Chapter 20

HIERARCHICAL BAYES MODEL

Greg M. Allenby, Ohio State University

Peter E. Rossi, University of Chicago

Introduction

Bayesian methods have become widespread in marketing. The past ten years have seen a dramatic increase in their use to develop new methods of analysis and models of consumer behavior. A challenge in the analysis of marketing data is that, at the individual-level, the quantity of relevant data is very limited. Respondents frequently become fatigued after providing 15-20 responses in a conjoint survey, and purchase histories greater than 20 observations are rare for all but a few product categories. The lack of data at the individual-level corresponding to a specific construct such as preferences, coupled with the desire to account for individual differences and not treat all respondents alike, results in severe challenges to the analysis of marketing data. Bayesian methods are ideally suited for analysis with limited data, and have resulted in new developments in modeling individual-level decision making, new characterizations of preferences and sensitivities across respondents, and models that include the analysis of a firm's decision in response to consumer demands.

The earliest impact of Bayesian methods in marketing was in the context of discrete choice models and its application to conjoint analysis (Allenby & Lenk, 1994; McCulloch & Rossi, 1994; Allenby & Ginter, 1995). These models were different in that they connected respondent-level parameters to i) a model of decision making; and ii) a model describing the distribution of preferences across respondents. Information from both sources (the individual's responses and the distribution of all responses) were used to learn about a particular respondent's preferences and sensitivity to variables such as prices. This resulted in a significant improvement in estimation, thus solving the limited data problem at the individual-level.

Since that time, new models have been proposed to measure the effects of purchase timing (Allenby, Leone, & Jen, 1999), satiation (Kim, Allenby, & Rossi, 2002), the presence of decision heuristics such as screening rules (Gilbride & Allenby, 2004), to name just a few. Human behavior is complex, and many models of behavior are currently being developed that reflect this complexity. While marketing has long recognized the importance of linking consumer needs to marketplace demand, it is just beginning to estimate extended models of behavior that include the relationship of needs to desired attributes or wants, wants to brand beliefs and consideration sets, and consideration sets to preference orderings and choice. These extended models are often conceptualized in a hierarchical manner, where movement from one model component to the next proceeds in a logical manner. Estimation of these new integrated models is not possible without Bayesian methods.

The nature and determinants of heterogeneity has also received much attention over the last ten years. Across dozens of studies, the distribution of heterogeneity has been shown to be better represented by a continuous, not a discrete distribution (e.g., from a finite mixture model) of heterogeneity (Allenby, Arora, & Ginter, 1998). This has important implications for analysis connected with market segmentation, where researchers often incorrectly assert the existence of a small number of homogeneous groups. Bayesian methods are being used to identify new basis variables that point to brand preferences (Yang, Allenby, & Fennell, 2002), new ways of dealing with respondent heterogeneity in scale usage (Rossi, Gilula, & Allenby, 2001), and new ways of characterizing social networks and their impact on demand (e.g., inter-dependent preferences, Yang & Allenby, 2003). These developments would not be possible modern Bayesian methods.

In this chapter, we provide an introduction to hierarchical Bayes models and overview of successful applications. Underlying assumptions are discussed in the next section, followed by an introduction to the computational arm of Bayesian analysis known as Markov chain Monte Carlo methods. A case study is then used to illustrate the use of Bayesian methods in the context of a conjoint study. A discussion of challenges in using hierarchical Bayes models follows, closing with an annotated bibliography of Bayesian models and applications.

Underlying Assumptions

The analysis of marketing data ranges from simple summaries of events (e.g., the average response) to analysis that attempts to uncover factors associated with, and is predictive of, the behavior of specific individuals. The desire to look behind the data requires models that reflect associations of interest. Consider, for example, an analysis designed to determine the influence of price on the demand for a product or service. If the offering is available in continuous units (e.g., minutes of cell phone usage), then a regression model (see Chapter 13) can be used to measure price sensitivity using the model:

$$y_t = \beta_0 + \beta_1 price_t + \varepsilon_t \qquad ; \qquad \varepsilon_t \sim Normal\left(0, \sigma^2\right) \qquad (1)$$

where y_t denotes demand at time "t", $price_t$ is the price at time t, and β_0, β_1 and σ^2 are parameters to be estimated from the data. The parameters β_0 and β_1 define the expected association between price and demand. Given the price at any time, t, one can compute $\beta_0 + \beta_1 price_t$ and obtain the expected demand, y_t. The parameter σ^2 is the variance of the error term ε_t, and reflects the uncertainty associated with relationship. Large values of σ^2 are associated with noisy predictions, and small values of σ^2 indicate an association without much uncertainty.

Individual-level demand, however, is rarely characterized by such a smooth, continuous association. The most frequently observed quantity of demand at the individual-level is zero (0), and the next most frequently observed quantity is one (1). Marketing data, at the individual-level, is inherently discrete and non-continuous. One approach to dealing with the discreteness of marketing data is to assume that the observed demand is a censored realization of an underlying continuous model:

$$y_t = \begin{cases} 1 & \text{if } \beta_0 + \beta_1 price_t + \varepsilon_t > 0 \\ 0 & \text{if } \beta_0 + \beta_1 price_t + \varepsilon_t \le 0 \end{cases} \quad ; \quad \varepsilon_t \sim Normal\left(0, \sigma^2\right) \tag{2}$$

Since individuals typically like to pay less for a good than more, the price coefficient (β_1) is usually negative and demand is zero for high values of price. As price falls, the likelihood of nonzero demand increases.

Equation (2) is an example of a model that allows a researcher to understand the data beyond that which is possible with graphical methods or cross-tabulation of the data. Graphical methods (e.g., scatterplots) can be used by researchers to detect the presence of a relationship between demand and price, but cannot be used to quantify the relationship. Cross-tabs provide an approach to understand the relationship between variables that take on a discrete number of values, but are difficult to use when one of the variables is continuous, such as price. The advantage of obtaining parameter estimates of β_0, β_1 and σ^2 in equation (2) is that they provide a quantification of the relationship between demand and price that can be used to explore the effect of any hypothetical price, not just a few or those that were observed in the past.

Until recently, marketing practitioners have not made widespread use of models to get behind the data and quantify relationships. An exception is in conjoint analysis, were models similar to equation (2) are being used to quantify the value (i.e., part-worth) of attributes and benefits of product offerings. Marketing academic are currently developing new models for the analysis of marketing data, including demand data from the marketplace and data from questionnaires. These models, often written in hierarchical form, offer new insight into consumer behavior and its correlates.

Hierarchical Models

Consider equation (2) where observed demand is thought of as a censored realization of an underlying, continuous process. The use of censoring mechanisms to deal with the discreteness of marketing data can be written as a hierarchical model by introducing a latent variable, z_t:

$$y_t = \begin{cases} 1 & if \;\; z_t > 0 \\ 0 & if \;\; z_t \leq 0 \end{cases} \tag{3}$$

$$z_t = \beta_0 + \beta_1 price_t + \varepsilon_t \qquad ; \qquad \varepsilon_t \sim Normal\left(0, \sigma^2\right) \tag{4}$$

Hierarchical models make use of a property call conditional independence. For the above model, the latent variable, z_t, is sufficient for making inferences about β_0, β_1 and σ^2. If we were able to observe z_t directly, no additional information would be revealed about these parameters by the discrete data y_t. In other words, inferences about β_0, β_1 and σ^2 are independent of y_t given z_t. All information about the parameters flows through the latent variable (z_t).

It is often useful to think of models hierarchically, where the story told by the model is elaborated with additional models, or equations. Equation (3) can be used to describe a scenario where a purchase is made ($y_t = 1$) if the value of an offering is sufficiently large ($z_t > 0$). Equation (4) then relates value to price, allowing it change as price changes. Further elaboration could include equations for the price coefficient, possibly describing when an individual is expected to be price sensitive and when they are not expected to be sensitive to price. Alternatively, additional covariates could be included in equation (4) to explain other environmental and personal factors. As a third example, one could think of data from multiple respondents modeled with equations (3) and (4), and the distribution of coefficients (β_0,β_1) distributed in the population according to a distribution (e.g., bivariate normal) whose parameters are to be estimated (i.e., a random-effects model).

In marketing, hierarchical models have been used to describe i) the behavior of specific respondents in a study and ii) the distribution of responses among respondents. The former is a model of within-unit behavior over time, and the later reflects cross-sectional variation in model parameters, often referred to as the distribution of heterogeneity. Marketing data often takes on a panel structure with multiple responses (e.g., purchases) per respondents, which allows estimation of parameters associated with each model component. An illustration of such an analysis is provided below.

Bayesian Analysis

Hierarchical Bayes models are hierarchical models analyzed using Bayeisan methods. Bayesian methods are based on the assumption that probability is operationalized as a degree of belief, and not a frequency as is done in classical, or frequentist, statistics. Most researchers in marketing have been trained to think about statistics in terms of frequencies. When computing a sample mean or test statistic, for example, many of us think of multiple realizations of a dataset that could lead to variability of the statistic. Even though the statistic is fixed for the data under investigation, we admit the possibility that other realization of the data could have been obtained. Assuming that the model under investigation is true, we compute the expected variability of the statistic.

To a Bayesian, such calculations are difficult to justify. There are two reasons for this. First, the researcher is usually interested in determining whether or not a particular model is correct, and the assumption that the model under investigation is true seems circular – why assume what you are trying to prove? In addition, the researcher has not observed the multiple realizations of the data required to construct measures of uncertainty. The researcher has only observed one dataset.

