.

Choice modeling is an effective tool for determining what product attributes individuals prefer but it has proven less successful at explaining the heterogeneity in consumer preferences. Explaining preference heterogeneity includes identifying covariates that serve as drivers of preference and enable targeting and promotion activities. The use of hierarchical Bayes in choice modeling allows for both individual-level attribute part-worth utilities and aggregate-level preference heterogeneity parameters. Part-worth estimates tell us what attributes consumers prefer. Parameters describing preference heterogeneity are conditioned on covariates that help explain cross-sectional variation in the part-worths.

Finding covariates that are predictive of part-worths has proven difficult. The primary benefit when using a random effect distribution of heterogeneity has been accounting for unexplained heterogeneity. Using discrete variables describing possible drivers of preference, such as demographics and psychographics, as covariates is standard. However, survey data are typically used as covariates where the number of covariates makes it impractical to include interactions. Additionally, we have growing access to new sources of discrete multivariate data outside of surveys, including text, that we expect will be a rich source of information for explaining choice yet incorporating it isn't obvious. We propose modeling this discrete multivariate data as part of the choice model in order to uncover covariates that can better explain preference heterogeneity.

In this paper we develop an expanded hierarchical Bayesian choice model where covariates for the upper level are from a grade of membership model (Woodbury et al., 1978; Erosheva et al., 2007). The grade of membership model is related to latent Dirichlet allocation, which serves as a touchstone within topic modeling (Blei et al., 2003). Both are part of a larger class of models known as mixed membership models that provide individual-level, low-dimensional representations of discrete multivariate data by accounting for interactions or co-occurrence (Airoldi et al., 2014). We propose modeling discrete variables describing potential drivers of preference where interaction among drivers will help further explain preference heterogeneity. We apply our model within the robotic vacuums category and find we can both explain preference heterogeneity and predict choice better than traditional models using observed covariates directly.

This paper contributes to efforts at using mixed membership models to improve marketing models. The application of this class of models to marketing contexts is still in its infancy. Extant research has focused on latent Dirichlet allocation (LDA), using product reviews and online forums to inform market structure (Lee and Bradlow, 2011; Netzer et al., 2012) and to identify preferences for product features (Archak et al., 2011). Most recently, Tirunillai and Tellis (2014) use LDA to conduct brand analysis while Büschken and Allenby (2016) develop a sentence-constrained LDA to better predict review ratings. However, mixed membership models have yet to be employed in the context of choice modeling. We believe this paper provides an important first step in this regard.

The remainder of the paper will be organized as follows. We specify our model in Section 3.2. We detail our empirical application in Section 3.3. In Section 3.4, we compare results from our proposed model, with covariates uncovered using the grade of membership model, and alternative models where standard discrete covariates are used. We discuss implications of and extensions to this research in Section 3.5.

3.2 Model Specification

3.2.1 Hierarchical Bayesian Choice Model

Hierarchical Bayesian choice models allow for the estimation of both individual and aggregate-level preference parameters, even in the presence of few observations per individual (Rossi and Allenby, 2003; Rossi et al., 2005). Decision tools associated with choice modeling make use of individual-level preference parameter estimates to forecast the results of various product policies while aggregate-level parameter estimates are employed to explain the source of individual preferences.

The likelihood in hierarchical Bayesian choice modeling is typically assumed to be a multinomial logit model such that the probability of individual n choosing product alternative j is a function of the attributes x_j that compose the given alternative and the part-worths or individual-level preferences β_n for the attributes:

$$Pr(y_n = j|\beta_n) = \frac{\exp(x_j'\beta_n)}{\sum_{p=1}^{P}\exp(x_p'\beta_n)} \tag{3.1}$$

where there are a total of P alternatives to consider. The distribution of heterogeneity, or upper level, models preference heterogeneity in the individual-level β_n's. The distribution of heterogeneity is typically assumed to be multivariate normal and is characterized as:

$$\beta_n = \Gamma' z_n + \xi_n, \quad \xi_n \sim \mathrm{N}(0, V_\beta) \tag{3.2}$$

where z_n is a vector of covariates for individual n and Γ is a matrix of coefficients that maps variation in z_n to variation in β_n. The mean of the distribution of heterogeneity $\Gamma' z_n$ is where the analyst can specify individual-specific covariates z_n that explain variation in the part-worths. Information is shared through the estimates of Γ and the heterogeneity covariance matrix V_β to estimate individual-level β_n's (Rossi et al., 2005).

The directed acyclic graph (DAG) in Figure 3.1 provides a visual representation of the hierarchical Bayesian choice model. The DAG utilizes plate notation, where a plate represents replication for the enclosed variables. In the DAG, white nodes represent parameters to be estimated, grey nodes represent fixed hyper-parameters, and black nodes represent observed data.

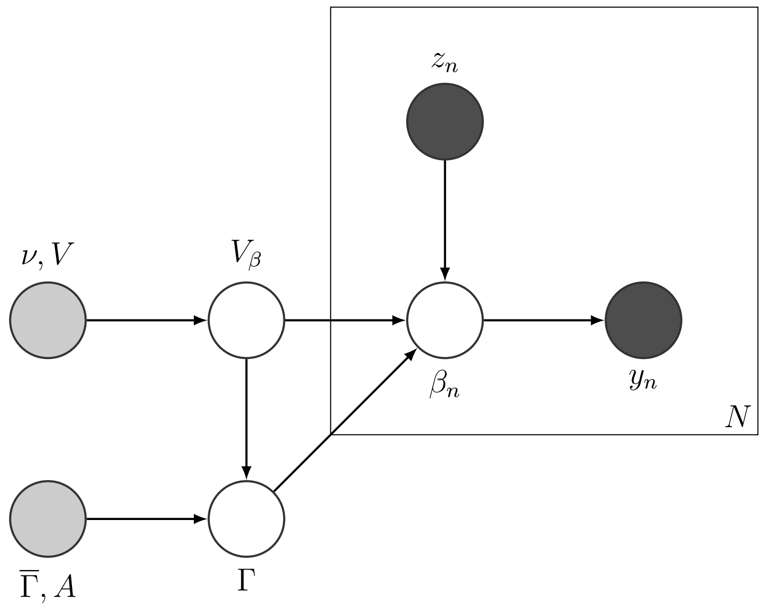

Figure 3.1: Hierarchical Bayesian Choice Model

From the use of plate notation in Figure 3.1, we can see that the hierarchical Bayesian choice model has both aggregate and individual levels. To be clear, at the aggregate level, $\overline{\Gamma}$ is the mean and A is the precision matrix for a conjugate normal prior on Γ and ν and V are the degrees of freedom and scale matrix for a conjugate inverse Wishart prior on V_β. At the individual level, y_n is a vector of observed choices and z_n are the observed covariates for individual n. We can see that the covariates $\{z_n\}_{n=1}^N$ are chosen independent of the model specification. As discussed, the covariates $\{z_n\}_{n=1}^N$ are the key to our ability to explain preference heterogeneity. We will use DAGs, beginning with Figure 3.1, to help motivate the proposed model.

Following Figure 3.1, the joint posterior distribution of the standard hierarchical Bayes choice model is:

$$\begin{aligned} &p(\{\beta_n\}_{n=1}^N, \Gamma, V_\beta | \{y_n\}_{n=1}^N, \overline{\Gamma}, A, \nu, V) \\ &\qquad\qquad \propto \left[\prod_{n=1}^N p(y_n|\beta_n) p(\beta_n|\Gamma, V_\beta)\right] p(\Gamma|V_\beta, \overline{\Gamma}, A) p(V_\beta|\nu, V) \end{aligned} \tag{3.3}$$

where $\prod_{n=1}^N p(y_n|\beta_n)$ is the likelihood, $\prod_{n=1}^N p(\beta_n|\Gamma, V_\beta)$ is the distribution of heterogeneity, and $p(\Gamma|V_\beta, \overline{\Gamma}, A)$ and $p(V_\beta|\nu, V)$ are the priors (Rossi et al., 2005). The known design matrix X and covariates $\{z_n\}_{n=1}^N$ are suppressed in Equation (3.3).

A variety of covariates have been employed to explain preference heterogeneity in the choice modeling literature. For example, Allenby and Ginter (1995b) used demographic variables, Lenk et al. (1996) included expertise, and Chandukala et al. (2011b) specified consumer needs to explain variation in β_n. However, explaining preference heterogeneity has not met with much success generally (Rossi et al., 1996; Horsky et al., 2006).

One unresolved issue is that discrete covariates are often employed without a practical way to include interactions. The problem is one of dimensionality. The number of interaction terms is J choose M, where J is the number of covariates and M is the number of desired interactions. For example, with $J = 30$ covariates and $M = 2$, there are 435 possible two-way interactions, to say nothing of higher-level interactions where $M > 2$. While Chandukala et al. (2011b) employ variable selection to determine which covariates matter, we are interested in a model general enough to account for interactions from traditional survey data as well as accommodate new sources of discrete data.

We propose using a non-standard model that accounts for the interaction or co-occurrence of variables to uncover covariates from discrete multivariate data for use in a choice model's random effect distribution of heterogeneity. Specifically, we propose combining a hierarchical Bayesian choice model with a grade of membership model to uncover covariates that account for interactions in order to explain preference heterogeneity better than using observed covariates directly. We first detail the grade of membership and the class of mixed membership models before specifying our expanded choice model.

3.2.2 The Grade of Membership Model

The grade of membership (GoM) model was developed to classify disease patterns using discrete patient-level clinical data (Woodbury et al., 1978; Clive et al., 1983). It has since been applied to modeling survey data (Erosheva et al., 2007; Gross and Manrique-Vallier, 2014). In these applications, each respondent answers a battery of survey questions with categorical responses. The research interest is to identify the patterns of interaction or co-occurrence in the categorical responses across respondents along with how each respondent relates to the patterns of co-occurrence. The GoM model characterizes these patterns of co-occurrence as profiles of archetypal respondents. Each respondent is a partial member of each of the profiles based on how similar their responses are to each pattern of co-occurrence.

Assume we have a collection of J discrete variables each with n_j categorical responses. The probability of respondent n selecting the lth category for question j is a function of the profiles λ describing the patterns of response co-occurrence across respondents and respondent n's membership vector g_n describing their partial membership in each profile:

$$Pr(w_{n,j} = l | g_n, \lambda) = \sum_{k=1}^{K} g_{n,k} \lambda_{j,k}(l) \tag{3.4}$$

where there are K profiles and K is specified by the analyst. The membership vector g_n for the nth respondent is constrained so that each element is non-negative and $\sum_{k=1}^{K} g_{n,k} = 1$. The λ is composed of $J \cdot K$ total vectors $\lambda_{j,k}$ each of length n_j that specify how likely each categorical response l is for question j for a hypothetical respondent that is only a member of profile k. Each $\lambda_{j,k}$ is also constrained with non-negative elements so that $\sum_{l=1}^{n_j} \lambda_{j,k}(l) = 1$.

To illustrate, consider responses to a battery of select-all-that-apply questions (i.e., pick any/J) such that $n_j = 2$ for all $J = 30$. Each respondent selects or indicates a subset of the $J = 30$ statements or items that apply to them in answer to the question: "What benefits does cereal provide that are important to you?" Figure 3.2(a) displays the items selected (i.e., $w_{n,j} = 1$) for a given respondent n together with their membership vector g_n. Figure 3.2(b) displays $\lambda_{j,k}(1)$ describing $K = 3$ aggregate-level profiles in terms of the likelihood of selecting each of the $J = 30$ items. Note that since $n_j = 2$ for all $J = 30$, each $\lambda_{j,k}$ is a vector with two elements such that $\lambda_{j,k}(0)$ is the complement of the values listed in Figure 3.2(b). Thus $\lambda_{j,k}(0) + \lambda_{j,k}(1) = 1$ for each $\lambda_{j,k}$.