An example is used to illustrate the issues. Consider a laboratory testing setting where a test for a heart attack is being developed. There are two states of nature for the patient: heart attack (H+) and no heart attack (H-). Likewise, there are two outcomes of the test: positive (T+) and negative (T-). The laboratory physician is critically interesting in the sensitivity Pr(T+|H+) and specificity Pr(T-|H-) of the test, where "Pr" denotes probability and the vertical bar, "|" is the symbol for conditional probability and means "given that". Large values of sensitivity, Pr(T+|H+), indicate that the test is sensitive to detecting the presence of a heart attack given that one actually occurred. Large values of specificity, Pr(T-|H-), indicate that the test is also good at detecting the absence of a heart attack given that it did not occur. Measures of the sensitivity and specificity of the test are developed by applying the test to multiple patients that are known to have suffered from heart attacks, and also multiple patients that are know to be healthy. The laboratory physician is using statistics in a traditional, frequentist manner – the status of the patient is assumed known and variability in the results, leading to outcome probabilities, is due to repeated samples.

Now consider the problem from the viewpoint of a clinician (e.g., internal medicine physician) examining a specific patient. The patient's history is taken which leads the suspicion of a heart attack, and a test is order to help determined if this is actually true. The lab returns the value of the test result (T+ or T-) and, based on this information, the clinician would like to determine whether or not the patient has, in fact, had a heart attack. In other words, the clinician wants to know Pr(H+|T+) if a positive test result is reported, not Pr(T+|H+). Moreover, the clinician has just one test result for the patient, not many. The clinician's inferences must be based on small samples (one test result in this example), and should not rely on characterizations based on hypothetical outcomes across multiple, imaginary test results.

Bayes theorem is used to move from Pr(T|H) to Pr(H|T). Suppose the test result is positive (T+). By the rules of conditional probability we have:

$$\Pr(H+\,|\,T+) = \frac{\Pr(H+,T+)}{\Pr(T+)} = \frac{\Pr(T+\,|\,H+)\times\Pr(H+)}{\Pr(T+)} \tag{5}$$

and

$$\Pr(H-\,|\,T+) = \frac{\Pr(H-,T+)}{\Pr(T+)} = \frac{\Pr(T+\,|\,H-)\times\Pr(H-)}{\Pr(T+)} \tag{6}$$

or, taking ratios:

$$\begin{aligned}&\frac{\Pr(H+\,|\,T+)}{\Pr(H-\,|\,T+)} = \frac{\Pr(T+\,|\,H+)}{\Pr(T+\,|\,H-)}\times\frac{\Pr(H+)}{\Pr(H-)}\\ &\textit{posterior odds} = \textit{likelihood ratio}\times\textit{prior odds}\end{aligned} \tag{7}$$

The expression on the left side of the equal sign is the posterior odds of a heart attack given a positive test result, the first factor to the right of the equal sign is the likelihood ratio, and the second factor on the right is the prior odds. Bayes theorem is used to move from the likelihood, which conditions on presence of the heart attack, to a statistics that is directly relevant to the clinician and allows her to update their prior belief about the presence of the heart attack. The numerator of the likelihood ratio is the sensitivity of the laboratory test, and the denominator is equal to one minus the specificity, statistics that are readily available from laboratory studies. For a test with sensitivity of $Pr(T+|H+) = 0.80$ and specificity $Pr(T-|H-) = 0.70$, the likelihood ratio is $0.80/(1-0.70) = 2.67$, indicating that the odds of the patient having a heart attack is 2.67 times more likely given a positive test result relative to the clinician's prior odds. Thus, Bayes theorem takes a large-sample concept like sensitivity and specificity and transforms it into a statistic so that inference can be made about a single patient. In addition, it combines these measures with prior beliefs expressed in the form of probabilities.

Bayes theorem, like any theorem in probability, is simply a device for keeping track of uncertainty. It does this by the laws of conditional probability. It provides a means of moving from probability statements about the outcome of events assuming we know how the world works, to statements about how we think the world might work based on what we observed in the data. It conditions on the observed data, and yields exact finite-sample inference that is not based on asymptotic, hypothetical outcomes that haven't been observed by the researcher. Bayesian analysis treats all unobservables the same, whether they are parameters, hypothesized relationships or confidence intervals – all are derived from the same theorem based on the concept of conditional probability.

Despite its elegance, the reason that Bayesian methods have not caught on until recently is due to the complexity of the calculations involved with implementing Bayes theorem in all but the simplest of models. For example, the model described by equations (1) and (2) involves continuous error terms and cutoff values that were difficult to quantify, until recently.

The HB Revolution

Quantitative analysis in marketing makes use of models with parameters, and these parameters are the object of analysis in hierarchical Bayes models, not just the presence or absence of an effect. Probability distributions are used to quantify prior beliefs about the parameters (e.g., the price coefficient, β_1, in equation (4)), which is updated with the information from the data to yield a posterior distribution. Bayes theorem is expressed as:

$$Posterior \propto Likelihood \times Prior \tag{8}$$

where the proportionality sign, "$\propto$", replaces the equal sign, "=", in equation (7) because the proportionality constant, Pr(T+), does not cancel out.

Prior to the computational breakthrough known as Markov chain Monte Carlo (MCMC), the implementation of Bayes theorem involved multiplying probability densities for the prior by the probability expression for the likelihood to arrive at a posterior distribution of the parameter. To illustrate the complexity involved, consider a simple regression model:

$$y_t = \beta x_t + \varepsilon_t \quad ; \qquad \varepsilon_t \sim Normal(0, \sigma^2) \tag{9}$$

where prior distributions for the regression coefficients are typically assumed to be distributed according to a normal distribution:

$$\pi(\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2}(\beta - \bar{\beta})^2\right] \tag{10}$$

and, similarly, a prior distribution is assumed for the variance term, $\pi(\sigma^2)$. Equation (10) is the formula for a bell-shaped curve centered at $\bar{\beta}_1$ and standard deviation equal to σ. This prior distribution was combined with the likelihood, which reflected the information contained in the data about the parameter:

$$\pi(y_t \mid \beta, \sigma^2, x_t) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2} (y_t - \beta x_t)^2 \right] \tag{11}$$

The posterior distribution of the parameter β is obtained by multiplying the prior distribution (equation 10) by the likelihood (equation 11) and viewing the resulting product to be a function of the unknown parameter:

$$\begin{aligned} &\pi(\beta, \sigma^2 \mid y_t, x_t) \propto \pi(y_t \mid \beta, \sigma^2, x_t) \times \pi(\beta) \times \pi(\sigma^2) \\ &\textit{posterior} \propto \textit{likelihood} \times \textit{priors} \end{aligned} \tag{12}$$

This explains why Bayes theorem, while conceptually elegant, was been slow to develop in marketing and other applied disciplines – the analytic calculations involved were too difficult to perform in all but the simplest of problems.

The emergence of Markov chain Monte Carlo (MCMC) methods has eliminated this analytic bottleneck. Rather than deriving the analytic form of the posterior distribution, MCMC methods substitute a set of repetitive calculations that, in effect, simulate draws from this distribution. These Monte Carlo draws are then used to calculate statistics of interest such as parameter estimates and confidence intervals. The idea behind the MCMC engine that drives the

HB revolution is to set up a Markov chain that generates draws from posterior distribution of the model parameters. The Markov chain for the model described by equations (10) and (11) is:

1. Draw β given the data $\{y_t,x_t\}$ and the most recent draw of σ^2
2. Draw σ^2 given the data $\{y_t,x_t\}$ and the most recent draw of β
3. Repeat

and the Markov chain for the model described by equations (3) and (4) is:

1. Draw z_t given the data $\{y_t,x_t\}$ and most recent draws of other model parameters
2. Draw β_0 given given z_t and most recent draws of other model parameters
3. Draw β_1 given given z_t and most recent draws of other model parameters
4. ….
5. Repeat

While idea behind MCMC methods is simple, its implementation requires the derivations of the appropriate (conditional) distributions for generating the draws so that the Markov chain converges to the posterior distribution. These distributions are derived using Bayes theorem, in a manner similar to the description above for equations (10) and (11). Fortunately, many tools exist to assist the researcher in generating the draws from more complicated models. As a result, the approach has wide application.

Illustration

An advantage of estimating hierarchical Bayes (HB) models with Markov chain Monte Carlo (MCMC) methods is that it yields estimates of all model parameters, including estimates of model parameters associated with specific respondents. In addition, the use of MCMC methods facilitate the study of functions of model parameters that are closely related to decisions faced by management. The freedom afforded by MCMC to explore the parameter space was one of the first properties exploited in the marketing literature (Allenby & Ginter, 1995). In this study, HB+MCMC is used to explore extremes (i.e., tails) of multivariate distributions, providing insight into actions that can be taken to profitably grow a firm's base of customers.

Background

Organizations excel when they understand and respond to their customers more effectively than their rivals. To succeed in a competitive environment, organizations need to identify which customers are *most* likely to buy new products and services, and which customers are *most* likely to buy (or switch) due to changes in pricing, distribution and advertising strategies. Organizations can design products and programs which are *most* likely to elicit an immediate, strong market response and direct them to the individuals who are *most* likely to respond favorably.

This process involves understanding *extremes*. Consider, for example, the task of identifying an optimal product offering and assessing its potential demand. This task is often based on the results of a conjoint study which determines the value of the product attributes. Products are introduced into the market if a profitable level of demand exists, where a large portion of this demand comes from customers who currently use an existing product. The customers who are most likely to switch to the new offering are those who *most* value its unique attribute level or combination of attribute levels (relative to existing products), or those who *least* value the properties of their current product. In other words, the switchers are those individuals who are most *extreme* in their preferences for the product attribute levels.

Model

In this illustration we use a hierarchical Bayes random-effects logit model that pools the data and retains the ability to study the preferences and characteristics of specific individuals. The model is developed within the context of a choice setting where respondents are asked to select their most preferred option, where choice probabilities (Pr) are related to attribute-levels, x, using the expression:

$$\Pr(i)_h = \frac{\exp[x_i'\beta_h]}{\sum_j \exp[x_j'\beta_h]} \tag{13}$$

where h indexes the respondents, i and j index the choice alternatives, x is a vector of attribute-levels that describe the choice alternative, β_h is a vector of regression coefficients that indicate the part-worths of the attribute-levels, and $\Pr(i)_h$ is the probability that respondent h selects the i^{th} choice alternative. The logit model maps a continuous variable ($x_i'\beta$) onto the (0,1) interval that correspond to a choice probability.