Using Figure 3.2, we can see how profiles emerge based on what items co-occur. For example, if item 11 "I want to make sure my family has breakfast in the morning" and item 3 "My kids will eat cereal for breakfast" are selected together frequently across respondents, this pattern may be part of a profile describing concern with breakfast for children. In Figure 3.2(a), the membership vector g_n describes the partial membership respondent n has in each of the $K = 3$ profiles – "Kids Breakfast," "Healthy Snack," and "Source of Fiber" – where the number of profiles $K = 3$ has been specified by the analyst and the weight given to each profile is determined by how

(a) Respondent n's Responses and Membership Vector g_n

What benefits does cereal provide that are important to you?
Item 1: It's a helpful way to get a serving of milk at the same time
Item 2: Cereal is a good source of fiber
Item 3: My kids will eat cereal for breakfast
Item 4: Cereal isn't just for breakfast, it's a good snack anytime
Item 11: I want to make sure my family has breakfast in the morning
Item 15: Cereal is easy to prepare
g_n "Kids Breakfast" 0.60 "Healthy Snack" 0.20 "Source of Fiber" 0.20

(b) Aggregate-Level Profiles Defined by the Probability of Each Item $\lambda_{j,k}(1)$

$\lambda_{j,k}(1)$	"Kids Breakfast"	"Healthy Snack"	"Source of Fiber"
Item 1	0.67	0.70	0.34
Item 2	0.22	0.85	**0.95**
Item 3	**0.97**	0.13	0.04
Item 4	0.32	**0.92**	0.10
⋮	⋮	⋮	⋮
Item 30	0.04	0.13	0.14

Figure 3.2: Modeling Pick Any/J Data with a GoM Model

similar respondent n's response pattern matches each of the aggregate-level profiles. For this particular respondent, they are primarily a member of the "Kids Breakfast" profile, with a weight of 0.60, while still being a partial member of the remaining two profiles. The membership vector g_n has non-negative elements and is constrained to equal 1.

The aggregate-level values $\lambda_{j,k}(1)$ in Figure 3.2(b) describe how likely it is for each item to occur within each profile. The profiles are composed of all $J = 30$ items with the item that is most likely within each profile in bold. Based on common response patterns across respondents, the profiles describe archetypal or extreme respondents, ones that in this case are either concerned wholly with cereal for "Kids Breakfast," a "Healthy Snack," or a "Source of Fiber," where the profile names have been determined by the analyst based on which items differentiate each profile. Thus each membership vector g_n describes where a respondent n is located within a convex hull defined by the profiles. These profiles account for the co-occurrence or interaction of the discrete items while reducing the dimensionality from J to K.

With this illustration in mind, we can apply Equation (3.4) to show that the probability of respondent n selecting item 1 "It's a helpful way to get a serving of milk at the same time" is a function of g_n, their partial membership in each profile, and $\lambda_{1,k}(1)$, how likely it is for item 1 to be selected in each profile. This results in a probability of $(0.60)(0.67) + (0.20)(0.70) + (0.20)(0.34) = 0.61$. Erosheva et al. (2007) use this relationship to introduce a latent profile assignment for each item.

Assuming that the J responses are conditionally independent given each membership vector g_n, the GoM likelihood is:

$$p(\{w_n\}_{n=1}^N | \{z_n\}_{n=1}^N, \lambda) p(\{z_n\}_{n=1}^N | \{g_n\}_{n=1}^N) = \prod_{n=1}^N \prod_{j=1}^J \prod_{k=1}^K g_{n,k}^{I(z_{n,j}=k)} \lambda_{j,k}(w_{n,j})^{I(z_{n,j}=k)} \tag{3.5}$$

where z_n is a J-dimensional vector of latent profile assignments for respondent n, following notation typical to data augmentation (Tanner and Wong, 1987; Rossi and Allenby, 2003). Note that the latent variables z_n in Equation (3.5) are different from the observed covariates specified in Equation (3.2). Both $p(\{w_n\}_{n=1}^N | \{z_n\}_{n=1}^N, \lambda)$ and $p(\{z_n\}_{n=1}^N | \{g_n\}_{n=1}^N)$ are multinomial distributions.

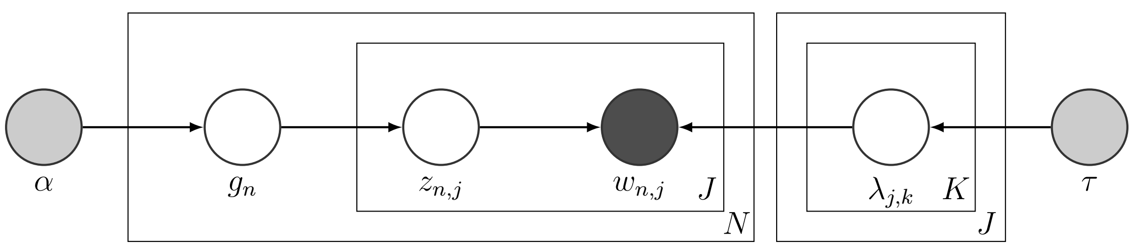

Figure 3.3: The Grade of Membership Model

The DAG in Figure 3.3 provides a visual representation of the GoM model. The plate notation demonstrates the three model levels: item, respondent, and aggregate. The aggregate-level λ describing profiles is homogeneous while the respondent-level membership vectors g_n are heterogeneous. To be clear, α and τ are both hyperparameters for conjugate Dirichlet priors on g_n and λ. Following Figure 3.3, the joint posterior distribution of the grade of membership model is:

$$p(\{z_n\}_{n=1}^N, \{g_n\}_{n=1}^N, \lambda|\{w_n\}_{n=1}^N, \alpha, \tau) \propto \left[\prod_{n=1}^N p(w_n|z_n, \lambda)p(z_n|g_n)p(g_n|\alpha)\right] p(\lambda|\tau) \tag{3.6}$$

where $\prod_{n=1}^N p(w_n|z_n, \lambda)p(z_n|g_n)$ is the likelihood and $\prod_{n=1}^N p(g_n|\alpha)$ and $p(\lambda|\tau)$ are priors.

In the marketing literature, it has been argued that identifying extreme responses is important for designing and promoting successful new products (Allenby and Ginter, 1995b). For example, extreme response behavior can be used to more efficiently target prospects with a high probability of adopting an innovation. Conceptualizing consumer heterogeneity as a continuous distribution of preferences has been shown to aid in the identification of extreme responses (Allenby et al., 1998; Allenby and Rossi, 1998). The GoM model represents discrete response behavior as a continuous proximity to a limited number of extreme profiles. Given that marketers often search for a limited number of product offerings for reasons of efficiency or resource limitations, a concept of heterogeneity that expresses differences among consumers in the space of a small number of extreme response profiles is appealing. We utilize the GoM model given this characterization of heterogeneity, which includes the respondent-level membership vectors g_n, in the development of our proposed model.

Relationship with Finite Mixture Models

Having a respondent-level membership vector g_n that consists of non-negative, real-valued latent variables that sum to one is the distinctive feature of mixed membership models, the class of models that includes the GoM and LDA. Contrast this with the general form of a finite mixture model (Kamakura and Russell, 1989):

$$p(x_n) = \sum_{k=1}^{K} g_k p_k(x_n) \tag{3.7}$$

where x_n is response data for respondent n. We see that the finite mixture model has a membership vector g_k at the aggregate level while the GoM model in Equation (3.4) has a membership vector $g_{n,k}$ at the individual level. This feature is common to all mixed membership models and illustrates why they are often referred to as individual-level mixture models.

Finite mixture models are a special case of mixed membership models (Erosheva et al., 2007; Galyardt, 2014). However, our use of the GoM within the class of mixed membership models is different than the typical use of finite mixture models in choice modeling. Instead of specifying a mixture of distributions of heterogeneity, we are interested in using the respondent-level membership vector g_n to serve as covariates that can further explain preference heterogeneity.

Relationship with Factor Analysis

Factor analysis is another related model and has long been a standard approach in marketing for dimension reduction (Stewart, 1981). The basic assumption is that a set of variables can be reduced to one or more latent constructs called factors. The data are assumed to arise in the following fashion:

$$x_{n,j} = c_j + \sum_{k=1}^{K} \zeta_{n,k}\lambda_{j,k} + \eta_{n,j}, \quad \eta_{n,j} \sim \mathrm{N}(0,1) \tag{3.8}$$

where c_j is a constant vector, $\zeta_{n,k}$ is a respondent-level vector of factor scores, and the collection of $\lambda_{j,k}$ is a matrix of aggregate-level regression coefficients known as factor loadings. The form of factor analysis in Equation (3.8) is similar to that of the GoM model in Equation (3.4), with factor scores in place of the membership vector and factor loadings in place of the profiles. Erosheva (2002) even demonstrates that the GoM model is equivalent to a binary factor analysis with an identity link function. However, there are key differences in the two approaches.

Factor analysis and GoM models differ in terms of their underlying assumptions, modeling objectives, and the type of data each method can process (Manton et al., 1994; Marini et al., 1996). First, standard factor analysis, as demonstrated in Equation (3.8), assumes continuous data. Even using a cut-point model, which assumes the observed data are discrete indicators of latent continuous variables, the underlying constructs (i.e., factors) are still considered to be continuous. On the other hand, the GoM model assumes both discrete data and discrete underlying constructs (i.e., profiles).

Second, the objective of factor analysis is to uncover latent constructs underlying a set of *variables*. The objective of the GoM model is to both uncover profiles representing extreme characterizations of *respondents* and measure each respondent's proximity to these profiles. In other words, the GoM model has the description of respondents and respondent heterogeneity as the objects of inference. Finally, unlike factor analysis, the GoM model can handle a combination of multinomial, ordinal, and other discrete multivariate data. For more detail on the comparison between factor analysis and the GoM model, see Appendix D.

3.2.3 Hierarchical Bayesian Choice Model with a GoM Model

The proposed model combines a hierarchical Bayesian choice model with a GoM model in order to use discrete multivariate data to uncover covariates that explain preference heterogeneity. A related concept is presented in the form of a supervised latent Dirichlet allocation (sLDA). In the sLDA topic model, each collection of discrete data (i.e., document, in the context of topic modeling) is paired with and used to be predictive of a response, such as using movie reviews to predict movie ratings (Blei and McAuliffe, 2007). We employ the same kind of pairing between a collection of discrete data and response, however our response is part-worth utility parameters and the collection of discrete data is from a battery of survey questions.

The individual-level choice model remains multinomial logit, as specified in Equation (3.1), and the distribution of heterogeneity remains multivariate normal as in Equation (3.2). Since there is a separate g_n for each respondent in the GoM model in Equation (3.4), we use these membership vectors as covariates to explain heterogeneity in the part-worths β_n (i.e., $\beta_n = \Gamma' g_n + \xi_n$). Thus the likelihood of the combined model is:

$$\begin{aligned} p(\{y_n\}_{n=1}^N, \{w_n\}_{n=1}^N | \{\beta_n\}_{n=1}^N, \Gamma, V_\beta, \{z_n\}_{n=1}^N, \{g_n\}_{n=1}^N, \lambda) \\ = \prod_{n=1}^N p(y_n|\beta_n) p(\beta_n|g_n, \Gamma, V_\beta) p(w_n|z_n, \lambda) p(z_n|g_n). \end{aligned} \tag{3.9}$$

Figure 3.4 illustrates the proposed hierarchical Bayesian choice model with a GoM model. From the DAG we can see that the proposed model is a three-level model where only the categorical responses w_n and choices y_n for each respondent are observed. The homogeneous profiles λ account for the interaction or co-occurrence among items and provide for the dimension reduction we need to use this collection of discrete data as covariates in the model of preference heterogeneity.

Figure 3.4 combines the DAGs in Figure 3.1 and Figure 3.3 to illustrate that the membership vector g_n serves as the link between the choice model and the GoM model. Thus g_n is informed by both the categorical responses w_n and the chosen

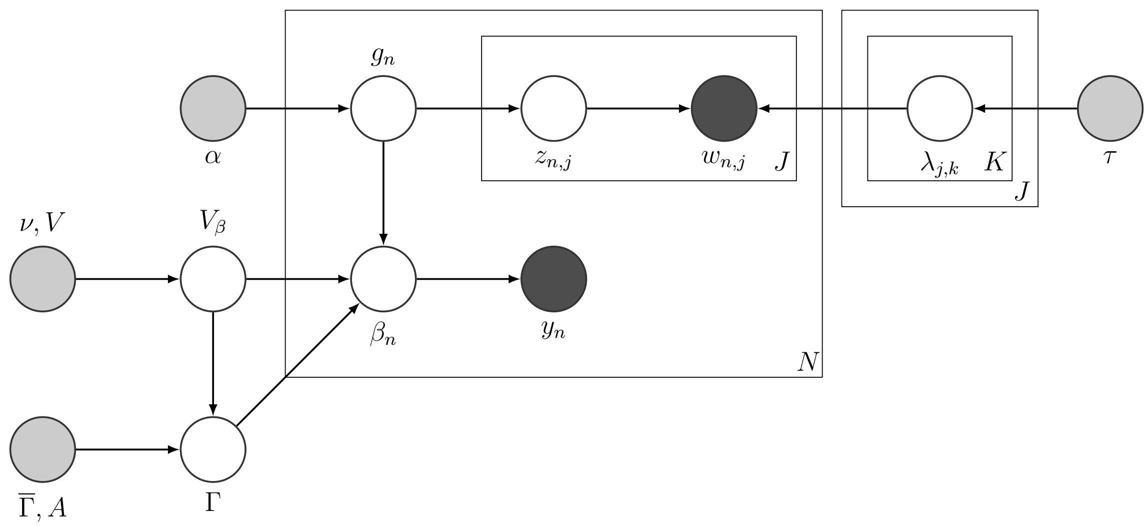

Figure 3.4: Hierarchical Bayesian Choice Model with a GoM Model

alternatives y_n. The proposed model is more complete than a model where g_n is estimated separately from choice since estimating all the parameters in the expanded model allows us to properly account for the uncertainty in g_n. Because g_n is identified when informed by w_n alone in the GoM model, g_n is also identified in the proposed model when identified by both w_n and y_n.

Following Figure 3.4, the joint posterior distribution of the proposed model is:

$$
\begin{aligned}
p(\{\beta_n\}_{n=1}^N, \Gamma, V_\beta, \{z_n\}_{n=1}^N, \{g_n\}_{n=1}^N, \lambda | \{y_n\}_{n=1}^N, \overline{\Gamma}, A, \nu, V, \{w_n\}_{n=1}^N, \alpha, \tau) \\
\propto \left[\prod_{n=1}^N p(y_n|\beta_n) p(\beta_n|g_n, \Gamma, V_\beta) p(w_n|z_n, \lambda) p(z_n|g_n) p(g_n|\alpha) \right] \\
\times\, p(\Gamma|V_\beta, \bar{\Gamma}, A) p(V_\beta|\nu, V) p(\lambda|\tau)
\end{aligned}
\tag{3.10}
$$

where $\prod_{n=1}^N p(y_n|\beta_n)p(\beta_n|g_n, \Gamma, V_\beta)p(w_n|z_n, \lambda)p(z_n|g_n)$ is the likelihood and $\prod_{n=1}^N p(g_n|\alpha)$, $p(\Gamma|V_\beta, \bar{\Gamma}, A)$, $p(V_\beta|\nu, V)$, and $p(\lambda|\tau)$ are the priors. A complete list of the variables in Equation (3.10) are detailed in Table 3.1.

Table 3.1: Variables in Hierarchical Bayesian Choice Model with a GoM Model

Choice Variables	Description
N	number of respondents
H	number of choice tasks for each respondent n
P	number of alternatives in each choice task
L	number of attribute levels in each choice task
y_n	H-dim vector of choices for respondent n
β_n	L-dim vector of part-worths for respondent n
Γ	$K \times L$ matrix representing the mean of the random effects distribution of heterogeneity
V_β	$L \times L$ covariance matrix of the random effects distribution of heterogeneity
GoM Variables	**Description**
K	number of profiles
J	number of categorical questions
n_j	number of categorical responses for question j
w_n	J-dim vector of respondent n's categorical responses
z_n	J-dim vector of respondent n's profile assignments
g_n	K-dim membership vector for respondent n
λ	collection of probability distributions $\lambda_{j,k}$ over the n_j response options for each question j and profile k

We validated our proposed model by generating data where $K = 2$, $N = 200$, $J = 13$, $n_j = 2$ for all J, $H = 50$, $P = 4$, and $L = 5$ and recovering parameter values. Details on generating data and the estimation procedure are provided in Appendices E and F. Each true parameter value was within or near the bounds of a 95% credible interval. We display the aggregate-level posterior means of Γ and λ in Figure 3.5. The posterior means line up along the diagonal, indicating parameter recovery. Note that the λ estimates are constrained to be within the $0 - 1$ bounds.

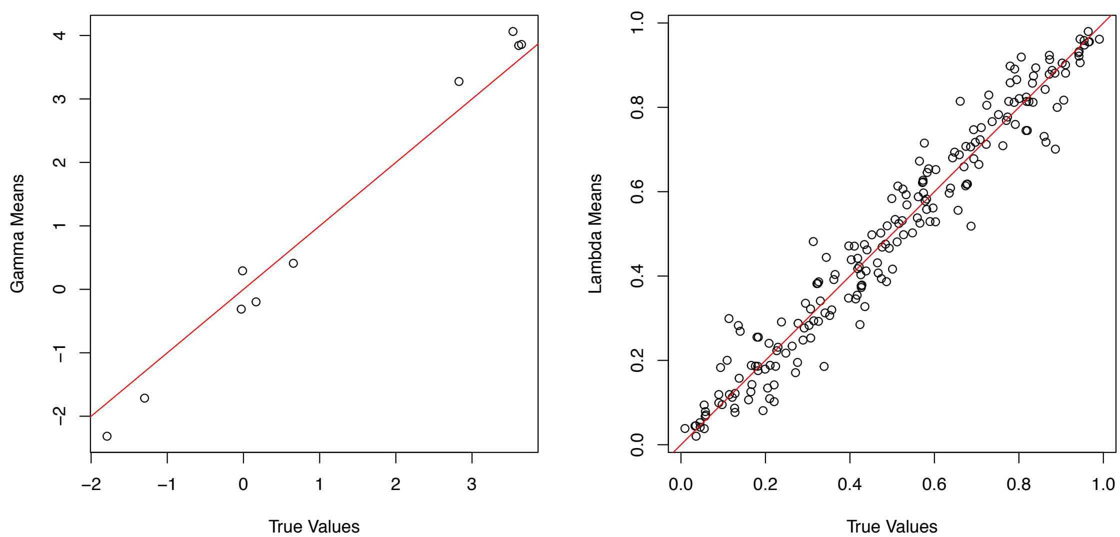

Figure 3.5: Simulation Study Results

3.3 Empirical Application

We use data from a national survey of preferences regarding robotic vacuums. A total of 332 respondents were carefully screened to ensure that the product options under consideration were relevant to them. In particular, qualified respondents had to own a robotic vacuum, currently be shopping for their first robotic vacuum, or might consider a robotic vacuum sometime in the next five years.

Before the conjoint experiment, respondents were asked to detail why the product was relevant to them or anyone in their household by selecting from a list of 11 statements on cleaning that robotic vacuums might help address. Respondents were also asked to select from among a list of 7 statements that described problems with robotic vacuums. The combined list of 18 statements regarding cleaning and robotic vacuums is provided in Table 3.2. Thus our discrete data consists of two possible categories (i.e., $n_j = 2$ for all $J = 18$) where not selecting an item is coded as a 0 and selecting an item is coded as a 1.

Standard models using this discrete data as observed covariates in the random effects distribution of heterogeneity don't have a practical way to include interactions, even though interactions should be expected. For example, we would expect that respondents who select statement 5 "I worry about germs and dirt on my floor and carpet" also select statement 10 "I spend over two hours per week cleaning" and that this interaction would have an impact on explaining preferences in the random effects distribution of heterogeneity. However, if we were to include two-way interactions, we would add an additional 153 covariates, to say nothing of the dimensionality introduced by higher-level interactions.

Table 3.2: Statements on Cleaning and Robotic Vacuums

No.	Item
1	I enjoy coming home to a clean house.
2	I don't feel relaxed when I know my home isn't clean.
3	I worry about pet hair and dander in the home.
4	I have trouble keeping the floor beneath my furniture clean.
5	I worry about germs and dirt on my floor and carpet.
6	I get anxious about having guests when my home is dirty.
7	I don't like going to someone's home that is dirty.
8	I don't like touching dirty things.
9	I don't spend much time cleaning.
10	I spend over two hours per week cleaning.
11	I have a cleaning person who cleans for me.
12	Robotic vacuums are too expensive.
13	Robotic vacuums are too complicated to program, set up, and operate.
14	Robotic vacuums often need to be "rescued" because they get stuck.
15	Robotic vacuums need to have their trash containers changed too often.
16	Robotic vacuums don't do a good enough job cleaning the floor and carpet.
17	Robotic vacuums don't spend enough time on really dirty spots on the floor.
18	Robotic vacuums scare household pets.

If these were your only options, which would you choose?

9 / 16

Brand	Samsung	Black & Decker	iRobot	Neato	NONE: I wouldn't choose any of these.
Cleaning Performance	85%	85%	70%	85%	
Capacity	Before every use	Before every use	Before every use	Before every 2-3 uses	
Navigation	Smart	Smart	Smart	Random	
Programming	App	App	App	Base Unit	
Virtual Borders	No	No	No	Yes	
Price	$299	$599	$399	$499	

Figure 3.6: Example Choice Task

After selecting from applicable statements on cleaning and robotic vacuums, respondents proceeded through a series of 16 choice tasks where they were asked to select which of five product alternatives they most preferred, including an outside option to not select any of the given alternatives. Figure 3.6 is a screenshot of one of these choice tasks. Each alternative was composed of seven separate attributes for a total of 12 estimable attribute levels, excluding the reference levels in red detailed in Table 3.3.