Heterogeneity is incorporated into the model with a random-effects distribution whose mean is a function of observable covariates (z), including an intercept term:

$$\beta_h = \Gamma z_h + \xi_h \qquad \xi_h \sim MVN(0, V_\beta) \tag{14}$$

where Γ is a matrix of regression coefficients, which affects the location of the distribution of heterogeneity given z_h. Γ is therefore useful for identifying respondents that, on average, have part-worths that are different from the rest of the sample.

The covariance matrix V_β characterizes the extent of unobserved heterogeneity. Large diagonal elements of V_β indicate substantial unexplained heterogeneity in part-worths, while small elements indicate that the heterogeneity is accurately captured by Γz_h. Off-diagonal elements of V_β

indicate patterns in the evaluation of attribute-levels. For example, positive covariations indicate pairs of attribute-levels which tend to be evaluated similarly across respondents. Product offerings composed of these attribute-levels will be more strongly preferred by certain individuals (i.e. more extreme preferences will exist).

Equations (13) states that choice probabilities are determined by attributes of the offering (x) and part-worths of the respondent (β_h). Equation (14) then links a respondent's part-worths to attributes of the respondent (z_h) and coefficients that describe the population of respondents (Γ and V_β). Thus, inferences about a specific respondent's part-worths is a function of that respondent's data and the distribution of part-worths in the sample. The model can be written in hierarchical form as follows:

$$y \mid x, \beta \tag{15}$$

$$\beta \mid z, \Gamma, V_\beta \tag{16}$$

$$\Gamma \mid a, A \tag{17}$$

$$V_\beta \mid w, W \tag{18}$$

where equations (17) and (18) are prior distributions of the hyper-parameters, and the analyst provides values for (a,A) and (w,W). The Markov chain for the model described by equations (15) through (18) is:

1. Generate draws of β_h (one respondent at a time) given $\{y_{i,h},x_i,x_j\}$ and other model parameters. Repeat for all respondents.
2. Generate a draw of Γ given the set of respondent-level parameters $\{\beta_h\}$ and V_β.
3. Generate a draw of V_β given $\{\beta_h\}$ and Γ
4. Repeat

The conditional independence property of hierarchical model simplifies the distributions in steps 2 and 3 in the above recursion. Given the individual-level coefficients $\{\beta_h\}$, the conditional distributions of Γ and V_β do not depend on the actual choices. All information from the data relevant to these parameters comes through the individual-level coefficients, $\{\beta_h\}$. As a result, the form of the distribution used to generate the draws in steps 2 and 3 are relatively simple.

The Markov chain generates draws of all model parameters $\{\beta_h\}$, Γ, and V_β, the number of which can be large. It is not uncommon in conjoint analysis for there to be hundreds of respondents (h), each with part-worth vectors (β_h) of dimension in the tens. Models with thousands of parameters can easily be estimated with hierarchical Bayes models. The estimation of models of such high dimension was unthinkable a short time ago, and as a result these methods constitute a breakthrough in statistical science.

Data

Data were obtained by a regional bank wishing to offer credit cards to customers outside of its normal operating region, labeled as "out-of-state" hereafter. As part of a larger study of assessing the needs and feasibility of making such an offer, a conjoint study was conducted over the telephone with 946 current customers who provided demographic information. The bank and the attribute levels are disguised in this case to protect the proprietary interests of the cooperating firm. Specific numeric attribute levels were used in the actual study. Credit card attributes and attribute levels are presented in Table 20.1.

Table 20.1

Description of the Data

Sample Size	946 Respondents
	14,799 Observations
Attributes and Attribute-Levels	
1. Interest Rate	High, Medium, Low fixed Medium variable
2. Rewards	The reward programs consisted of annual fee waivers or interest rebate reductions for specific levels of card usage and/or checking account balance. Four reward programs were considered.
3. Annual Fee	High, Medium, Low
4. Bank	Bank A, Bank B, Out-of-State Bank
5. Rebate	Low, Medium, High
6. Credit line	Low, High
7. Grace period	Short, Long
Demographic Variables	Age (years) Annual Income ($000) Gender (=1 if female, =0 if male)

Preferences were obtained from a tradeoff study in which respondents were asked to choose between two credit cards that were identical in every respect except for two attributes. For example, respondents were asked to state their preference between the following offerings:

> The first card has a medium fixed annual interest rate and a medium annual fee, and
>
> The second card has a high fixed annual interest rate and low annual fee.

Each respondent provided responses to between 13 to 17 paired-comparisons involving a fraction of the attributes. A respondent trading-off interest rates and annual fees, for example, did not choose between rebates and credit lines. As a result it was not possible to obtain fixed-effect estimates of the entire vector of part-worths for any specific respondent. Moreover, even if all attribute levels were included for each respondent, constraints on the length of the questionnaire preclude collecting a sufficient number of trade-offs for accurate estimation of individual respondent part-worths. As noted above, this data limitation is less important in random-effect models which pool information across respondents. In all, a total of 14,799 paired-comparisons were available for analysis.

Results

Age, income and gender are mean-centered in the analysis so that the intercept of Γ can be interpreted as the average part-worth for the survey respondents. Figure 20.1 displays the series of draws of these elements of Γ. The Markov chain was run a total of 20,000 iterations, and plotted is every 20th draw of the chain. The figure indicates that chain converged after about 6000 iterations. Unlike traditional methods of estimation, the draws from the Markov chain converge in distribution to the true posterior distribution, not to point estimates. Convergence is determined by draws having the same mean value and variability over iterations.

Figure 1. Average Respondent Part-Worths

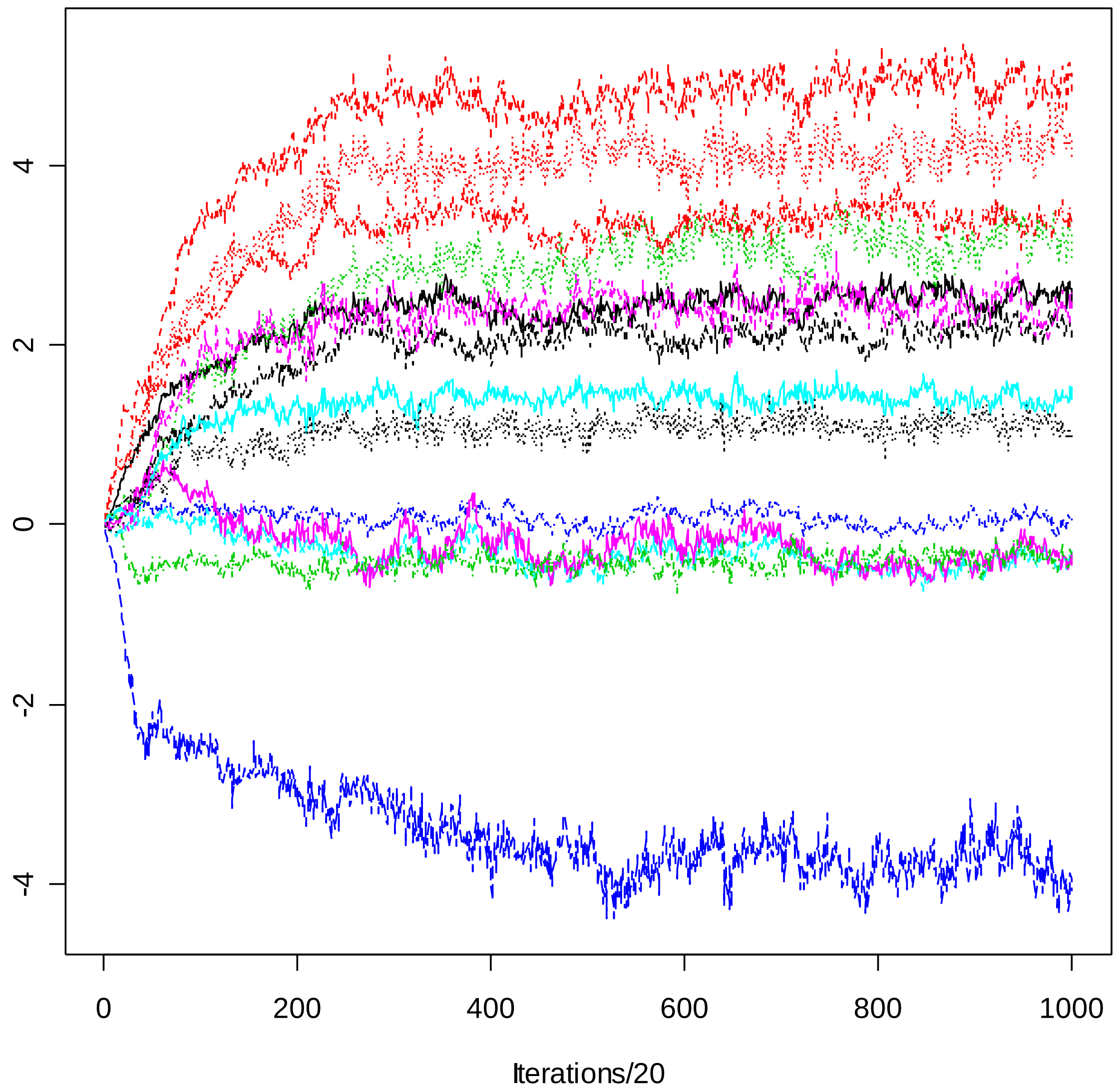


Table 20.2 reports posterior means for the elements of Γ. The column labeled "Intercept" corresponds to the coefficients displayed in Figure 20.1. The intercept estimates indicate that, on average, respondent penalize out-of-state banks by 3.758 utiles. This penalty can be overcome by offering low fixed interest rates (relative to high fixed interest), low annual fees, and long grace periods. In addition, older respondents assign less importance to changes in interest rates and other product attributes, while richer respondents are more likely to respond to the same incentives. For example, consider the part-worth estimates for the attribute level "low fixed interest rate". The coefficient for age is -0.025, indicating that an additional 50 years of age is associated with a reduction of 1.25 in the estimated part-worth. Viewing this as an adjustment to the fixed effect coefficient, we see that the part-worth is nearer zero for older individuals. Similarly, the estimated income coefficient for "high rebate" is 0.021, implying that an additional $50,000 in annual income is associated with an increase of 1.05 in the estimated part-worth. The results indicate that younger, high income respondents assign less penalty to out-of-state banks. In addition, the coefficients for gender indicate that females are particularly responsive to lower annual fees.