Besides brand and price, we see that the attributes were defined in terms of features, including the vacuum's performance (i.e., what percentage of dirt and debris it picks up), capacity (i.e., how often it needs to be emptied), the type of navigation (i.e., does it change directions by just bumping into things or is it "smart" and able to scan and determine an optimal path), where it can be programmed, and whether

Table 3.3: Attribute Levels

Attributes	Levels				
Brand	Outside Option	Neato	iRobot	Samsung	Black & Decker
Performance	70%	85%			
Capacity	Every use	Every 2-3 uses			
Navigation	Random	Smart			
Programming	Base unit	App			
Virtual Borders	No	Yes			
Price	\$299	\$399	\$499	\$599	

Table 3.4: Data Summary

Choice Variables	Description
$N = 332$	total number of respondents
$H = 16$	number of choice tasks for each respondent n
$P = 5$	number of alternatives in each choice task
$L = 12$	number of attribute levels in each choice task
GoM Variables	**Description**
$J = 18$	number of categorical questions
$n_j = 2$	number of categorical responses for each question j

or not virtual borders can be set to keep the robotic vacuum away from certain areas of the home. A summary of the data using model notation is provided in Table 3.4.

3.4 Results

We report the results of three models. The Intercept model only includes an intercept in the upper level model (i.e., $\beta_n = \gamma + \xi_n$) and serves as a baseline. The Binary Covariates model includes all 18 dummy-coded statements from Table 3.2 as covariates in the upper level model (i.e., $\beta_n = \Gamma' z_n + \xi_n$) and represents the typical way these discrete covariates would be used in practice. Finally, the Membership Vector model is our proposed model, which uses the membership vectors g_n from the grade of membership model as covariates for $K = 5$ profiles (i.e., $\beta_n = \Gamma' g_n + \xi_n$).

The number of profiles K is determined by the analyst. Following the review on model selection criteria by Joutard et al. (2007), we ran an isolated GoM model on the 18 statements in Table 3.2 and compared two measures of fit. The first is the Newton-Raftery approximation of the log marginal density (LMD) (Newton and Raftery, 1994), a standard Bayesian measure. The second is the deviance information criterion (DIC), developed by Spiegelhalter et al. (2002). Values closer to zero indicate improvement in fit for both measures. Figure 3.7 includes charts for the values of both LMD and DIC for models with $K = 2$ to $K = 18$. According to the LMD, $K = 5$ is best. According to the DIC, $K = 7$ is best. With the range of possible models narrowed, we ran the proposed model for $K = 5$ to $K = 7$. Comparing results to find profiles that are sufficiently differentiated and non-repeating, the model with $K = 5$ was deemed best.

The final 75 respondents were reserved as a hold-out sample, leaving 257 respondents for calibration. In addition, one choice task was held out from each respondent in the calibration sample for an additional measure of predictive fit. We ran each model for 50,000 iterations, saving every 50th draw, and using the final 20,000 iterations for inference. We checked for but found no substantial evidence of label switching.

Choice model LMD is used for in-sample fit. Out-of-sample fit is provided in terms of hit probabilities. A hit probability is the average posterior probability of a set of observed choices given a specific model. The hit probability is averaged over a

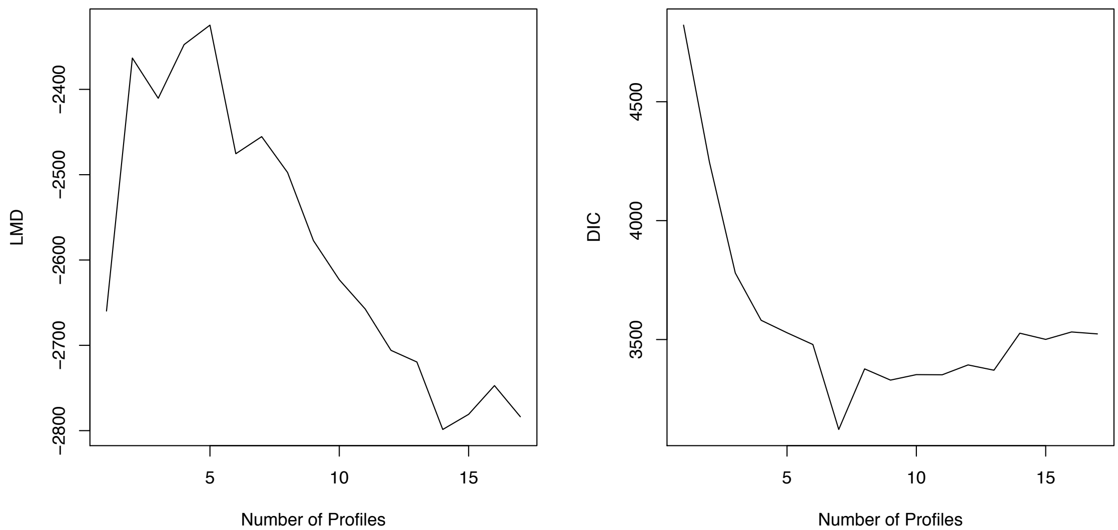

Figure 3.7: Specifying K

set of respondents N^*, observations H^*, and post-burn-in MCMC draws R^*. The hit probability for a given model M is:

$$\text{HP}(M) = \frac{1}{N^*} \sum_{n^*=1}^{N^*} \left[\frac{1}{H^*} \sum_{h^*=1}^{H^*} \left(\frac{1}{R^*} \sum_{r^*=1}^{R^*} Pr(j|\beta^M_{n^*,r^*}, X_{h^*})_{n^*} \right) \right] \tag{3.11}$$

where j is the observed choice from the design matrix X_{h^*} for each observed choice task H^* and $\beta^M_{n^*,r^*}$ are respondent n^*'s estimated coefficients for each of the R^* post-burn-in MCMC draws for model M. The two hit probabilities of interest are for the hold-out tasks from the calibration sample and the hold-out sample, respectively. For the calibration hold-out task hit probability, $N^* = 257$, $H^* = 1$, $R^* = 401$, and $\beta^M_{n^*,r^*}$ are available from each model M. For the hold-out sample hit probability, $N^* = 75$, $H^* = 16$, $R^* = 401$, and $\beta^M_{n^*,r^*}$ is drawn from the distribution of heterogeneity, $N(\Gamma^{M\prime}_{r^*} z^M_n, V^M_{\beta,r^*})$ for the two alternative models and $N(\Gamma^{M\prime}_{r^*} g^M_{n^*,r^*}, V^M_{\beta,r^*})$ for the proposed model. However, the covariates in the proposed model $g_{n^*}^M$ are generated as part of the model and thus are not available for the hold-out sample.

To address this, the observed choices y_{n^*} for the respondents in the hold-out sample were withheld while their observed categorical responses w_{n^*} were included to produce the covariates $g_{n^*}^M$ needed to compute the hit probability. Following Gelman et al. (2013), we treat the withheld observed choices y_{n^*} for the hold-out sample respondents as missing data and employ data augmentation to impute the missing observations at each iteration in the MCMC chain. This allows us to produce covariates $g_{n^*}^M$ for the hold-out sample that are informed by the complete model, including the hold-out sample's observed categorical responses w_{n^*} and the calibration sample's observed choices y_n and categorical responses w_n, and thus draw β_{n^*,r^*}^M to compute the hit probability. We perform this data augmentation for the hold-out sample respondents for each of the reported models.

Table 3.5 demonstrates that, across all measures of model fit, covariates uncovered with mixed membership modeling have more explanatory and predictive power than standard models using discrete covariates. Another alternative to the proposed model would be to include interactions directly. However, in running this alternative model, problems manifested themselves with only two-way interactions. First, the flexibility of the model induced by including so many covariates clearly allowed for overfitting. As we increased the number of iterations in the Markov chain, we continued to see an improvement in in-sample fit with no change in predictive fit and no sign of convergence. Second, the number of interactions would make interpretation infeasible. For these reasons we don't report the results of this model.

Table 3.5: Model Fit

Model	In-Sample LMD	Out-of-Sample Hit Prob.[1]	Hit Prob.[2]
Intercept $\beta_n = \gamma + \xi_n$	-2,441.048	0.654	0.371
Binary Covariates $\beta_n = \Gamma' z_n + \xi_n$	-2,424.537	0.651	0.300
Membership Vector $\beta_n = \Gamma' g_n + \xi_n$	-2,307.619	0.670	0.451

[1] Using calibration respondent hold-out tasks.

[2] Using hold-out sample respondents where $\beta^M_{n^*,r^*} \sim N(\Gamma^{M\prime}_{r^*} z^M_n, V^M_{\beta,r^*})$ or $N(\Gamma^{M\prime}_{r^*} g^M_{n^*,r^*}, V^M_{\beta,r^*})$.

The proposed model also improves inference regarding the drivers of preference heterogeneity. To illustrate, let's consider the posterior means of Γ from the Binary Covariates model. Table 3.6 displays the complete Γ matrix. The attribute levels are on the left and each column in the matrix is associated with the intercept or one of the statements from Table 3.2. The posterior means highlighted in red and green are more than two standard deviations below and above zero, respectively. This matrix should inform a marketer concerning the drivers of preference for promotion and targeting strategies. However, making sense of the significant values or considering how these items may interact is cumbersome.

Table 3.6: Binary Covariates Model Γ Estimates

Attribute Levels	Int.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Neato	1.52	0.03	0.22	-0.05	-0.31	2.19	-1.36	-0.78	-2.18	-0.74	-0.29	3.75	-1.03	1.65	-0.83	-0.66	0.56	-1.25	-0.17
iRobot	2.76	0.26	0.79	-0.31	-0.50	2.41	-1.73	-1.65	-2.53	-0.26	-0.26	3.27	-0.84	1.80	-0.95	-0.24	0.52	-0.69	0.22
Samsung	2.96	-0.46	0.42	-0.19	-0.23	2.27	-1.53	-1.02	-1.82	-0.76	-0.39	3.69	-0.95	1.37	-1.38	-0.83	1.00	-0.77	-0.27
Black & Decker	2.69	-0.42	0.21	-0.53	-0.10	1.99	-0.86	-1.70	-1.85	-0.55	-0.05	4.23	-1.81	2.77	-0.60	-0.58	0.82	-1.18	-0.07
Performance: 85%	0.91	0.43	0.29	0.47	0.68	-1.17	0.10	0.93	0.68	-0.25	0.21	-0.43	0.92	-0.31	0.83	0.34	0.92	-0.15	0.61
Capacity: Every 2-3 uses	0.47	-0.22	0.24	0.09	0.26	-0.02	0.21	0.07	0.03	-0.36	-0.15	-0.07	0.02	0.19	-0.10	0.14	0.43	-0.19	-0.14
Smart Navigation	0.25	0.47	-0.32	0.30	0.14	0.21	0.40	0.18	0.06	0.21	0.11	-0.34	-0.21	-0.14	-0.14	-0.22	-0.74	0.31	-0.38
App Programming	-0.50	0.14	0.48	-0.07	-0.12	0.57	-0.07	-0.10	-0.12	0.63	0.22	-1.15	0.05	0.34	-0.29	0.16	-0.38	-0.04	-0.21
Virtual Borders	1.01	-0.22	-0.29	0.23	0.23	-0.10	0.07	-0.40	0.12	0.02	-0.03	-0.18	-0.08	-1.18	0.30	0.21	-0.63	0.04	0.44
$399	-1.10	-0.22	0.07	-0.36	-0.40	0.43	0.82	-1.11	0.73	-1.43	0.03	1.90	-1.16	-0.71	-0.07	0.65	-0.22	-0.28	-0.41
$499	-3.04	-0.87	0.37	0.13	-1.29	1.22	1.14	-1.97	1.02	-2.54	-0.30	3.22	-2.93	-1.48	0.07	1.57	-0.87	0.33	-0.60
$599	-4.83	-0.96	0.81	0.61	-1.41	0.77	0.76	-1.75	0.90	-2.69	-0.29	4.28	-3.41	-2.56	-0.71	2.04	-1.64	0.16	-0.43

For example, we can use Table 3.6 to infer that respondents who are concerned about germs and dirt (i.e., statement 5 "I worry about germs and dirt on my floor and carpet") prefer any brand of robotic vacuum relative to the outside good while not being concerned about getting the highest level of performance. We might expect this is because they are cleaning frequently (e.g., statement 10 "I spend over two hours per week cleaning") and having a robotic vacuum is simply one part of a larger cleaning solution. Without a way to properly account for interactions, we aren't able to understand these more detailed explanations of preference heterogeneity.