Table 20.2

Posterior Mean of Γ

Attribute-Levels	Intercept	Age	Income	Gender
Medium Fixed Interest	2.513	-0.013	0.011	0.106
Low Fixed Interest	4.883	-0.025	0.021	0.324
Medium Variable Interest	3.122	0.002	0.025	-0.354
Reward Program 2	0.061	0.005	0.001	-0.248

Reward Program 3	-0.391	0.022	0.014	-0.224
Reward Program 4	-0.297	0.019	0.016	-0.243
Medium Annual Fee	2.142	-0.004	0.002	0.668
Low Annual Fee	4.158	-0.010	0.004	1.302
Bank B	-0.397	0.001	0.003	0.124
Out-of-State Bank	-3.758	-0.003	0.013	-0.054
Medium Rebate	1.426	-0.008	0.003	0.231
High Rebate	2.456	-0.014	0.021	0.379
High Credit Line	1.116	-0.010	-0.003	0.368
Long Grace Period	3.399	-0.020	0.019	0.296

Table 20.3 reports the covariance matrix (V_β) that characterizes the unexplained variability of part-worths across respondents. The diagonal elements of the matrix are large and indicate substantial unexplained heterogeneity in response. Off-diagonal elements of the covariance matrix indicate attribute-levels that tend to be evaluated similarly across respondents. Most of these estimates are large and significantly different from zero. The covariation between out-of-state bank and low annual fee, for example, is equal to 8.1. This translates to a correlation coefficient of 0.55. This positive covariance implies that respondents who prefer a low annual fee are those who are less sensitive to whether the bank is out-of-state. In making this interpretation, recall that the out-of-state attribute-level estimate is negative (Table 20.4), and a more positive evaluation implies it is closer to zero. Therefore, offering a credit card with low annual fee may be a particularly effective method of inducing usage by out-of-state customers, if this group is large enough. This issue is explored in more detail in the next section.

Table 20.3

Posterior Mean of V_β

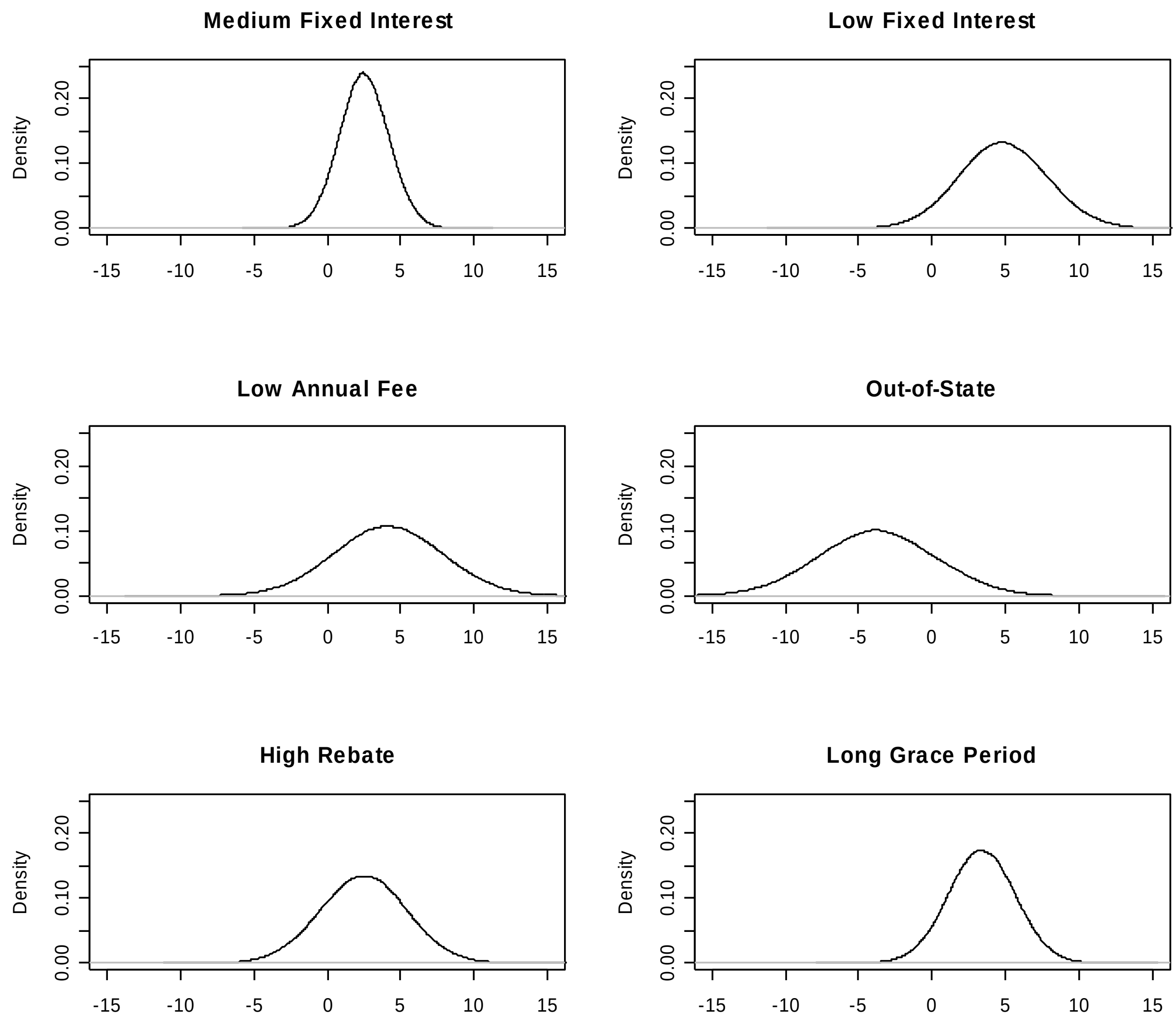
Attribute-Levels														
Medium Fixed Interest	2.8	4.5	4.3	0.0	0.2	0.5	1.1	2.0	0.3	1.2	0.7	1.4	0.9	1.4
Low Fixed Interest	4.5	8.8	8.2	0.0	0.4	0.9	1.8	3.4	0.5	2.1	1.2	2.5	1.2	2.4
Medium Variable Interest	4.3	8.2	10.2	-0.1	0.5	1.2	1.9	3.7	0.1	1.5	1.5	2.9	2.1	3.0
Reward Program 2	0.0	0.0	-0.1	1.2	0.2	-0.2	-0.3	-0.5	0.0	0.5	-0.3	-0.5	-0.4	-0.5
Reward Program 3	0.2	0.4	0.5	0.2	2.0	1.8	-0.2	-0.7	0.5	1.4	-0.4	-0.9	0.4	-1.0
Reward Program 4	0.5	0.9	1.2	-0.2	1.8	3.0	-0.3	-0.8	0.2	0.3	-0.6	-1.2	0.3	-1.1
Medium Annual Fee	1.1	1.8	1.9	-0.3	-0.2	-0.3	4.1	6.9	1.4	4.4	1.0	2.2	2.3	2.6
Low Annual Fee	2.0	3.4	3.7	-0.5	-0.7	-0.8	6.9	13.5	2.7	8.1	2.1	4.7	4.3	5.1
Bank B	0.3	0.5	0.1	0.0	0.5	0.2	1.4	2.7	3.5	5.6	1.2	2.5	2.0	0.9
Out-of-State Bank	1.2	2.1	1.5	0.5	1.4	0.3	4.4	8.1	5.6	15.9	2.3	4.8	3.9	2.0
Medium Rebate	0.7	1.2	1.5	-0.3	-0.4	-0.6	1.0	2.1	1.2	2.3	2.4	4.0	2.0	2.1
High Rebate	1.4	2.5	2.9	-0.5	-0.9	-1.2	2.2	4.7	2.5	4.8	4.0	8.5	3.6	4.5
High Credit Line	0.9	1.2	2.1	-0.4	0.4	0.3	2.3	4.3	2.0	3.9	2.0	3.6	6.3	2.6
Long Grace Period	1.4	2.4	3.0	-0.5	-1.0	-1.1	2.6	5.1	0.9	2.0	2.1	4.5	2.6	5.0

Distribution of Heterogeneity

The value of the hierarchical Bayes model lies in its ability to characterize heterogeneity in preferences while retaining its ability to study specific individuals. Figure 20.2 displays the distribution of heterogeneity for selected part-worths for all respondents. In contrast, Figure 20.3 displays the same distributions for one randomly selected respondent – respondent #250. Both figures were constructed from the individual-level part-worth draws, $\{\beta_h\}$, with Figure 20.2 using the draws over all respondents and Figure 20.3 using only those draws from one respondent.

The distribution of heterogeneity reported in Figure 20.2 illustrates that respondents are diverse in their preferences and sensitivities to features of credit cards. The mass of the distributions are located in regions with expected algebraic signs – e.g., a long grace period is expected to be preferred to a short grace period, and the distribution of part-worths for the long grace period is mostly positive. In addition, the spread of the distribution is large for the logit model (equation 14) where coefficient values greater than five are associated with large changes in the choice probabilities. The mean of the distributions of heterogeneity are reported in the column labeled "Intercept" in Table 20.2, and, as seen in Figure 20.1, the estimated mean of the distribution is precisely measured and different from zero. Heterogeneity of preferences is therefore an important aspect of the model, and ignoring its presence can lead to incorrect inferences.

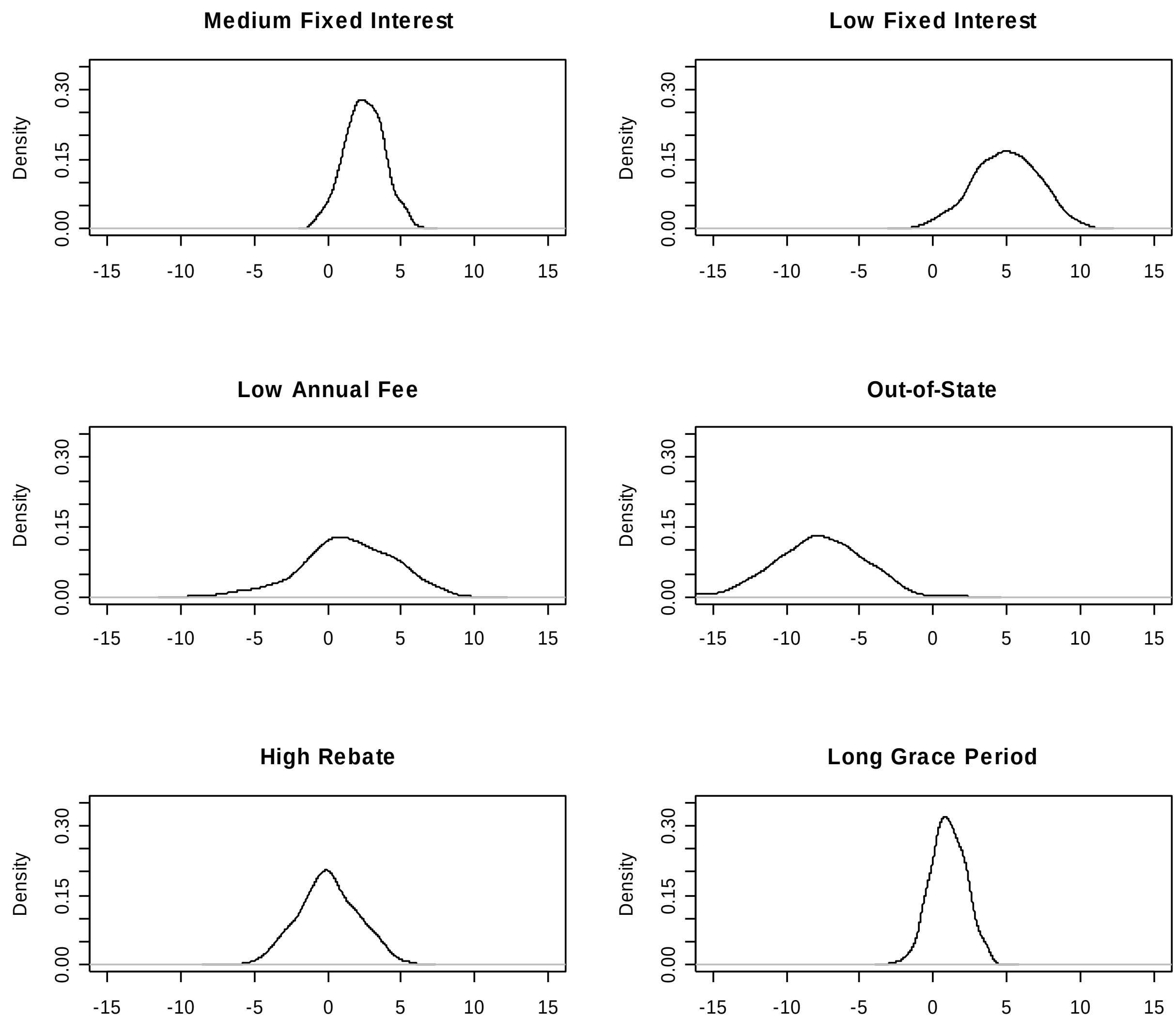
Figure 2. Distribution of Heterogeneity for Selected Part-Worths



The distributions displayed in Figure 20.3 indicate that there exists substantial uncertainty in drawing inferences about the part-worths of a specific respondent. An advantage of Bayes theorem is that it yields exact finite-sample estimates of the posterior distribution of respondent-level parameters. Some of the distributions displayed in Figure 20.3 have mass centered away from zero (e.g., medium fixed interest, low fixed interest, out-of-state, long grace period) and some of the distributions have mass centered near zero (e.g., low annual fee and high rebate). These distributions can be used to form confidence intervals for individual-level part-worth estimates, to generate point-estimates (e.g., the mean of the posterior distribution), and can be used in conjunctions with choice simulators to explore aspects of product reformulation. It conducting such analysis it is important to keep in mind that the individual-level part-worth estimates are not precisely estimated, and the use of point-estimates without acknowledging the amount of uncertainty will lead to over-confidence predictions of effect-sizes. This over-confidence can be avoided by using all the draws of $\{\beta_{250}\}$ to make marketplace predictions, not just point estimates.

The beauty of modern Bayesian analysis is that the construction of distribution means and confidence intervals, and characterizing the uncertainty associated with marketplace predictions, is easy. For example, the mean of the out-of-state distribution displayed in Figure 20.3 can be estimated by retaining the draws of β_{250} corresponding to the out-of-state part-worth, and simply computing the sample average for those draws from the portion of the converged Markov chain (e.g., iterations 10,000 through 20,000). Similarly, an estimate of the standard deviation of any posterior distribution is obtained by computing the sample standard deviation of the appropriate draws.

Figure 3. Part-Worth Distributions for Respondent 250



Focusing on Extremes

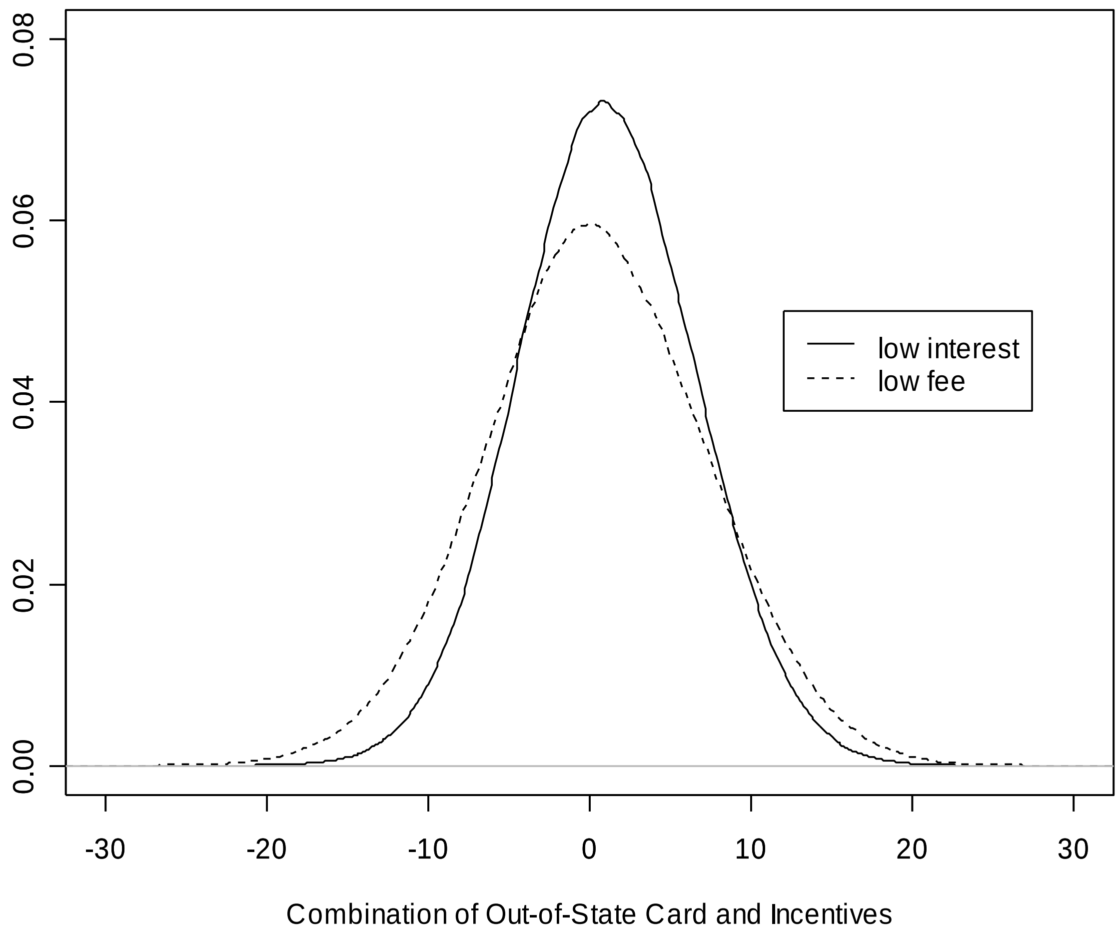
The objective of the conjoint study is to identify incentives that could be offered so that a regional bank can successfully offer credit cards to out-of-state customers. The fixed-effect estimates in Table 20.4 indicate that on average, respondents penalize out-of-state credit cards 3.758 utiles relative to their current bank, all other attributes assumed equal. In the following discussion we consider two different incentives to overcome this penalty: low fixed interest (versus high fixed interest) and low annual fee (versus high annual fee). The coefficient estimates in Table 20.2 indicate that on average each of these incentives is sufficient to overcome the out-of-state penalty. However we show below that the analysis of heterogeneity identifies important managerial differences in these incentives.