The proposed model accounts for such interactions by identifying differentiated respondent profiles. Table 3.7 details the profiles as described by the estimates of $\lambda_{j,k}(1)$. Since the respondents were qualified by owning or being interested in a robotic vacuum, it isn't surprising that every profile has statement 1 "I enjoy coming home to a clean house" occurring with high probability. Profile 1 is differentiated from the other models by statement 2 "I don't feel relaxed when I know my home isn't clean," statement 10 "I spend over two hours per week cleaning," and statement 5 "I worry about germs and dirt on my floor and carpet" occurring with high probability and statement 11 "I have a cleaning person who cleans for me" occurring with the lowest probability. We name this profile "Constantly Cleaning."

Profile 2 is differentiated by statement 12 "Robotic vacuums are too expensive," statement 9 "I don't spend much time cleaning," and statement 10 "I spend over two hours per week cleaning" occurring with high probability and statement 13 "Robotic vacuums are too complicated to program, set up, and operate" occurring with the lowest probability. We name this profile "Price Sensitive with Little Cleaning." Profile 3 is differentiated by statement 2 "I don't feel relaxed when I know my home isn't

Table 3.7: Membership Vector Model $\lambda_{j,k}(1)$ Estimates

No.	Statements	$\lambda_{j,1}(1)$	$\lambda_{j,2}(1)$	$\lambda_{j,3}(1)$	$\lambda_{j,4}(1)$	$\lambda_{j,5}(1)$
1	I enjoy coming home to a clean house.	0.75	0.65	0.87	0.89	0.96
2	I don't feel relaxed when I know my home isn't clean.	0.56	0.14	0.82	0.48	0.89
3	I worry about pet hair and dander in the home.	0.36	0.14	0.58	0.47	0.82
4	I have trouble keeping the floor beneath my furniture clean.	0.28	0.14	0.44	0.67	0.83
5	I worry about germs and dirt on my floor and carpet.	0.50	0.11	0.77	0.49	0.83
6	I get anxious about having guests when my home is dirty.	0.46	0.28	0.79	0.53	0.93
7	I don't like going to someone's home that is dirty.	0.19	0.18	0.80	0.51	0.91
8	I don't like touching dirty things.	0.16	0.12	0.75	0.18	0.87
9	I don't spend much time cleaning.	0.09	0.31	0.11	0.44	0.06
10	I spend over two hours per week cleaning.	0.51	0.26	0.65	0.41	0.87
11	I have a cleaning person who cleans for me.	0.07	0.04	0.14	0.04	0.08
12	Robotic vacuums are too expensive.	0.36	0.63	0.28	0.92	0.60
13	Robotic vacuums are too complicated to program, set up, and operate.	0.09	0.04	0.08	0.23	0.19
14	Robotic vacuums often need to be "rescued" because they get stuck.	0.21	0.26	0.25	0.35	0.86
15	Robotic vacuums need to have their trash containers changed too often.	0.27	0.17	0.21	0.15	0.50
16	Robotic vacuums don't do a good enough job cleaning the floor and carpet.	0.16	0.06	0.22	0.22	0.40
17	Robotic vacuums don't spend enough time on the really dirty spots on the floor.	0.15	0.26	0.18	0.20	0.18
18	Robotic vacuums scare household pets.	0.17	0.17	0.23	0.26	0.42

Table 3.8: Profile Names

No.	Profile Names
1	Constantly Cleaning
2	Price Sensitive with Little Cleaning
3	Anxious about Cleanliness
4	Price Sensitive with Difficulty Cleaning
5	Anxious and Suspicious

clean," statement 7 "I don't like going to someone's home that is dirty," and statement 6 "I get anxious about having guests when my home is dirty" occurring with high probability. We name this profile "Anxious about Cleanliness."

Profile 4, like profile 2, has statement 12 "Robotic vacuums are too expensive" occurring with high probability, but is further differentiated by statement 4 "I have trouble keeping the floor beneath my furniture clean" and statement 14 "Robotic vacuums often need to be 'rescued' because they get stuck." We name this profile "Price Sensitive with Difficulty Cleaning." Finally, profile 5, like profile 3, has statements 6, 7, and 2 occurring with high probability – statements describing being anxious about cleanliness – as well as, like profile 4, a high probability of statements 14 and 4, which describe difficulty cleaning along with a belief that robotic vacuums get stuck. We name this profile "Anxious and Suspicious."

Table 3.9 displays the matrix of estimated coefficients Γ that maps variability in the membership vectors to variability in the part-worths. Again, the posterior means highlighted in red and green are more than two standard deviations below and above zero, respectively. Note that the size of the coefficients is in part a function of the size of K and the sum-to-one constraint on g_n. As K increases in size, each element

Table 3.9: Membership Vector Model Γ Estimates

Attribute Levels	P1	P2	P3	P4	P5
Neato	26.49	-16.13	0.27	9.04	-20.17
iRobot	27.52	-11.49	1.84	9.86	-20.75
Samsung	24.28	-14.70	8.57	9.18	-21.94
Black & Decker	25.96	-15.93	4.19	10.59	-21.59
Performance: 85%	-1.19	-0.68	-2.15	2.00	20.47
Capacity: Every 2-3 uses	1.93	-0.62	0.18	0.19	1.70
Smart Navigation	0.27	3.10	-1.12	-0.01	2.04
App Programming	1.49	0.77	-1.23	-1.33	-0.06
Virtual Borders	0.00	4.87	-1.38	-1.17	1.31
$399	1.84	-0.30	3.51	-16.36	1.07
$499	3.26	-0.32	4.70	-36.66	0.37
$599	3.09	-0.93	8.40	-53.50	-2.58

of the membership vector g_n gets smaller and the coefficients of Γ get larger to map to the part-worth estimates.

Even taking this constraint into account, the coefficients are still larger than those produced by the standard model as represented in Table 3.6. This is because partial membership in these extreme profiles allows the distribution of preferences to move into the extremes. To illustrate, Figure 3.8 provides the marginal posterior distributions of heterogeneity following Equation (3.2) for the robotic vacuum brands in the study. The densities on the left correspond to each of the models in Table 3.5. M3, the proposed model, is different from M1 and M2 for all four brands. However, this difference results in more variance rather than a shift in the marginal posteriors. The densities on the right correspond to the five profiles from the proposed model, where each respondent has been assigned to the profile with the largest posterior weight

$g_{n,k}$. The pattern of the densities on the right demonstrate why we have an increased variance in the densities for M3 on the left – they are shifted away from zero in such a way so that when they are combined they compose a single distribution with longer tails. That said, the focus in interpreting the coefficients in Table 3.9 remains on their relative sign and magnitude.

As with Table 3.6, the matrix in Table 3.9 should inform a marketer concerning the drivers of preference for promotion and targeting strategies. However, using the proposed model, we are able to explain preferences in terms of the extreme profiles. For example, profile 1, "Anxious about Cleanliness" includes statements 5 "I worry about germs and dirt on my floor and carpet" and 10 "I spend over two hours per week cleaning" with high probability. With this profile we can answer what was only suggested from Table 3.6, that the more an individual is aligned with this profile, the more they prefer any brand of robotic vacuum while caring about a high-capacity robotic vacuum rather than one that performs the best. In other words, since they are cleaning often, they want a robotic vacuum with high capacity in order to effectively assist but not replace other cleaning efforts.

We can better inform targeting and promotion strategies using the proposed model. We can use the estimate of Γ as a roadmap for targeting by matching what respondents prefer with a more detailed explanation of what is driving those preferences. For example, for consumers above a certain threshold in their partial membership in profile 4 "Price Sensitive with Difficulty Cleaning," we know that pricing promotions should be especially effective since they have a need for robotic vacuums but are incredibly price sensitive. The dimension-reduction provided by employing a GoM model makes this plausible with the 12×5 Γ matrix in Table 3.9 compared with a

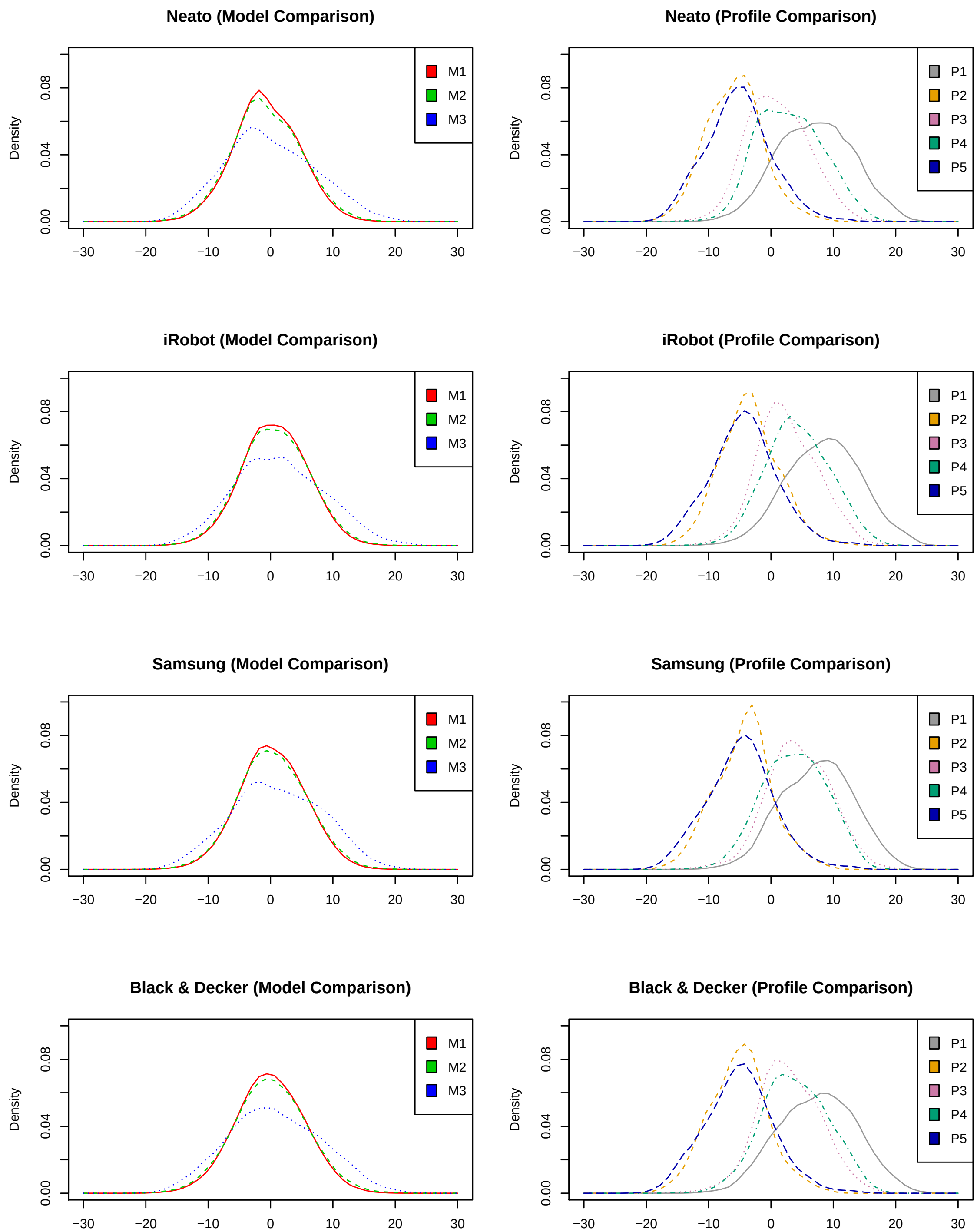

Figure 3.8: Marginal Posterior Distributions of Heterogeneity

similar task using the 12×19 Γ matrix in Table 3.6 from the alternative model or an even larger Γ matrix that includes interactions directly.