The covariance matrix of unobserved heterogeneity in Table 20.3 indicates that respondents who prefer low annual fee are less sensitive to the out-of-state status of the bank (i.e., covariance is equal to 8.1). This is because of the strong positive covariance of respondent evaluations for the low annual fee and out-of-state bank attribute-levels. Respondents who are more sensitive to the annual fee tend to have a higher (closer to zero) part-worth for the out-of-state attribute-level. In contrast, the covariance between low fixed interest and out-of-state bank is only 2.1 (correlation equal to 0.18), indicating that respondents who prefer low fixed interest have about average sensitivity to the bank's location. Since the response potential in Equation (13) is monotonically related to a respondent's utility for the attribute-levels, it is appropriate to compare the distributions of net utility arising from these incentives to overcome the out-of-state status of the bank.

The distribution of net utility for a particular product concept is a linear transformation of this distribution. When the distributions of net utility are summed across respondents, the resulting aggregate distribution summarizes the information available at the market level about preferences for the product. Figure 20.2 displays the aggregate distributions of respondent net utility for an out-of-state bank with low interest or with low annual fee. All other attributes are assumed to be equal to the first attribute-level reported in Table 20.1. The figure is constructed by adding together the elements of $\{\beta_h\}$ corresponding to the net utility (e.g., $\beta_{\text{low fee}} + \beta_{\text{out-of-state}}$) for each respondent and each iteration, and then constructing a distribution from these draws. The means for low interest (the solid line) is larger than the mean for low annual fee (the dashed line), consistent with the intercepts reported in Table 20.2. However, the large positive covariation between low annual fee and out-of-state bank results in a wider distribution.

For purposes of illustration, suppose that the target market is defined as those individuals with net utility greater than 10.0. Approximately 4.5% of the low interest distribution is above this value, in contrast to 7.5% for the low fee distribution. Even though the average utility for low interest is 0.7 units larger, the low fee distribution has almost twice the mass in the region of the distribution corresponding to respondents with strong preference. These individuals are more likely to represent profitable targets as indicated by Equation (13).

Figure 4. Distribution of Net Preference



Challenges to Using HB Models

Hierarchical Bayes models free researchers from computational constraints that allows researchers and practitioners to develop more realistic models of buyer behavior and decision making. Moreover, this freedom enables exploration of marketing problems that have proven elusive over the years, such as models for advertising ROI, sales force effectiveness, and similarly complex problems that often involve simultaneity. The promise of Bayesian statistical methods lies in the ability to deal with these complex problems, but the very complexity of the problems creates a significant challenge to both researchers and practitioners.

In most cases, there exists no off-the shelf programs for estimating complex HB models. The WinBugs software is one exception. Practitioners, however, have reported problems in applying WinBugs to datasets of the size encountered in commercial applications. Sawtooth Software has created two programs for estimating HB models, one for choice-based conjoint and one for OLS regression. However, these programs are limited to modeling relatively simple problems. In fact, these programs are really HB extensions to "standard" choice-based conjoint and regression models. While the availability of these programs from Sawtooth Software has been a major impetus behind the adoption of HB estimation for choice-based conjoint, the ability to estimate more complex models is limited by the need to customize programs for each new model.

The forthcoming book, Bayesian Statistics and Marketing, by Rossi, Allenby and McCulloch (2005) provides a comprehensive introduction to HB models from a theoretical and a practical perspective. The book contains software written in the R computing language for estimating most of the models discussed in this chapter. Moreover, the programs are written so that users can modify the software to adapt to their needs, and tutorials are provided to help the user understand how to write code in this environment.

Researchers must also adjust to some of the differences they will encounter in using HB methods. As noted earlier, HB estimation methods do not "converge" on a closed-form solution in the way that many of our classical estimation methods, such as multinomial logit, do. Practitioners will need to become comfortable with the fact that once the variance in the estimation stabilizes after several thousand "burn-in" iterations, there will still be considerable variation in the average parameter estimates. Another important difference with HB methods is that, instead of a point estimate of values for each respondent, we end up with a distribution of estimates for each respondent. While this is powerful in terms of understanding uncertainty, it adds to the complexity of the analysis, particularly in the case of market simulation.

In closing, the future of Bayesian methods in marketing is promising. Bayesian models give researchers the freedom to study the complexities of human behavior in a more realistic fashion than was previously possible. Human behavior is extremely complex. Unfortunately, many of the models and variables used in our analysis are not. Consider, for example, the linear model in equation (2). While this model has been the workhorse of much statistical analysis, particularly in conjoint analysis, it does not provide a true representation of how respondents encode, judge, and report on items in a questionnaire. The future holds many opportunities to look behind the responses in survey data, and other marketing data, to gain better insight into how individuals will act in the marketplace.

REFERENCES

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research, 35*, 384-389.

-----, & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research, 32*, 392-403.

-----, & Lenk. P. J. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of American Statistical Association, 89*, 1218-1231.

-----, & ----- (1995). Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice. *Journal of Business & Economic Statistics, 13*, 281-289.

-----, Leone, R. P., & Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, *94*, 365-374.

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, *23*, 391-406.

Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, *21*, 223-228.

McCulloch, R. E., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*, 217-228

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian statistics and marketing*. New York: Wiley (forthcoming).

-----, Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association, 96*, 20-31.

Yang, S., & Allenby. G. M. (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research*, *40*, 282-294.

-----, -----, & Fennell, G. (2002). Modeling variation in brand preference: The roles of objective environment and motivating conditions. *Marketing Science*, *21*, 14-31.

Annotated Bibliography of Bayesian Applications in Marketing

<u>Books</u>

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. New York: Wiley (forthcoming).

A comprehensive summary of the theory and application of Bayesian methods in marketing. Includes software in the R statistical language to implement commonly used models, and some of the more advanced models recently published in the marketing literature.

<u>Journal Articles</u>

Ainslie, A., & Rossi, P. (1998). Similarities in choice behavior across product categories. *Marketing Science, 17*, 91-106.

A multi-category choice model is proposed where household response coefficients are assumed dependent across category. The estimated distribution of heterogeneity reveals that price, display and feature sensitivity are not uniquely determined for each category, but may be related to household-specific factors.

Allenby, G. M., Shiveley, T., Yang, S. & Garratt, M. J. (2004). A choice model for packaged goods: Dealing with discrete quantities and quantity discounts. *Marketing Science*, *23*, 95-108.

A method for dealing with the pricing of a product with different package sizes is developed from utility-maximizing principles. The model allows for the estimation of demand when there exist a multitude of size-brand combinations.

Allenby, G. M., Leone, R. P., & Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, *94*, 365-374.

Customer inter-purchase times modeled with a heterogeneous generalized gamma distribution, where the distribution of heterogeneity is a finite mixture of inverse generalized gamma components. The model allows for structural heterogeneity where customers can become inactive.

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research, 35*, 384-389.

A normal component mixture model is compared to a finite mixture model using conjoint data and scanner panel data. The predictive results provide evidence that the distribution of heterogeneity is continuous, not discrete.

Allenby, G. M., Jen, L., & Leone, R. P. (1996). Economic trends and being trendy: The influence of consumer confidence on retail fashion sales. *Journal of Business & Economic Statistics, 14*, 103-111.

A regression model with autoregressive errors is used to estimate the influence of consumer confidence on retail sales. Data are pooled across divisions of a fashion retailer to estimate a model where influence has a differential impact on pre-season versus in-season sales.

Allenby, G. M., & Lenk, P. J. (1995). Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice. *Journal of Business & Economic Statistics, 13*, 281-289.

The logistic normal regression model of Allenby and Lenk (1994) is used to explore the order of the brand-choice process, and to estimate the magnitude of price, display and feature advertising effects across four scanner panel datasets. The evidence indicates that brand-choice is not zero order, and merchandising effects are much larger than previously thought.

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

A heterogeneous random-effects binary choice model is used to estimate conjoint part-worths using data from a telephone survey. The individual-level coefficients available in hierarchical Bayes models are used to explore extremes of the heterogeneity distribution, where respondents are most and least likely to respond to product offers.

Allenby, G. M., Arora, N., & Ginter, J. L. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research, 32*, 152-162.

Oridinal prior information is incorporated into a conjoint analysis using a rejection sampling algorithm. The resulting part-worth estimates have sensible algebraic signs that are needed for deriving optimal product configurations.

Allenby, G. M., & Lenk, P. J. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of American Statistical Association, 89*, 1218-1231.

A discrete choice model with autocorrelated errors and consumer heterogeneity is developed and applied to scanner panel dataset of ketchup purchases. The results indicate substantial unobserved heterogeneity and autocorrelation in purchase behavior.

Allenby, G. M. (1990). Hypothesis testing with scanner data: The advantage of Bayesian methods. *Journal of Marketing Research*, *27*, 379-389.

Bayesian testing for linear restrictions in a multivariate regression model is developed and compared to classical methods.

Allenby, G. M. (1990). Cross-validation, the Bayes theorem, and small-sample bias. *Journal of Business & Economic Statistics, 8*, 171-178.

Cross-validation methods that employ plug-in point approximations to the average likelihood are compared to formal Bayesian methods. The plug-in approximation is shown to over-state the amount of statistical evidence.

Andrews, R., Ansari, A., & Currim, I. (2002). Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *Journal of Marketing Research*, 39, 87-98

A simulation study is used to investigate the performance of continuous and discrete distributions of heterogeneity in a regression model. The results indicate that Bayesian methods are robust to the true underlying distribution of heterogeneity, and finite mixture models of heterogeneity perform well in recovering true parameter estimates.

Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research, 37*, 363-375.

Random-effect specifications for respondents and stimuli are proposed within the same linear model specification. The model is used to pool information from multiple data sources.

Ansari, A., Jedidi, K., & Jagpal, S. (2000). A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. *Marketing Science, 19*, 328-347.

Covariance matrix heterogeneity is introduced into a structural equation model, in contrast to standard models in marketing where heterogeneity is introduced into the mean structure of a model. The biasing effects of not accounting for covariance heterogeneity is documented.

Arora, N., & Allenby, G. M. (1999). Measuring the influence of individual preference structures in group decision making. *Journal of Marketing Research, 36*, 476-487.

Group preferences differ from the preferences of individuals in the group. The influence of the group on the distribution of heterogeneity is examined using conjoint data on durable good purchases by a husband's, wife's and their joint evaluation.

Arora, N., Allenby, G. M., & Ginter, J. L. (1998). A hierarchical Bayes model of primary and secondary demand. *Marketing Science, 17*, 29-44.

An economic discrete/continuous demand specification is used to model volumetric conjoint data. The likelihood function is structural, reflecting constrained utility maximization.

Blattberg, R. C., & George, E. I. (1991). Shrinkage estimation of price and promotional elasticities: Seemingly unrelated equations. *Journal of the American Statistical Association, 86*, 304-315.

Weekly sales data across multiple retailer in a chain are modeled using a linear model with heterogeneity. Price and promotional elasticity estimates are shown to have improved predictive performance.

Boatwright, P., Dhar, S., & Rossi, P. E. (2004). The role of retail competition, demographics and account retail strategy as drivers of promotional sensitivity. *Quantitative Marketing and Economics*, *2*, 169-190.

Explanatory variables are entered as a regression for the mean of the distribution of heterogeneity. The variables are found to account for 30 percent of the variation in promotion sensitivity, with retail competition variables being least important.

Boatwright, P., Borle, S., & Kadane, J. B. (2003). A model of the joint distribution of purchase quantity and timing. *J. Amer. Statist. Assoc., 98* 564-572.

Purchase quantity and timing are modeled as the marginal distribution of purchase timing and the distribution of purchase quantity conditional on purchase timing in their paper. The authors propose a hierarchical Bayes model that disentangles the weekly and daily components of the purchase timing.

Boatwright, P., McCulloch, R., & Rossi, P. E. (1999). Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *Journal of the American Statistical Association* , *94*, 1063-1073.

A common problem in the analysis of sales data is that price coefficients are often estimated with algebraic signs that are incompatible with economic theory. Ordinal constraints are introduced through the prior to address this problem, leading to a truncated distribution of heterogeneity.

Bradlow, E. T., Hu, Y., & Ho, T.-H. (2004). A learning-based model for imputing missing levels in partial conjoint studies (with discussion). *Journal of Marketing Research*, *41*, 269-396.

Alternative models of data imputation are investigated to assess respondent inference about missing attribute in partial-profile conjoint analysis. The results indicate that inference is based on prior knowledge about the category and design aspects of the study.

Bradlow, E. T., & Fader, P. S. (2001). A Bayesian lifetime model for the "Hot 100" *Billboard* songs. *Journal of the American Statistical Association, 96*, 368-381.

A time series model for ranked data is developed using a latent variable model. The deterministic portion of the latent variable follows a temporal pattern described by a generalized gamma distribution, and the stochastic portion is extreme value.

Bradlow, E. T., & Schmittlein, D. (1999). The little engines that could: Modeling the performance of World Wide Web search engines. *Marketing Science, 19*, 43-62.

A proximity model is developed for analysis of the performance of Internet search engines. The likelihood function reflects the distance between the engine and specific URL's, with the mean location of the URL's parameterized with a linear model.

Bradlow, E. T., & Rao, V. R. (2000). A hierarchical Bayes model for assortment choice. *Journal of Marketing Research, 37*, 259-268.

A statistical measure of attribute assortment is incorporated into a random-utility model to measure consumer preference for assortment beyond the effects from the attribute levels themselves. The model is applied to choices between bundled offerings.

Bucklin, R. E., & Sismeiro, C. (2003). A model of Web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, *40*, 249-267.

Heterogeneity is introduced into a Tobit model of number of pages viewed and number of visits to a web site.

Chaing, J., Chib, S., & Narasimhan, C. (1999). Markov chain Monte Carol and models of consideration set and parameter heterogeneity. *Journal of Econometrics*, *89*, 223-248.

Consideration sets are enumerated and modeled with a Dirichlet prior in a model of choice. A latent state variable is introduced to indicate the consideration set, resulting in a model of structural heterogeneity.

Chang, K., Siddarth, S., & Weinberg, C. B. (1999). The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects. *Marketing Science, 18*, 178-192.

A random utility model with reference prices is examined with and without allowance for household heterogeneity. When heterogeneity is present in the model, the reference price coefficient is estimated to be close to zero.

Chib, S., Seetharaman, P. B., & Strijnev, A. (2004). Model of brand choice with a no-purchase option calibrated to scanner-panel data. *Journal of Marketing Research*, *41*, 184-196.

An outside good is introduced into a discrete choice model, and related to household inventory and other covariates. In-sample and predictive fits are shown to be superior to nested logit and general GEV models.

DeSarbo, W. S., Fong, D. K. H., Liechty, J., & Chang, J. (2004). Evolutionary preference/utility functions: A dynamic perspective. *Psychometrika*. Forthcoming.

The paper introduces a Bayesian dynamic linear methodology that permits the detection and modeling of potential changes to the underlying preference/utility structure of the respondent. An illustration of revealed/stated preference analysis is given involving students' preferences for apartments and their underlying attributes and features.

DeSarbo, W., Kim, Y., & Fong, D. (1999). A Bayesian multidimensional scaling procedure for the spatial analysis of revealed choice data. *Journal of Econometrics*, *89*, 79-108.

The deterministic portion of a latent variable model is specified as a scalar product of consumer and brand coordinates to yield a spatial representation of revealed choice data. The model provides a graphical representation of the market structure of product offerings.

Edwards, Y., & Allenby, G. M. (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research*, *40*, 321-334.

Pick any/J data is modeled with a multivariate probit model, allowing standard multivariate techniques to be applied to the parameter of the latent normal distribution.

Identifying restrictions for the model are imposed by post-processing the draws of the Markov chain.

Frühwirth-Schnatter, S., Tüchler, R., & Otter, T. (2004). Bayesian analysis of the heterogeneity model. *Journal of Business and Economic Statistics*, *22*, 1, 2-15

Identification of the normal component mixture model for the distribution of heterogeneity is explored within the context of a conjoint model. Improved methods of identification are proposed, the issue of label switching is shown to be irrelevant to many important questions in marketing, and the importance of model parameterization is demonstrated.

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, *23*, 391-406.

Indicator functions are introduced in a discrete choice model to screen alternatives. The model identifies attribute-levels required by offerings for detailed consideration by consumers.

Huber, J., & Train, K. (2001). On the similiarity of classical and Bayesian estimates of individual mean partworths. *Marketing Letters, 12*, 259-269.

Classical and Bayesian estimation methods are found to be yield similar individual-level estimates. The classical methods condition on estimated hyper-parameters, while Bayesian methods account for their uncertainty.

Jedidi, K., Jagpal, S., & Manchanda, P. (2003). Measuring heterogeneous reservation prices for product bundles. *Marketing Science*, *22*, 107-130.

A latent utility model is parameterized in terms of reservation prices and estimated using MCMC methods. The distribution of heterogeneity is specified to allow for general patterns of substitution and complementarity.

Jen, L., Chou, C.-H., & Allenby, G. M. (2003). A Bayesian approach to modeling purchase frequency. *Marketing Letters*, *14*, 5-20.

A model of purchase frequency that combines a Poisson likelihood with gamma mixing distribution is proposed, where the mixing distribution is a function of covariates. The covariates are shown to be useful for customers with short purchase histories or have infrequent interaction with the firm.

Kalyanam, K., & Shiveley, T. S. (1998). Estimating irregular pricing effects: A stochastic spline regression approach. *Journal of Marketing Research, 35*, 16-29.

Stochastic splines are used to model the relationship between price and sales, resulting in a more flexible specification of the likelihood function.

Kalyanam, K. (1996). Pricing decision under demand uncertainty: A Bayesian mixture model approach. *Marketing Science, 15*, 207-221.

Model uncertainty is captured in model predictions by taking a weighted average where the weights correspond to the posterior probability of the model. Pricing decisions are shown to be more robust.

Kamakura, W. A., & Wedel, M. (1997). Statistical data fusion for cross-tabulation. *Journal of Marketing Research, 34*, 485-498.

Imputation methods are proposed for analyzing cross-tabulated data with empty cells. Imputation is conducted in an iterative manner to explore the distribution of missing responses.

Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, *21*, 223-228.

A choice model with interior and corner solutions is derived from a utility function with decreasing marginal utility. Kuhn-Tucker conditions are used to relate the observed data with utility maximization in the likelihood specification.

Lee, J., Boatwright, P., & Kamakura, W. (2003). A Bayesian model for prelaunch sales forecasting of recorded music. *Management Science*, *49*, 179-196.

The authors study the forecasting of sales for new music albums prior to their introduction. A hierarchical logistic shaped diffusion model is used to combine a variety of sources of information on attributes of the album, effects of marketing variables and dynamics of adoption.

Leichty, J., Ramaswamy, V., & Cohen, S. H. (2001). Choice menus for mass customization. *Journal of Marketing Research, 38*, 183-196.

A multivariate probit model is used to model conjoint data where respondents can select multiple items from a menu. The observed binomial data is modeled with a latent multivariate normal distribution.

Liechty, J., Fong, D. K. H., & DeSarbo, W. S. (2004). Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Science*. Forthcoming.

Heterogeneity is incorporated into a conjoint model around an aggregate dynamic trend.