Accounting for the co-occurrence or interactions among items is akin to segmenting the market. The blocks of significant attribute level coefficients in Table 3.9 are reminiscent of such segmentation solutions. Unlike mixture models, which are typical in clustering applications, where a respondent is assigned to a single category, mixed membership models like the GoM allow for the more realistic description of each respondent being a partial member of each profile. In our empirical application, it makes sense that consumers interested in robotic vacuums are not going to be constantly cleaning, anxious about cleanliness, skeptical of robotic vacuums, or price sensitive exclusively. Rather, each individual is a mix of all the profiles, with weights determined heterogeneously. Accounting for such differences improves our ability to conduct inference.

3.5 Discussion

In this paper we show that modeling interactions among discrete multivariate data does more to explain consumer preferences than the discrete covariates on their own. This is accomplished by combining a grade of membership model, part of the class of mixed membership models, with choice modeling to estimate membership vectors for use in a hierarchical Bayesian random effects distribution of heterogeneity.

Choice modeling remains an essential fixture of marketing research. However, finding covariates that are explanatory of preference heterogeneity has proven difficult. Our proposed model provides a novel way to account for interactions, and provide dimension reduction, for survey data that explain variation in part-worth utilities.

The empirical application utilizes typical survey response data to demonstrate the use of the proposed model. However, with growing access to unstructured collections of discrete data, we see this approach as an important step to utilizing such data, including text, to improve choice modeling.

Latent Dirichlet allocation, as another kind of mixed membership model, performs in a similar way to the GoM. Text data results in the same kind of sparse matrix as the multinomial data used in the GoM model, with LDA proceeding with words instead of items or statements and a single document for each individual. The dimension reduction using text is even more dramatic when starting with potentially thousands of unique words in the count matrix. However, the amount of data needed to run LDA with words composing the collection of discrete data is significant due to the large number of words in any given vocabulary. Without enough data, there are a variety of developments in topic modeling that are ripe for application within marketing, including using Dirichlet process priors (Ferguson, 1973; Antoniak, 1974) as a kind of distribution of heterogeneity over topic proportions. We leave the practical problems of using text in the place of traditional survey questions as an extension to this research.

Another extension relates to estimating the optimal size of K. While there isn't a consensus as to which measure of model fit provides the gold standard for determining the size of K, there are a number of extant methods for navigating across possible model dimensions that could be employed to include K as a parameter in the model (Green, 1995; Green et al., 2015). The technical details of how to incorporate such methods into the proposed model is left for future research.

More generally, we see the use of mixed membership models as a model-based approach to classifying consumers that yields a more realistic description of the individual as being a mixture of various extreme consumer profiles. This paper serves as a step toward fulfilling a broader need to provide more complete descriptions and explanations of consumer preference heterogeneity.

Bibliography

Airoldi, Edoardo M, David Blei, Elena A Erosheva, Stephen E Fienberg. 2014. *Handbook of Mixed Membership Models and Their Applications*. 1st ed. Chapman & Hall/CRC.

Allenby, Greg M, Neeraj Arora, James L Ginter. 1998. On the Heterogeneity of Demand. *Journal of Marketing Research* **35**(3) 384–389.

Allenby, Greg M, James L Ginter. 1995a. The effects of in-store displays and feature advertising on consideration sets. *International Journal of Research in Marketing* **12** 67–80.

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

Allenby, Greg M, Peter E Rossi. 1998. Marketing Models of Consumer Heterogeneity. *Journal of Econometrics* **89**(1-2) 57–78.

Antoniak, Charles E. 1974. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics* **2**(6) 1152–1174.

Archak, Nikolay, Anindya Ghose, Panagiotis G Ipeirotis. 2011. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science* **57**(8) 1485–1509.

Bagozzi, Richard P, Utpal Dholakia. 1999. Goal Setting and Goal Striving in Consumer Behavior. *Journal of Marketing* **63** 19–32.

Bhat, Chandra R. 1995. A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological* **29**(6) 471–483.

Blei, David M, Jon D McAuliffe. 2007. Supervised topic models. *Neural Information Processing Systems*.

Blei, David M, Andrew Y Ng, Michael I Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* **3** 993–1022.

Büschken, Joachim, Greg M Allenby. 2016. Sentence-Based Text Analysis for Customer Reviews. *Marketing Science* (forthcoming).

Chandukala, Sandeep R, Jeffrey P Dotson, Jeff D Brazell, Greg M Allenby. 2011a. Bayesian Analysis of Hierarchical Effects. *Marketing Science* **30**(1) 123–133.

Chandukala, Sandeep R, Yancy D Edwards, Greg M Allenby. 2011b. Identifying Unmet Demand. *Marketing Science* **30**(1) 61–73.

Chapman, Randall G, Richard Staelin. 1982. Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model. *Journal of Marketing Research* **19**(3) 288–301.

Clive, Jonathan, Max A Woodbury, Ilene C Siegler. 1983. Fuzzy and Crisp Set-Theoretic-Based Classification of Health and Disease. *Journal of Medical Systems* **7**(4) 317–332.

Dellaert, Benedict G C, Bas Donkers, Arthur Van Soest. 2012. Complexity Effects in Choice Experiment–Based Models. *Journal of Marketing Research* **49**(3) 424–434.

Erosheva, Elena A. 2002. Grade of Membership and Latent Structure Models With Application to Disability Survey Data. *Ph.D. thesis, Department of Statistics, Carnegie Mellon University* .

Erosheva, Elena A, Stephen E Fienberg, Cyrille Joutard. 2007. Describing Disability Through Individual-Level Mixture Models for Multivariate Binary Data. *The Annals of Applied Statistics* **1**(2) 346–384.

Fennell, Geraldine, Greg M Allenby. 2014. Conceptualizing and Measuring Prospect Wants: Understanding the Source of Brand Preference. *Customer Needs and Solutions* **1**(1) 23–39.

Ferguson, Thomas S. 1973. A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics* **1**(2) 209–230.

Fiebig, Denzil G, Michael P Keane, Jordan J Louviere, Nada Wasi. 2010. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science* **29**(3) 393–421.

Galyardt, April. 2014. Interpreting Mixed Membership. *Handbook of Mixed Membership Models and Their Applications*. Chapman & Hall/CRC, 39–65.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, Donald B Rubin. 2013. *Bayesian Data Analysis*. Third edition ed. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Green, Peter J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **82**(4) 711–732.

Green, Peter J, Krzysztof Łatuszyński, Marcelo Pereyra, Christian P Robert. 2015. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing* **25**(4) 835–862.

Griffin, Abbie, John R Hauser. 1993. The Voice of the Customer. *Marketing Science* **12**(1) 1–27.

Gross, Justin H, Daniel Manrique-Vallier. 2014. A Mixed-Membership Approach to the Assessment of Political Ideology from Survey Responses. *Handbook of Mixed Membership Models and Their Applications*. Chapman & Hall/CRC, 119–139.

Gutman, Jonathan. 1982. A Means-End Chain Model Based on Consumer Categorization Processes. *Journal of Marketing* **46**(2) 60–72.

Horsky, Dan, Sanjog Misra, Paul Nelson. 2006. Observed and Unobserved Preference Heterogeneity in Brand-Choice Models. *Marketing Science* **25**(4) 322–335.

Johnson, Valen E, James H Albert. 2006. *Ordinal Data Modeling*. Springer Science & Business Media.

Joutard, Cyrille, Edoardo M Airoldi, Stephen E Fienberg, Tanzy M Love. 2007. Discovery of Latent Patterns with Hierarchical Bayesian Mixed-Membership Models and the Issue of Model Choice. *Data Mining Patterns: New Methods and Applications*. IGI Global, Hershey, PA, USA, 1–36.

Kamakura, Wagner A, Gary J Russell. 1989. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research* **26**(4) 379–390.

Kim, Dong Soo, Roger A Bailey, Nino Hardt, Greg M Allenby. 2014. Benefit-Based Conjoint Analysis. *Working Paper* 1–40.

Kim, Yeung Jo, Jongwon Park, Robert S Wyer Jr. 2009. Effects of Temporal Distance and Memory on Consumer Judgments. *Journal of Consumer Research* **36**(4) 634–645.

Lavidge, Robert J, Gary A Steiner. 1961. A Model for Predictive Measurements of Advertising Effectiveness. *Journal of Marketing* **25**(6) 59–62.

Lee, Sik-Yum. 2007. *Structural Equation Modeling: A Bayesian Approach*, vol. 711. John Wiley & Sons.

Lee, Thomas Y, Eric T Bradlow. 2011. Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research* **48**(5) 881–894.

Lenk, Peter J, Wayne S DeSarbo, Paul E Green, Martin R Young. 1996. Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science* **15**(2) 173–191.

Louviere, Jordan J, Deborah Street, Richard Carson, Andrew Ainslie, J R Deshazo, Trudy Cameron, David Hensher, Robert Kohn, Tony Marley. 2002. Dissecting the Random Component of Utility. *Marketing Letters* **13**(3) 177–193.

Luo, Lan, P K Kannan, Brian T Ratchford. 2008. Incorporating Subjective Characteristics in Product Design and Evaluations. *Journal of Marketing Research* **45**(2) 182–194.

Manton, Kenneth G, Max A Woodbury, H Dennis Tolley. 1994. *Statistical Application Using Fuzzy Sets*. Wiley, New York.

Marini, Margaret Mooney, Xiaoli Li, Pi-Ling Fan. 1996. Characterizing Latent Structure: Factor Analytic and Grade of Membership Models. *Sociological Methodology* **26** 133–164.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, Moshe Fresko. 2012. Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science* **31**(3) 521–543.

Netzer, Oded, Olivier Toubia, Eric T Bradlow, Ely Dahan, Theodoros Evgeniou, Fred M Feinberg, Eleanor M Feit, Sam K Hui, Joseph Johnson, John C Liechty, James B Orlin, Vithala R Rao. 2008. Beyond conjoint analysis: Advances in preference measurement. *Marketing Letters* **19**(3-4) 337–354.

Newton, Michael A, Adrian E Raftery. 1994. Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(1) 3–48.

Rossi, Peter E, Greg M Allenby. 2003. Bayesian Statistics and Marketing. *Marketing Science* **22**(3) 304–328.

Rossi, Peter E, Greg M Allenby, Robert E McCulloch. 2005. *Bayesian Statistics and Marketing*. J. Wiley and Sons.

Rossi, Peter E, Robert E McCulloch, Greg M Allenby. 1996. The Value of Purchase History Data in Target Marketing. *Marketing Science* **15**(4) 321–340.

Salisbury, Linda Court, Fred M Feinberg. 2010. Alleviating the Constant Stochastic Variance Assumption in Decision Research: Theory, Measurement, and Experimental Test. *Marketing Science* **29**(1) 1–17.

Spiegelhalter, David J, Nicola G Best, Bradley P Carlin, Angelika Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4) 583–639.

Stewart, David W. 1981. The Application and Misapplication of Factor Analysis in Marketing Research. *Journal of Marketing Research* **18**(1) 51–62.

Strong, E K Jr. 1925. Theories of selling. *Journal of Applied Psychology* **9**(1) 75–86.

Swait, Joffre, Jordan J Louviere. 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* **30**(3) 305–314.

Tanner, Martin A, Wing Hung Wong. 1987. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* **82**(398) 528–540.

Tirunillai, Seshadri, Gerard J Tellis. 2014. Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research* **51**(4) 463–479.

Vakratsas, Demetrios, Tim Ambler. 1999. How Advertising Works: What Do We Really Know? *Journal of Marketing* **63**(1) 26–43.

van Osselaer, Stijn M J, Chris Janiszewski. 2012. A Goal-Based Model of Product Evaluation and Choice. *Journal of Consumer Research* **39**(2) 260–292.

Woodbury, Max A, Jonathan Clive, Arthur Garson Jr. 1978. Mathematical Typology: A Grade of Membership Technique for Obtaining Disease Definition. *Computers and Biomedical Research* **11**(3) 277–298.

Yang, S, Greg M Allenby, Geraldine Fennell. 2002. Modeling Variation in Brand Preference: The Roles of Objective Environment and Motivating Conditions. *Marketing Science* **21**(1) 14–31.

Appendix A: Generating Data for Chapter 2

We generate data using the extended model of brand choice to demonstrate parameter recovery in a simulation experiment. First, set the true values of Δ, V_β, and γ. Then, for each respondent h:

1. Generate a vector of M needs z_h by rounding random draws from a continuous uniform distribution.

2. Generate a matrix of $N \times M$ brand beliefs B_h by rounding random draws from a continuous uniform distribution.

3. Draw $\beta_h \sim \text{MVN}(\Delta' w_h, V_\beta)$. Recall that $\beta_h' = [\beta_{zh}, \beta_{ph}]$.

4. Compute a brand intercept for each alternative j $\beta_{0jh} = \sum_{m=1}^{M} \beta_{mzh} b_{mjh}$.

5. Create β_{0ph} by concatenating β_{0h} and β_{ph}.

6. For each benefit choice task $s = 1, \ldots, S$:

   (a) Generate the design matrix $X_s^{L_1}$.

   (b) Compute latent utility for each alternative: $U_{sh} = X_s^{L_1} \beta_{zh} + \epsilon_{sh}$ where $\epsilon_{sh} \sim EV(0, 1)$ and is generated by $-\log(-\log(Unif(0, 1)))$.

   (c) Rank each alternative according to latent utility.

7. For each brand-price choice task $t = 1, \ldots, T$:

(a) Generate the design matrix $X_t^{L_2}$ with full-rank alternative-specific constants for the N brands.

(b) Calculate:

$$\sigma_{jh} = \exp\left[\gamma \cdot I\left(\sum_{m=1}^{M} z_{mh} \geq 1\right) \cdot I\left(\sum_{m=1}^{M} b_{mjh} z_{mh} = \sum_{m=1}^{M} z_{mh}\right)\right].$$

(c) Compute latent utility for each alternative: $U_{th} = X_t^{L_2}\beta_{0ph} + \epsilon_{th}$ where $\epsilon_{th} \sim EV(0, \sigma_{jh})$ and is generated by $-\log(-\log(Unif(0,1))) \cdot \sigma_{jh}$.

(d) Rank each alternative according to latent utility. Account for the possibility that up to three of the N brands aren't in the respondent's consideration set and so aren't included in $X_t^{L_2}$.

Appendix B: Estimation Procedure for Chapter 2

Extended models of behavior are characterized by conditional independence. We employ two likelihoods that are connected with the definition of the brand intercepts as the conditioning argument. It is also important to note that β_h is heterogeneous and γ is homogeneous. We proceed with estimation using a random-walk Metropolis-Hastings algorithm as follows. For each iteration in the Markov chain:

1. Initialize the log likelihood value at 0 for the given iteration.

2. Draw the candidate value γ^c where the old value γ^d is initialized at 0 and $\gamma^c = \gamma^d + \epsilon, \epsilon \sim \text{Normal}(0, s_\gamma)$, where s_γ is the step-size specified for γ.

3. For each respondent $h = 1, \ldots, H$:

   (a) Draw the candidate values β_h^c where the old values β_h^d are initialized at 0 and $\beta_h^c = \beta_h^d + \epsilon, \epsilon \sim \text{MVN}(0, s_\beta \cdot V_\beta)$, where s_β is the step-size specified for all β_h.

   (b) Compute each candidate brand intercept $\beta_{0(j)h}^c = \sum_{m=1}^{M} \beta_{mzh}^c b_{m(j)h}$.

   (c) Create the candidate coefficient vector for the brand-price tradeoff β_{0ph}^c by concatenating β_{0h}^c and β_{ph}^c.

   (d) Compute the log likelihood of the benefit evaluation L_1 with the exploded multinomial logit with y_{L_1h} and X^{L_1} for β_{zh}^d and β_{zh}^c using Equation (2.4).

(e) Compute the log likelihood of the brand-price evaluation L_2 with the exploded multinomial logit with y_{L_2h}, X^{L_2}, and γ^d for β^d_{0ph} and β^c_{0ph}. For responses where $\sigma_{jh} \neq 1$, we no longer have a closed form and must solve for the choice probabilities using numerical integration as in Equation (2.5). We employ the composite Simpson's rule (i.e., three-point quadrature):

$$\begin{aligned} Pr(j)_h &\approx \int_{-c\sigma_{jh}}^{c\sigma_{jh}} \left[\prod_{k=1}^{K} F([V_{jh} - V_{kh} + \epsilon_{jh}]/\sigma_{kh}) \right] f(\epsilon_{jh}/\sigma_{jh}) d\epsilon_{jh} \\ &\approx \frac{\Delta\epsilon_{jh}}{3} \left(y_1 + 4 \sum_{n=2,4,\ldots}^{N} y_n + 2 \sum_{n=3,5,\ldots}^{N-1} y_n + y_{N+1} \right). \end{aligned}$$

Each y represents the height of the corresponding point in the support:

$$y_n = \left[\prod_{k=1}^{K} F([V_{jh} - V_{kh} + x^n_{jh}]/\sigma_{kh}) \right] f(x^n_{jh}/\sigma_{jh}) \text{ for all } k \neq j$$

where the $EV(0, \sigma)$ CDF and PDF are:

$$\begin{aligned} F(x/\sigma) &= \exp\left[-\exp\left(\frac{-x}{\sigma} \right) \right] \\ f(x/\sigma) &= \frac{1}{\sigma} \exp\left[-\frac{x}{\sigma} - \exp\left(\frac{-x}{\sigma} \right) \right]. \end{aligned}$$

The integer c determines how much of the support we cover, N is an even integer, and $\Delta\epsilon_{jh} = (c\sigma_{jh} + c\sigma_{jh})/N$ is the fixed width of each subinterval. After conducting a sensitivity analysis, we set $c = 10$ and $N = 50$. Applying this to the heteroscedastic exploded multinomial logit, and taking into account that we only need to use numerical integration when $\sigma_{(i)h} \neq 1$, we have:

$$Pr(U_{(1)} > U_{(2)} > \cdots > U_{(K)})_h$$

$$\approx \prod_{i=1}^{K-1} \int_{-c\sigma_{jh}}^{c\sigma_{jh}} \left[\prod_{k=i}^{K} F([V_{(i)h} - V_{(k)h} + \epsilon_{(i)h}]/\sigma_{(k)h}) \right] f(\epsilon_{(i)h}/\sigma_{(i)h}) d\epsilon_{(i)h}$$

$$\approx \prod_{i=1}^{K-1} \left[\frac{\exp[V_{(i)h}]}{\sum_{k=i}^{K} \exp[V_{(k)h}]} \right]^{\phi}$$

$$\times \left[\frac{\Delta\epsilon_{(i)h}}{3} \left(y_1 + 4 \sum_{n=2,4,\ldots}^{N} y_n + 2 \sum_{n=3,5,\ldots}^{N-1} y_n + y_{N+1} \right) \right]^{1-\phi}$$

where $\phi = I(\sigma_{(i)h} = 1)$ and each y now takes into account the rank of the data:

$$y_n = \left[\prod_{k=i}^{K} F([V_{(i)h} - V_{(k)h} + x^n_{(i)h}]/\sigma_{(k)h}) \right] f(x^n_{(i)h}/\sigma_{(i)h}) \text{ for all } k \neq i \; .$$

Note that Chapman and Staelin (1982) recommend an exploded depth of 3. Therefore, even if $K > 4$, only consider up to $K = 4$. Finally, normalize the choice probabilities for the given choice task to ensure the probabilities sum to 1 before selecting the choice probability for the chosen alternative to use in computing the likelihood.

(f) Compute the log of the distribution of heterogeneity over all β^d_h and β^c_h according to Equation (2.8), which we denote as $p(\beta^d_h)$ and $p(\beta^c_h)$.

(g) Accept β^c_h with probability (after exponentiation the log likelihoods and the log of the random-effects distributions):

$$\alpha = \min \left(1, \frac{L_1(\beta^c_{zh}) \times L_2(\beta^c_{0ph}, \gamma^d) \times p(\beta^c_h)}{L_1(\beta^d_{zh}) \times L_2(\beta^d_{0ph}, \gamma^d) \times p(\beta^d_h))} \right).$$

(h) Compute the total log likelihood for γ^c by summing the log likelihood of the benefit evaluation with β^d_{zh} using Equation (2.4) and the log likelihood of the brand-price evaluation with y_{L_2h}, X^{L_2}, and β^d_{0ph} for γ^c using Equation (2.5).

4. Draw Δ and V_β using $B_{\beta^d_h} = W\Delta + \eta$ where $\eta \sim \text{MVN}(0, V_\beta)$.

5. Accept γ^c with probability (after exponentiating the total log likelihoods):

$$\alpha = \min\left(1, \frac{\text{Total Likelihood}(\gamma^c) \times p(\gamma^c)}{\text{Total Likelihood}(\gamma^d) \times p(\gamma^d)}\right)$$

   where the Normal prior on γ is denoted $p(\gamma)$. Note that the total likelihood of γ^d is equivalent to the total likelihood from the respondent-level step. In other words, we accept γ^c only if it leads to an improvement in the likelihood of the model overall.

6. Print the acceptance rate every 5 iterations for β^c_h, the number of accepted candidate draws of β_h divided by the number of respondents, and for γ^c, the number of accepted candidate draws of γ divided by 5. The acceptance rate for γ^c resets after every 5 iterations.

7. The step sizes s_β and s_γ will be tuned during the first third of iterations if the acceptance rates are greater than .60 or less than .20.