Lenk, P., & Rao, A. (1990). New models from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Science, 9*, 42-53.

The non-linear likelihood function of the Bass model is combined with a random-effects specification across new product introductions. The resulting distribution of heterogeneity is shown to improve early predictions of new product introductions.

Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M.(1996). Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science, 15*, 173-191.

Fractionated conjoint designs are used to assess ability of the distribution of heterogeneity to "bridge" conjoint analyses across respondents to impute part-worths for attributes not examined.

Manchanda, P., Rossi, P. E., & Chintagunta, P. (2004). Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*, *41*, 467-478.

A sales response model is developed where the level of marketing mix variable is partially determined by the expected gain of the allocation. The model is applied to pharmaceutical sales data where dependence between the level of detailing and sales effectiveness is assessed.

Manchanda, P., Ansari, A., & Gupta, S. (1999). The “shopping basket”: A model for multicategory purchase incidence decisions. *Marketing Science, 18*, 95-114.

Multi-category demand data are modeled with a multivariate probit model. Identifying restrictions in the latent error covariance matrix require use of a modified Metropolis-Hastings algorithm.

Marshall, P., & Bradlow, E. T. (2002). A unified approach to conjoint analysis models. *Journal of the American Statistical Association*, *97*, 674-682.

Various censoring mechanisms are proposed for relating observed interval, ordinal and nominal data to a latent linear conjoint model.

McCulloch, R. E., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*, 217-228

The multinomial probit model is estimated using data augmentation methods. Approaches to handling identifying model identification are discussed.

Moe, W., & Fader, P. (2002). Using advance purchase orders to track new product sales. *Marketing Science, 21*, 347-364.

A hierarchical model of product diffusion is developed to forecasting new product sales. The model features a mixture of Weibulls as the basic model with a distribution of heterogeneity over related products. The model is applied to data on music album sales.

Montgomery, A., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Predicting online purchase conversion using Web path analysis. *Marketing Science*. Forthcoming.

A dynamic Multinomial Probit model with auto-regressive lags and a hidden Markov switching structure is used to model individual navigation paths through web pages, where the pages have been grouped into different categories. The model demonstrates that incorporating path specific data in this manner results in improved predictive power.

Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science, 16*, 315-337.

Bayesian Hierarchical Models applied to store-level scanner data. The model specification involves store-level demographic variables. Profit opportunities for store-level pricing are explored using constraints on the change in average price.

Montgomery, A. L., & Bradlow, E. T. (1999). Why analyst overconfidence about the functional form of demand models can lead to overpricing. *Marketing Science, 18*, 569-583.

The specification of a function form involves imposing exact restrictions in an analysis. Stochastic restrictions are introduce via a more flexible model specification and prior distribution, resulting in less aggressive policy implications.

Montgomery, A. L., & Rossi, P. E. (1999). Estimating price elasticities with theory-based priors. *Journal of Marketing Research, 36*, 413-423.

The prior distribution is used to stochastically impose restrictions on price elasticity parameters that are consistent with economic theory. This proposed approach is compared to standard shrinkage estimators that employ the distribution of heterogeneity.

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science, 18*, 115-136.

Alternative information sets are explored for making new product forecasts in domestic and international markets using a Poisson model for attendance with log-normal heterogeneity.

Otter, T., Tüchler, R., & Frühwirth-Schnatter, S. (2004). Capturing consumer heterogeneity in metric conjoint analysis using Bayesian mixture models. *International Journal of Research in Marketing*, *21*, 285-297

Finite mixture models are compared to hierarchical Bayes models with continuous distributions of heterogeneity. Theoretical and practical implications for segment identification are explored.

Pammer, S. E., Fong, D. K. H., & Arnold, S. F. (2000). Forecasting the penetration of a new product--A Bayesian approach. *Journal of Business and Economic Statistics***,** *18*, 428-435.

Prior information are incorporated into the data analysis for predicting the penetration of a new product into a market. The penetration curve is assumed to be a nondecreasing function of time and may be under shape constraints. An example on forecasting the penetration of color TV using the information from black-and-white TV is provided in the paper.

Putler, D. S., Kalyanam, K., & Hodges, J. (1996). A Bayesian approach for estimating target market potential with limited geodemographic information. *Journal of Marketing Research, 33*, 134-149.

Prior information about correlation among variables is combined with data on the marginal distribution to yield a joint posterior distribution.

Rossi, P. E., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science, 22*, 304-328.

The paper reviews the essence of the Bayesian approach and explains why it is particularly useful for marketing problems.

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association, 96*, 20-31.

Consumer response data on a fixed point rating scale are assumed to be censored outcomes from a latent normal distribution. Variation in the censoring cutoffs among respondents allow for scale use heterogeneity.

Rossi, P. E., McColloch, R., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science, 15*, 321-340.

The information content of alternative data sources is evaluated using an economic loss function of coupon profitability. The value of a household's purchase history is shown to be large relative to demographic information and other information sets.

Rossi, P. E., & Allenby, G. M. (1993). A Bayesian approach to estimating household parameters. *Journal of the Marketing Research, 30*, 171-182.

Individual-level parameters are obtained with the use of an informative, but relatively diffuse prior distribution. Methods of assessing and specifying the amount of prior information are proposed.

Sandor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research, 28*, 430-444.

The information from an experiment involving discrete choice models depends on the experimental design and the values of the model parameters. Optimal designs are determined with an information measure that is dependent on the prior distribution.

Seetharaman, P. B., Ainslie, A., & Chintagunta, P. (1999). Investigating household state dependence effects across categories. *Journal of Marketing Research, 36*, 488-500.

Multiple scanner panel datasets are used to estimate a model of brand choice with state dependence. Individual-level estimates of state dependence effects are examined among categories.

Shively, T. A., Allenby, G. M., & Kohn, R. (2000). A nonparametric approach to identifying latent relationships in hierarchical models. *Marketing Science, 19*, 149-162.

Stochastic splines are used to explore the covariate specification in the distribution of heterogeneity. Evidence of highly non-linear relationships is provided.

Sismeiro, C., & Bucklin, R. (2004). Modeling purchase behavior at an e-commerce Web site: A task-completion approach. *Journal of Marketing Research*, *41*, 306-323.

A sequence of binary probit models are used to model the joint distribution of tasks (e.g., locating products, adding items to a shopping cart, etc.) as marginal and conditional distributions. Variables predictive of task completion are identified.

Steenburgh, T. J., Ainslie, A., & Engebretson, P. (2002). Massively categorical variables: Revealing the information in zipcodes. *Marketing Science*, *22*, 40-57.

The effects associated with massively categorical variables, such as zip codes, are modeled in a random-effects specification. Alternative loss functions are examined for assessing the value of the resulting shrinkage estimates.

Talukdar, D., Sudhir, K., & Ainslie, A. (2002). Investing new production diffusion across products and countries. *Marketing Science, 21*, 97-116.

The Bass diffusion model is coupled with a random effects specification for the coefficients of innovation, imitation and market potential. The random effects model includes macro-economic covariates that have large explanatory power relative to unobserved heterogeneity.

Ter Hofstede, F., Wedel, M., & Steenkamp, J. (2002). Identifying spatial segments in international markets. *Marketing Science, 21*, 160-177.

The distribution of heterogeneity in a linear regression model is specified as a conditional Guassian field to reflect spatial associations. The heterogeneity specification avoids the assumption that the random-effects are globally independent.

Ter Hofstede, F., Kim, Y., & Wedel, M. (2002). Bayesian prediction in hybrid conjoint analysis. *Journal of Marketing Research, 34*, 253-261.

Self-state attribute-level importance and profile evaluations are modeled as joint outcomes from a common set of part-worths. The likelihoods for the dataset differ and include other, incidental parameters that facilitate the integration of information to produce improved estimates.

Van Heerde, H. J., Mela, C. F., & Manchanda, P. (2004). The dynamic effect of innovation on market structure. *Journal of Marketing Research*, *41*, 166-183.

The effects of product innovation on market structure is investigate with a Bayesian state-space model where parameters follow an autoregressive process with intervention variables. Innovation is found to make th existing brands appear more similar.

Wedel, M., & Zhang, J. (2004). Analyzing brand competition across subcategories. *Journal of Marketing Research*, *41*, 448-456.

A sales model is developed capable of assessing cross-category price effects specific to individual stock keeping units. Variability across categories and stores is modeled using multiple random-effects distributions.

Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science, 19*, 297-312.

A multi-level model of attention and memory response is used to investigate the effect of brand, pictorial and text attributes of print advertisements. Information in the data are integrated through a multi-layered likelihood specification.

Yang, S., Chen, Y., & Allenby, G. M. (2003). Bayesian analysis of simultaneous demand and supply (with discussion). *Quantitative Marketing and Economics*, *1*, 251-304.

Consumer demand and prices are treated as dependent variables in an analysis that incorporates respondent heterogeneity using scanner panel data. The analysis investigates alternative supply-side models of behavior.

Yang, S., & Allenby, G. M. (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research*, *40*, 282-294.

The distribution of heterogeneity is modeled using a spatial auto-regressive process, yielding inter-dependent draws from the mixing distribution. Heterogeneity is related to multiple networks defined with geographic and demographic variables.

Yang, S., Allenby, G. M., & Fennell, G. (2002). Modeling variation in brand preference: The roles of objective environment and motivating conditions. *Marketing Science*, *21*, 14-31.

Intra-individual variation in brand preference is documented and associated with variation in the consumption context and motivations for using the offering. The unit of analysis is shown be at the level of a person-occasion, not the person.

Yang, S., & Allenby, G. M. (2000). A model for observation, structural, and household heterogeneity in panel data. *Marketing Letters, 11*, 137-149.

Structural heterogeneity is specified as a finite mixture of non-nested likelihoods, and covariates are associated with the mixture point masses.