Appendix C: Hit Probabilities for Ranked Data

Hit probabilities are the average predicted choice probabilities of making the observed choices in the hold-out sample given a specific model. These predicted probabilities are averaged over respondents H, hold-out observations S, and post-burn-in MCMC draws R^*. The hit probability for model M is:

$$\mathrm{HP}(M) = \frac{1}{H}\sum_{h=1}^{H}\left[\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{R^*}\sum_{r=1}^{R^*} P(j|\beta_{hr}^{M}, X_s)_h\right)\right]$$

where j is the observed choice from the design X_s for each choice scenario S in the hold-out sample and β_{hr}^{M} are respondent h's estimated coefficients for each of the R^* post-burn-in MCMC draws for model M.

Applying this to ranked data, the observed choice becomes a sequence of choices:

$$\mathrm{HP}(M) = \frac{1}{H}\sum_{h=1}^{H}\left[\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{R^*}\sum_{r=1}^{R^*} Pr(U_{(1)} > U_{(2)} > \cdots > U_{(K)}|\beta_{hr}^{M}, X_s)_h\right)\right].$$

Appendix D: Comparing Factor Analysis and the GoM

To detail the differences between factor analysis and the grade of membership (GoM) model, it is useful to write down a factor analytic model for binary data in the form of a cut-point model (Lee, 2007). In this model, observed responses from respondent n, w_n, are generated as follows:

$$w_{n,j} = 1 \text{ if } z_{n,j} > 0 \ \forall j \in \{1, \ldots, J\}, \quad z_n \sim N(\Lambda\zeta_n, \Sigma) \tag{D.1}$$

where w_n is a $J \times 1$ vector of observed binary responses, z_n is a $J \times 1$ vector of latent continuous responses, ζ_n is a $K \times 1$ vector of factor scores, the $J \times K$ matrix Λ indicates factor loadings, Σ indicates a $J \times J$ covariance matrix of the z's, and the threshold 0 is a fixed, arbitrarily chosen cut-point. The model in Equation (D.1) can easily be extended to ordinal data via additional cut-points (Johnson and Albert, 2006). Apart from the normal errors specification of this model and the subsequent need for identification constraints because of its scale invariance, this is a standard factor model and equivalent to a model in which multiple observed responses are regressed on unobserved factor scores.

The probability of observing a single response $w_{n,j}$, given this specification, can be expressed as an integral over the z space:

$$\begin{aligned} Pr(w_{n,j} = 1|\zeta_n, \Lambda) &= \int_0^\infty p(z_{n,j}|\{z_{n,-j}\}, \zeta_n, \Lambda, \Sigma)dz \\ &= \int_0^\infty N(\zeta'_{n,k}\gamma_{j,k}, \sigma^2_{n,j})dz \\ Pr(w_{n,j} = 0|\zeta_n, \Lambda) &= 1 - \int_0^\infty N(\zeta'_{n,k}\gamma_{j,k}, \sigma^2_{n,j})dz \end{aligned} \tag{D.2}$$

where $\sigma^2_{n,j}$ is the univariate variance of $z_{n,j}$, conditional on $\{z_{n,-j}\}$. In the case of conditionally independent regression errors, $\sigma^2_{n,j} = \sigma^2_n$. Equation (D.2) expresses the probability of observing a given response as the integral over the latent z space truncated at 0, given unobserved unit-level factor scores and across-unit factor loadings.

The GoM model expresses the probability of observing response l as an individual-level, multinomial mixture model of K profiles in which each response option for each question has profile-specific multinomial choice probabilities that are mixed over unit-specific weights (Erosheva et al., 2007):

$$Pr(w_{n,j} = l|g_n, \lambda) = \sum_{k=1}^{K} g_{n,k}\lambda_{j,k}(l) \tag{D.3}$$

in which the g_n indicate respondent-level weights of the profiles and $\lambda_{j,k}(l)$ is the probability of observing response l to question j given exclusive membership to profile k. As with every mixture model, the GoM model imposes the following constraints on the weights: $0 \leq g_{n,k} \leq 1$ and $\sum_{k=1}^{K} g_{n,k} = 1$. In the case of binary responses, the

GoM model is simply:

$$\begin{aligned}
Pr(w_{n,j} = 1 | g_n, \lambda) &= \sum_{k=1}^{K} g_{n,k} \lambda_{j,k}(1) \\
Pr(w_{n,j} = 0 | g_n, \lambda) &= \sum_{k=1}^{K} g_{n,k} (1 - \lambda_{j,k}(1))
\end{aligned} \tag{D.4}$$

where the g_n are subject to the same constraints.

Comparing the factor model in Equation (D.2) to the GoM model in Equation (D.4) reveals several differences between the two models. First, the GoM model is an individual-level mixture model of K latent profiles. The factor model is essentially a linear multivariate regression model. This leads to a different interpretation of g_n, compared to the latent factor scores ζ_n. The g_n represent convex weights over a multidimensional latent space whereas the factor scores are the set of latent sources of observed responses, each of which is unidimensional and which contribute to the observed responses in a linear fashion. The important difference lies in what the underlying construct is. In the GoM model, a profile is defined as a set of response probabilities across *all* J questions and their response options. In factor analysis, a factor is assumed to exist independently from the measurements.

Second, and because of its mixture model property, the GoM model allows us to capture response heterogeneity in two ways. First, it allows for unit-level latent scores g_n, which captures differences among respondents. Second, it captures heterogeneity in responses through profile-specific response probabilities $\lambda_{j,k}$. Individual response behavior is described in terms of the similarity of individual and profile-specific response probabilities. An individual's response behavior more similar to one of the profiles across *all* responses is expressed by a higher weight of that profile for that individual.

Third, factor analysis makes specific assumptions concerning the distribution of observed responses. More specifically, factor analysis assumes that for the data in Equation (D.2), the z are distributed multivariate normal. The GoM model, in comparison, makes no assumption about the joint distribution of observed responses.

This suggests that whether or not the GoM model is to be preferred over a factor model is essentially a question of which model is an adequate description of respondent heterogeneity in a particular application. The GoM model describes heterogeneity as similarity between individuals and extreme profiles. The number of extreme profiles in the GoM model is defined a priori and can be used to reduce the dimensionality of the response space. Factor analysis is often used for the same purpose, but it lacks the property of locating individuals in the convex space spanned by extreme response behavior.

Appendix E: Generating Data for Chapter 3

We generate data using the proposed hierarchical Bayesian choice model with a GoM model to demonstrate parameter recovery in a simulation experiment. First, fix the number of profiles K, the number of respondents N, the number of categorical questions J, and the categorical levels for each question n_j (e.g., $n_j = 2$ for all J in the case of pick any/J data), the number of choice tasks H, the number of alternatives in each choice task P, and the number of attribute levels L. Next, set the true values of Γ, V_β, and λ. We set α and τ to be vectors of 1 for the Dirichlet priors, creating a uniform distribution on the respective simplex to mirror our lack of information regarding partial membership and profile composition.

For each respondent n, we proceed as follows:

1. Draw $g_n \sim \text{Dirichlet}(\alpha)$, the membership vector.
2. For each of the $j = 1, \ldots, J$ questions:
   (a) Draw $z_{n,j} \sim \text{Multinomial}(g_n)$, a profile assignment.
   (b) Draw $w_{n,j} \sim \text{Multinomial}(\lambda_{j,k=z_{n,j}}(1), \ldots, \lambda_{j,k=z_{n,j}}(n_j))$, a categorical response drawn from the appropriate entry in $\lambda_{j,k}$ indexed by j and $k = z_{n,j}$.
3. Draw $\beta_n \sim \text{Normal}(\Gamma' g_n, V_\beta)$.
4. For each of the $h = 1, \ldots, H$ choice tasks:

(a) Generate a design matrix $X_{n,h}$.

(b) Compute latent utility for each alternative p:

$$U_{n,h,p} = X_{n,h,p}\beta_n + \varepsilon_{n,h,p};\ p = 1, \ldots, P$$

where $\varepsilon_{n,h,p} \sim \mathrm{EV}(0, 1)$.

(c) Let $y_{n,h} = \arg\max_p \left(\{U_{n,h,p}\}_{p=1}^{P}\right)$.

Appendix F: Estimation Procedure for Chapter 3

We employ a Markov chain Monte Carlo estimation procedure with both random-walk Metropolis-Hastings and Gibbs steps. A Gibbs sampler similar to that detailed in Erosheva et al. (2007) is used to estimate the GoM portion of the proposed model. However, since g_n is included in the distribution of preference heterogeneity, we use a random-walk Metropolis-Hastings algorithm to estimate g_n's that are predictive of the part-worths. The remaining choice model portions of the proposed model utilize standard estimation methods. We proceed with estimation as follows for R iterations:

1. For each of the $n = 1, \ldots, N$ respondents:

   (a) For each of the $j = 1, \ldots, J$ questions, draw $z_{n,j}$ using $g_{n,k}$, the partial membership respondent n has in each profile k, and $\lambda_{j,k}(w_{n,j})$, the probability of the chosen response $w_{n,j}$ for each profile k:

   $$z_{n,j} = \arg\max_k \left(\text{Multinomial}(p_1, \ldots, p_K)\right), \text{ where } p_k \propto g_{n,k}\lambda_{j,k}(w_{n,j}).$$

   (b) Draw g_n^{new} using a random-walk Metropolis-Hastings step where g_n^{old} is initialized at $1/K$ for all K elements and $g_n^{new} = \text{Dirichlet}(g_n^{old} \times s_p)$, where s_p is the specified step size. The larger s_p is, the closer g_n^{new} will be to g_n^{old}.

(c) Accept g_n^{new} with probability

$$\alpha_{\text{accept}} = \min\left(1, \frac{\left[\prod_{j=1}^{J} p(z_{n,j}|g_n^{new})\right] p(g_n^{new}|\alpha)p(\beta_n^{old}|g_n^{new}, \Gamma, V_\beta)p(g_n^{old}|g_n^{new})}{\left[\prod_{j=1}^{J} p(z_{n,j}|g_n^{old})\right] p(g_n^{old}|\alpha)p(\beta_n^{old}|g_n^{old}, \Gamma, V_\beta)p(g_n^{new}|g_n^{old})}\right)$$

where the random-walk proposal density $p(g_n^{old}|g_n^{new})$ and $p(g_n^{new}|g_n^{old}) \sim$ Dirichlet.

(d) Draw β_n^{new} using a random-walk Metropolis-Hastings step where β_n^{old} is initialized at 0 and $\beta_n^{new} = \beta_n^{old} + \epsilon, \epsilon \sim N(0, V_\beta \times s_\beta)$, where s_β is the specified step size. The smaller s_β is, the closer β_n^{new} will be to β_n^{old}.

(e) Accept β_n^{new} with probability

$$\alpha_{\text{accept}} = \min\left(1, \frac{p(y_n|X_n, \beta_n^{new})p(\beta_n^{new}|g_n^{old}, \Gamma, V_\beta)}{p(y_n|X_n, \beta_n^{old})p(\beta_n^{old}|g_n^{old}, \Gamma, V_\beta)}\right)$$

where the multivariate normal random-walk proposal density cancels out.

2. Draw Γ and V_β using $B = \Gamma' G + \Xi$ where G is a matrix with each g_n^{old} as a row vector, B is a matrix with each β_n as a row vector, and $\Xi \sim N(0, V_\beta)$.

3. Draw λ using counts of the augmented variable z:

$$\lambda_{j,k} \sim \text{Dirichlet}(p_{k,1}, \ldots, p_{k,n_j}), \text{ where } p_{k,l} \propto 1 + \sum_{n=1}^{N} I(z_{n,j} = l)$$

for $j = 1, \ldots, J$ and $k = 1, \ldots, K$.